ISSN 1304-7191
E-ISSN 1304-7205

sigma

Journal of Engineering and Natural Sciences

Mühendislik ve Fen Bilimleri Dergisi

Volume 44
Number 2
Year 2026

YTÜ PRESS

www.sigma.yildiz.edu.tr

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

EDITOR-IN-CHIEF

Ahmet Selim DALKILIÇ
Department of Mechanical Engineering, Yıldız Technical University, İstanbul, Turkey

ASSOCIATE EDITORS

İhsan KAYA
Department of Industrial Engineering, Yıldız Technical University, İstanbul, Turkey

Nergis ARSU
Department of Chemistry, Yıldız Technical University, İstanbul, Turkey

SECTION EDITORS

Ali ERDOĞMUŞ
Department of Chemistry, Yıldız Technical University, İstanbul, Turkey

Alper YILMAZ
Department of Bioengineering, Yıldız Technical University, İstanbul, Turkey

Azmi Seyhun KIPÇAK
Department of Chemical Engineering, Yıldız Technical University, İstanbul, Turkey

Banu MANSUROĞLU
Department of Moleculer Biology and Genetics, Yıldız Technical University, İstanbul, Turkey

Emel AKYOL
Department of Chemical Engineering, Yıldız Technical University, İstanbul, Turkey

Eyüp DEBİK
Department of Enviromental Engineering, Yıldız Technical University, İstanbul, Turkey

Fatih AKYOL
Department of Metallurgical and Materials Engineering, Yıldız Technical University, İstanbul, Turkey

Gülhayat GÖLBAŞI ŞİMŞEK
Department of Statistics, Yıldız Technical University, İstanbul, Turkey

Hakan YILMAZER
Department of Metallurgy and Malterials Engineering, Yıldız Technical University, İstanbul, Turkey

Hasan KÖTEN
Department of Mechanical Engineering, İstanbul Medeniyet University, İstanbul, Turkey

Hayriye SUNDU PAMUK
Department of Physics, Kocaeli University, Kocaeli, Turkey

Mesut AKGÜN
Department of Chemical Engineering, Yıldız Technical University, İstanbul, Turkey

Metin TÜLÜ
Department of Chemistry, Yıldız Technical University, İstanbul, Turkey

Muhammet Kasım ŞENER
Department of Chemistry, Yıldız Technical University, İstanbul, Turkey

Müslüm ARICI
Department of Mechanical Engineering, Kocaeli University, Kocaeli, Turkey

Nader JAVANI
Department of Mechanical Engineering, Yıldız Technical University, İstanbul, Turkey

Osman Nuri UÇAN
Department of Electrical and Electronics Engineering, Altınbaş University, İstanbul, Turkey

Ömer Said TOKER
Department of Food Engineering, Yıldız Technical University, İstanbul, Turkey

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

Necmiye Özlem ÜNVERDİ
Department of Electronics and Communication Engineering, Yıldız Technical University, İstanbul, Turkey
Selçuk ÇEBİ
Department of Industrial Engineering, Yıldız Technical University, İstanbul, Turkey
Sevil YÜCEL
Department of Bioengineering, Yıldız Technical University, İstanbul, Turkey
Vildan ÇETKİN
Department of Mathematics, Kocaeli University, Kocaeli, Turkey
Yasemin BUDAMA KILINÇ
Department of Bioengineering, Yıldız Technical University, İstanbul, Turkey

REGIONAL EDITORS

Abdelraheem Mahmoud ALY
South Valley University, Egypt
Abdullahi YUSUF
Federal University Dutse, Nigeria;
Biruni University Istanbul, Turkey
Adem KILICMAN
Department of Mathematics, University Putra, Malaysia
Amin SHAHSAVAR
Kermanshah University of Technology, Iran
Chandramohan V.P.
National Institute of Technology Warangal, India
Gobinath RAVINDAN
SR University, India
Hijaz AHMAD
University of Engineering and Technology Peshawar, Khyber Pakhtunkhwa, Pakistan
Md. Sabir HOSSAIN
Chittagong University of Engineering and Technology, Bangladesh
Mohammad Reza SAFAEI
Department of Civil and Enviromental Engineering, Florida International University, USA
Mostafa Safdari SHADLOO
Insa of Rouen, France
Muthu Poosan MOOPANAR
Department of Mathematics, National Institute of Technology, Warangal, India
Oscar CASTILLO
Tijuana Institute of Technology, Mathematics Division, Mexico
Pravin KATARE
Marathwada Mitra Mandal's College of Engineering, India
Ravindar KUMAR
Lovely Professional University, India
Sania QURESHI
Department of Basic Sciences and Related Studies, Mehran University of Engineering and Technology, Pakistan
Siamak HOSEINZADEH
Department of Planning, Design, and Technology of Architecture, University of Roma, Italy
Shengda ZENG
School of Mathematics and Statistics, Yulin Normal University, Yulin, China;
Faculty of Mathematics and Computer Science, Jagiellonian University in Krakow, Krakow, Poland
Taha M. RABABAH
Department of Nutrition and Food Technology, Jordan University of Science and Technology, Jordan

E-ISSN 1304-7205

Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

ADVISORY EDITORS

Bestami ÖZKAYA
Department of Civil Engineering, Yıldız Technical University, İstanbul, Turkey

Cengiz KAHRAMAN
Department of Industrial Engineering, İstanbul Technical University, İstanbul, Turkey

Erol TUTUMLUER
Department of Civil Engineering, University of Illinois Urbana-CHampaign, USA

Hari Mohan SRIVASTAVA
Department of Mathematics and Statistics, University of Victoria, Canada

Mohsen SHEIKHOLESLAMI
Babol Noshirvani University of Technology, Iran

Melike Elif BİLDİRİCİ
Department of Economics, Yıldız Technical University, İstanbul, Turkey

Somchai WONGWISES
Department of Mechanical Engineering, King Mongkut's University of Technology, Thonburi, Thailand

Umut Rıfat TUZKAYA
Department of Industrial Engineering, Yıldız Technical University, İstanbul, Turkey

Zekai ŞEN
İstanbul Medipol University, İstanbul, Turkey

Zeynep IŞIK
Department Civil Engineering, Yıldız Technical University, İstanbul, Turkey

SECRETARY

Senol BİLGİN
Yıldız Technical University, İstanbul, Turkey

Abstracting and Indexing:
- Directory of Open Access Journals (DOAJ)
- EBSCO Host Online Research Databases
- Emerging Sources Citation Index (ESCI)
- Scopus
- MIAR
- Scilit
- OUCI
- Research4Life
- Ideal Online

Journal Description: The journal is supported by Yıldız Technical University officially, and is a blind peer-reviewed free open-access journal, published bimontly (February-April-June-August-October-December).

Owner: Yıldız Technical University

Editor-in-Chief: Ahmet Selim Dalkiliç

Language of Publication: English

Frequency: Bimontly (6 issues per year)

Publication Type: Online e-version

Publisher: Kare Publishing

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

CONTENTS

Research Articles

747 **Process parameter optimization in fused deposition modeling of quad bike fixtures for improved mechanical properties and efficiency**
Mohansing PARDESHI, Pankaj BELDAR, Keshav NANDURKAR, Vishnu WAKCHAURE

768 **Electrooxidation-based on-site disinfectant generation using IrO_2-RuO_2 coated titanium anodes: Investigation and development**
Yeşim DEDE SAĞSÖZ, Alper Erdem YILMAZ, Beyhan KOCADAĞİSTAN, Donya AKBARZADEH YAZDİ

781 **The effect of different light intensities on the growth of *Chlorella Vulgaris* in advanced biologically treated poultry slaughterhouse wastewater**
Gamze KATIRCIOĞLU SINMAZ, İsmail Ayhan ŞENGİL

794 **Optimization of fibre laser parameters for micro-texturing in piston ring manufacturing**
Atulkumar PATIL, Pankaj BELDAR, Prashant KUSHARE, Vishnu WAKCHAURE, Uddhav SHIRSAT

820 **Exploring fractional dynamics of fourth order parabolic partial differential equations with power law Kernel**
Shelly ARORA, Wenxiu MA, Sukhjit SINGH DHALIWAL, Atul PASRIJA

843 **Finite element modeling and nonlinear analysis of tire deformation using hyperelastic material models**
Ashok MACHE, Shivraj KOTHEKAR, Aniket SALVE, Nishant KULKARNI

856 **Deep learning-based approach for automated detection of lumpy skin disease**
Rohit SINGH, Ramraj DANGI, Lokesh MALVIYA, Akshay JADHAV

869 **Bioconservative effect of *Cotula cinerea* Delile on Klila cheese**
Rym Nouria BENAMARA, Fatima SAHNOUNI, Asmaa BELGHARBI, Kada IBRİ, Smail KHALED, Mirvette BENFRÉHA BENYELLES, Sihem ZITOUNI, Hadjer ZEMMOURI

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

880 **Multi-zone commercial market building HVAC control strategy based on reinforcement learning algorithm models**
Ganesh MURADE, Bhanu Pratap SONI, Ankit Kumar SHARMA

894 **Narrowband and highly linear LNA design for 4.5 GHZ on 180NM CMOS**
Serdar KAYA, Alparslan Çağrı YAPICI, Hayrullah YILDIZ, Mübarek KAYA, Mustafa D. CERCI

908 **Missense single nucleotide polymorphisms analysis using bioinformatics tools in *SURF1* and *NDUFV1* genes associated with leigh syndrome**
Aleyna ÖZTÜRK, Ömer Faruk KARASAKAL, Muhsin KONUK

919 **Driving style recognition with machine learning for intelligent control of vehicles**
Ahmet BEŞKARDEŞ, Yakup HAMEŞ

954 **A study on numerical solutions of a fractional-order model for CAR T-cell therapy in leukemia using the laplace-adomian decomposition method**
Rezaul KARİM, M. Ali AKBAR, M. A. BKAR PK, Pinakee DEY

970 **A hybrid deep learning framework with attention mechanism for anti money laundering in cryptocurrency transactions**
Emine CENGİZ, Murat GÖK

983 **Investigating the effectiveness of IS-DOS prevention mechanisms for mitigating DOS attacks in vanet**
Harsh Pratap SINGH, Nagendra SINGH, Sai Kiran Naik BANOTH, Vedika PATIL, Neha ARYA, Neeraj SHRIVASTAVA

994 **Solving Hamiltonian systems based on a data-driven deep learning algorithm**
Tayfun ÜNAL, Ayten İrem IŞIK, Ünver ÇİFTÇİ

1004 **Selection of optimal coating materials for bamboo reinforcement in concrete using entropy-fuzzy topsıs model**
Vivek PAHUJA, Pradeep Kumar GHOSH, Anil KUMAR

1016 **Genetic algorithm based smart irrigation system for agricultural fields using linear regression and IOT to ensure proper utilization and saving of natural water resources**
Abhijit Dnyaneshwar JADHAV, Archana A. CHAUGULE, Harish U. TIWARI, Prashant G. AHIRE

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

1029 **The effects of the feature reduction by machine learning algorithms and feature decomposition by fuzzification on the mathematical model success criterion**
Filiz KANBAY, Melike GÖKSU

1041 **Synthesis of compounds containing Cyano Anchor Groups, production of dye-sensitive solar cells and invesigation of the antibacterial properties of these compounds**
Fatma AYTAN KILIÇARSLAN, Melda ALTIKATOĞLU YAPAÖZ, İbrahim ERDEN

1048 **Evaluation of reliability and performance in multi-subsystem serial system: Analyzing K-out-of-n redundancy in complex series-parallel configuration using Copula methodology**
Kabiru Garba IBRAHIM, Ibrahim YUSUF, Hamisu Ismail AYAGI

1069 **A comprehensive approach to decision making in uncertain environments using picture fuzzy hypersoft topology**
Muhammad IMRAN HARL, Muhammad SAEED, Muhammad HARIS SAEED

1082 **New ranking for picture fuzzy sets and novel way to address picutre fuzzy transportation problem**
K. HEMALATHA, VENKATESWARLU. B

1094 **Bayesian computational methods for estimation of q-Weibull distribution**
Ibrahim SADOK

1111 **Sustainability-financial assessment indicators for renewable energy systems using fuzzy swara/waspas approach**
Razieh KESHAVARZFARD, Mohsen ABOLHASANI

1127 **Dynamic mechanical, morphological, and thermal characteristics of polypropylene/red mud polymer composites**
Beril EKER, Özlem YAĞCI, Elif ULUTAŞ, Munir TAŞDEMIR, Volkan Murat YILMAZ

1138 **Assessment of future renewable energy use and CO_2 emissions of Türkiye and other countries in terms of climate targets of international agreements: An estimate with anfis**
Abdullah Zübeyr ŞEKERCİ, Berk AYVAZ

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

1153 **Optimized mix framework for accelerated carbonation concrete by ground granulated blast furnace slag and silica fume**
Ramatai B. SOMWANSHI, Abhijeet R. NARDEY, Dr. Padmanabh Arun GADGE

1172 **Investigation of correlation between standard penetration test blow counts and internal friction angles for loess soils**
Murat GÜLEN

1187 **Analyzing challenges to maximize the efficiency of industrial processes using carbon capture and hydrogen energy production from methane**
Kareem A. ABDELKAREEM, Ahmed El-AWADY

1219 **Hyperparameter optimization web tool: Hyperopt**
Fatma ALTINSOY, Muhammed Maruf ÖZTÜRK

1240 **Numerical simulation of magnetohydrodynamic casson Williamson hybrid nanofluid flow accounting thermal stratification**
Jintu Mani NATH, Ashish PAUL, Tusar Kanti DAS

1251 **Evaluation and effectiveness of student's awareness on outcome-based education perception and implementation in teaching/learning activities a cross-sectional study**
Vennila J., Basker P., Venkatachalam M., Prakash V. S., Dafik, Devi P.

1261 **Leveraging unexplored regional dynamics and temporal interdependencies in crop yield prediction: A graph theory approach**
Mamta KUMARI, Suman, Devendra PRASAD

1283 **A novel generalized lifetime *q*-distribution**
Nurgül OKUR, Kaoubara DJONG-MON

1299 **Selection of longitudinal fin profiles for electronic cooling using Delphi AHP-TOPSIS method**
Priyakhi SARMAH, Nishita GOGOI, Dikshita DUTTA, Dikshamoni SAIKIA, Tuhin DESHAMUKHYAL

1313 **Tribological analysis of textured ring-cylinder liner pair under varying operating conditions**
Atulkumar PATIL, Pankaj BELDAR, Vishnu WAKCHAURE, Prashant KUSHARE

E-ISSN 1304-7205

Sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

1328 **Solving multi-depot vehicle routing problem with incoterms: A case study for export logistics**
Gökçe ÖZDEN-GÜRCAN, Suna KÖSE

1344 **Current research aspects in minimum quantity lubrication for turning operations: Cutting fluid as mono and hybrid nanofluid**
Altaf NALBANDH, Neeraj CHAVDA[1,], Rakesh BUMATARIA

1369 **Effect of window glass area on optimum insulation thickness and environmental analysis**
Meral ÖZEL, Serhat ŞENGÜR

1383 **The role of building information modeling in enhancing public-private partnership project outcomes: Insights from Jordan and Malaysia**
Ali Al-SHAMAYLEH, Sharifah ZAKARIA, Nur RAHIM

1399 **Assessing the impact of natural language processing-driven clustering on engineering students' growth in mentoring programs**
Pankaj BELDAR

1419 **Enhancing dynamic performance of turbine-generator systems with optimal SVC control using hybrid CSA-GWO technique in series-compensated networks**
Sangeeta BAMBA, SK GUPTA

1440 **Exploring transient potential steps on rotating disk electrodes with steady-state concentration profiles in single electron transfer reactions**
V. SREELATHA DEVI, A. KAVITHA, M L. SURESH, D. VIJAYABALAN, T. LOGANATHAN, R. MANIMUTHUKUMAR

1456 **Arabic sign language recognition using enhanced optimization: Enhancing communication for people with hearing disabilities**
Mahmoud ROKAYA, Hossam MESHREF, Mehedi MASUD, Ibrahim GAD, Abdulqader M. ALMARS, Basel MAHMOUD, Ammar ALQARNI, El-Sayed ATLAM

1481 **Deep neural network multilayer perceptron for Parkinson's disease diagnosis using vocal data**
Nalineekumari ARASAVALI, Vinod Kumar MOGADALA, LS Narasimharao POTHANA, Sasibhushanarao GOTTAPU

E-ISSN 1304-7205

sigma
Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Volume 44 Number 2 Year 2026 - April

1491 **Dam-break analysis with infoworks ıcm model for earthquake-damaged Sultansuyu Dam: Different scenarios and computational costs**
Burak ÇIRAĞ, Alp Ertuğrul ÖZER, Rıdvan KARAGÖZ, Bilal UTANCİK, Alaaddin Özgüray AYDIN, Mahmut FIRAT

Review Articles

1509 **Recent advancements in refractory materials and digital innovations in precision manufactuing**
Mustafa M. MANSOUR

1526 **Systematic review and analysis on recent techniques of photo sketch matching**
Vishakha SEHDEV VERMA, Nidhi MALIK, Anshul BHATIA

1540 **The impact of digital design on Façade performance: A literature review**
Ayça AKKAN ÇAVDAR, Nilhan VURAL

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 747–767, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2008

Research Article

Process parameter optimization in fused deposition modeling of quad bike fixtures for improved mechanical properties and efficiency

Mohansing PARDESHI[1], Pankaj BELDAR[2,*], Keshav NANDURKAR[1], Vishnu WAKCHAURE[1]

[1]Department of Mechanical Engineering, Amrutvahini College of Engineering, Savitribai Phule Pune University, 411007, India
[2]Department of Mechanical Engineering, K.K.Wagh Institute of Engineering Education and Research, Savitribai Phule Pune University, 411007, India

ARTICLE INFO

Article history
Received: 30 October 2024
Revised: 30 December 2024
Accepted: 30 January 2025

Keywords:
3D Printing; ASTM Tensile Test Specimen; ATV QUAD Bike; Taguchi

ABSTRACT

This paper will involve optimization of the parameters of the 3D printing process of ASTM D638 test samples with Polylactic Acid (PLA) to produce fixtures to support a quad bike chassis that supports 155 kg. The experiment measures the effects of layer height, printing rate, nozzle size and infill density on production time, weight, cost and mechanical characteristics, such as hardness, tensile strength and roughness of the surface. The findings indicate that higher layer height, speed and nozzle diameter lead to less time spent on manufacturing with the most significant effect on manufacturing time being the layer height and the speed of printing. The dimensional accuracy of specimen 1 and 6 was close to 100%. The best surface roughness was topographically at a layer height of 0.2 mm, infill density of 100, nozzle diameter of 0.4 mm, and a print speed of 80 mm /s. When it comes to hardness, the most positive results were reached due to a layer height of 0.08 mm and a nozzle diameter of 0.2 mm. The peak tensile strength was obtained using the layer height of 0.4 mm, infill density of 80 percent, nozzle diameter of 0.15 mm, and a printing speed of 120 mm/s. The originality of this assignment consists in its use in automotive engineering where specially designed 3D-printed fixtures are used instead of standard metal parts, which are lightweight and economical. This study is very insightful and can be applied to improve quality and efficiency of 3D printing in manufacturing of quad bike fittings and further include more additive manufacturing.

Cite this article as: Pardeshi M, Beldar P, Nandurkar K, Wakchaure V. Process parameter optimization in fused deposition modeling of quad bike fixtures for improved mechanical properties and efficiency. Sigma J Eng Nat Sci 2026;44(2):747–767.

INTRODUCTION

The traditional manufacturing techniques, e.g. machining, casing and injection molding, have many issues, which not only affect the production efficiency but also the product quality. Quality variation is one of the foremost issues because it may be challenging to ensure a similar level of standards in the course of many production cycles. More current developments in the optimization of 3D printing parameters have

***Corresponding author.**
*E-mail address: pankajrbell@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

used advanced technologies to improve the performance of materials. Indicatively, Alafaghani et al. [1] said Moreover, traditional approaches also tend to lead to the material waste that can also serve as both the cause of an increase in the cost of production, as well as an even greater effect on the ecology. The equipment costs of these processes are also high especially when the production is in small batches or some parts are to be produced to order and so they are economically impractical to use in some applications. The rapid development of the 3D printing technology has put an immense burden on optimization of the put forth process parameters to enhance material performance. Research, including the one Kerekes et al. [2], has emphasized the significance of uniformity of processes and dimensional accuracy in 3D printing particularly in teaching students about ASTM tensile testing standards. The impact of the printing conditions on the mechanical characteristics of ABS composites was also highlighted by Ansari et al. [3]. The example here is tooling, molds and dies materials utilized in injection molding or casting, such materials may prove to be prohibitively-costly when it comes to production in low quantities. On the contrary, 3D printing, or additive manufacturing, has been found to provide flexibility and affordability. In contrast to the normal methods, 3D printing manufactures the parts or parts of the components according to a layered building method and thus the means of material is deposited where it is required. This minimizes wastage of material and makes it a more environmentally friendly and cost effective process especially where it comes to complex and customized parts. Moreover, 3D printing allows creating more complex geometries, which are hard to develop or even impossible to perform with conventional methods, e.g. highly detailed, lightweight, and customized parts of All-Terrain Vehicles (ATVs). Nevertheless, even though it has benefits, 3D printing still offers drawbacks in the form of no systematic approach to making process parameters optimal to generate sub-optimal print quality, inefficient production time, and components that are not strong enough. This is more so important in applications where there are tight tolerances and fits where the interaction between known parts should be precise. Many investigations have been done on how the process parameters including the layer height, the printing speed, the nozzle diameter and the infill density affect the 3D printed parts. To illustrate, research conducted by Kanu et al. [4] it was revealed that special attention to these variables is required to obtain the required material performance. Also, Kananathan et al. [5] addressed the subject of coconut wood-PLA composites demonstrating the fact that natural fiber composites can be used to enhance the properties of materials and provide predictive methods of optimization. A later study by Rahimizadeh et al. [6] and Anbalagan et al. [7] describes the importance of uniformity and dimensional accuracy whereas the same study and other researchers test how filament production and the material properties impact mechanical strength. Nonetheless, these efforts have not yielded the quest to have a systematic and comprehensive methodology of optimizing these parameters. Majority of the available methods involve commercial programs, which might not be flexible and adapting to special situations particularly when dealing with custom design components. Other scholars, have used the Taguchi technique to optimize the parameters of a 3D printing process, although this technique is still known to be in need of scalability. Furthermore, Response Surface Methodology (RSM) has been applied to model and optimize 3D printing parameters Pernica et al. [8], but frequently it lacks a degree of modeling the full scope of complex material behavior and geometry complexity common to more complex roles in which automotive engineering can be found.

There is a gap in literature given the absence of a framework to optimize 3D printing with high-level applications such as the ATV quad bikes. The production of ATV quad bikes needs parts with high precision of mechanical properties and dimensional limits, including customized fixtures with enhanced performance, handling, and ergonomics, but in this case, to reach these advantages, the process parameters should be fine-tuned with the quality of the parts and their strength.

This work can be motivated by the growing need of components of low weight, with specifications customized in the automotive industry, particularly in the ATV quad bikes, which demand components with particular mechanical properties, and of high precision. The existing methods of optimization of 3D printing are not flexible and adaptive to this application. This work has value in the fact that it introduces a more systematic and flexible optimization framework with Python programming that provides a more customized method to optimizing 3D printing processes. The main aim of the research is not only to optimize the part quality but also provides a new prospect of sustainable and economics production of the 3D printing with specific orientation on the production of the test specimen of ASTM D638 with the PLA (polylactic acid).

The paper has started with a thorough analysis of 3D printing processes and their key parameters such as the height of the layer, printing speed, the diameter of the nozzle, and the infill density. The purpose of this investigation was to find out how they affect the print quality, production time, and property of the materials. The successive production run operations were done in an iterative manner to progressively change these parameters to find the best combinations to get desired balance of the material efficiency, shorter production time and greater component strength. Then the printed components were put to strict mechanical tests to determine the hardness, surface roughness, and dimensional accuracy. The results of these tests yielded necessary information on the influence of various parameter settings on the quality of the parts overall and their performance. Lastly, statistical analysis of data obtained during these tests and production runs was performed and this included the Taguchi method. The result of this analysis was the creation of a predictive model, which helps to select

the specific values and parameters of 3D printing to ensure the attainment of certain goals, including maximizing force production, minimizing the time spent to produce, or high dimensional accuracy.

The study seeks to close the knowledge gap in the area of systematic utilization of the parameter optimization in 3D printing that will lead to the promotion of the total efficiency and effectiveness of the process, especially when the utilization of the PLA (Polylactic Acid) is applied. The research will help to improve the work of the 3D printing technology and its usage in the creation of the high-quality and reliable components by creating a strong optimization model. Demonstrates that the parameters of filament production and printing conditions have a direct effect on the strength and properties of 3D-printed materials that further highlights the necessity to optimize the parameters in detail.

Employed use of the Taguchi method in order to optimize continuous fiber reinforced polymer composite, and gain significant improvement in their performance. Vambol et al. [9] used the parameter swarm optimization to maximize its parameters. This tendency is observed in these studies in terms of the more complex methods of optimization of printing parameters associated with the goal of improving the properties of materials, which means that the sphere of additive manufacturing is becoming more sophisticated.

The development of multi- material and composite structures in 3D printing has given them signs of improvement in the properties of materials and versatility of use. Arora et al. [10] explored the application of PLA and ABS multi-material additive manufacturing and showcased advantages and disadvantages of using two separate materials together. In their study on the mechanical properties of the biocompatible materials employed to work in the dental sphere, Markiz et al. [11] demonstrated that the material selection is critical to the optimal results in medical purposes. The studies [12-14] shows the various hyperparameters tunning in fused deposition methods on wide ranges. As also emphasized in the studies by Cahyati et al. [15] and Hamat et al. [16], the quality and structural integrity of the 3D-printed parts is also dependent on compliance with the international standards. Sola et al. [17] have also emphasized on the development of sustainable natural fibre composites to convey specific functions with reference to

Table 1. Literature summary

Research area	Key findings	Identified gaps	Relevant papers
Material-process interaction	Studies show material properties strongly influence bonding and layer adhesion in FDM.	Lack of studies on the interaction between fiber-reinforced filaments and complex quad bike fixture geometries.	Kananathan et al. [5], Rahimizadeh et al. [6]
Geometry and design complexity	Complex geometries affect stress distribution and weight optimization in functional components.	Minimal exploration of topology optimization for lightweight, high-strength quad bike fixtures.	Pernica et al. [8], Vambol et al. [9]
Printing parameters and strength	Variations in nozzle temperature, layer height, and infill density affect tensile strength.	Limited focus on parameter optimization for high-load applications like quad bike fixtures.	Arora et al. [10], Markiz et al. [11]
Environmental factors	Temperature and humidity significantly impact part performance and dimensional accuracy.	Insufficient focus on UV resistance and long-term exposure effects for outdoor applications.	Darsin et al. [12], Yermurat et al. [13], Shergill et al. [14]
Durability and fatigue resistance	Fatigue resistance is crucial for load-bearing components in dynamic conditions.	Limited empirical studies on fatigue life specific to quad bike fixtures fabricated using FDM.	Hamat et al. [16], Sola et al. [17]
Testing and standardization	Standardized testing methods improve comparability across studies and industries.	Inadequate guidelines for testing custom-designed quad bike fixtures manufactured using FDM.	Rodriguez et al. [18], Akessa et al. [19]
Printing parameters and strength (Advanced)	Process optimization significantly improves mechanical performance under high loads.	Need for validation in real-world quad bike fixture applications.	Ahmad et al. [20], Baharudin et al. [21], Pyl et al. [22]
Sustainability in 3D printing	Recycling materials and minimizing waste improve sustainability.	Lack of case studies addressing sustainable FDM for high-performance automotive applications.	Tanikella et al. [23], Mazlan et al. [24], Sekar et al. [25], Zakernah [26]
Customization and scalability	Customization enhances usability but complicates manufacturing scalability for large-scale needs.	Few studies address the scalability of custom fixtures for mass production.	Kumar et al. [27], Zakaria et al. [28], Shaik et al. [29]

the contribution of eco-friendly materials to the development of 3D printing technologies. Rodriguez et al. [18] used response surface methodology to maximize the FDM parameters to enhance tensile strength.

The brief outline of literature review is presented in the Table 1.

Conventionality of the current testing procedures and quality controls in 3D printing is paramount to the quality and reliability of the materials used. whereas Tientcheu et al. [30] overviewed issues of tensile testing and demanded better standardization. These initiatives point to the commitment that has been in development of better testing methods and high standards in additive manufacturing.

This study is founded on what other studies have researched, as it focuses on the key gaps in optimization of the 3D printing parameters and the mechanical properties of PLA material. This study is an innovative one compared to previous literature which to a large extent depended on the use of commercial package such as Minitab to perform Taguchi analysis and ANOVA because it uses Python as its programming language to perform these analyses. This optimization of the results through hyperparameter optimization of the programming process has been possible and has taken a new standard in the results of optimizing parameters in the field of 3D printing research.

This research concentrates on four main parameters of the process, including: layer height, printing speed, nozzle diameter, and infill density with the aim of making the optimal set of findings with the aid of Taguchi (Python) and ANOVA and test regression. In order to ascertain the credibility of mechanical property testing, ASTM D638 standard specimens were set up to carry out tensile testing. This stringent strategy does not only increase credibility of the findings, but also allows reproducibility of the research in comparable ones.

The originality of this study is the fact that it is applied to auto engineering, and in this case, the 3D-printed fixtures are designed specifically to replace the customary metal fixtures on the ATV Quad bikes. This study constitutes a pioneer in the use of additive manufacturing in high-performance uses in that it shows that it is possible to create lightweight, cost-effective pieces using this technology. This approach, which is driven by Python and has its focus on sustainability and customization, makes the current study a groundbreaking project, which opens the field to the further study of multi-material composites and greener manufacturing approaches.

MATERIALS AND METHODS

The aim of the paper was to optimise the parameters of 3D printing process, starting with the choice of the most important parameters like the layer height, speed of the printing process, the diameter of nozzle and density of the infill which strongly affect the quality and the performance of the printed parts. A typical CAD model of test samples of ASTM D638 was prepared and subjected to Cura slicing software to create G-code to print. The Taguchi L9 orthogonal array was utilized to prepare nine test specimens in total, which permitted the intended assessment of a number of combinations with regard to the parameters. Measures of key parameters, such as manufacturing time, roughness of the surface, hardness and end result tensile strength, were taken and evaluated to determine the best parameters in the process. Finally, the optimal settings were chosen to produce a product, which would give an ultimate combination of quality, efficiency, and performance and provide the best mechanical properties.

Taguchi Array Designs

There are many approaches toward specifics design of experiments in accordance to certain production or procedure, and it is essential to select the most appropriate one to guarantee efficiency and accuracy of outputs. Since many methods were considered, the Taguchi experimental design proves to be the most popular in the whole world, which provides some specific benefits. It is highly useful because its ability to be resistant to noise factors can help it overcome real-world variability. Also, the Taguchi experiments have fewer runs than other designs of experiment (DOE) models, e.g. full factorial designs or response surface methodology, and therefore, save on time and resources. The simplicity and easy implementation aspect of the method makes it available to more people such as engineers and technicians. Moreover, Taguchi also gives a systematic approach to the experimentation process, including the process of AVC and analytical result, thus providing a comprehensive and properly carried out studies, which will produce credible results. Taguchi is in line with the objective of most industrial processes by concentrating more on quality enhancement, as well as their capacity to achieve multiple goals at the same time. Since this is a project, based on project focus and objectives, followed by theoretical and experimental analysis using Minitab statistical package, the Taguchi Array Design is the best tool in undertaking design of experiment.

The Taguchi Method was invented by Genichi Taguchi and is a statistical method that aims at enhancing the quality of the product in a least variation and lowest cost. This action starts with the clear definition of problem or goal of the experiment and definition of the factors that can affect the output [5]. Key factors are picked and the levels are determined and tested, these factors are either controllable such as machine settings or uncontrollable such as the environmental conditions. The experimental design is then designed to take advantage of the experimental factors, wherein an orthogonal array is used to design the experiment to be able to test all combinations of factor levels with the minimum amount of runs. This design matrix is then used to carry out the experiments and the observation of each combination is scrupulously recorded. The Taguchi Method does have an important element, namely, the

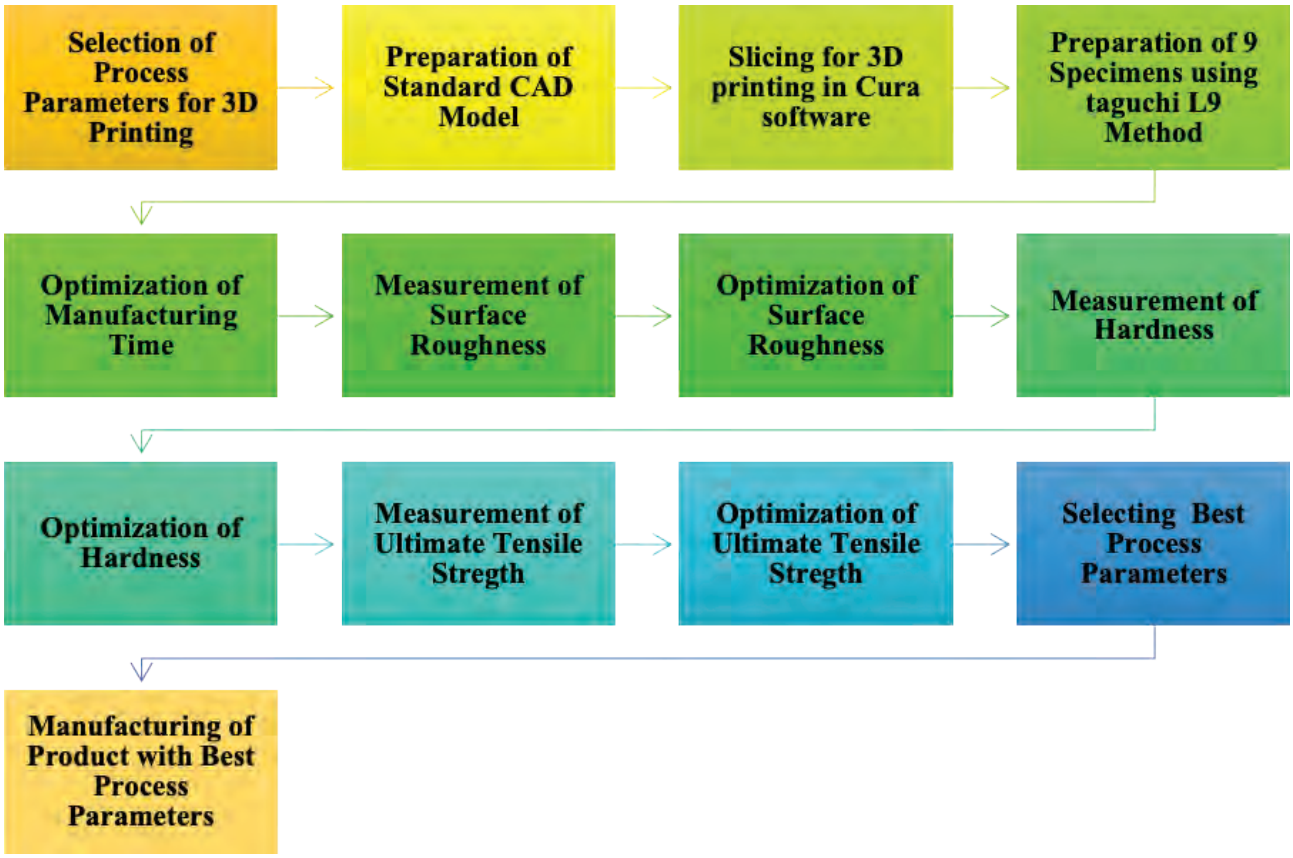


Figure 1. Methodology.

computation of the signal to noise ratio (S/N) at individual experimental runs that examines the quality features of an output. Taguchi recommends that we use the S/N ratio as compared to conventional means and variances because it is a more sound measure of performance.

Methodology for this research work is as follows shown in Figure 1

"larger the better" optimum condition shown by Equation 1.

$$\text{S/N ratio} = -10\log_{10}\left(1/n \sum_{i=1}^{n}\left(\frac{1}{y2}\right)\right) \quad (1)\ [30]$$

"smaller the better" condition is illustrated by Equation 2.

$$\text{S/N ratio} = -10\log_{10}\left(1/n \sum_{i=1}^{n} y2\right) \quad (2)[30]$$

where y denotes the unique response of the experimental data and n the number of replications.

Table 2 reveals the Process Parameters under Taguchi Optimization with the parameter optimization criteria

Table 2. Process parameters for Taguchi optimization

Sr. No.	Input parameters				Output parameters	Condition for S/N Ratio [30]
1	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Manufacturing Time (min)	Smaller is Better
2	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Surface Roughness of Bed Side (R_a) µm	Smaller is Better
3	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Surface Roughness of Top Side (R_a) µm	Smaller is Better
4	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Hardness (HB)	Larger is Better
5	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Ultimate Tensile Strength (N/mm^2)	Larger is Better

Table 2 shows the Process Parameters for Taguchi Optimization with the criteria for optimizing parameters.

- Optimization: Identify the optimal factor levels that result in the highest S/N ratio or the desired response. The goal is to find settings that are robust against variations and produce high-quality products consistently.
- Confirmation Experiment: Validate the results obtained from the optimal settings through confirmation experiments. Ensure that the optimal settings are indeed effective under practical conditions.
- Implementation and Control: Implement the recommended factor settings in the production process. Continuously monitor and control the process to ensure quality improvement and consistency.

The key principles of the Taguchi Method include efficiency (achieving results with minimal resources and experiments) and quality improvement (minimizing variability and achieving optimal performance).

Parameters and Materials Details

In this chapter, the materials and parameters of the 3D printing process are thoroughly examined. By investigating a wide range of 3D printing methods and analyzing various materials, the project seeks to understand the intricate relationship between materials and printing parameters, ultimately enhancing efficiency, dimensional accuracy, and component strength. Key factors include layer height, which refers to the thickness of each deposited layer; smaller heights yield finer details but extend print time. Print speed, the rate at which the printer's extruder moves, can expedite printing, but excessively high speeds may compromise quality and precision. Infill density, expressed as a percentage, indicates the amount of internal structure within a print; higher densities enhance strength but increase material use and print duration. Bed temperature is crucial for materials like ABS and PETG to ensure proper adhesion and minimize warping. Support structures, added to prevent overhangs from collapsing, are temporary scaffolds during printing. The cooling fan rapidly cools the material post-extrusion to solidify it quickly, particularly beneficial for PLA. Additionally, print bed adhesion is improved through various methods, such as using adhesives or heated beds. Nozzle size, which affects both layer resolution and print speed, varies; smaller nozzles provide detail but increase print time, while larger nozzles offer speed at the expense of detail. Lastly, travel speed, the rate of movement when not extruding, should be optimized to improve overall print efficiency. Specifications for PLA are detailed in Table 3.

Table 3. PLA specifications

Features	Amount
Name	Polylactic acid
Crystallinity	37%
Chemical Formula	(C3H4O2)n
Tensile modulus	2.7-16 Gpa
Density	1.210-1.430 g.cm^3
Melting point	150-160 °C
Glass transition	60-65 °C
Injection mold temperature	178-240 °C

Optimization Optimize the factors based on which you obtain maximally a high S/N ratio or some intended response. This aims at identifying the environment that can withstand fluctuations and yield quality products at all times.

Confirmation Experiment: Verification: Defense the findings of the optimal settings by confirmation experiments. Make sure that the optimal settings are effective indeed in real-life situations.

Implementation and Control: Realize the suggested setting factors in the production process. Continuous monitoring and controlling of the process will be done to allow quality enhancement and consistency.

The most important concepts of the Taguchi Method are efficiency (the attainment of results with the small quantity of resources and experiments) and quality improvement (the reduction of variability and optimal functioning).

In this chapter, the 3D printing process materials and parameters are analyzed in detail. The project aims to explore a diverse choice of 3D printing techniques and explore different materials, attempting to comprehend both of the complex interdependence between materials and printing parameters and ultimately achieve the best efficiency, dimensional quality, and component strength. The main ones are the height of the layers deposited, or its thickness; the smaller it is, the tiniest details will be produced but at the cost of a long printing process. The printing can be fastened by the speed at which the extruder travels in the printer, however, too high speeds can significantly affect quality and accuracy. The density of infill used as a percentage represents the penetration of internal structure in a print; higher densities will provide increased strength, but will also consume more material and require additional time to print. Bed temperature plays a vital role in such materials as ABS and PETG in order to achieve proper adhesion and to reduce the level of warping. The support structures consist of the scaffolds that are added to avoid collapsing of the overhangs; these are temporary structures at the time of printing. The cooling fan is fast to cool the material after extrusion to harden it fast especially in PLA. Moreover, print bed adhesion can be enhanced in a number of ways, including the adhesives or heated beds. Nozzle size, that is determined by layer resolution and print speed, varies, smaller nozzles give detail, but larger ones give speed at the price of detail, larger nozzles. Finally, the speed of travel, the speed at which nothing is being extruded, ought to be maximized to enhance the general efficiency of prints.

Specimen Design

In this research, the size of the specimens played a decisive role in terms of consistency and reliability in conducting tests. The size of the specimen was standardized to correspond with the industry standards so that all the samples looked the same way. Also, the principles of designing, were carefully abided by to make the specimens as efficient and easy to make as possible. These regulations included different aspects like selection of materials, geometrical limitations and bearing capacities. In addition, the ASTM specimen design was undertaken to comply with the internationally recognized tests. The reason behind this design was that it would offer comparability with existing studies and also make mechanical property appraisals accurate.

Experiments Design

The paper utilizes Taguchi Design of Experiments (DOE) design, which is the powerful statistical tool that is generally accepted to be effective in optimization of any process, together with the usage of Python as experimental design and analysis software. Which parameters are chosen is a crucial issue in any optimization study since it determines the quality and effectiveness of the work. Although there is a wide range of variables that could influence the 3D printing process such as temperature, type of material being used, rate at which the material cools and the orientation at which it is going to be built, in this study, four important parameters are being examined: layer height, printing speed, infill density, and nozzle diameter. These parameters have been selected because they play a key role in defining the quality, strength and dimensional accuracy of the parts being printed as demonstrated in vast literature in additive manufacturing. Besides, shrinking the selection to a few parameters makes the design and analysis of the experiment easier and more manageable which is a less expensive undertaking. These parameters are interdependent, so that simultaneous optimization of the parameters can result in greater benefits in both print quality and efficiency through synergy. Moreover, these parameters are usually manipulated by the 3D printing operators to optimize the printing process based on some solid needs in order to solve the real-life problems encountered by the additive manufacturing technology users and its manufacturers.

Parameters Overview

- Layer Height (mm): it is the y distance between layers of filament in the printing process. It directly affects the surface prettiness, the resolution and the print time of the model.
- Printing Speed (mm/s): This is the rate at which the material gets extruded and deposed to build platform. It influences the time of print and possibly the quality and accuracy of the end product or object that is printed.
- While it might seem contrary, Infusion Density is actually a percent that displays the percentage of the internal structure that has been filled during printing of the item. It influences the part strength, weight and material consumption of the part.
- Nozzle Diameter (mm): The nozzle diameter controls the width of extruded filament, and has an impact on the details, degree of accuracy and printing resolution.
- Experimental Design
- This project uses experimental design of Taguchi L9 layout which contains nine experimental runs. The study generated and tested ASTM D638 test specimens in order to determine the mechanical characteristics of the 3D-printed components. These specimens are also meant to determine the tensile strength and other relevant properties giving a standard way through which the performance of the printed materials can be compared.
- The choice to use the L9 approach or the approach that presupposes 9 experimental runs was proven on the basis of a number of technical and practical considerations:
- Resource Constraints: The resource-intensive nature of the experimental trials, e.g., that would be needed in projects like L18 or L27 design, would be a huge strain on resources, e.g., in form of time, material and machine usage. Since the project is too limited and has a scope, the approach of In Taguchi, L18 and L27 signify certain orthogonal arrays applied in the experimental design. Experiments with a number of combinations are considered with L18 orthogonal array up to 18 combinations which will be appropriate to analyze 6 factors potentially, each with 2 or 3 levels. This array assists in streamlining experiments because of minimizing the amount of runs required and still give useful information about factor effects and interactivity. Conversely, the L27 orthogonal array supports up to 27 combinations or 13 factors with each of the levels having a maximum of 3. Such array provides a more detailed analysis, which means that more complicated interactions between factors can be considered by the researchers using more experimental runs. An array of both cases is characteristic of the Taguchi method, which streamlines the experimental practice and improves the validity of findings. L9 method provides a compromise between resources efficiency and thoroughness of the experiment [31].
- Statistical Rigor: The size of the L9 orthogonal array (though smaller than other designs) is appropriate to be able to reliably identify the main effects but some of the interaction effects. This can be performed to robustly subject to statistical analysis and inference and avoid excessive experimentation [32]
- Time and Machine Prejudice: The L9 strategy is quite consistent with the time limitations presented by the project timeframes and access to 3D printing machinery. The mechanism of data collection is enhanced by minimizing the extent of experimental runs that sole reduce the time required to optimize the process in addition to yielding the required data as the entire process is streamlined to suit other needs of the project Ansari AA [3].

- Literature Support: A literature search on the optimization of the 3D printing process revealed that most success studies have used L9 designs or orthogonal arrays of the same.

Slicing of Specimens

Slicing in 3D printing involves converting a 3D model into printable layers. Cura software, frequently utilized for this purpose, enables users to adjust settings like layer thickness and infill density.

Machine specifications are shown in Table 4 as follows-

Table 4. Machine specifications

Machine specifications	Details
Manufactures Name	Make3D
Name of the Equipment	Pratham 5.0 Monster 3D Printer
Technology	FDM
Build Size	500 x 500 x 500 mm
Extruder Type	Single/Dual
Nozzle Size	0.4 mm
Dimensional Tolerance	0.1 to 0.05 mm
Print Speed	50 to 120 mm/sec
Layer Resolution	0.08 / 0.1 / 0.2 / 0.3 / 0.4 mm
Filament Diameter	1.75 mm

RESULTS AND DISCUSSION

Taguchi Analysis for Optimizing Time

The Taguchi Analysis used in the research aims at optimization of manufacturing time of 3D-printed parts by examining the effect of multiple factors: the layer height, the printing speed, nozzle diameter, and the infill density. This data set consists of nine experimental tests with differentiated combinations of these parameters, manufacturing time, weights and cost of each test. The Table 5 introduced the process parameters that were regulated in a systematic way to determine the best settings in order to minimize the production time. Plots of the main effect as in Figure 2 give a visual impression of the impact of the various parameters on the manufacturing time giving an idea of the most contributing factors. In order to support these results, ANOVA analysis was performed, whose findings are outlined in Table 6. The analysis revealed that the parameters selected were statistically significant. Moreover, the contribution of each of the parameters to the improvement of time is presented showing the percentage influence of each of the factors on the total manufacturing efficiency.

Cost Calculation Breakdown for sample 5:

1. Material Cost Calculation:
   - **Weight**: 11.42 gm
   - **Filament Cost per gm**: ₹1.00 (assumed to align with table cost)

   Material Cost=Weight×Cost per gm=11.42 gm×₹1.00/gm=₹11.42
2. Manufacturing Time Cost Calculation:
   - **Manufacturing Time**: 29 minutes = 0.483 hours (29 / 60)
   - **Machine Cost per hour**: ₹70 (adjusted to align with total cost)

   Time Cost=Manufacturing Time (hours)×Cost per hour=0.483 hours×₹70/hour=₹33.81
3. Additional Costs:

   To match the table value, we need to account for any additional fixed or overhead costs. Let's calculate the additional costs needed:

   Additional Costs=Total Cost−(Material Cost+Time Cost)
   =₹45.68−(₹11.42+₹33.81)
   =₹0.45

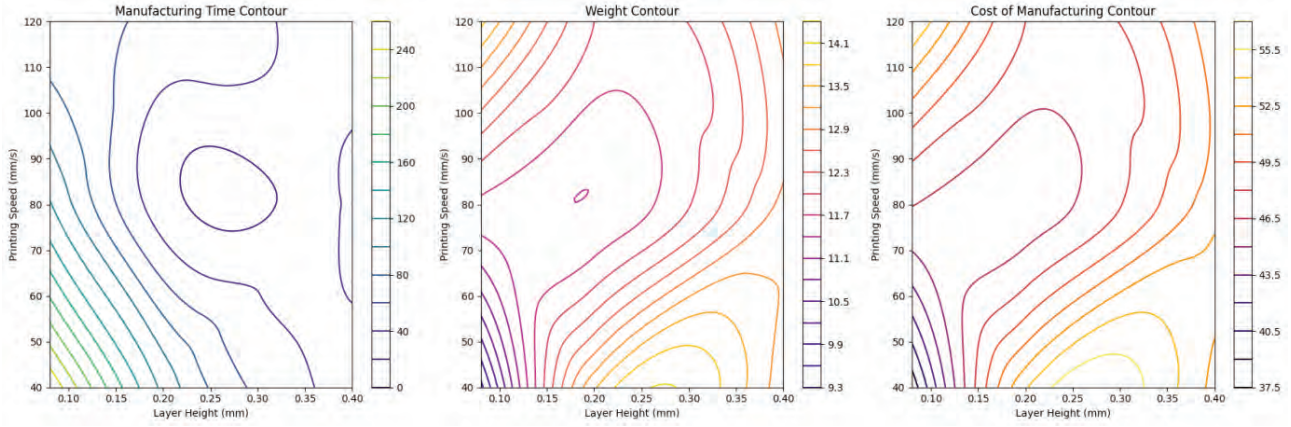
   Total Cost Calculation:

   Adding the material cost, manufacturing time cost, and additional costs:

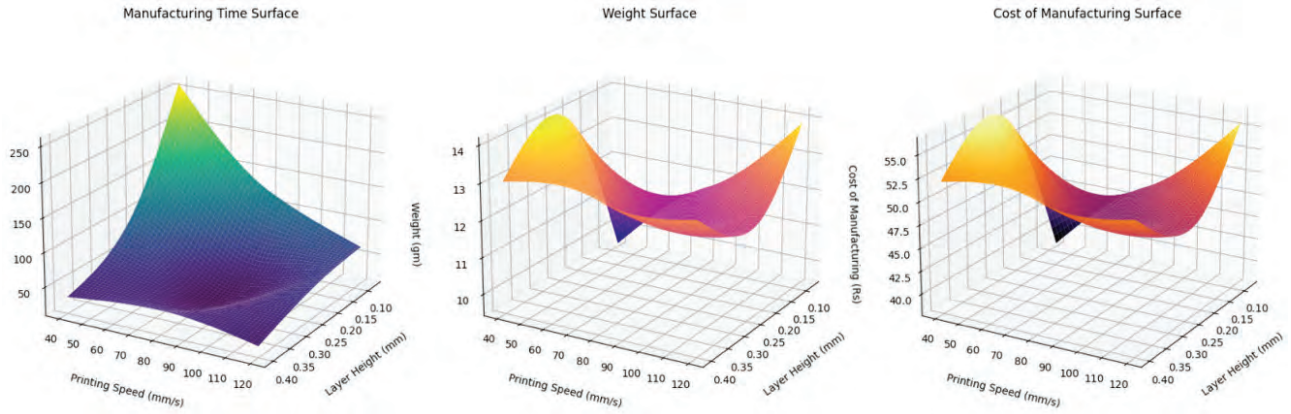
   Total Cost=Material Cost+Time Cost+Additional Costs
   =₹11.42+₹33.81+₹0.45
   =₹45.68

Table 5. Process parameters optimizing time

SR. No.	Layer Height (mm)	Printing Speed (mm/s)	Nozzle Diameter (mm)	Infill Density (%)	Manuf. Time (min)	Weight (gm)	Cost of Manufacturing (Rs)
1	0.08	40	0.15	60	258	9.5	38
2	0.08	80	0.2	80	122	11.64	46.56
3	0.08	120	0.4	100	64	13.89	55.56
4	0.2	40	0.2	100	115	13.45	53.8
5	0.2	80	0.4	60	29	11.42	45.68
6	0.2	120	0.15	80	57	12	48
7	0.4	40	0.4	80	33	13	52
8	0.4	80	0.15	100	44	13	52
9	0.4	120	0.2	60	25	13	52



a)



b)

Figure 2a. Counter plot, **2b.** Surface plots of all variables

Figure 2a and 2b shows the Counter plot and Surface plots of all variables respectively.

The findings indicate the interaction between the process parameters and the effect on manufacturing time, weight, and cost. Reduced layer heights (e.g. 0.08 mm) are capable of producing names of the fine detail, but reduce the speed of production by a great deal, at the cost of surface quality. submitted by increased layer heights (0.4 mm) is able to produce names more quickly but at the cost of surface quality. Increased printing speeds and nozzle size do not only shorten the time, but can affect the precision of parts. Infused density has a strong association with weight and cost where the higher the density, the stronger it is at a higher cost. The best settings will be determined by the requirements of the application with the low layer height and average infilling when dealing with precise parts and higher speeds and nozzle sizes when dealing with structural components that require efficiency. These trade-offs are very important in targeting the desirable results in FDM processes. * Layer Height: The results of the analysis show that the greater the layer height, the shorter the manufacturing time it will have. An example is where the maximum time was spent when the layer height was 0.08 mm (258 minutes), and the minimum time was spent when the layer height was 0.4 mm (25 minutes). This shows that the time spent in the manufacturing process may be tremendously reduced by raising the layer height. * Printing Speed: The faster the speed of printing, the slower the manufacturing time will be. The results indicate a strong tendency with swift printing speed being linked to a short manufacturing time with the lowest time being at the high speed (120 mm/s). * Nozzle Diameter: The effect of nozzle diameter on manufacturing time is not as great as it is in the rest. Although changing the nozzle diameter to 0.15 mm and then altering it to 0.4 mm has a certain influence on time, it is not as high as layer height and printing speed. As an illustration, a particular nozzle diameter of 0.4 mm gave shorter manufacturing time in some instances, which is more subtle. * Infills Density: The higher the manufacturing time is increased by increasing the infills density. To illustrate, components having an infill density of 100% would require more time to make in comparison to parts with an infill density of 60%. This implies that the denser the infill structure, the longer the production time so the importance of the infill density being optimized to balance between time and material consumption below Figure 3. Shows the Main effect plots for optimizing time.

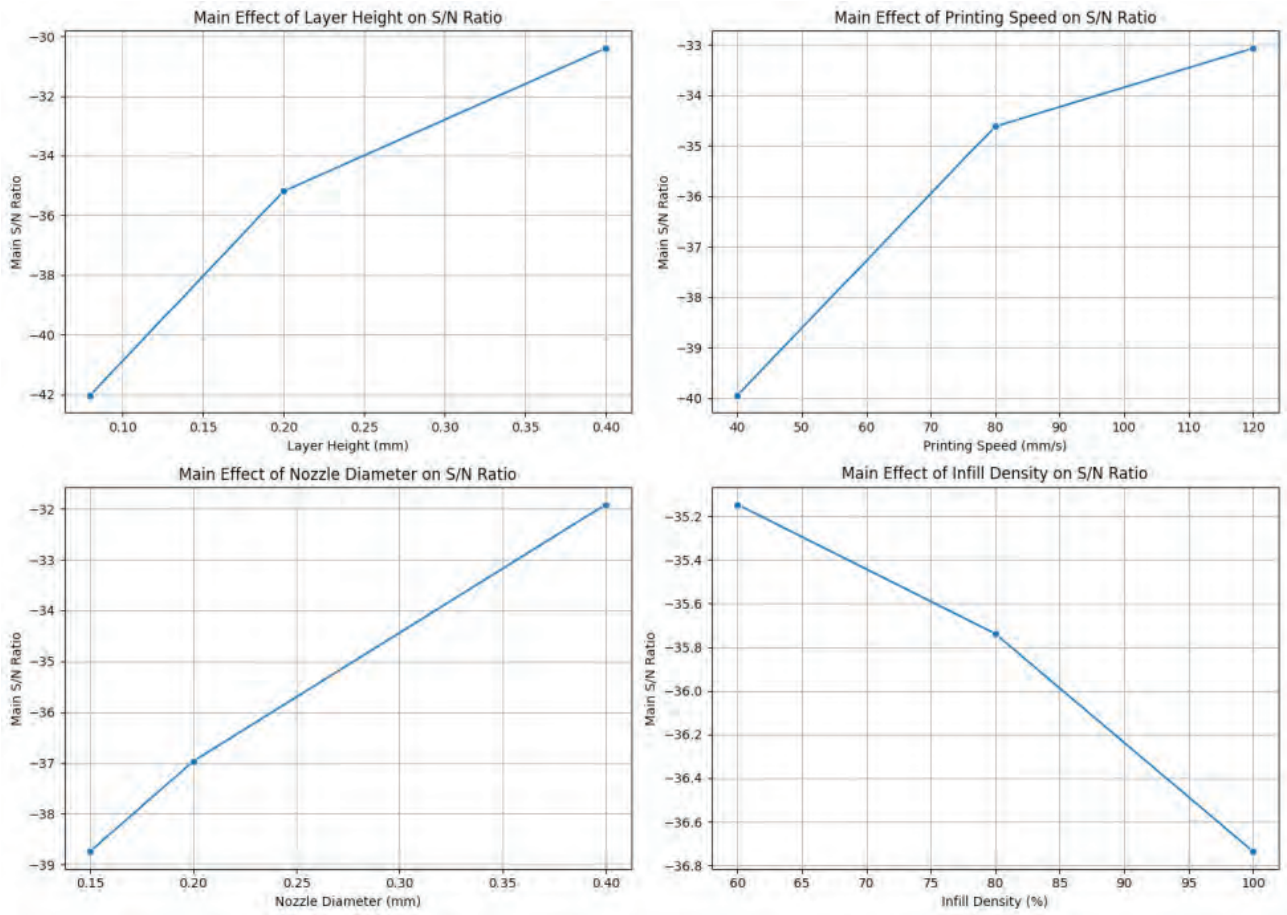


Figure 3. Main effect plots for optimizing time.

Regression Equation

$$\textbf{Manufacturing Time (min) = 399.41 + (-382.90 * Layer Height (mm)) + (-1.08 * Printing Speed (mm/s)) + (-290.51 * Nozzle Diameter (mm)) + (-0.85 * Infill Density (\%))} \quad (3)$$

Regression Equation 03 indicates that increasing layer height, printing speed, nozzle diameter, or infill density reduces manufacturing time. Specifically:

- **Layer Height (mm)**: Thickness of each printed layer in millimeters.
- **Printing Speed (mm/s)**: Speed at which the printer head moves during printing, measured in millimeters per second.
- **Nozzle Diameter (mm)**: Diameter of the printer nozzle, in millimeters.
- **Infill Density (%)**: The percentage of the volume inside the printed part that is filled with material.

The equation helps in optimizing production processes by adjusting these factors to achieve faster manufacturing times.

The ANOVA table of manufacturing time gives a closer examination of the difference in the manufacturing times as a result of the various parameters: layer height, printing speed, nozzle diameter and infill density. The results were a breakdown as follows:

* Layer Height (mm): The low p-value and high F-value of the layer height demonstrate that the manufacturing time was significantly affected. The layer height also adds around 40 percent of the total variance in the manufacturing time. This implies that production time is affected significantly by changes in the height of the layers.

* Printing Speed (mm/s): The contribution that printing speed has to manufacturing time is also significant as it amounts to around 25%. The F-value shows the moderate significance, whereas the p-value is slightly less than the traditional significance value of 0.05, meaning that printing speed is a somewhat worthy condition when determining manufacturing time.

* Nozzle Diameter (mm): Nozzle diameter is adding some 19 percent to the difference in manufacturing time. Whereas it has a slightly below the 0.05 p-value, meaning that it is not as statistically significant as any of the

Table 6. Anova table for manufacturing time

Parameters	sum_sq	df	F	PR(>F)	% Contribution
Q("Layer Height (mm)")	17779.5918	1	12.90034	0.022927	39.94696
Q("Printing Speed (mm/s)")	11266.6667	1	8.174754	0.045969	25.3138
Q("Nozzle Diameter (mm)")	8628.66667	1	6.2607	0.066614	19.38678
Q("Infill Density (%)")	1320.16667	1	0.957873	0.383139	2.966133
Residual	5512.90816	4	-	-	12.38633

previously mentioned parameters, its effect on the manufacturing time remains significant.

* Infill Density (%): The infill density causes the minimal impact on time of manufacturing with approximately 3 percent of the total variance. It has a high p-value so it cannot be said to be significant at the traditional level of 0.05 that results imply that infill density is not an important factor with its effect on production duration. The most significant factors that can influence the manufacturing time is the layer height and printing speed in which they contribute large portions to the variance. The diameter of nozzles is significant with a minor part being played by infill density. Such lessons can help in optimization activities aimed at minimizing manufacturing time. Surface roughness optimization Taguchi analysis. Below Table 6 shows the Anova table for manufacturing time

Taguchi Analysis for optimizing Surface Roughness

Surface roughness testing is needed to measure surface quality and texture produced using 3D printing. The Mitutoyo Surface Roughness Tester SJ-120 is an accurate device utilized to do so. In this research, there was a surface roughness tester that was utilized to measure the surface roughness of the samples. Table 7 provides the main process parameters to operate in improving the surface roughness and provides the values used in the experiments. The great implications of these parameters on surface roughness are illustrated in the main effects plots in Figure 4. Main effect plots for optimizing surface roughness that gives emphasis on the impacts of each parameter on the result. Process parameters for optimizing surface roughness shows in Table 7. The results of the analysis of variance (ANOVA) of the surface roughness of the top side are displayed in Table 8, whereas the values of the surface roughness of the bed side of the samples are presented in Table 9.

Measurement conditions used for the testing's are: -

1. Testing Standard – ISO1997
2. Profile – R
3. Filter – Gauss
4. N – 5
5. M speed – 0.25 mm/s
6. Range – 360 μ

The table of ANOVA of the surface roughness on the top demonstrates the degrees of variation in the impact of the different parameters on the surface roughness. The most influential factor is nozzle diameter with the contribution being about 59.30 to the variance of surface roughness and a fairly large F-value as well as p-value approaching the significance level ($p = 0.054$). This means that the nozzle diameter has significant influence on the surface roughness, but this is not significant at the traditional threshold of 0.05.

Speed of printing affects it moderately, with an F-value and a p-value of 6.46% to the variance, indicating that it has a small but significant effect on the variance, yet it is not

Table 7. Process parameters for optimizing surface roughness

SR. No.	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Infill density (%)	Surface roughness of top side (Ra) μm	Surface roughness of bed side (Ra) μm
1	0.08	40	0.15	60	0.88	0.055
2	0.08	80	0.2	80	1.135	0.035
3	0.08	120	0.4	100	1.713	0.021
4	0.2	40	0.2	100	1.367	0.021
5	0.2	80	0.4	60	3.335	0.1
6	0.2	120	0.15	80	0.992	0.165
7	0.4	40	0.4	80	2.121	0.071
8	0.4	80	0.15	100	1.097	0.038
9	0.4	120	0.2	60	0.038	0.346

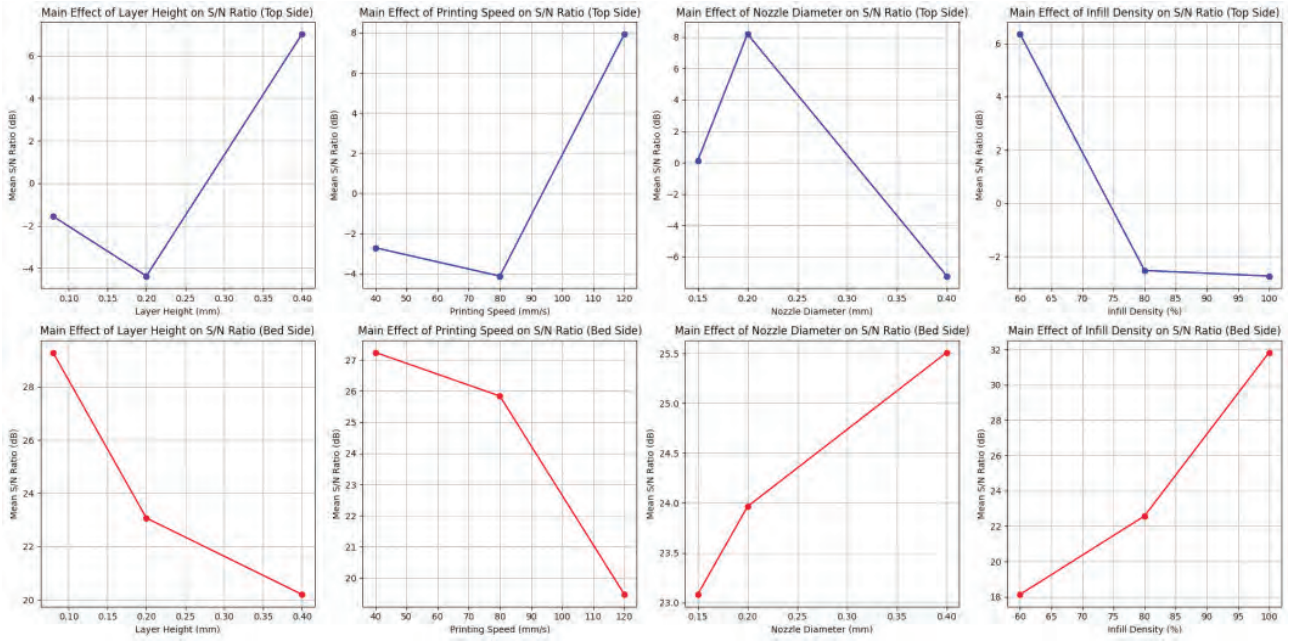


Figure 4. Main effect plots for optimizing surface roughness.

Table 8. Anova table for surface roughness of top side

Parameters	sum_sq	df	F	PR(>F)	% Contribution
Q("Layer Height (mm)")	0.114926	1	0.207203	0.672581	1.686023
Q("Printing Speed (mm/s)")	0.440104	1	0.793475	0.423389	6.456558
Q("Nozzle Diameter (mm)")	4.041782	1	7.287035	0.054125	59.295051
Q("Infill Density (%)")	0.000963	1	0.001736	0.968766	0.014123
Residual	2.218615	4	-	-	32.548246

very high to be taken as statistically significant. Conversely, layer height and infill density have a small impact on surface roughness. The impressiveness of the contribution of layer height is only 1.69 and has a large p-value meaning that it does not have a strong influence and infill density adds just 0.01 and has a large p-value which means that it does not affect much in this case.

The residual variance that explains the variance that is around 32.55 percent of the total variance represents the variance that cannot be explained by the variables under investigation. To sum up, nozzle diameter is the most important parameter influencing the roughness of the surface with other factors of the layer height, printing speed, and infill density having comparatively small impact.

Table 9. Surface roughness of bed side

Parameters	sum_sq	df	F	PR(>F)	% Contribution
Q("Layer Height (mm)")	0.019264	1	6.637755	0.061562	21.91536
Q("Printing Speed (mm/s)")	0.024704	1	8.512244	0.043348	28.10422
Q("Nozzle Diameter (mm)")	0.002785	1	0.95957	0.382759	3.168139
Q("Infill Density (%)")	0.02954	1	10.17857	0.033204	33.6058
Residual	0.011609	4	-	-	13.20649

The ANOVA table of the surface roughness of the bed side creates some insight on the contribution of various parameters to the roughness seen at the table. Here's a detailed discussion: Infill Density is found to be the most significant factor with a value of about 33.61 to the variance in surface roughness. The highest F-value (10.18) and 0.033 p-value show that the parameter is having a statistically significant impact on the surface roughness at the level of 0.05. The immense contribution indicates that a change in the infill density can have profound effect on the roughness of the bed side. Printing Speed is also playing a major role with the contribution of approximately 28.10. The F-value of 8.51 and p-value of 0.043 shows that it is an important variable that determines surface roughness. This means that variation in the speed of printing can influence the bed side surface roughness though, the influence is not so big as the infill density. The contribution of Layer Height to the variance is 21.92% with a F-value of 6.64 to a p-value of 0.062, which is just above the traditional value of 0.05. This indicates that a moderate amount of influence on surface roughness is affected by the layer height, although the impact is not as statistically significant as that of infill density or printing speed. The least contributing parameter among the parameters in the study is Nozzle Diameter which only adds 3.17 percent to the variance. The F-value of 0.96 and great p-value (0.383) show that nozzle diameter does not have a significant impact on surface roughness of the bed side. The remaining variance that explains about 13.21 percent of total variance is the unexplained variability which can be explained by the effects of the parameters under the study. However, infill density, printing speed are the most important considerations in the roughness of the surface on the bed side and the rest of the considerations make moderate or little effect, including layer height and nozzle diameter.

Equation 4 and 5 shows the Surface Roughness of Top Side and Bed Side

$$\text{('Surface Roughness of Top Side (Ra) µm} = 0.6440 + (-0.8561 * \text{Layer Height (mm)}) + (-0.0068 * \text{Printing Speed (mm/s)}) + (6.2043 * \text{Nozzle Diameter (mm)}) + (-0.0006 * \text{Infill Density (\%)})\text{'} \quad (4)$$

$$\text{'Surface Roughness of Bed Side (Ra) µm} = 0.2083 + (0.3505 * \text{Layer Height (mm)}) + (0.0016 * \text{Printing Speed (mm/s)}) + (-0.1629 * \text{Nozzle Diameter (mm)}) + (-0.0035 * \text{Infill Density (\%)})\text{'} \quad (5)$$

On the upper side, surface rugaeiness. * Layer Height (mm): -0.8561 is the coefficient which means that the higher is the layer, the smaller the surface roughness of the top side will be. It is not surprising since a larger layer height may lead to stronger layer lines and, therefore, roughness. * Printing Speed (mm/s): The coefficient of -0.0068 indicates that there is a very small impact of printing speed on the roughness of the surface. It means that the printing speed variation changes do not have a significant effect on the roughness of the top surface. * Nozzle Diameter (mm): The coefficient of 6.2043 would mean high nozzle diameter causes considerable change in the top roughness of the surface. This could be as a result of bigger extrusions resulting in greater surface defects. Bytes Infused On The Average: The value of -0.0006 shows that as infill density increases, roughness decreases slightly. Nonetheless, the influence is extremely insignificant, which indicates that the infill density does not influence strongly the top surface roughness. Surface Roughness of Bed Side. * Layer Height (mm): The coefficient that is positive with the value of 0.3505 implies that a rise in the layer height elevates the surface roughness of the bed side. The implication of this is that the upper layers increase the surface texture and ruggedness on the bed side. * Printing Speed (mm/s): The coefficient 0.0016 demonstrates a very insignificant positive influence of printing speed on the surface roughness. This means that the change in the speed of printing does not significantly affect the roughness of the bed side. * Nozzle Diameter (mm): The negative coefficient of extracts -0.1629, hence indicates that an increase in nozzle diameter brings about a reduction of the surface roughness of the bed side. Bigger nozzles would perhaps contribute to a smoother finish since it is easier to control extrusion. and Infill Density (%): The value -0.0035 of the negative coefficient implies that an increase in the roughness with an increase in the infill density is low. The effect it has is slight but it indicates some possible benefit of denser infill on the quality of the surface at the bed side.

Optimization Hardness Taguchi Analysis

Hard testing testing is critical analysis descriptive technique that determines mechanical characteristics and quality of 3D printed components. This is a test where resistance of material to indentation or penetration is measured and this gives an idea on the strength and the durability of the material. Rockwell hardness tester was used in hardness testing of specimens. Parameters of process involved in optimizing hardness are presented in Table 10 and the exact values and settings are used in the experiments. The results of analysis of variance (ANOVA) to optimize hardness can be observed in Table 11 and it provides an idea of the statistical significance of every parameter to analyze the results. The primary impacts of the said parameters on the hardness can be seen in Figure 5 that presents the main effect plot of hardness and identifies the impact of each factor on the hardness measured. The shape of the hardness test samples is evaluated by standard. The shape and dimensions of the test samples of hardness utilized in most materials, including additive manufacturing research, are usually influenced by guidelines including the cell of microhardness testing, e.g. ASTM E384, or the cell of Rockwell hardness testing, e.g. ASTM D785.

Table 10. Process parameters for optimizing hardness

SR. No.	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Infill density (%)	Hardness (HB)
1	0.08	40	0.15	60	620
2	0.08	80	0.2	80	630
3	0.08	120	0.4	100	628
4	0.2	40	0.2	100	632
5	0.2	80	0.4	60	610
6	0.2	120	0.15	80	615
7	0.4	40	0.4	80	621
8	0.4	80	0.15	100	630
9	0.4	120	0.2	60	607

Table 11. ANOVA table for optimizing hardness

Parameters	sum_sq	df	F	PR(>F)	% Contribution
Q("Layer Height (mm)")	53.882086	1	3.449392	0.136835	7.87494
Q("Printing Speed (mm/s)")	88.166667	1	5.644203	0.076341	12.88568
Q("Nozzle Diameter (mm)")	11.52381	1	0.737725	0.438816	1.68422
Q("Infill Density (%)")	468.16667	1	29.97082	0.005418	68.42319
Residual	62.482993	4	-	-	9.131974

The smart ANOVA table of hardness provides information on the influence of various parameters on the hardness of the material. The following is an analysis of the data rooted on the provided information: * Infill Density (%): The highest influence on hardness is taken by infill density (it contributes 68.42% to the variability in the results of the hardness measurements). The fact that the F- value is high (29.97) and the p-value is low (0.0054) depicts that the effect of infill density changes is emphatic on the hardness of the material. This implies that infill density must be well regulated to meet optimality of hardness. * Printing Speed (mm/s): The hardness variability has a contribution of 12.89% to the speed of printing. F-value is 5.64 which is relatively large but p-value of 0.0763 is marginally beyond the standard significance value of 0.05. This implies that printing speed can have significant impact on hardness but not significantly at 5 percent level. However, it may be applicable to real-world application. * Layer Height (mm): The layer height is adding 7.87% to the hardness variability. F-value of 3.45 is also less than that of printing speed and the p-value of 0.1368 exceeds the 0.05 threshold implying that layer height does not exert significant effect on hardness at 5 percent level. Its impact, as it can be measured is not that potent as infill density and printing speed. * Nozzle Diameter (mm): The most minimal contribution of 1.68 is made by the nozzle diameter. The high p-value of 0.4388 showed that nozzle diameter has no statistically significant influence on hardness and the F-value of 0.74. It implies that the hardness of the material does not depend much on the changes in the nozzle diameter. To conclusion, the infill density is the most influential parameter which influences hardness significantly and makes an important contribution. Hardness also depends on the speed of printing to a smaller degree but not significantly at the 5 per cent level. The hardness is insensitive to layer height and insensitive to nozzle diameter, with another aspect least affected being nozzle diameter. The findings present the role played by infill density in the optimization of the desired hardness in the material.

Regression Equation for Hardness (HB) is given in equation 06

$$\mathbf{HB = 600.60 + (-18.54) * Layer\ Height + (-0.10) * Printing\ Speed + (-10.48) * Nozzle\ Diameter + (0.44) * Infill\ Density} \quad (6)$$

Layer Height: The higher the Layer Height, the lesser the Hardness is, which decreases, by 18.54 HB used each millimeter. This implies that denser layers can make the material softer. * Printing Speed: Increasing printing speed reduces the Hardness by 0.10 HB units/mm/s, indicating that there is the possibility of using faster printing in terms of Hardness. Nozzle Diameter: Larger nozzle diameters found a decrease in Hardness of 10.48 HB units/millimeter, which underscores the fact that bigger nozzles provide soft prints. * Infill Density: A 1 per cent increase in Hardness is

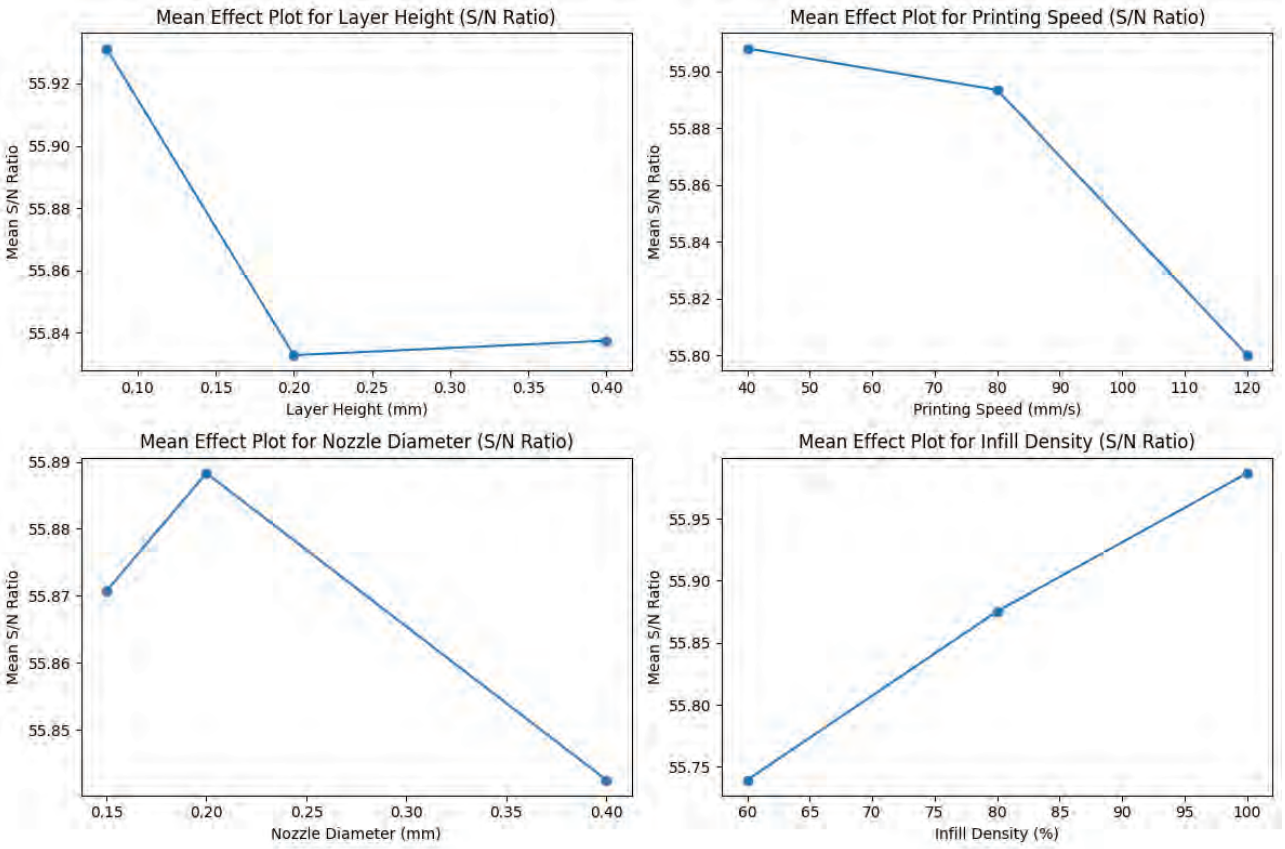


Figure 5. Main effect plot for hardness.

related to a 0.44 HB unit increase, that is, the denser is the infill, the harder is the specimen.

Taguchi Analysis for optimizing Ultimate Tensile Strength

Evaluation of tensile strength is an important test used to determine the mechanical properties and effectiveness of components that are manufactured through 3D printing. This test ensures the ability of a material to withstand tension or the pulling forces without breaking. Universal testing machine (UTM) is a versatile device that can be used to perform tensile strength tests on various forms of material, including those that are used in 3D printing. The prepared specimens were subjected to Tensile Testing with the help of a Universal Tensile Tester which was manufactured by Shanta Engineering. The reference samples that are to be tested are represented in Figure 6 and this shows a visual representation of the materials that are to be tested. Table 12 provides the process parameters of tensile strength optimization whereby the details of the conditions and variables are outlined to be implemented in the experimentation. Figure 7 below shows the main effect plot of ultimate tensile strength which illustrates the effect of each parameter on tensile strength. The outcomes of the analysis of variance (ANOVA) of ultimate tensile strength are presented in Table 13 that can provide a complete picture of the level of statistical significance and the impact of each parameter on tensile strength results. The testing system can take large loads of about 500 kg per unit, and hence can be used in numerous applications. It has a displacement resolution of 0.2 mm and this enables movement to be measured accurately. The column clearance is 380 mm and it offers a lot of space which can be used to test bigger specimens. Its cross head can be moved up to 900 mm with no grips or fixtures, which makes it flexible in the kind of testing of various kinds of samples. With an A.C. servo drive and variable speeds (1 mm/min to 500 mm/ min) are used, it can be adjusted to a large variety of testing requirements. The system needs 0.5 horsepower, which is not very high and can be used in the typical laboratory conditions. Figure 7 shows the Main effect plot for ultimate tensile strength

The ANOVA outcomes of Ultimate Tensile Strength (UTS) indicate that there is a substantial amount of information about the effect of different printing parameters on the material performance. The most powerful criteria is Layer Height (mm) that measures 90.41 percent of the variation in UTS. The F-statistic of this parameter is remarkably high (324.37) and the p-value of the parameter is extremely low (0.000056), meaning that the impact is very strong and statistically significant. The large percentage of

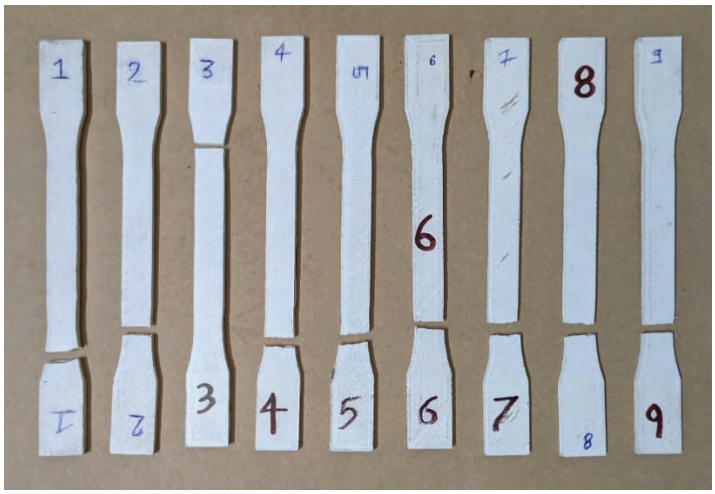


Figure 6. Tested standard specimens.

Table 12. Process parameters for optimizing tensile strength

Sr. No.	Layer height (mm)	Printing speed (mm/s)	Nozzle diameter (mm)	Density (%)	Ultimate tensile strength (N/mm²)
1	0.08	40	0.15	60	14.23
2	0.08	80	0.2	80	17.81
3	0.08	120	0.4	100	19.62
4	0.2	40	0.2	100	24.22
5	0.2	80	0.4	60	25.86
6	0.2	120	0.15	80	32.13
7	0.4	40	0.4	80	35.71
8	0.4	80	0.15	100	41.96
9	0.4	120	0.2	60	41.67

Contribution points to its dominant position in the determination of UTS, where maximizing of the height of the layer is important in maximizing the strength of the material. Printing Speed (mm/s) has also an influence on UTS but to a smaller extent as it explains 7.35% variation. This parameter (F-statistic 26.39, p-value 0.006806) is statistically significant, which indicates that the changes in printing speed do affect UTS, but the effect is much smaller when compared to the impact of layer height. Conversely, there is little effect of Nozzle Diameter (mm) and Infill Density (%) on UTS. The nozzle diameter is insignificant in the change in UTS (0.80) and the F-statistic is 2.86 and the p-value stands at 0.166192. On the same note, infill density explains only 0.32 percent of the variation, F-statistic of 1.16 and p-value of 0.341904, which indicates that it contributes insignificantly to UTS. The remainder or the amount of UTS not covered by the parameters in this model is the residual variance that returns the amount of variability that is not taken care of in this model 1.11%. This left-over shows that although the model is able to explain a significant amount of variability in UTS, there might be other things or other sources of variation that are not represented in the analysis. In a nutshell, the height of the layers is the most essential aspect that enhances UTS and printing speed also counts in the matter. Nevertheless, nozzles size, and infill density have insignificant influence, and unused variance implies that a further study on the subject might be necessary to capture other factors that do not create an anthropomorphic difference in UTS.

Regression Equation of UTS is given in equation 07:

$$\textbf{UTS} = 5.24 + (69.62) * \textbf{Layer Height} + (0.08) * \textbf{Printing Speed} + (-7.99) * \textbf{Nozzle Diameter} + (0.03) * \textbf{Infill Density} \quad (7)$$

- Layer Height significantly improves UTS, with each millimeter increase boosting UTS by 69.62 N/mm².
- Printing Speed has a minimal positive effect on UTS.

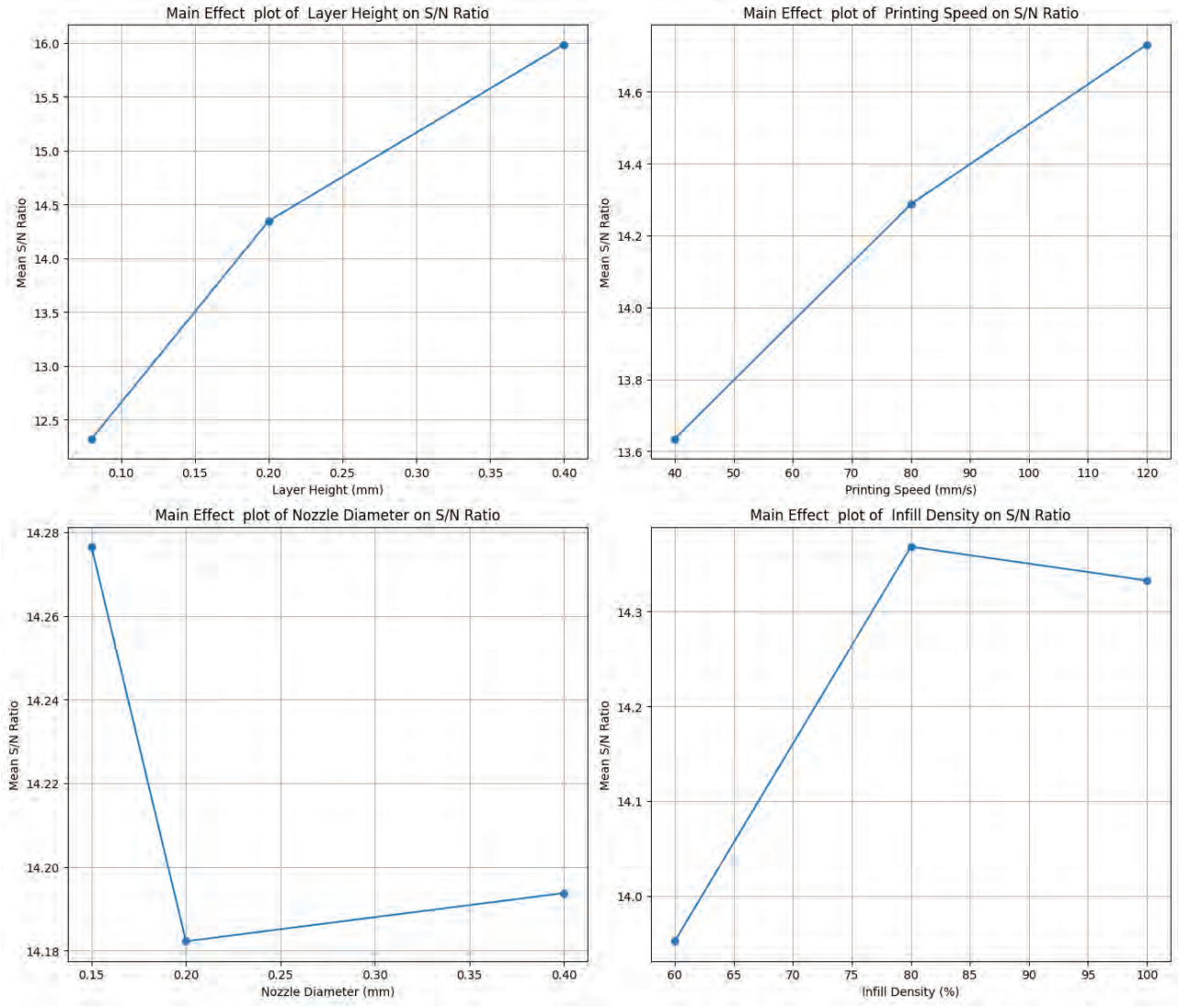


Figure 7. Main effect plot for ultimate tensile strength.

Table 13. ANOVA table for ultimate tensile strength

Parameters	sum_sq	df	F	PR(>F)	% Contribution
Q("Layer height (mm)")	760.0172	1	324.37385	0.000056	90.41041
Q("Printing speed (mm/s)")	61.8246	1	26.386617	0.006806	7.354554
Q("Nozzle diameter (mm)")	6.696021	1	2.857849	0.166192	0.796548
Q("Infill density (%)")	2.720267	1	1.161004	0.341904	0.323598
Residual	9.372115	4	-	-	1.114891

- Nozzle Diameter negatively impacts UTS; larger diameters decrease strength by 7.99 N/mm² per millimeter.
- Infill Density slightly increases UTS with each percent rise contributing 0.03 N/mm².

Layer height is the most influential factor, while nozzle diameter detracts from strength.

Dimensional Accuracy

Dimensional accuracy testing is crucial and first step in evaluating quality and precision of 3D printed parts. One common approach to perform this test is using a digital vernier caliper. The equipment used for this test is Mitutoyo Digital vernier caliper 300mm.

Table 14. Dimensional accuracy

3D Printed ASTM Specimen					% Accuracy			
SR. No.	Full length (mm)	Thickness (mm)	Parellel section width (mm)	Grip section width (mm)	Accuracy of full length (%)	Accuracy of thickness (%)	Accuracy of parallel section width (%)	Accuracy of grip section width (%)
1	172.9100	4.0900	13.0800	19.1350	99.9653	103.6670	100.6150	100.7110
2	172.9800	3.8466	13.3500	19.3400	99.9884	96.1667	102.6920	101.7900
3	172.7500	3.8644	12.8300	19.0750	99.8555	96.5833	98.6923	100.3950
4	172.8400	3.5866	13.0800	19.0650	99.9075	89.6667	100.6150	100.3420
5	173.0200	3.7200	13.1100	19.0200	100.0120	93.0000	100.8460	100.1050
6	173.1800	3.9933	13.4000	19.2950	100.1040	99.8333	103.0770	101.5530
7	172.9200	3.4600	13.2000	19.2900	99.9538	86.5000	101.5390	101.5260
8	173.0500	3.5833	13.3000	19.3150	100.0290	89.5833	102.3080	101.6580
9	172.8700	3.7100	13.1600	19.1300	99.9249	92.7500	101.2310	100.6840

The detailed measurements and results for dimensional accuracy are provided in Table 14**,** which lists the recorded values for each dimension.

The dimensional accuracy of the 3D printed ASTM specimens generally meets high standards, with full length and grip section width showing excellent consistency, ranging from 99.855% to 100.104% and 100.1053% to 101.6579%, respectively. However, thickness accuracy varies more significantly, from 86.5% to 103.6667%, indicating potential inconsistencies in the printing process. The parallel section width accuracy is fairly consistent, ranging from 98.69231% to 103.0769%. These variations suggest that while the overall printing process is reliable, improvements in calibration and process control are needed to enhance thickness consistency and achieve more uniform accuracy across all dimensions.

Using optimum settings and regression equations following parts are manufactured with the help of 3D printing for quad bike as shown in Figure 8.

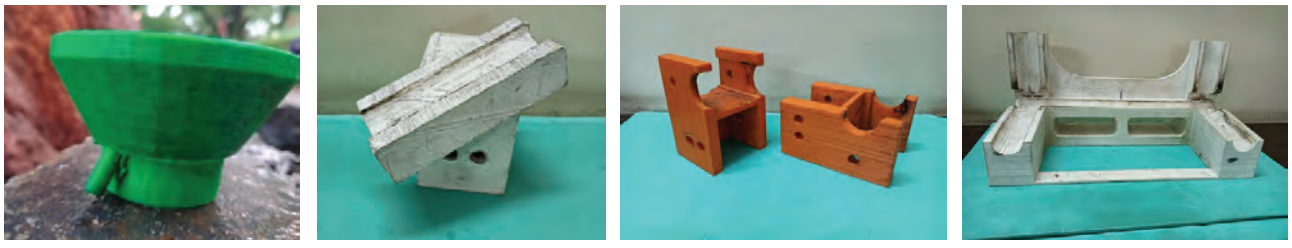
Figure 8 displays the 3D printed parts of the fixture for the quad bike, including: (a) the oil tank cap with funnel, (b) the fixture for the camber link assembly, (c) the welding fixture for the rear chassis' left and right mount position arm, and (d) the welding fixture for the rear chassis. This figure provides a comprehensive view of the various components designed and printed for the quad bike fixture.

Figure 9 shows the fixture positions for quad bike chassis during assembly.

Relations with the Current Research

The present research is closely related to several major spheres of current research and integrates fundamental knowledge from multiple domains. Previous studies by Kanu et al. [4] and Anbalagan et al. [7] highlighted the significant influence of environmental factors such as temperature and humidity on material efficiency, which must be carefully considered during the printing process. Furthermore, the work of Pernica et al. [8] and Vambol et al. [9] emphasized that complex geometries and manufacturability constraints play a crucial role in determining the structural performance of printed components.

In addition, Arora et al. [10] demonstrated that non-conventional printing conditions can lead to substantial variations in material behavior, affecting the final part quality. The multi-scale modeling approaches discussed



a) b) c) d)

Figure 8. 3D printed parts of fixture for Quad bike a) oil Tank Cap with funnel, b) fixture for camber link assembly, c) welding fixture for rear chassis' left and right mount position arm and d) welding fixture for rear chassis

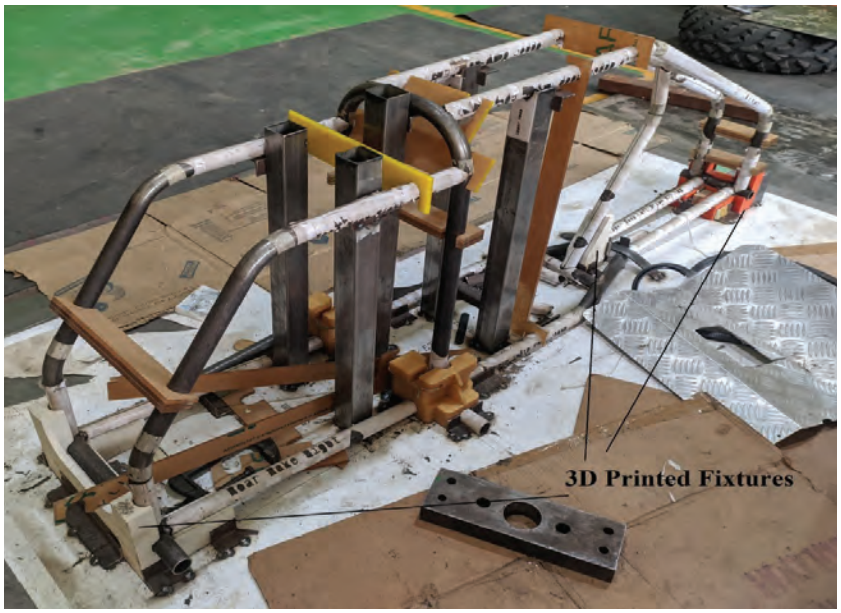


Figure 9. Fixture positions for quad bike chassis during assembly.

by Cahyati et al. [15] further support the need to integrate microstructural and macro-level performance analysis. Similarly, Hamat et al. [16] and Rodriguez et al. [18] contributed to the understanding of testing methodologies and practical manufacturability considerations.

The influence of printing parameters on mechanical properties has been extensively studied by Ahmad et al. [20] and Baharudin et al. [21], confirming that parameter optimization significantly enhances performance and consistency. Supporting this, Pyl et al. [22] investigated long-term durability and fatigue resistance, demonstrating the importance of material selection and process conditions in improving lifecycle performance. Sustainability aspects in additive manufacturing were addressed by Tanikella et al. [23], emphasizing the growing importance of recycling and environmental impact reduction.

Advanced modeling and simulation approaches proposed by Zakarneh et al. [26] further enhance predictive capabilities for material behavior. The importance of customization and user-specific design was highlighted by Kumar A et al. [27], showing its direct impact on performance optimization. Additionally, Zakaria S et al. [28] developed advanced testing procedures that improve the accuracy and reliability of material evaluation. Finally, Tientcheu SWT et al. [30] explored complex interactions between different materials and composites, confirming their significant influence on overall performance outcomes.

CONCLUSION

This paper has reviewed the effect of 3D printing parameters on manufacturing time, surface quality, hardness, and tensile strength through Taguchi and regression analysis using Python. The main conclusions are discussed as follows:

Manufacturing Time: Layer height, the speed of printing, and the nozzle diameter of the printing increased in significant amounts shortening manufacturing time. The impact was determined through regression analysis, and the major factors affecting it were the height of the layer and printing speed. The equation of manufacturing time indicates that it is strongly correlated with these aspects in a negative way.

Dimensional Accuracy: Both Specimen 1 and 6 had almost a perfect dimensional accuracy indicating the need to ensure accurate choices of parameters yield high-quality outputs.

Surface Roughness: The layer height of 0.2 mm, density of 100 percent, nozzle diameter of 0.4 mm, and printing speed of 80 mm/min the minimum level of roughness on the surface was attained. The regression analysis proved that the most important parameter that affects the roughness is layer height.

Hardness: The best hardness was attained using a layer height of 0.08 mm, nozzle diameter of 0.2 mm, 100% density and a print speed of 40 mm/min. The hardness equation stressed layer height and density, which is very crucial.

Tensile Strength: The maximum tensile strength was recorded when the layer height was 0.4 mm, the 80% density, nozzle diameter was 0.15 mm, and the print speed was 120 mm/min, which illustrates the trade-offs between the speed, structure strength and density.

The 3D-printed fixtures fabricated had a great impact on the welding accuracy of the Quad Bike chassis, and decreased the assembly time, which contributed to the overall stability of the product. This critical review highlights

the significance of parameterizing the processes to implement the desired results, which should represent a powerful scheme of optimizing manufacturing tasks using FDM.

Constraints and Bondage.

The study has given meaningful information on optimization of the manufacturing parameters, though limited to some extent. The study is limited to the range of parameters under test that might not represent all possibilities of variation. Also, the study mainly addresses mechanical properties and does not consider all the other aspects including degradation of the material overtime or the effect of the environmental conditions under long term use.

Future studies would be able to build on these results by considering a wider selection of parameters and materials. The durability and performance of the optimized parameters in real-world applications is also long-term, but a more detailed analysis would be conducted to gain a clearer insight into their efficiency. Moreover, it might be possible to involve the use of more advanced modeling methods and simulations that will provide more profound understanding of the relationships between different factors and their influence on performance. The study may also be extended in future to discuss how young technologies and materials may be incorporated in order to widen the choice and application of the research results.

ABBREVIATIONS

ATV	All-Terrain vehicle
ANOVA	Analysis of variance
ASTM	American Society for Testing and Materials
CAD	Computer aided design
3D	Three dimensional entities
PLA	Polylactic acid

ACKNOWLEDGEMENT

We acknowledge the 3D printing lab facility provided by AICTE IDEA Lab K.K.Wagh Institute of Engineering Education and Research Nashik and Amrutvahini College of Engineering Sangamner, Research Center of Savitribai Phule Pune University, Pune. for preparing specimens and testing of the fixtures.

AUTHORSHIP CONTRIBUTIONS

Authors' contributions- MP: conceptualization, PB: resources, methodology, investigation, visualization writing original and revised draft, KN: supervision and project administration, VW: investigation, visualization writing original and revised draft.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Alafaghani A, Qattawi A. Investigating the effect of fused deposition modeling processing parameters using Taguchi design of experiment method. J Manuf Process 2018;36:164–174. [CrossRef]

[2] Kerekes TW, Lim H, Joe WY, Yun GJ. Characterization of process–deformation/damage property relationship of fused deposition modeling (FDM) 3D-printed specimens. Addit Manuf 2019;25:532–544. [CrossRef]

[3] Ansari AA, Kamil M. Performance Study of 3D Printed Continuous Fiber-Reinforced Polymer Composites Using Taguchi Method. J Mater Eng Perform 2023;32:9892–9906. [CrossRef]

[4] Kanu RC, Hale C, Piper PON. The use of 3D printing to introduce students to ASTM standards for testing tensile properties of acrylonitrile-butadiene-styrene (ABS) plastic material. In: ASEE Annual Conference and Exposition, Conference Proceedings; New Orleans,USA. 2016. [CrossRef]

[5] Kananathan J, Samykano M, Kadirgama K, Ramasamy D, Rahman MM. Comprehensive investigation and prediction model for mechanical properties of coconut wood–polylactic acid composites filaments for FDM 3D printing. Eur J Wood Wood Prod 2022;80:75–100. [CrossRef]

[6] Rahimizadeh A, Kalman J, Henri R, Fayazbakhsh K, Lessard L. Recycled glass fiber composites from-wind turbine waste for 3D printing feedstock: effects of fiber content and interface on mechanical performance. Materials 2019;12:3929. [CrossRef]

[7] Anbalagan A, Launchbury EJ, Kauffman M, Pazhani A, Xavior MA. Investigation on CFRP 3D printing build parameters and their effect on topologically optimised complex models. Mater Today Proc 2023. [CrossRef]

[8] Pernica J, Vodak M, Sarocky R, Sustr M, Dostal P,Cerny P, et al. Mechanical Properties of Recycled Polymer Materials in Additive Manufacturing. Manuf Technol 2022;22:200–203. [CrossRef]
[9] Vambol O, Kondratiev A, Purhina S, Shevtsova M. Determining the parameters for a 3D-printing process using the fused deposition modeling in order to manufacture an article with the required structural parameters. East-Eur J Enterp Technol. 2021;2. [CrossRef]
[10] Arora A, Hasan MMF. Flexible oxygen concentrators for medical applications. Sci Rep 2021;11. [CrossRef]
[11] Markiz N, Horváth E, Ficzere P. Influence of printing direction on 3D printed ABS specimens. Prod Eng Arch 2020;26. [CrossRef]
[12] Darsin M, Mauludy RR, Hardiatama I, Fachri BA, Ramadhan ME, Parningotan D. Analysis of the effect 3D printing parameters on tensile strength using Copper-PLA filament. SINERGI 2022;26. [CrossRef]
[13] Yermurat B, Secgin O, Tasdemir V. Multi-material additive manufacturing: investigation of the combined use of ABS and PLA in the same structure. Mater Test 2023. [CrossRef]
[14] Shergill K, Chen Y, Bull S. What controls layer thickness effects on the mechanical properties of additive manufactured polymers. Surf Coat Technol 2023;475:130131. [CrossRef]
[15] Cahyati S, Putra RA. The Effect of Infill Pattern and Density Parameters on Tensile Strength of Polymer Materials in 3d Printing. J Rekayasa Mesin 2023;14. [CrossRef]
[16] Hamat S, Ishak MR, Salit MS, Yidris N, Ali SAS, Hussin MS, et al. Tensile Properties of 3D Printed Recycled PLA Filament: A Detailed Study on Filament Fabrication Parameters. J Adv Res Appl Mech 2023;110. [CrossRef]
[17] Sola A, Chong WJ, Simunec DP, Li Y, Trinchi A, Wen C, et al. Open challenges in tensile testing of additively manufactured polymers: A literature survey and a case study in fused filament fabrication. Polym Test 2023;117:107859. [CrossRef]
[18] Rodriguez O, Pence MA, Rodríguez-López J. Hard Potato: A Python Library to Control Commercial Potentiostats and to Automate Electrochemical Experiments. Anal Chem 2023;95. [CrossRef]
[19] Akessa AD, Adugna YW, Tucho WM. The influences of international standards on structural performance and fracture behaviour of 3D printed long fiber composite structures. IOP Conf Ser Mater Sci Eng 2023;1294. [CrossRef]
[20] Ahmad MN, Yahya A. Effects of 3D Printing Parameters on Mechanical Properties of ABS Samples. Designs 2023;7:136. [CrossRef]
[21] Baharudin ME, Saad MS, Zakaria MZ, Nor AM, Talib MHA. FDM Parameters Optimization for Improving Tensile Strength using Response Surface Methodology and Particle Swarm Optimization. J Adv Res Appl Sci Eng Technol 2024;38. [CrossRef]
[22] Pyl L, Kalteremidou KA, Hemelrijck DV. Exploration of specimen geometry and tab configuration for tensile testing exploiting the potential of 3D printing freeform shape continuous carbon fibre-reinforced nylon matrix composites. Polym Test 2018;71:318–328. [CrossRef]
[23] Tanikella NG, Wittbrodt B, Pearce JM. Tensile strength of commercial polymer materials for fused filament fabrication 3D printing. Addit Manuf 2017;15:40–47. [CrossRef]
[24] Mazlan MA, Mustar MF, Abdullah AH, Zakaria NAC, Hashim NM, Pangesty AI. Effects of Printing Parameters on the Mechanical Strength of Thermoplastics 3D Printed Specimens. J Mech Eng 2023;SI12:101–117. [CrossRef]
[25] Sekar V, Noum SYE, Putra A, Sivanesan S, Sheng DDCV. Fabrication of light-weighted acoustic absorbers made of natural fiber composites via additive manufacturing. Int J Lightweight Mater Manuf 2022;5:520–527. [CrossRef]
[26] Zakarneh B, Al-Ramahi N, Mahmoud M. Challenges of teaching English language classes of slow and fast learners in the United Arab Emirates universities. Int J High Educ 2020;9. [CrossRef]
[27] Kumar A, Chhabra D. Multidisciplinary topology and material optimization approach for developing patient-specific limb orthosis using 3D printing. Rapid Prototyp J 2023;29:1757–1771. [CrossRef]
[28] Zakaria S, Stighfarrinata R, Maghfiroh AM. Optimasi Parameter Proses 3d Printing Terhadap Kuat Tarik Filament Petg Menggunakan Metode Taguchi. JUSTI J Sist Dan Tek Ind 2023;3. [CrossRef]
[29] Shaik YP, Naidu NK, Yadavalli VR, Muthyala MR. The Comparison of the Mechanical Characteristics of ABS Using Three Different Plastic Production Techniques. OALib 2023;10:e10097. [CrossRef]
[30] Tientcheu SWT, Djouda JM, Bouaziz MA, Lacazedieu E. A review on fused deposition modeling materials with analysis of key process parameters influence on mechanical properties. Int J Adv Manuf Technol 2024;130:2119–2158. [CrossRef]
[31] Pastrnak A, Henriquez A, Saponara VL. Parametric study for tensile properties of molded high-density polyethylene for applications in additive manufacturing and sustainable designs. J Appl Polym Sci 2020;137. [CrossRef]
[32] John J, Devjani D, Ali S, Abdallah S, Pervaiz S. Optimization of 3D printed polylactic acid structures with different infill patterns using Taguchi-grey relational analysis. Adv Ind Eng Polym Res 2023;6:62–78. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 768–780, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2009

Research Article

Electrooxidation-based on-site disinfectant generation using IrO_2-RuO_2 coated titanium anodes: Investigation and development

Yeşim DEDE SAĞSÖZ[1,*], Alper Erdem YILMAZ[1], Beyhan KOCADAĞİSTAN[1], Donya AKBARZADEH YAZDİ[1]

[1]*Department of Environmental Engineering, Ataturk University, Erzurum, Türkiye*

ARTICLE INFO

Article history
Received: 16 October 2024
Revised: 23 December 2024
Accepted: 30 January 2025

Keywords:
Active Chlorine; Disinfectant; Electrooxidation; On-Site Generation

ABSTRACT

Disinfection is a critical process for preventing microbial contamination of water and surfaces, thereby protecting public health. Increasing hygiene needs and the need for a safe water supply necessitate effective and sustainable disinfectant generation. The study covers the examination and development of on-site disinfectant generation by electrooxidation method. With this process, disinfectants can be generated at the site of use in the required quantities. In addition, the fact that the generated disinfectant can be used without storage causes its disinfectant activity to be high, which makes disinfection more effective. An electro-oxidation reactor with 100 × 100 × 200 mm dimensions was designed, utilizing F-type titanium cathodes and IrO_2-RuO_2 coated F-type titanium anodes. The effects of experimental parameters, such as pH (5-8), flow rate (20-100 mL/min), current intensity (5-25 A), and NaCl concentration (1-5 M), were investigated. Based on the obtained data, the most efficient active chlorine generation occurred at pH 7. As the flow rate increased, residence time in the reactor decreased, leading to reduced energy consumption. Considering both energy efficiency and disinfectant generation, the optimum flow rate was determined as 40 mL/min. Current intensity up to 20 A increased disinfectant generation; however, the expected rise at 25 A was not observed. Thus, the optimum current intensity was selected as 20 A. Salt concentrations up to 3 M enhanced disinfectant generation, but higher concentrations did not yield further improvements. Under optimum conditions (pH 7, 40 mL/min flow rate, 20 A current intensity, 3 M salt concentration), 5450 mg/L of disinfectant was generated. Additional experiments to assess higher current intensities and saturated salt concentrations demonstrated 7750 ppm disinfectant generation at 50 A, placing this value among the highest reported in the literature. This study evaluates a broad range of parameters to determine optimal conditions, demonstrating that the disinfectant yield from this reactor is applicable for both laboratory-scale and field applications.

Cite this article as: Dede Sağsöz Y, Yılmaz AE, Kocadağistan B, Akbarzadeh Yazdi D. Electrooxidation-based on-site disinfectant generation using IrO_2-RuO_2 coated titanium anodes: Investigation and development. Sigma J Eng Nat Sci 2026;44(2):768–780.

***Corresponding author.**
*E-mail address: yesim.dede@ogr.atauni.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Disinfection is an important process used to eliminate harmful microorganisms and prevent their spread [1]. Considering the decreasing global water resources [2], disinfection is a vital process to ensure health and safety standards, especially in drinking water and wastewater treatment plants [3]. The disinfection process is used for surface cleaning in homes, public areas, the health sector, and various industries as well as drinking water and wastewater treatment plants [4,5]. Today, chlorine and chlorine compounds are the most widely used disinfectants in chemical disinfection applications [4,6]. The main reason for this situation is the high disinfectant efficiency of chlorine [7,8]. In addition, the easy availability of chlorine, its low cost, and its application with simple techniques are some of the reasons why it is frequently preferred [9,10].

Chlorine, especially in the form of hypochlorite, has been used for cleaning and disinfection for years [11]. Sodium hypochlorite (NaOCl) is known as a highly effective disinfectant in the elimination of harmful microorganisms [12,13]. Another advantage that makes NaOCl stand out is that it shows high effectiveness even at low doses. Even at a concentration of only a few ppm, it can inactivate most harmful microorganisms [14,15]. Another advantage of NaOCl is that it breaks down quickly and easily after application. In this case, the risk of environmental accumulation is eliminated. While providing effective disinfection, the negative effects on the environment and human health are also quite low [14,16].

Due to its high activity, NaOCl is widely used as a disinfectant in household cleaning, health institutions, various industrial facilities, and drinking water treatment plants [17,18]. However, some problems may arise in its storage and preservation. NaOCl in solid form is chemically unstable. It tends to deteriorate under atmospheric conditions [19]. NaOCl in liquid form is more stable than its solid form, but it must be stored at low temperatures. If it is not stored under appropriate conditions, it decomposes and causes the release of chlorine gas, which is harmful to human and environmental health [20]. To prevent these problems, the use of on-site disinfectant generation (OSDG) technologies in disinfectant generation offers a great advantage in terms of sustainability [21].

In classical disinfectant generation, disinfectant is transported from the centers where it is produced to the places where it will be used. The transportation distance is usually long distances. This situation both increases the carbon footprint and reduces the effectiveness of the disinfectant if it is not transported under appropriate conditions. In addition, failure to provide appropriate storage conditions in the places where it will be used causes the disinfectant to deteriorate and become an environmental risk [22]. OSDG offers a reliable and environmentally friendly alternative, especially in areas where hygiene is critical.

In OSDG, the electrooxidation process (EOP) is widely used [23]. EOP is a process that stands out with its short reaction time, versatility, environmental compatibility, low investment cost, high degree of automation potential, ease of operation, and safety [24, 25]. This method is an important process widely used in the removal of many organic and inorganic pollutants, as well as OSDG [26]. EOP is essentially a process based on the treatment of organic substances by direct and indirect oxidation using insoluble anode electrodes [26, 27]. The stability and catalytic activity of the anode electrode in EOP is a critical factor that determines the reactor's operating life and oxidation capacity [28, 29]. Therefore, it is very important to select the anode electrode appropriate to the experimental conditions in EOP studies. The anode types commonly used in EOP are platinum [30], graphite [31], stainless steel [32], dimensionally stable anodes (Ti/PtO_2 -IrO_2, Ti/RuO_2 -TiO_2, Ti/RuO_2 -IrO_2, and Ti/IrO_2 -Ti_2O_5) [33,34], boron-coated diamond [35], ruthenium [36], and titanium [37]. Boron-coated diamond, lead, graphite, and pure titanium anodes are anode electrodes with high catalytic activity [38].

Apart from the electrode type, the distance between the electrodes, flow rate, current density, electrolyte type and concentration, feed water pH value, temperature, and placement of the electrodes are important [39,40].

The basic working principle of EOP and OSDG is the generation of active chlorine compounds from salt water solution. Salt solutions that can be used in this process are $MgCl_2$, KCl , HCl and NaCl. The most commonly used electrolyte type among these is NaCl [41,42]. The reasons why NaCl is widely preferred can be listed as; low cost, easy availability, practical transportation and easy dissolution in water [43]. In addition, the formation of NaOCl, which is widely used as a strong oxidant and disinfectant, is another reason why NaCl is preferred. Chlorine and chlorinated disinfectants (NaOCl, OCl^-, HOCl) are obtained in the EOP and OSDG process where NaCl and pure water solution are used [23, 44]. In addition, this process allows the generation of disinfectants in the required amounts and concentrations and eliminates the need for storage [45, 46]. Therefore, it provides a more environmentally friendly and low-cost solution compared to traditional disinfection methods [47,48]. However, depending on the purpose and frequency of use, the electrodes to be used and the reactor design as well as the optimum conditions must be determined appropriately. Since the electrodes used wear out over time, the selection of long-lasting electrodes contributes to the process [49].

The mechanism of on-site disinfectant generation by electrooxidation occurs through the reactions given in equations 1-7 [50].

Reactions occurring at the anode;

$$2Cl^- \rightarrow Cl_{2(g)} + 2e^- \quad (1)$$

Reactions occurring at the cathode;

$$2H_2O + 2e^- \rightarrow 2OH^- + H_2 \quad (2)$$

Reactions occurring in solution;

$$Cl_2 + H_2O \rightarrow HOCl + H^+ + Cl^- \quad (3)$$

$$HOCl + OH^- \rightarrow H_2O + OCl^- \quad (4)$$

$$Na^+ + OH^- \rightarrow NaOH \quad (5)$$

$$Cl_2 + 2NaOH \rightarrow NaOCI + NaCl + H_2O \quad (6)$$

The final reaction is;

$$NaCl + H_2O \rightarrow NaOCl + H_2 \quad (7)$$

Although many studies exist on this subject, further research is needed to enhance the efficiency and controllability of the process. In this context, this study focuses on the research and development of the OSDG process via EO. An EO cell has been designed, and experimental parameters have been established for the operation of the process. The effects of initial pH, salt solution concentration, flow rate, and current intensity on the process have been examined. As a result of these examinations, the conditions for achieving optimum efficiency in the process have been determined.

MATERIALS AND METHODS

EOP and OSDG; 5 pieces of 80 x 180 mm F type titanium mesh cathodes and 5 pieces of 80 x 180 mm IrO_2-RuO_2 coated F type titanium mesh anodes have been provided. The main reason for preferring titanium coated anodes for EOP and OSDG process in the study is that these anode types have high catalytic activity and their resistance is suitable for salty environment conditions [51,52]. The reason for choosing mesh anodes is that they provide low energy loss and homogeneous current distribution [53]. Total surface area of the plates has been calculated as 2880 cm^2 . Surface area of the electrodes has an important role in terms of reaction efficiency and disinfectant generation capacity. Increasing surface area of the electrodes supports increasing reaction rate and disinfectant generation. However, increasing surface area increases energy consumption and initial investment cost. For these reasons, establishing a balance between efficiency and energy consumption is very important. After the electrodes have been supplied, the EO cell has been designed with dimensions of 100*100*200 mm and the distance between the electrodes has been fixed as 5 mm. The distance between the electrodes is a parameter that affects the interaction between the reactor efficiency, energy consumption and the electrode surface in the electron oxidation process. The reason for fixing the distance between the electrodes as 5 mm in the study is that it is a common value used to optimize these parameters [54,55]. After the electrodes have been placed in the reactor, steel screws and silicone gaskets have been used to close and seal the reactor. The durability and chemical resistance of the silicone gasket have been tested in our previous studies. However, it needs to be replaced at certain intervals for long-term use.

Solutions prepared with NaCl and pure water have been used for the experiments. There is an inlet and outlet system in the cell where NaCl solution is given at a constant concentration and flow rate. The disinfectant solution produced is taken from the outlet line. The inlet flow rate is provided in a constant and controlled manner by a peristaltic pump. The solution taken from the outlet line is collected in a graduated container. Possible deviations of the flow rate have been observed during a full cycle. As a result of the observations, no significant deviations have been generally detected in the outlet flow rates. A DC power supply (Chrom a 62024P-40-120) capable of producing 40 V-120 A power has been used in the experiments.

The analysis of disinfectants taken from the exit line has been carried out by Thermo It has been done with a scientific orion star portable pH/Ion meter. The device has been calibrated at certain intervals. The dilution method has been applied while analyzing the samples. In order to reduce the error rate of the measurements, three repetitive analyses have been made for each output sample. In general, similar results have been obtained, but a more reliable value has been obtained by averaging the results obtained. In experimental studies, the temperature and pH parameters of the feed water and the produced disinfectant have been measured with a WTW brand multi-parameter meter. This device has been calibrated at certain intervals and measurement security has been ensured.

Some experimental conditions have been repeated at regular intervals to increase reliability, and generally consistent results have been obtained.

The experimental setup where OSDG with EOP takes place is shown in Figure 1.

The experimental parameters studied in the experimental setup in Figure 1 are given in Table 1. With the help of the fixed variables specified in Table 1, the given parameters have been examined and the optimum experimental conditions have been determined.

The purpose of the study, previous studies, and the limitations of the equipment and materials used have been considered when determining the parameter ranges in this experimental study.

Equation 8 has been used to calculate the energy consumption in experimental studies [26].

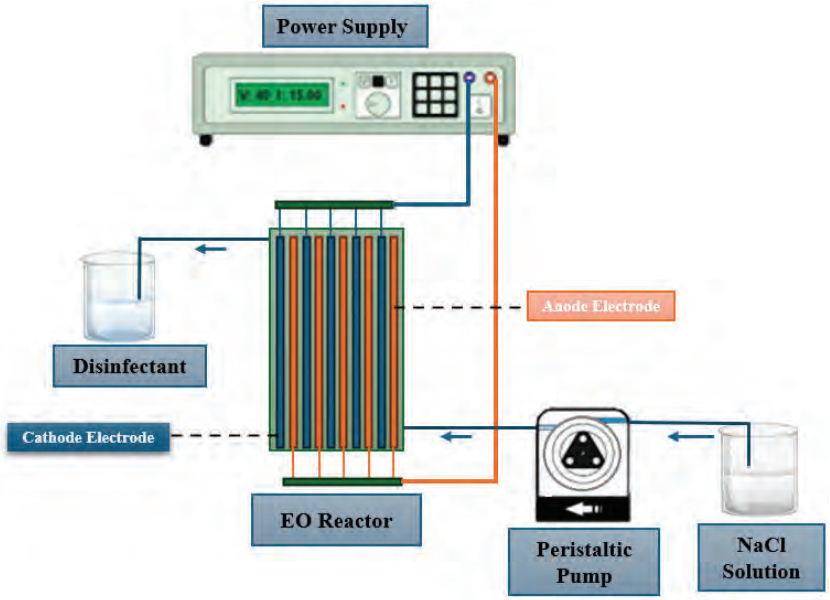


Figure 1. The experimental setup where OSDG with EOP.

Table 1. Parameters affecting the OSDG process in the EO cell

Parameters examined	Parameter range	Variables held constant
Initial pH value	5, 6, 7 and 8	10 A current intensity, 1 M feed water NaCI concentration, 50 ml/min flow rate
Flow rate, (ml/min)	20, 30, 40, 50 and 100	10 A current intensity, 1 M feed water NaCI concentration, Optimum pHi
Current intensity (A)	5, 10, 20, 25, 30, 40 and 50	Optimum pHi is 1 M feed water NaCl concentration
Feed water NaCl concentration, (M)	1, 2, 3, 4, 5 and saturation concentration	Optimum current intensity, Optimum pHi

$$E\ (kW\text{-}h\ /m^3) = (I^*V^*h)/\ v \quad (8)$$

Here, E represents energy consumption. (kW-h /m^3), I constant current intensity (A), V is the applied potential difference value (Volt), h is the retention time (h) depending on the flow rate, and v (m^3) is the electrochemical reactor volume.

The applied potential difference value has been read directly from the power supply. The value read during the experiment has been used in experimental calculations.

RESULTS AND DISCUSSION

The effects of initial pH value, salt concentration, flow rate, and current intensity on the EOP and OSDG process have been examined by considering the parameter ranges given in Table 1. The results obtained are available in the subheadings.

Effect of Initial pH Value on System Parameters

The effects of the initial pH value of the solution in the OSDG process have been examined with EOP. The initial pH value is a critical parameter affecting the chemical reactions and disinfectant formation in the OSDG process [56]. The optimum initial pH value must be determined to provide the desired effectiveness and efficiency in disinfectant generation.

The experiments have been conducted by maintaining a constant current intensity of 10 A and a sodium chloride (NaCl) concentration of 1 M, with the flow rate set at 50 mL/min. During the experiments, the ambient temperature has been measured as 19°C. The initial pH values of the feed solutions have been adjusted to 5, 6, 7, and 8. No external intervention has been made to the pH values throughout the experiments. The variations in pH observed in the outlet water are attributed to the reactions occurring within the EO cell.

The effluent pH values and energy consumption depending on the initial pH value are given in Figure 2.

When Figure 2 is examined, it is seen that the effluent pH values and energy consumption give similar results even though the initial pH values are different. Therefore, it is possible to say that the initial pH value has no effect on

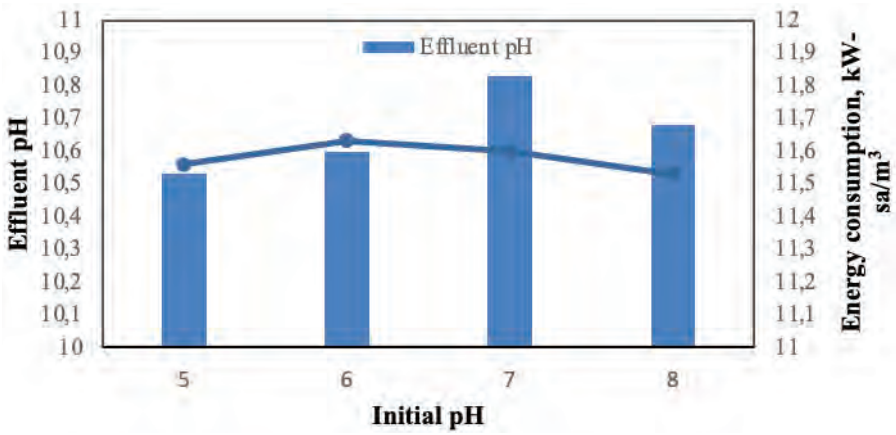


Figure 2. Effect of initial pH value on effluent pH and energy consumption.

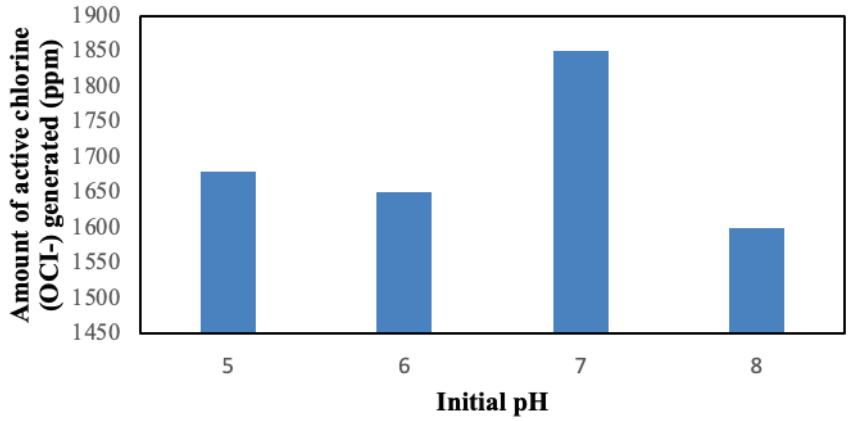


Figure 3. Effect of initial pH value on active chlorine (OCl^-) generation as NaOCl.

energy consumption and effluent pH values. In addition, it is seen that the output pH values are basic. The reason for this is that hypochlorite ions (OCl^-) consume hydrogen ions (H^+) while reacting with water. In addition, the OH^- ions released cause the pH to increase.

The effect of the initial pH value on disinfectant generation (NaOCl) is given in Figure 3.

When Figure 3 is examined, it is seen that when the initial pH value is 7, the generation of active chlorine (OCl^-) is higher than other initial pH values. This reveals the fact that the initial pH values affect disinfectant generation.

It is known that chlorine gas is released in OSDG with EOP. When the pH increases above 6.5, the hydrolysis reaction of chlorine accelerates. Therefore, at conditions above pH 6.5, a large part of the chlorine dissolves in the anode diffusion layer by hydrolysis [57]. In addition, the highest oxidation capacity of chlorine occurs at neutral pH levels [58,59]. This information explains the increased active chlorine efficiency when the initial pH value is 7.

Another important parameter affecting the efficiency of the OSDG carried out by EOP is the HOCI (hypochlorous acid) concentration. Around pH 5, the dominant active chlorine species in the medium is HOCI [60]. As pH increases, HOCI rapidly ionizes and this accelerates the reactions. Therefore, the highest active chlorine generation after pH 7 occurred at pH 5.

When the initial pH value increased above 7, a decrease in active chlorine efficiency has been observed. Similarly, Le et al. [61] reported in their study comparing pH 7 and pH 8 that the active chlorine efficiency at pH 8 has been almost half of the efficiency obtained at pH 7.

Effect of Flow Rate on System Parameters

The effect of the feed solution flow rate on the EOP, particularly in OSDG, is a complex issue influenced by several factors. Generally, the feed flow rate is a critical parameter affecting the efficiency of the EOP, energy consumption, and the quality of the disinfectant. Optimizing the flow rate is essential to ensure effective mass transfer, proper electrode reaction, and minimized energy usage,

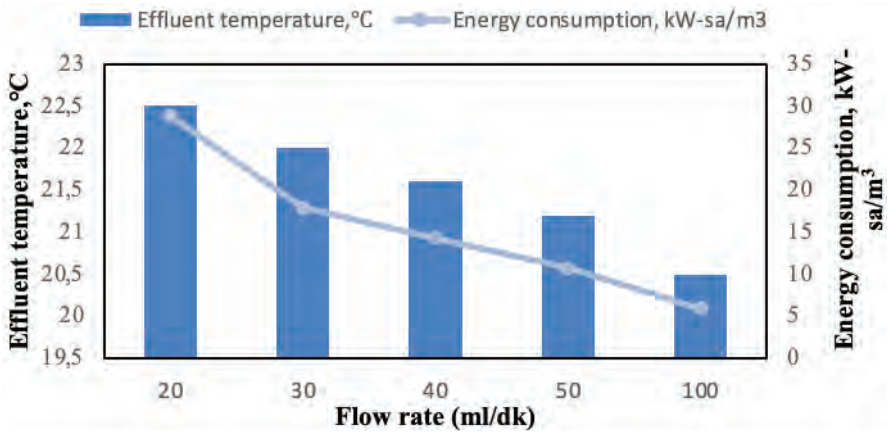


Figure 4. Effluent temperature values and energy consumption depending on flow values.

while also maintaining the desired properties of the end product.

The study has investigated five different flow rates: 20, 30, 40, 50, and 100 ml/min. Since the reactor volume is approximately 1200 ml, the holding times corresponding to each flow rate have been calculated as 60, 40, 30, 24, and 12 min, respectively.

While examining the effects of flow rate on the process, the salt solution concentration has been kept constant at 1 M, the initial pH value is 7, and the current intensity is 10 A. The ambient temperature has been determined at 19 °C. The effluent temperature values and energy consumption depending on the flow rate values are given in Figure 4.

When Figure 4 is examined, it can be observed that the effluent temperature and energy consumption decrease as the flow rate increases. This phenomenon is primarily attributed to the reduction in residence time within the reactor. A shorter residence time results in lower energy consumption, as less electrical energy is converted into thermal energy. Moreover, a reduced residence time decreases the resistance within the system, thereby diminishing the overall heat generation. Consequently, the temperature increase in the solution is less pronounced.

The effect of flow rate changes on disinfectant generation is given in Figure 5.

The increase in flow rate in the EO cell has shortened the residence time in the reactor, making it challenging to achieve reactions with the desired efficiency. In experiments, when the flow rate has been set to 20 ml/min, approximately 2160 mg/L of active chlorine has been produced for a solution with an initial pH of 7 after subjecting a 1 M concentration solution to an electrochemical reaction for 60 minutes. Increasing the flow rate to 100 ml/min under the same conditions has reduced the residence time in the reactor to 12 minutes, resulting in a decrease in active chlorine generation to 1400 mg/L. Similarly, Hsu (2003) has determined that higher flow rates decrease active chlorine concentration in the effluent [62].

The active chlorine concentration measured at 40 ml/min in the study has reached 2080 ppm, which is close to the value at 20 ml/min. To optimize time and energy,

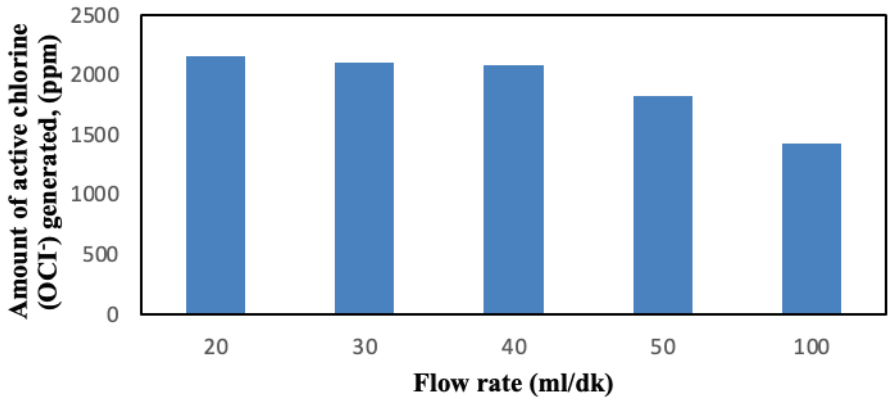


Figure 5. Effect of flow rate changes on active chlorine (OCl^-) generation as NaOCl.

subsequent trials have been conducted at flow rates of 40, 50, and 100 ml/min.

Effect of Current Intensity on System Parameters

In OSDG with EOP, current intensity plays a critical role in initiating and directing redox reactions between electrodes. Thus, it is essential to investigate the effects of current intensity on disinfectant generation.

In this part of the study, the effect of current intensity on system parameters has been investigated. Current intensities have been set at 5, 10, 15, 20, and 25 A, while flow rates have been maintained at 40, 50, and 100 ml/min. The initial pH has been adjusted to 7, and the ambient temperature has been measured at 18℃. Table 2 has shown the effluent temperature and energy consumption values for the different current intensities.

When Table 2 is examined, it can be seen that the temperature values and energy consumption have increased with the rising current intensity. The primary reason for this is that higher current intensities have increased the system resistance. Under a constant electrical conductivity, an increase in current intensity results in the generation of higher potential difference values. The rise in potential difference values, in turn, has led to an increase in energy consumption.

Figure 6 presents the active chlorine (OCl^-) concentrations produced at varying current intensities and flow rates.

When Figure 6 is examined, it can be observed that increasing current intensity has significantly enhanced disinfectant generation at the evaluated flow rates. A review of the literature indicates that efficiency increases as the applied amperage intensity to inlet waters with the same initial pH and flow rate rises[63]. The results obtained at 20 A and 25 A values appear to be quite close to each other. Specifically, when disinfectant generation is analyzed at 40 ml/min for 20 A and 25 A, the concentrations

Table 2. Effluent temperature and energy consumption values according to changing current intensities

Effluent temperature values corresponding to changing current intensities (°C)			
	40 ml/min	**50 ml/min**	**100 ml/min**
5 A	19.8	18.7	19.3
10 A	22.3	21.2	20.5
15 A	23.7	22.2	21.9
20 A	26.2	24.5	22.3
25 A	28.1	24.9	22.9
Energy consumption values for varying current intensities (kW-h/m³)			
	40 ml/min	**50 ml/min**	**100 ml/min**
5 A	6.48	5.27	2.64
10 A	14.29	10.74	5.89
15 A	23.81	19.50	9.75
20 A	34.33	27.27	14.13
25 A	46.25	35.58	18.50

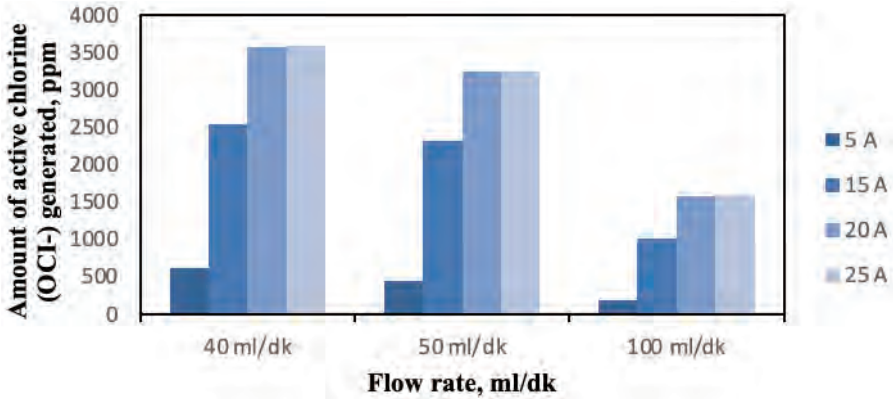


Figure 6. Effect of current intensity and flow rate changes on active chlorine (OCl^-) generation as NaOCl.

are found to be 3576 ppm and 3580 ppm, respectively. At a flow rate of 50 ml/min, the disinfectant generation values at 20 A and 25 A are 3244 ppm and 3240 ppm, respectively. Similarly, at 100 ml/min, the generation values at 20 A and 25 A are recorded as 1580 ppm and 1590 ppm, respectively. Therefore, since the results at 20 A and 25 A with a 1 M salt concentration are similar, it appears more advantageous to select a current intensity of 20 A in terms of energy savings.

Effect of Salt Concentration on System Parameters

In the OSDG with EOP, salt concentration is an important parameter affecting product efficiency and reaction speed. High salt concentrations increase disinfectant generation efficiency by accelerating electrochemical reactions. Optimal salt level reduces by-products and NaCl consumption, improving product quality and reducing costs.

To investigate the effects of salt concentration on disinfectant generation, experiments have been conducted using 1, 2, 3, 4, and 5 M NaCl, with an initial pH value set at 7, flow rates of 40, 50, and 100 ml/min, and a constant current intensity of 20 A. The ambient temperature has been measured in the range of 17.0-17.3 °C. Table 3 presents the effluent temperature and energy consumption values at different salt concentrations.

When Table 3 is examined, the change in effluent water temperature with the increase in salt concentration is observed. Higher salt concentrations have increased the electrical conductivity of the solution while decreasing the electrical resistance in the system, thereby limiting the conversion of electrical energy into thermal energy. As a result, the temperature rise in the electrochemical reactor has been mitigated with increasing salt concentrations. Additionally, the increase in salt concentration has led to a partial decrease in energy consumption. Even at the lowest salt concentration, high electrical conductivity has been achieved, resulting in similar potential difference values within the system. This situation has prevented a significant change in energy consumption.

Table 3. Effluent temperature and energy consumption values at different salt concentrations

Effluent temperature values at different salt concentrations (°C)			
	40 ml/min	**50 ml/min**	**100 ml/min**
1M	26.2	24.5	22.3
2M	25.4	23.6	21.5
3M	22.1	21.5	20.1
4M	21.7	19.4	18.1
5M	21.2	18.8	17.6
Energy consumption values at different salt concentrations (kW- h /m³)			
	40 ml/min	**50 ml/min**	**100 ml/min**
1M	33.25	27.33	13.67
2M	32.92	26.93	13.59
3M	32.75	26.73	13.55
4M	32.33	26.65	13.53
5M	31.87	26.59	13.51

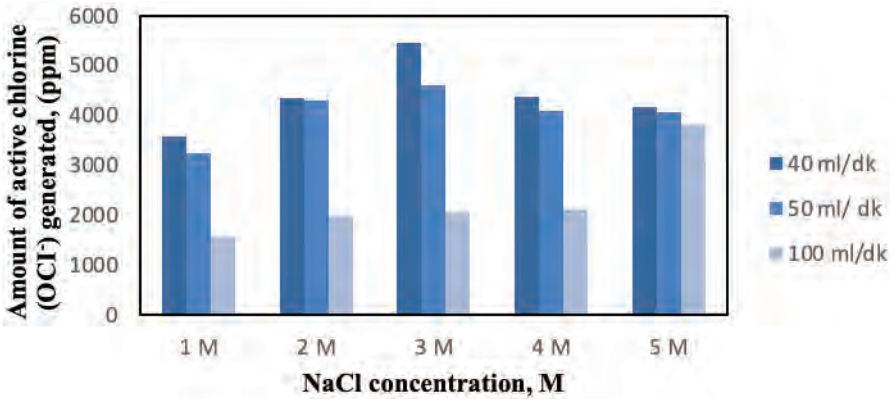


Figure 7. Effect of salt concentration and flow rate variations on active chlorine (OCl^-) generation as NaOCl.

Figure 7 shows the active chlorine (OCl^-) concentrations corresponding to different salt concentrations and flow rates.

When Figure 7 is examined, the increase in salt concentration from 1 M to 3 M has led to an increase in disinfectant concentration at all flow rates. However, the rise in salt concentration from 3 M to 5 M has resulted in a relative decrease in active chlorine formation. The highest active chlorine yield of 5450 ppm has been achieved at a 3 M salt concentration and a flow rate of 40 mL/min. In the experiments conducted with a 5 M solution, the effect of the flow rate diminished, resulting in similar active chlorine concentrations across all three flow rates. This situation has positively impacted the generation of relatively more active chlorine at higher flow rates.

Effect of Saturation Salt Concentration and High Current Intensities on Disinfectant Generation

Examining the effects of salt concentrations reaching saturation levels and high current intensities on disinfectant generation is critically important for determining optimum operating conditions and understanding the relationships between energy consumption and efficiency. In the conducted studies, it has been observed that increasing salt concentration positively affects the generation of relatively more active chlorine at high flow rates. Therefore, this stage of the study has been carried out at a flow rate of 100 ml/min, with an initial pH value of 7 and a saturated salt concentration of 5.8 M NaCl. Current intensities have been set at 30, 35, 40, and 50 A, while the ambient temperature has been measured at 19 °C. Figure 8 presents the effluent temperature values and energy consumption at different current intensities.

The increase in current intensity caused an increase in energy consumption and effluent temperature values.

Figure 9 presents the concentrations of active chlorine (OCl^-) generated at different current intensities.

Figure 9 illustrates that increasing current intensity has led to an increase in active chlorine generation. At a current intensity of 50 A, the active chlorine concentration has reached up to 7750 ppm, which is among the highest values reported in the literature for OSDG with EOP [44,64].

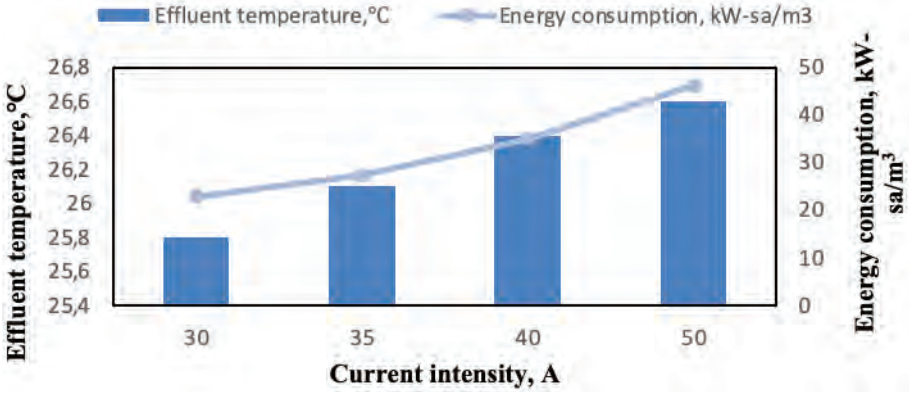


Figure 8. Effluent temperature values and energy consumption at different current intensities.

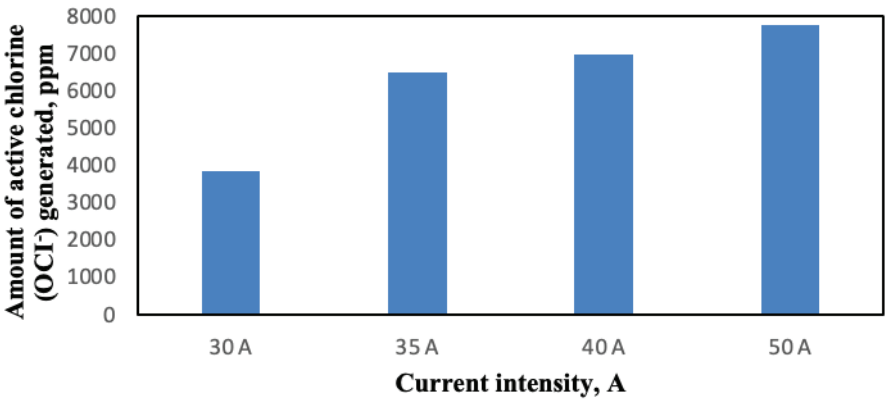


Figure 9. Active chlorine (OCl^-) generation as NaOCl concentrations formed at different current intensities.

CONCLUSION

In this study, on-site disinfectant generation using the electro-oxidation method was investigated. An electro-oxidation reactor was designed with 5 F-type titanium cathodes (80 × 180 mm) and 5 IrO_2-RuO_2 coated F-type titanium anodes (80 × 180 mm). The EO cell was constructed with dimensions of 100 × 100 × 200 mm, and the distance between the electrodes was fixed at 5 mm. The effects of experimental parameters such as pH, flow rate, current intensity, and NaCl concentration on disinfectant production were examined. The active chlorine generated as a disinfectant was in the form of OCl^- (sodium hypochlorite, NaOCl). The conclusions obtained from the study are as follows.

- ✓ The pH values were set between 5-8 and the system was operated without intervention. It was shown that the initial pH values had minimal effect on the effluent pH values. The system tended to a basic environment during the reaction. Active chlorine generation was found to be most efficient at pH 7.
- ✓ In the study, current intensities between 5 and 25 A were examined and it was observed that increasing current intensity increased disinfectant production. However, it was determined that the effect of increasing current intensity up to 20 A on disinfectant production was seen more clearly, but the expected increase was not seen at 25 A. Therefore, the optimum current intensity was determined as 20 A in order to ensure energy efficiency.
- ✓ The flow rates were set between 20 and 100 mL/min and it was observed that the energy consumption decreased as the residence time in the reactor decreased with increasing flow rates. However, the increase in flow rates caused a decrease in the amount of disinfectant produced. Considering both high disinfectant production and low energy consumption, a flow rate of 40 mL/min was determined as optimum.
- ✓ Salt concentration was examined between 1 and 5 M and it was concluded that increasing salt concentration increased disinfectant production. However, it was observed that increasing salt concentration up to 3 M increased disinfectant production visibly, and this increase was not as expected at higher salt concentrations. Therefore, 3 M salt concentration was determined as the optimum salt concentration.
- ✓ Optimum conditions were determined as 3 M salt concentration, 20 A current intensity, pH 7, and 40 mL/min flow rate. Under these conditions, 5450 mg/L disinfectant production was obtained. Under the determined optimum conditions, 32.75 kW-h/m^3 energy consumption occurred.
- ✓ Since it was determined that increasing the current intensity and salt concentration positively affected the disinfectant production, another study was conducted to determine the maximum production capacity of the system. Experiments were conducted at saturated salt concentration, pH 7, 100 mL/min flow rate, and 30, 35, 40, and 50 A current intensities. Since a current intensity above 50 A would negatively affect the system, no current intensity above this intensity was studied. 7750 ppm disinfectant production was achieved at a current intensity of 50 A and this value was recorded as a very high production amount according to the literature.
- ✓ The disinfectant amounts obtained are above the values determined for microbial applications in most areas. Therefore, it is possible to use a reactor of the same size in laboratory and field applications. It is possible to revise and use the process appropriately according to the required disinfectant concentrations and amounts.
- ✓ This process prevents disinfectants from being transported from generation areas to long distances where they will be used. In addition, being in a form that can be used after generation eliminates the need for storage. This eliminates possible risks to human and environmental health during transportation and storage. This enhances the sustainability of the process.

ACKNOWLEDGEMENTS

This study was supported by the Ataturk University Scientific Research Projects Coordination Unit under Project ID 8774, titled "Electrochemical Methods for On-site Disinfectant Generation." (Elektrokimyasal Yöntemlerle Yerinde Dezenfektan Üretiminin İncelenmesi). The data used in this study were obtained from the thesis numbered 849152 (YÖK ID), prepared within the scope of the 100/2000 Doctoral Program supported by the Council of Higher Education (YÖK).

ABBREVIATIONS

EO	Electrooxidation
EOP	Electrooxidation process
OSDG	On-site disinfectant generation
NaOCI	Sodium hypochlorite
OCl^-	Hypochlorite

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Bloomfield SF. Importance of disinfection as a means of prevention in our changing world hygiene and the home. GMS Krankenhhyg Interdiszip 2007;13:2.

[2] Asif M, Mahmood L, Mukwana K, Hussain S, Bashir I, Nasar M, Selele M, Abdulrahim M, Ahmed Z, Hasnain M, Abubakar A. Groundwater implications of water pollution in Pakistan: A review on sustainable environmental management of solid waste. J Pollut Monit Eval Stud Control 2024;3:47–58.

[3] Hand S, Cusick RD. Electrochemical disinfection in water and wastewater treatment: Identifying impacts of water quality and operating conditions on performance. Environ Sci Technol 2021;55:3470–3482. [CrossRef]

[4] Gallandat K, Kolus RC, Julian TR, Lantagne DS. A systematic review of chlorine-based surface disinfection efficacy to inform recommendations for low-resource outbreak settings. Am J Infect Control 2021;49:90–103. [CrossRef]

[5] Alarcon P, Marco-Jimenez F, Horigan V, Ortiz-Pelaez A, Rajanayagam B, Dryden A, Simmons H, Konold T, Marco C, Charnley J, et al. A review of cleaning and disinfection guidelines and recommendations following an outbreak of classical scrapie. Prev Vet Med 2021;193:105388. [CrossRef]

[6] Clayton GE, Thorn RMS, Reynolds DM. The efficacy of chlorine-based disinfectants against planktonic and biofilm bacteria for decentralised point-of-use drinking water. npj Clean Water 2021;4:48. [CrossRef]

[7] Kye-Hwan B, Sang HH, Jang-won Y, Si HP, Sang-Do H. Efficacy of chlorine-based disinfectants (sodium hypochlorite and chlorine dioxide) on Salmonella Enteritidis planktonic cells, biofilms on food contact surfaces and chicken skin. Food Control 2021;123:107838. [CrossRef]

[8] Akgiray Ö. Use of chlorine in the treatment of water and wastewater. Plumbing J 2003.

[9] Son H, Cho M, Kim J, Oh B, Chung H, Yoon J. Enhanced disinfection efficiency of mechanically mixed oxidants with free chlorine. Water Res 2005;39:721–727. [CrossRef]

[10] Jeong J, Kim JY, Cho M, Choi W, Yoon J. Inactivation of Escherichia coli in the electrochemical disinfection process using a Pt anode. Chemosphere 2007;67:652–659. [CrossRef]

[11] Kim SK, Shin DM, Rhim JW. Designing a high-efficiency hypochlorite ion generation system by combining cation exchange membrane aided electrolysis with chlorine gas recovery stream. J Membr Sci 2021;630:119318. [CrossRef]

[12] Lantagne DS. Sodium hypochlorite dosage for household and emergency water treatment. Am Water Works Assoc 2008;100:106–119. [CrossRef]

[13] Wilheim N, Kaufmann A, Blanton E, Lantagne D. Sodium hypochlorite dosage for household and emergency water treatment: Updated recommendations. J Water Health 2018;16:112–125. [CrossRef]

[14] Bernardi AO, Stefanello A, Garcia MV, Parussolo G, Stefanello RF, Moro CB, Copetti MV. Efficacy of commercial sanitizers against fungi of concern in the food industry. LWT Food Sci Technol 2018;97:25–30. [CrossRef]

[15] Visconti V, Coton E, Rigalma K, Dantigny P. Effects of disinfectants on inactivation of mold spores relevant to the food industry: A review. Fungal Biol Rev 2021;38:44–46. [CrossRef]

[16] Ma X, Bibby K. Free chlorine and monochloramine inactivation kinetics of Aspergillus and Penicillium in drinking water. Water Res 2017;120:265–271. [CrossRef]

[17] Lin H, Zhu X, Wang Y, Yu X. Effect of sodium hypochlorite on typical biofilms formed in drinking water distribution systems. J Water Health 2017;15:218–227. [CrossRef]

[18] Bhasin A, Jain M, Mishra B, Loomba P, Sharma A, Aggrawal T. Effect of sodium hypochlorite on biofilm-producing organisms isolated from a hospital drinking water. J Microbiol Infect Dis 2013;13:17–24. [CrossRef]

[19] Urben PG. Bretherick's Handbook of Reactive Chemical Hazards. 7th ed. Academic Press; 2007.

[20] Kirihara M, Okada T, Sugiyama Y, Akiyoshi M, Matsunaga T, Kimura Y. Sodium hypochlorite pentahydrate crystal: A convenient and environmentally benign oxidant for organic synthesis. Org Process Res Dev 2017;21:1925–1937. [CrossRef]

[21] Hobro S, Mourad AK. Developing chlorine-based antiseptic by electrolysis. Sci Total Environ 2020;709:136108.

[22] Soman AR, Sundararaj G. Accidental release of chlorine from a storage facility and an on-site emergency mock drill: A case study. Sci World J 2015;483216. [CrossRef]

[23] Parra DMC, Mahecha FC, Pinzon AFR, Bustos DR, Llorente LAT, Jimenez MFJ. A prototype for on-site generation of chlorinated disinfectant for use in rural aqueducts. Water Sci Eng 2023;1–8.

[24] GilPavas E, Dobrosz-Gómez I, Gómez-García MA. Optimization of sequential chemical coagulation – electro-oxidation process for the treatment of an industrial textile wastewater. J Water Process Eng 2018;22:73–79. [CrossRef]

[25] Bampos G, Petala A, Frontistis Z. Recent trends in pharmaceuticals removal from water using electrochemical oxidation processes. Environments 2021;8:85. [CrossRef]
[26] Yılmaz AE, Dede Sağsöz Y, Cengiz İ, Sakarya M. The investigation of flow rate effect on leachate treatment by electrooxidation process. Sigma J Eng Nat Sci 2020;11:73–82.
[27] Bhatt P, Engel BA, Shivaram KB, Turco RF, Zhou Z, Simsek H. Treatment and optimization of high-strength egg-wash wastewater effluent using electrocoagulation and electrooxidation methods. Chemosphere 2024;347:140632. [CrossRef]
[28] Montañés MT, García-Gabaldón M, Giner-Sanz JJ, Mora-Gómez J, Pérez-Herranz V. Effect of the anode material, applied current and reactor configuration on the atenolol toxicity during an electrooxidation process. Heliyon 2024;10:e27266. [CrossRef]
[29] Zhuang S, Duan N, Xu F. Synergistic strategy of solute environment and phase control of Pb-based anodes to solve the activity-stability trade-off. J Hazard Mater 2024;470:134119. [CrossRef]
[30] Bakaraki Turan N, Sari Erkan H, Ilhan F, Onkal Engin G. Decolorization of textile wastewater by electrooxidation process using different anode materials: Statistical optimization. Water Environ Res 2022;94:e1683. [CrossRef]
[31] Köktaş İY, Gökkuş Ö, Kariper IA, Othmani A. Tetracycline removal from aqueous solution by electrooxidation using ruthenium-coated graphite anode. Chemosphere 2023;315:137758. [CrossRef]
[32] Li W, Xiong M, Liang X, Wang D, Zhu H, Pan Y. An electrochemical way to generate amphiphiles from hydrazones for the synthesis of 1,2,4-triazole scaffold cyclic compounds. ChemistryOpen 2022;11: e202100268. [CrossRef]
[33] Ukundiman Z, Omwene PI, Gengec E, Can OT, Kobya M. Electrooxidation as post-treatment of ultrafiltration effluent in a landfill leachate MBR treatment plant: Effects of BDD, Pt and DSA anode types. Electrochim Acta 2018;286:252–263. [CrossRef]
[34] Zhang K, Wang R, Wang H, Li M, Zhao P, Wang Y, et al Electrooxidation of chlorophene and dichlorophen by reactive electrochemical membrane: Key determining factors of removal efficiency. Environ Res 2024;241:117612. [CrossRef]
[35] Carneiro JF, Aquino JM, Silva AJ, Barreiro JC, Cass QB, Rocha-Filho RC. The effect of the supporting electrolyte on the electrooxidation of enrofloxacin using a flow cell with a BDD anode: Kinetics and follow-up of oxidation intermediates and antimicrobial activity. Chemosphere 2018;206:674–681. [CrossRef]
[36] Bakaraki Turan N, Sari Erkan H, İlhan F, Onkal Engin G. Highlighting the cathodic contribution of an electrooxidation post-treatment study on decolorization of textile wastewater effluent pre-treated with a lab-scale moving bed-membrane bioreactor. Environ Sci Pollut Res Int 2021;28:25972–25983. [CrossRef]
[37] Mousazadeh M, Khademi N, Kabdaşlı I, Rezaei S, Hajalifard Z, Moosakhani Z, et al. Domestic greywater treatment using electrocoagulation-electrooxidation process: Optimisation and experimental approaches. Sci Rep 2023;13:15852. [CrossRef]
[38] Najafinejad MS, Chianese S, Fenti A, Iovino P, Musmarra D. Application of electrochemical oxidation for water and wastewater treatment: An overview. Molecules 2023;28:4208. [CrossRef]
[39] Anglada A, Urtiaga A, Ortiz I. Contributions of electrochemical oxidation to wastewater treatment: Fundamentals and review of applications. J Chem Technol Biotechnol 2009;84:1747–1755. [CrossRef]
[40] Cornejo OM, Murrieta MF, Aguilar ZG, Rodríguez ZF, Márquez AA, León MI, et al. Recent advances in electrochemical flow reactors used in advanced oxidation processes: A critical review. Chem Eng J 2024;496:153935. [CrossRef]
[41] Hricova D, Stephan R, Zweifel C. Electrolyzed water and its application in the food industry. J Food Prot 2008;71:1934–1947. [CrossRef]
[42] Ampiaw RE, Yaqub M, Lee W. Electrolyzed water as a disinfectant: A systematic review of factors affecting the production and efficiency of hypochlorous acid. J Water Process Eng 2021;43:102228. [CrossRef]
[43] Özel Çelik E, Özel Akdemir Ü, Çelik H. Liquid, gas chlorine and on-site generation in drinking water facilities: Design consideration and comparison of operating costs. J Çukurova Univ Fac Eng Arch 2017;32:81–89.
[44] Saleem M, Chakrabarti MH, Hasan DB, Islam MS, Yussof R, Hajimoloma SA, et al. On-site electrochemical production of sodium hypochlorite disinfectant for a power plant utilizing seawater. Int J Electrochem Sci 2012;7:3929–3938. [CrossRef]
[45] Abdul-Wahab SA, Al-Weshahi MA. Brine management: Substituting chlorine with on-site produced sodium hypochlorite for environmentally improved desalination processes. Water Resour Manag 2009;23:2437–2454. [CrossRef]
[46] Ghernaout D, Ghernaout B. From chemical disinfection to electrodisinfection: The obligatory itinerary?. Desalination Water Treat 2010;16:156–175. [CrossRef]
[47] Isa MH, Kutty SRM, Yusoff HAM, Bashir MJK, Farooqi IH. Electrochemical production of free available chlorine. In: International Conference on Emerging Technologies in Environmental Science and Engineering; October 26–28, Aligarh Muslim University, Aligarh, India. 2009.
[48] Rajab M, Heim C, Letzel T, Drewes JE, Helmreich B. Electrochemical disinfection using boron-doped diamond electrode: The synergetic effects of in situ ozone and free chlorine generation. Chemosphere 2015;121:47–53. [CrossRef]

[49] Mirshafiee A, Nourollahi M, Shahriary A. Application of electrooxidation process for treating wastewater from petrochemical with mixed metal oxide electrode. Sci Rep 2024;14:1–14. [CrossRef]
[50] Boal AK. The Applications and Scientific Mechanisms of Disinfectants Generated On-Site Using Salt, Water and Electricity. The National Environmental Services Centre; 2018.
[51] Shinde SR, Apte S. Correlation analysis of the experimental data for in-situ chlorine generation in electrochlorination. J Environ Eng 2023;149:4. [CrossRef]
[52] Shmychkova O, Girenko D, Velichenko. Cl^-/ClO^- process on SnO_2-based electrodes in low-concentrated NaCl solutions. Electrochem Sci Adv 2021;2:1–15. [CrossRef]
[53] Petersen MA, Sale TC, Reardon KF. Electrolytic trichloroethene degradation using mixed metal oxide coated titanium mesh electrodes. Chemosphere 2007;67:1573–1581. [CrossRef]
[54] Fil BA, Günaslan S. Comparison of different anodes and investigation of energy consumption in the treatment of real wastewater by electrooxidation. Int J Environ Res 2022;16:79. [CrossRef]
[55] Özyonar F, Korkmaz MU. Sequential use of the electrocoagulation-electrooxidation processes for domestic wastewater treatment. Chemosphere 2022;290:133172. [CrossRef]
[56] Yan P, Daliri EB, Oh DH. New clinical applications of electrolyzed water: A review. Microorganisms 2021;9:136. [CrossRef]
[57] Spasojević M, Krstajić N, Spasojević P, Ribić-Zelenović L. Modelling current efficiency in an electrochemical hypochlorite reactor. Chem Eng Res Des 2014;93:591–601. [CrossRef]
[58] Moura DC, Araújo CKC, Zanta CLPS, Salazar R, Martínez-Huitle CA. Active chlorine species electrogenerated on $Ti/Ru_{0.3}Ti_{0.7}O_2$ surface: Electrochemical behavior, concentration determination and their application. J Electroanal Chem 2014;731:145–152. [CrossRef]
[59] Panizza M, Cerisola G. Direct and mediated anodic oxidation of organic pollutants. Chem Rev 2009;109:6541–6569. [CrossRef]
[60] Forghani F, Park JH, Oh DH. Effect of water hardness on the production and microbicidal efficacy of slightly acidic electrolyzed water. Food Microbiol 2015;48:28–34. [CrossRef]
[61] Le GT, Ta NT, Pham TQ, Dao YH. Response surface analysis of fenobucarb removal by electrochemically generated chlorine. Water 2019;11:899. [CrossRef]
[62] Hsu SY. Effects of water flow rate, salt concentration and water temperature on efficiency of an electrolyzed oxidizing water generator. J Food Eng 2003;60:469–473. [CrossRef]
[63] Zaviska F, Drogui P, Pablo G. Statistical optimization of active chlorine production from a synthetic saline effluent by electrolysis. Desalination 2012;296:16–23. [CrossRef]
[64] Sánchez Aldana D, Corral NO, Rocha Gutiérrez AB, Ballinas Casarrubias L, Pérez-Domínguez EJ, Nevárez Moorillon GV, Soto Salcido LA, Ortega-Hernández S, Cardenas-Félix G, González Sánchez G. Hypochlorite generation from a water softener spent brine. Water 2018;10:1733. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 781–793, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2010

Research Article

The effect of different light intensities on the growth of *Chlorella Vulgaris* in advanced biologically treated poultry slaughterhouse wastewater

Gamze KATIRCIOĞLU SINMAZ[1,*], İsmail Ayhan ŞENGİL[1]

[1]*Department of Environmental Engineering, Sakarya University, Sakarya, 54050, Türkiye*

ARTICLE INFO

Article history
Received: 17 October 2024
Revised: 02 January 2025
Accepted: 13 February 2025

Keywords:
Chlorella Vulgaris; FAME; Light Intensity; Lipid; Wastewater

ABSTRACT

Determining the optimal light intensity is essential for the growth of *Chlorella vulgaris* and the production of suitable biomass for biodiesel. This study aims to investigate the cultivation of *Chlorella vulgaris* under optimal light intensity to enhance biodiesel production while advancing the biological treatment of poultry slaughterhouse wastewater. To fill the relevant gap in the literature, the suitability of the biologically treated wastewater for microalgae growth was investigated at 20, 60, 100, and 140 µmolphoton/m²s light intensities with alkaline pH adjustment and pH shocking methods. Optimal phosphate and ammonium removal (up to 100%) was achieved using the alkaline pH method at 100 and 140 µmolphoton/m²s. In terms of biodiesel production, the most suitable result was obtained with the pH shocking method and the lipid ratio was calculated as 20% in the dry biomass of the C5.2 samples at 100 µ molphoton/m²s.The highest chlorophyll-a concentration (3.32 mg/L) was measured in the same sample at 140 µmol photons/m²s. Fatty acid methyl ester species obtained predominantly C16 and C18 fatty acids at all light intensities. According to the results, *Chlorella vulgaris* grown at 100 µmolphoton/m²s is optimal for biodiesel production due to its high lipid content and fatty acid methyl ester types. Kinetic analyses showed that phosphate and ammonium removal predictability with *Chlorella vulgaris* followed first-order kinetics with R^2 values of 96% and 85%, respectively. Under optimal light intensity, integrating *Chlorella vulgaris* cultivation into a biological slaughterhouse wastewater treatment system enables advanced treatment with real-world applicability.

Cite this article as: Katırcıoğlu Sınmaz G, Şengil İA. The effect of different light intensities on the growth of *Chlorella Vulgaris* in advanced biologically treated poultry slaughterhouse wastewater. Sigma J Eng Nat Sci 2026;44(2):781–793.

INTRODUCTION

The continued rise in CO_2 levels in the atmosphere is disquieting. The most crucial advantage of biodiesel obtained from oilseed plants and microalgae is that the carbon dioxide in the atmosphere is used via photosynthesis. According to the rapid growth rate, high lipid content and long-chain fatty acid methyl ester profile of microalgae, they are considered as a suitable source for biodiesel

***Corresponding author.**
*E-mail address: gkatircioglu@sakarya.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

production [1]. The fatty acid methyl ester composition of microalgae determines the quality of the produced biodiesel [2]. For this reason, studies on biodiesel production from microalgae have gained speed in recent years.

Microalgal growth and lipid production are directly affected by abiotic conditions such as essential nutrients, light, temperature, CO_2, pH, and growth phase [3]. Light intensity is remarkably significant for developing microalgae culture in the presence of sufficient nutrients. Since microalgae photosynthesize and grow by carbon fixation in the air with the presence of light, more biomass is expected to be formed as the light intensity increases [4]. Nonetheless, excessive light intensity inhibits microalgae growth [5]. While the microalgae biomass increased with the increase in light intensity, lipid production mainly increased depending on the microalgae species and other environmental conditions [6,7] and decreased in some studies [8,9] and unaffected by light intensity [10]. While developing microalgae culture, using wastewater as a nutrient provides both-sided benefits by providing advanced treatment of wastewater compared to traditional treatment methods and obtaining biodiesel from the developed microalgae biomass [11]. Microalgae take the nutrients necessary for their growth from wastewater in NH_4^+ and PO_4^{3-} forms [12]. As the light intensity increases, the removal of nitrogenous and phosphorous compounds in wastewater increases [13]. While ammonium can be removed from wastewater with high efficiency at low light intensity (14 µmolphoton / m^2s), it is also possible to remove phosphate with lower efficiency [14]. Light is one of the most important parameters that *Chlorella vulgaris* needs during photosynthesis to develop. However, light's wavelength and light intensity are essential for the lipid content and fatty acid methyl ester (FAME) composition of the microalgae biomass [3].

While some microorganisms and microalgae can live in a consortium, there is competition with some organisms [12]. When biologically treated wastewater is used as a nutrient medium, eliminating wastewater from harmful microorganisms is very important for developing microalgae. For *Chlorella ellipsoidea* cultures grown using secondary treatment domestic wastewater effluent, wastewater was filtered and sterilized before the experiment [15]. In another study where *Chlorella sp.* was grown using domestic wastewater treatment plant effluent, filtration and UV radiation methods were used to solve the rotifer and bacteria problem [16]. In a study in which *Nannochlropsis salina* was grown in an artificial nutrient medium supplemented with rotifers, the hydrodynamic cavitation method was used to remove rotifers, and a high rate of rotifer removal efficiency was obtained [17]. In addition to these methods, lower cost and large scale applicable methods should be developed. In a previous study, pH shocking method was used to remove harmful organisms from biologically treated slaughterhouse wastewater. The pH shocking time and period, which give the most suitable result for microalgae development, have been determined [18]. In a recent study, the pH of the culture medium was adjusted to alkaline in order to destroy the microorganisms in biologically treated poultry slaughterhouse wastewater that inhibit *Chlorella vulgaris* growth. *Chlorella vulgaris* culture was developed successfully at pH 10.5, which is unsuitable for the survival of rotifers [19].

Studies on utilizing poultry slaughterhouse wastewater for microalgae cultivation have gained attention. These wastewaters have undergone various pretreatment processes and have been used to grow different microalgae species. In studies on microalgae cultivation, poultry slaughterhouse wastewaters have been used in various forms: untreated [20], autoclaved [21], filtered and autoclaved [22], or pretreated with acid precipitation [23]. Research gaps that need to be addressed include developing sustainable and cost-effective methods to eliminate harmful microorganisms in biologically treated slaughterhouse wastewater and investigating the suitable light intensity for growing *Chlorella vulgaris*. Therewithal, there is a gap in the literature regarding the effect of light intensity on the cultivation of *Chlorella vulgaris* with biologically treated poultry slaughterhouse wastewater

This study aims to determine the optimal light intensity for developing *Chlorella vulgaris* culture in biologically treated poultry slaughterhouse wastewater. At the same time, it was aimed to grow *Chlorella vulgaris* culture suitable for biodiesel production in biologically treated poultry slaughterhouse wastewater. In all light intensity studies, it was planned to investigate the sufficient nutrient supply of slaughterhouse wastewater for microalgal development with *Chlorella vulgaris* cultures grown in BG-11 nutrient medium as a control group. In order to monitor microalgae growth, total time dependent biomass and chlorophyll-a determinations were made. At the same time, the amounts of NH_4^+ and PO_4^{3-} were measured over time in order to follow the advanced treatment of slaughterhouse wastewater. Kinetic model calculations were made for microalgae growth and advanced treatment of slaughterhouse wastewater.

MATERIALS AND METHODS

Culture Condition

Within the scope of the study, it was aimed to examine the effect of *Chlorella vulgaris* on the advanced treatment of biologically treated poultry slaughterhouse wastewater and at the same time to compare the pH shocking and alkali culture (fixed pH at 10.5) methods used in the removal of harmful microorganisms in the slaughterhouse wastewater. *Chlorella vulgaris* is grown in the laboratories of the Environmental Engineering Department of Sakarya University and fed regularly with sterilized BG-11 nutrient medium. BG-11 is extensively used as a nutrient medium for the growth of blue-green algae species. In 1 liter of BG-11 nutrient medium contains 1.5 g $NaNO_3$, 0.075 g $MgSO_4.7H_2O$, 0.04 g K_2HPO_4, 0.036 g $CaCl_2.2H_2O$, 0.02

Table 1. Biologically treated poultry slaughterhouse wastewater

Parameter	Value
COD (mg/L)	98
Oil & Gres (mg/L)	20
NH_4^+-N (mg/L)	96.4
NO_3^--N (mg/L)	20.5
PO_4^{3-}-P (mg/L)	1.82
pH	6.98

g Na_2CO_3, 0.006 g citric acid, 0.006 g ferric ammonium citrate, 0.001 g EDTA (disodium magnesium salt), and 1 ml trace element mix [24]. *Chlorella vulgaris* culture at a rate of 10% by volume has been inoculated into the samples. The slaughterhouse wastewater was collected as 24 hours composite after the biologically treatment of poultry slaughterhouse wastewater at Sakarya – Turkiye (Table 1). A 24-hour composite biologically treated poultry slaughterhouse wastewater sample was obtained from the biological treatment plant outlet of a private chicken meat producer. Microalgae growth experiments were started within 2 hours after the biologically treated poultry wastewater sample was taken. Experiments were carried out with at least three replicates and average values are given for all results.

With the intention of developing *Chlorella vulgaris* in biologically treated slaughterhouse wastewater, the cultures were growth at room temperature with continuous white light illumination of 4 different light intensities, 20 - 60 - 100 - 140 μmol photons $m^{-2}s^{-1}$ in 1 L erlenmeyer flasks for 12 days. Cultivation was carried out continuously on a shaker at 150 rpm. For each light intensity, four different samples were prepared. These samples were named as BG11, BG11pH, C5.2 and pH10.5. Two of the samples contain BG11 as a nutrient medium, while two contain biologically treated slaughterhouse wastewater. BG11 nutrient medium provides all the nutrients required for the growth of *Chlorella vulgaris*. In order to compare the effect of biologically treated poultry slaughterhouse wastewater on the growth of *Chlorella vulgaris*, samples containing BG11 nutrient medium were also studied. Some rotifers and bacteria harm microalgal growth in the biologically treated poultry slaughterhouse wastewater that used in the study. Rotifers, in particular, consume microalgae, and therefore, algae death occurs in a short time. Two different methods were applied to inhibit predators without damaging the microalgae cells; pH stabilization at alkaline pH (pH10.5) and pH shocking. pH methods were preferred in order to cause minimum damage to microalgae life (not to add extra chemicals) and to present a different method as a contribution to the literature.

- BG11 is a control sample and its pH has not been interfered with.
- The pH of the C5.2 sample was adjusted to pH 2 for 5 minutes every 2 days to eliminate harmful microorganisms in the biologically treated slaughterhouse wastewater, adjusted to pH 8 at the end of the 5th minute. This process is called pH shocking.
- BG11pH sample was prepared to examine the effect of pH shocking on microalgae grown in nutrient medium containing BG-11. The pH shock process was carried out under the same conditions as the C5.2 sample.
- pH10.5 sample was prepared to determine the growth of microalgae in the Biologically treated slaughterhouse wastewater in alkaline condition (pH10.5) and to compare with pH shocking conditions.

Predatory microorganisms at low and high pH levels cannot survive in biologically treated water. However, to not adversely affect the microalgae's life, the pH change is carried out within 5 minutes with this method called pH shocking [18]. Also *Chlorella* vulgaris microalgae can live in alkaline conditions up to pH 11 [19].

Analytical Methods

Microalgae culture has been developed in the biologically treated water of poultry slaughterhouse and BG-11 nutrient medium. In order to determine the nutrient removal in samples, ammonium ($N\text{-}NH_4^+$), phosphate ($P\text{-}PO_4^{3-}$) analyses have been conducted at beginning and every four days [25]. Before the nutrient analysis, the samples were centrifuged at 14000 rpm for 15 minutes at 4°C in order to separate the grown biomass from the samples. Microalgae growth has been monitored by total suspended solids (mg/L) and, chlorophyll-a at beginning and every four days [25,26]. NH_4 and PO_4^{3-} test kits were purchased from MERCK. Ammonium ($N\text{–}NH_4$) analogous to the method: APHA 4500-NH_3 F), phosphate ($P\text{–}PO_4$) analogous to method APHA 4500-P C analyses have been conducted with spectrophotometric methods (APHA 2017). The MERCK brand Pharo 3600 spectrophotometer has been used for the analyses.

Lipid Extraction and Transesterification

Developed biomass has been harvested by sedimentation. The harvested biomass has been dried entirely at 40°C in the vacuum oven. In order to get efficient results from lipid extraction, the cell wall has been broken down in an autoclave with 50 ml of distilled water containing 0.5% nitric acid at 121°C for 30 min following the method by [27]. The total lipid of dried microalgae biomass has been extracted with a mixture of chloroform and methanol (2:1 v/v) for 3h. The solvent was centrifuged at 3000 rpm for 5 min and dried in a vacuum oven at 40°C. The weight of the extracted total lipid was verified gravimetrically [28,29].

The transesterification method that used was modified from [29]. Acidic esterification method was chosen for transesterification and was carried out in a water bath with sulfuric acid and methanol. The fatty acid methyl esters formed as a result of esterification were transferred to the

n-heptane phase for injection to the gas chromatograph. The fatty acid profile was determined on a gas chromatograph (Shimadzu - QP 2010 GC/MS) equipped with a GC/MS detector and a highly polar Rt2560 capillary column (100 m x 0.25 mm x 0.20 µm) [30].

Kinetic Model

Kinetic model calculations were made according to reaction time and reaction rate constant. The regression coefficient values were calculated according to the kinetic model calculations. For the growth of microalgae, first-order and second-order kinetic models were examined in the kinetic calculations for biomass and chlorophyll-a increases. The k and R^2 values were calculated for the investigated kinetic models. The kinetic equations (Eq.1) and (Eq. 2) are as given below [31]:

$$1^{st} \text{ order kinetics: } \ln\frac{C_0}{C} = k.t \quad (1)$$

$$2^{nd} \text{ order kinetics: } \frac{1}{C} - \frac{1}{C_0} = k.t \quad (2)$$

Co: the initial concentration of biomass

C: concentration of biomass in t time

k: reaction rate constants.

First-order and second-order kinetic models were also examined in the kinetic calculations for removing PO_4^{3-} and NH_4^+ from the biologically treated poultry slaughterhouse wastewater by microalgae.

RESULTS AND DISCUSSION

Nutrient Removal from Wastewater Depending on Light Intensity

Chlorella vulgaris cultures have been developed in BG11, BG11pH, C5.2, and pH10.5 samples for 12 days at 20, 60, 100, and 140 µmolphoton/m^2s light intensities. Since the initial ammonium and phosphate amounts of the samples containing BG11 and the biologically treated poultry slaughterhouse wastewater samples were different, the results were examined in terms of the removal rates calculated with the values measured at the end of the 12th day.

The lowest NH_4^+ removal was obtained at 20 µmolphoton/m^2s compared to other light intensities. However, 77.95% NH_4^+ removal was achieved in the biologically treated poultry slaughterhouse wastewater sample, C5.2 at 20 µmolphoton/m^2s (Figure 1). For all light intensities, the NH_4^+ removal at the end of the 12th day was almost 100% for pH 10.5 sample. In the samples containing BG11, NH_4^+ removal increased as the light intensity increased. In the pH-shocked BG11pH sample, more NH_4^+ removal was obtained than the control sample BG11.

PO_4^{3-} removal increased with the increase of light intensity up to 100 µmol photon/m^2s for C5.2 and pH 10.5 samples. Since the removal amounts were relatively high, it tended to be fixed at 100 and 140 µmol photon/m^2s. Another reason for this fixation trend is that microalgae growth is similar and at high concentrations at 100 and 140 µmol photon/m^2s. 100% PO_4^{3-} removal was achieved for C5.2 and pH 10.5 samples at 100 and 140 µmol photon/m^2s (Figure 2).

As the light intensity increased, nutrient removals increased, and more than 70 percent nutrient removal was obtained even at the lowest light intensity in samples containing wastewater. For this reason, in addition to the purpose of cultivating *Chlorella vulgaris* culture in the biologically treated poultry slaughterhouse wastewater for biodiesel production, high efficiency was obtained at low light intensities in order to increase nutrient removal. In a study performed at 14 µmol/m^2s at low light intensity, 93.5% of ammonium and 11.1% of phosphate were removed with anaerobic wastewater treatment secondary effluent after 8 days of development of *Chlorella vulgaris* [14]. In another study, 79.63% nitrate and 100% phosphate were removed

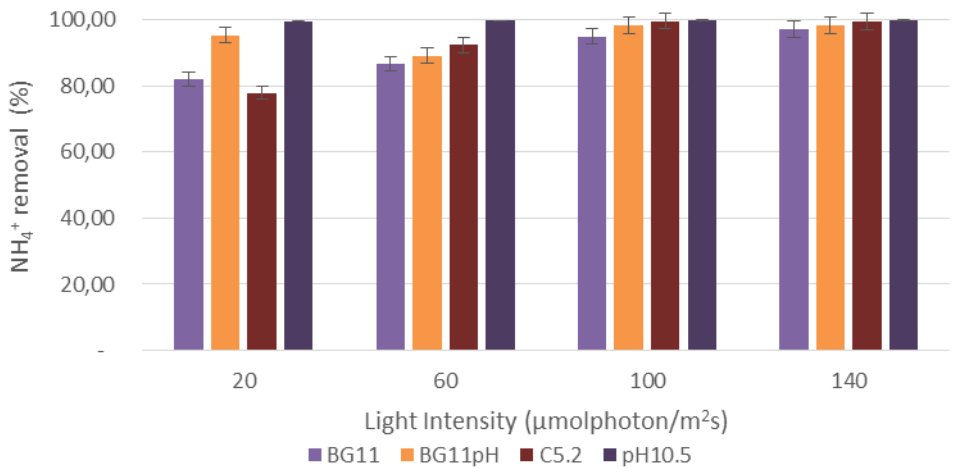


Figure 1. NH_4^+ removal at 12th day depending on light intensity.

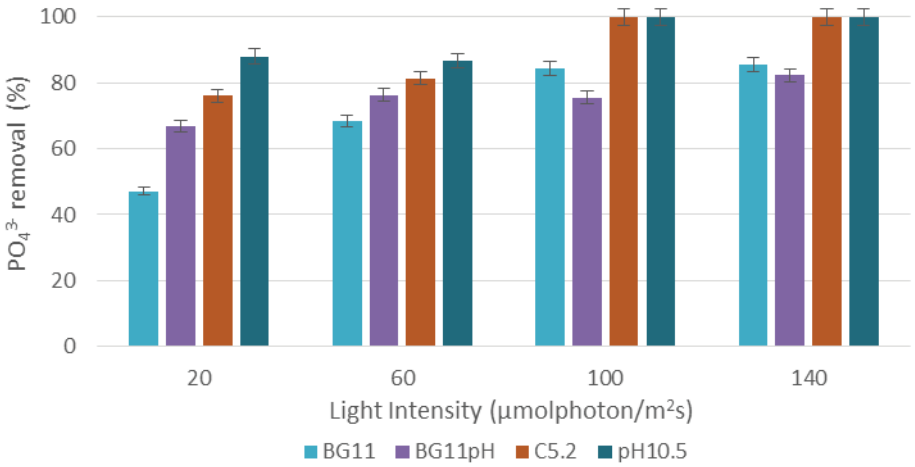


Figure 2. PO_4^{3-} removal at 12th day depending on light intensity.

from the domestic wastewater treated with *Chlorella vulgaris* after a 14-day culture period at a light intensity of 3500 lux (≈ 50 µmolphoton/m²s) [32]. In a study investigating the effect of light intensity in wastewater with different types of microalgae, it was found that as light intensity increased, total nitrogen (TN) removal for *Chlorella vulgaris* decreased, while total phosphorus (TP) removal increased. At the end of 8 days, it was calculated that total nitrogen and total phosphorus were removed by 75% with all microalgae species [7]. In a study conducted by Solmaz and Işık, nutrient removal results were obtained as 50.08% for NH_4-N and 4.54% for PO_4-P from synthetic wastewater with a culture of a mixture of different microalgae species [33].

In a study, total nitrogen and total phosphorus removal in pretreated slaughterhouse wastewater with mixed microalgae culture using a photobioreactor were obtained as 70.2% and 96.2%, respectively, after 7 days [22]. In another study, nitrate and phosphate removal efficiencies were obtained as 18% and 48%, respectively, with *Chlorella pyrenoidosa* culture in autoclaved and diluted poultry slaughterhouse wastewater [21]. When the studies in the literature are examined; Although nutrient removal rates are related to microalgae growth, it is understood that they vary with initial concentration, culture time, and microalgae species [7,14,32]. It has also been proven that as the light intensity increases, microalgae growth increases up to a point, and that very high light intensity will be inhibitory for microalgae growth and nutrient removal [5].

Chlorella Vulgaris Growth Depending on Light Intensity

In experiments carried out at different light intensities (20, 60, 100 and 140 µmolphoton/m²s) for microalgae growth, *Chlorella vulgaris* biomass grown in BG11, BG11pH, C5.2 and pH 10.5 samples and chlorophyll-a pigment contents of these samples were compared. Biomass graphs were drawn for all samples depending on time and light intensity and are given in Figure 3. As the light intensity increased in all samples, the amount of biomass increased continuously for 12 days.

According to the time-dependent measurement results for 12 days, less biomass increase was measured in the C5.2 sample compared to the control sample BG11 at all light intensities. Biomass growth of BG11pH was slightly slower compared to BG11. At the end of the 12th day at 140 µmolphoton/m²s, the amounts of biomass formed in BG11pH and BG11 samples were measured in very close amounts as 1030 mg/L and 1180 mg/L, respectively. This situation indicates the negative effect of pH shock on microalgae. However, it is possible to say that pH shocking for the destruction of harmful microorganisms in the biologically treated poultry slaughterhouse wastewater is negligible in terms of microalgae growth. pH 10.5 was the sample in which the least and slow microalgae growth was detected for each light intensity among all samples, and the microalgae biomass grown was relatively small compared to C5.2. The 20 µmolphoton/m²s light intensity was insufficient for microalgae growth in all samples.

Samples containing BG11 increased their biomass from around 200 mg/L to over 1000 mg/L at 100 and 140 µmolphoton/m²s light intensities at the end of the 12th day. The biomass amount of the C5.2 sample was measured as 1000 mg/L at 100 µmolphoton/m²s and 860 mg/L at 140 µmolphoton/m²s at the end of the experiment. This is due to the difference in wastewater and artificial nutrient media content and the presence of competitive microorganisms in the biologically treated poultry slaughterhouse wastewater. As in this study, in studies in the literature, high light intensity has an inhibitory effect on algae growth after a point [3,13]. Measurement of chlorophyll-a amount depending on time and light intensity is an essential parameter for monitoring microalgae growth in samples [34]. Chlorophyll-a amounts measured for all samples at 4 different light intensities for 12 days are shown in Figure 4. The highest amount of

chlorophyll-a was obtained with the control sample, BG11. After 12 days, chlorophyll-a was measured as 4.40 and 4.19 mg/L for BG11 and BG11 pH, respectively. Increases and decreases in chlorophyll-a amounts in BG11pH and C5.2 samples indicate the effect of pH shock and algae adaptation. In samples containing biologically treated poultry slaughterhouse wastewater, the highest chlorophyll-a 3.32 mg/L for C5.2 at 140 µmolphoton/m^{2}s; For pH 10.5, it was measured as 2.83 mg/L. Chlorophyll-a amounts due to microalgae deaths decreased in all samples at 20 µmolphoton/m^{2}s

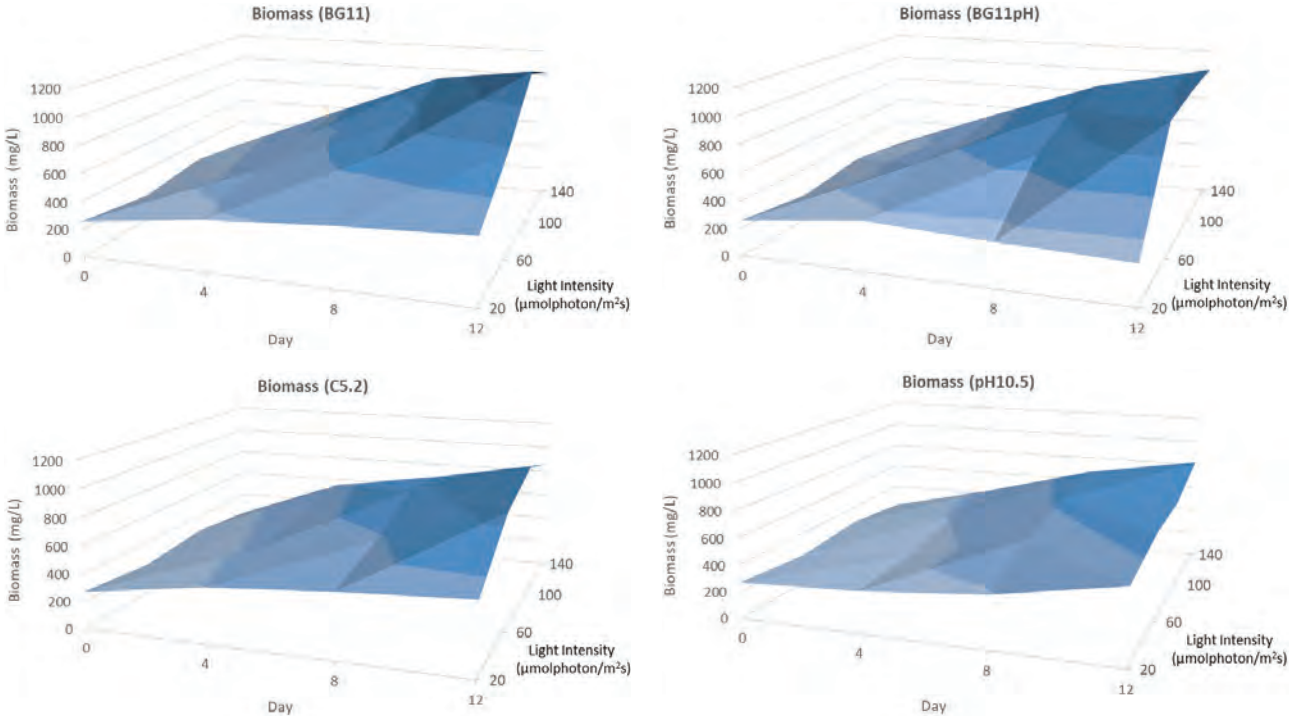


Figure 3. Biomass growth depending on light intensity.

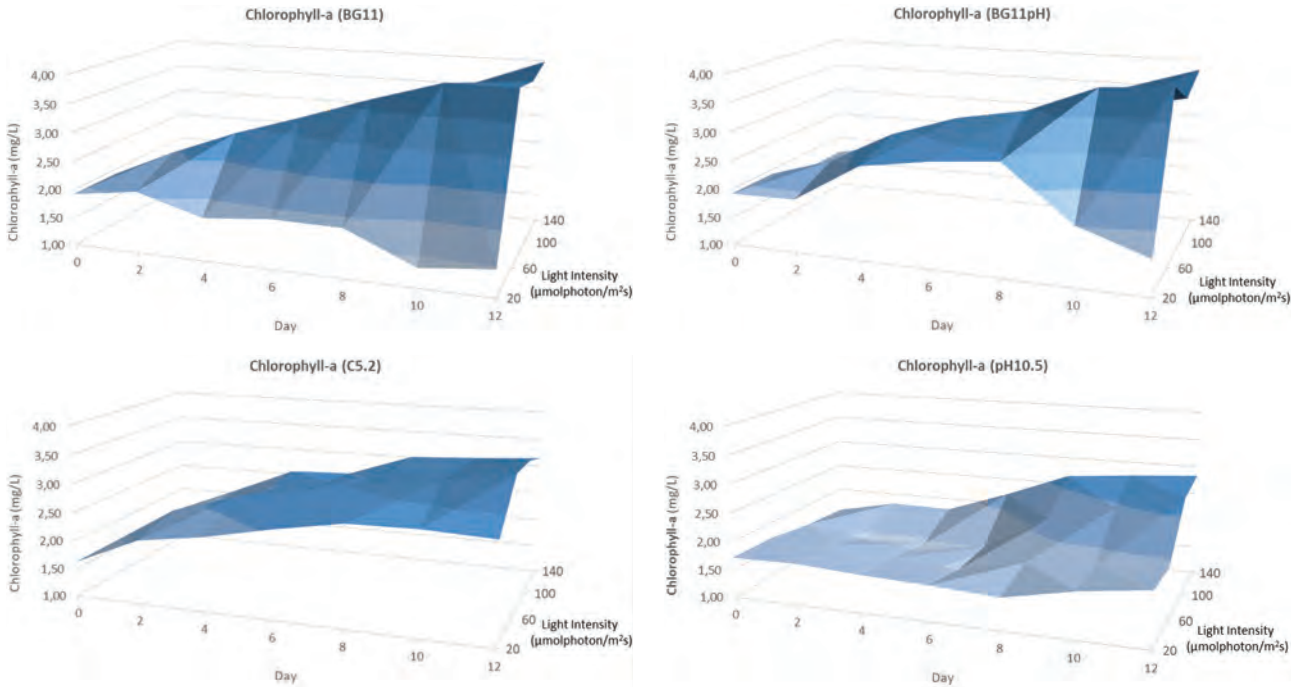


Figure 4. Chlrophyll-a growth depending on light intensity.

and were insufficient for microalgae growth under these experimental conditions.

The amount of chlorophyll-a represents viable and healthy algal cells. For this reason, it is understood by comparing the results that it is different from the total biomass amount and increases slightly. Increasing light intensity significantly increases chlorophyll-a synthesis in microalgae cells [35]. A study examining the effect of light intensity stated that chlorophyll began to degrade at high light intensities. In the same study, *Chlorella sp.* For continuous light, medium-intensity light (200 μmolphoton/m^2s) was stated to be the best for biomass increase [34].

Total Lipid and Fatty Acid Methyl Esters Formed Depending on Light Intensity

Table 2 shows that the total lipid production for all samples increased as the light intensity increased. According to the results obtained, it is clear that the pH 10.5 method is less suitable for lipid production. C5.2 sample containing the biologically treated poultry slaughterhouse wastewater developed by pH shocking method was compared with BG11 control sample, and matching results were obtained regarding lipid production. The effect of pH shocking on microalgae growth is revealed by comparing the BG11pH sample with the control sample. Comparing the total lipid ratios obtained for the BG11pH sample with the control sample BG11, it formed 1-3% less lipid at low light intensities and less than 1% at high light intensities.

The highest total lipid content in the C5.2 sample was measured with 0.1142 g at 140 μmolphoton/m^2s light intensity. The lowest total lipid amount was calculated as 0.0886 g at 20 μmolphoton/m^2s for C5.2. With the increase in light intensity, the lipid ratios of the dry biomass obtained increased. In studies in the literature, While microalgae biomass increased with increasing light intensity, lipid production increased primarily due to microalgae species and other environmental conditions [6,7], decreased in some studies [8,9], and was not affected by light intensity in some studies [10].

The lipid ratios in the dry biomass of the C5.2 sample were obtained as 19.5% and 20.5% of the dry biomass at 100 and 140 μmolphoton/m^2s light intensities, respectively. These lipid ratios obtained are slightly higher than BG11. The highest lipid content of 14.8% was obtained in the biomass developed in slaughterhouse wastewater with different microalgae species [20]. pH 10.5 has the lowest lipid content in dry biomass at all light intensities. High light intensity causes a decrease in lipid content. In a study conducted by [34] in which the effect of light intensity was examined, it was stated that chlorophyll-a began to deteriorate at high light intensities, but the lipid content of the biomass increased. In a study, it was explained that light intensities higher than 150 μmolphoton/m^2s in a sample given over a photobioreactor, depending on the microalgae species, prevent microalgae growth [5]. For this reason, it is not possible to obtain microalgae growth with the desired efficiency at very high light intensities.

Low light intensity, on the other hand, is more suitable for lipid deposition, because at high light intensity microalgae tend to proliferate rather than increase lipid [8]. For this reason, it is vital to determine the optimum light intensity for lipid production. Light intensities higher than 140 μmolphoton/m^2s were not studied in this study due to the decrease in light intensity compared to 100 μmolphoton/m^2s in fatty acid methyl esters at 140 μmolphoton/m^2s and the inhibitory property of high light intensity. Total lipid amounts and lipid ratios in dry biomass were higher for each light intensity in the C5.2 sample than in the BG11 control sample. This situation can be explained by the increased lipid ratio of *Chlorella vulgaris* under stress with pH shocking. In addition, it can be explained by the high concentration of nitrogen and phosphorus nutrient content of poultry slaughterhouse wastewater compared to BG11 nutrient medium [16,36].

FAME obtained as a result of lipid transesterification. The FAME profile obtained from Chlorella vulgaris consists of palmitic acid (C16:0), stearic acid, palmitoleic acid (C16:1), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) (Table 2). As a result of the increase in light intensity, there were differences in the composition and concentration of fatty acid methyl esters.

The total amount of lipids obtained was also low in parallel with the microalgae growth, chlorophyll-a pigment analysis, and biomass measurement results at 20 μmolphoton/m^2s. Likewise, the concentration of fatty acid methyl esters and the detected FAME species are also quite low compared to the concentrations obtained at other light intensities. For this reason, low illumination was found insufficient for this study in terms of microalgae growth and the use of developed biomass as biofuel raw material. The sample, which was constantly kept at pH 10.5, gave insufficient results regarding the obtained FAME species and concentrations compared to the pH shocking method.

The highest concentrations of saturated fatty acids (C16:0 and C18:0) were obtained in the C5.2 sample at 60 μmolphoton/m^2s, and saturated fatty acids constitute 73.39% of the total fatty acids. At 140 μmolphoton/m^2s, the highest light intensity, the C5.2 sample has a higher lipid ratio than the control sample. Saturated fatty acids C16:0 and C18:0, and unsaturated fatty acids C18:2 and C18:3 formed at 100 and 140 μmolphoton/m^2s. Fatty acid methyl ester content from microalgae biomass is an important parameter for evaluating the suitability of microalgae for biodiesel production.

The presence of C16 and C18 fatty acid methyl esters is an indicator of quality biodiesel [37]. C18 fatty acids obtained from microalgae; they exist mainly in unsaturated form (C18:1, C18:2, and C18:3) [38]. C16-C18 saturated and unsaturated fatty acids obtained in this study overlap with those obtained from microalgae in the literature.

Table 2. Ratios of total lipid (g) and FAME (%) grown in total microalgae depending on light intensity

Light Intensity (μmolphoton/m²s)	Sample	Total lipid (g)	Lipid ratio in dry biomass (%)	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3
20	BG11	0.078	11.986	0.032					
	BG11pH	0.072	12.966	0.022					
	C5.2	0.089	16.407	0.120	0.010	0.010		0.049	
	pH10.5	0.056	13.455	0.045				0.012	
60	BG11	0.105	13.690	0.227	0.009	0.012	0.007	0.019	
	BG11pH	0.111	16.609	0.272	0.030	0.060		0.075	
	C5.2	0.103	15.785	0.208	0.021	0.043		0.070	
	pH10.5	0.056	12.790	0.047				0.010	
100	BG11	0.087	17.283	0.211		0.047		0.076	0.034
	BG11pH	0.107	16.612	0.152		0.026		0.052	0.019
	C5.2	0.109	19.517	0.184		0.033		0.074	0.023
	pH10.5	0.079	18.769	0.067		0.025		0.021	
140	BG11	0.121	18.311	0.147		0.034		0.050	0.009
	BG11pH	0.119	17.633	0.162		0.042		0.074	0.019
	C5.2	0.114	20.429	0.158		0.030		0.064	0.017
	pH10.5	0.052	9.756	0.013					

The suitability of the obtained fatty acids for biodiesel production was evaluated by the European biodiesel quality standard (EN 14214). As stated in EN 14214, the main limiting factor is considered to be the ratio of C18:2 fatty acids to total fatty acids [39]. C18:3 linolenic acid, should not exceed 12% in fatty acids; It was below this ratio at all light intensities in the C5.2 sample (Table 3). While linolenic acid (C18:3) did not form in the C5.2 sample at 20 and 60 μmolphoton/m²s light intensities, it was calculated as 7.325% and 6.320%, respectively, at 100 and 140 μmolphoton/m²s light intensities.

The obtained biodiesel is not desired that it consists of only saturated or only unsaturated fatty acids in order to establish the appropriate balance of properties such as combustion, oxidative stability, and viscosity. Ideally, quality biodiesel is expected to consist of a mixture of long-chain saturated fatty acids (SFA), short-chain mono (MUFA), and polyunsaturated (PUFA) fatty acids [1,39,40]. In this case, there is no objection to European standards to using the obtained biodiesel as fuel.

As given in Table 3, the biodiesel obtained from *Chlorella vulgaris* grown in the biologically treated water of the poultry slaughterhouse provides the necessary conditions for a quality biodiesel. Appropriate lipid ratio and FAME content for biodiesel production were obtained from the C5.2 sample at all light intensities. In terms of lipid ratio of dry biomass, fatty acid methyl ester amounts and species, the most suitable results were obtained for microalgae growth in the C5.2 sample at 100 μmolphoton/m²s light intensity.

Table 3. SFA, MUFA and PUFA ratios (%) in C5.2 sample at different light intensity

	Light intensities (μmolphoton/m²s)			
FAME	20	60	100	140
C16:0	0.12	0.208	0.184	0.158
C16:1	0.01	0.021		
C18:0	0.01	0.043	0.033	0.03
C18:2	0.049	0.07	0.074	0.064
C18:3			0.023	0.017
Total	0.189	0.342	0.314	0.269
%SFA	68.783	73.392	69.108	69.888
%MUFA	0.015	0.029	0	0
%PUFA	25.926	20.468	30.892	30.112
% C18:3	0	0	7.325	6.320

Kinetic Model Calculations

Microalgae growth kinetic model calculations were made depending on the time of microalgae developed in C5.2 sample at optimum light intensity. Similarly, the removal kinetic models were calculated for NH_4^+ ve PO_4^{3-} in the C5.2 sample at optimum light intensity.

Microalgae Growth Kinetics

The kinetic equations were calculated by Equation 1 and Equation 2 [31]. The k and R^2 values obtained from the kinetic calculations are given in Table 4.

Table 4. R^2 and k values for biomass and Chlorophyll-a obtained from kinetic calculations

Kinetic Model	Biomass		Chlorophyll-a	
	k	R^2	k	R^2
1st order	-0.1397	0.8553	-0.07	0.8647
2nd order	-0.0003	0.8499	-0.032	0.7852

According to the kinetic calculations, the increase in biomass and chlorophyll-a gave results consistent with the 1st order kinetics. As seen in Figure 5, the experimental results obtained and the theoretical results calculated according to the 1st-order kinetic model calculated for biomass increase are compatible. However, after the 8th day, the experimental concentration value of the biomass tends to stabilize. This situation is because microalgae growth has entered a stagnation phase with nutrient reduction due to nutrient depletion [41]. The microalgae growth entered the stagnation phase after the logarithmic increase phase. At the beginning of the stagnation phase, the increase in biomass tends to stabilize.

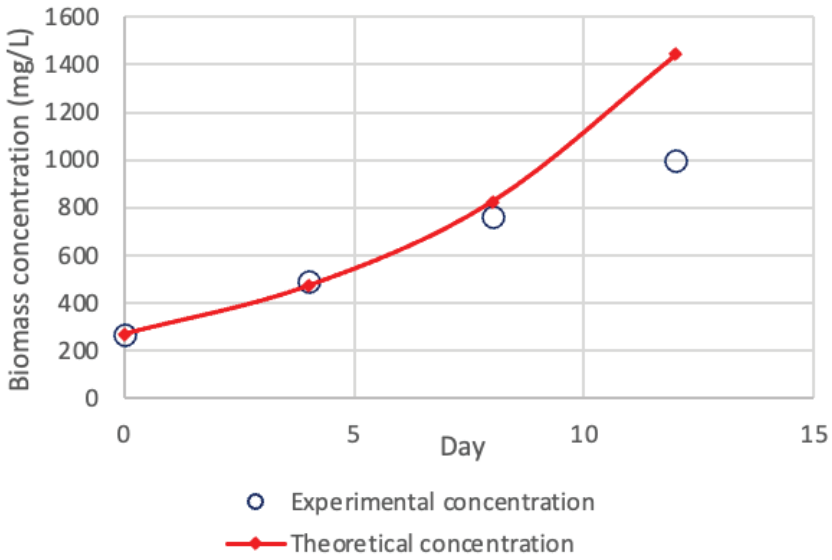


Figure 5. Comparison of theoretical and experimental biomass amounts obtained from kinetic calculations. Light intensity:100 µmolphoton/m^2s.

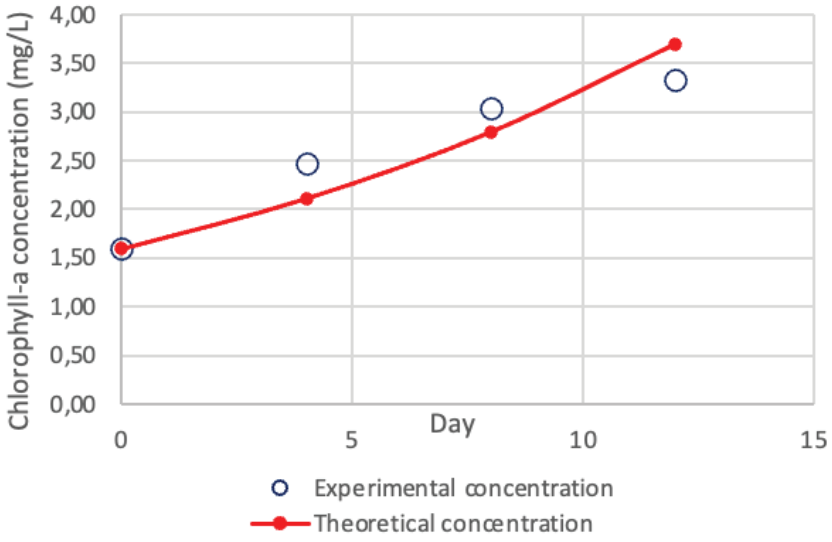


Figure 6. Comparison of theoretical and experimental chlorophyll-a amounts obtained from kinetic calculations. Light intensity:100 µmolphoton/m^2s.

Experimental and theoretical results in chlorophyll-a increase calculated according to the 1st order kinetic model are close to each other (Figure 6). On the 12th day, the experimental concentration of chlorophyll-a decreased compared to the theoretically calculated value due to the depletion of the nutrient in the medium. As for biomass, the experimental concentration of chlorophyll-a tends to stabilize at the end of the experiment. The compatibility of biomass and chlorophyll-a increases, proving the accuracy of the 1st-order kinetic models.

Nutrient Removal Kinetics from the Biologically Treated Poultry Slaughterhouse Wastewater

Kinetic model calculations were made according to reaction time and reaction rate constant. The regression coefficient values were calculated according to the kinetic model calculations. First-order and second-order kinetic models were examined in the kinetic calculations for removing PO_4^{3-} and NH_4^+ from the biologically treated poultry slaughterhouse wastewater by microalgae. The k and R^2 values were calculated for the investigated kinetic models. The kinetic equations were calculated by Equation 1 and Equation 2 [31]. The k and R^2 values obtained from the kinetic models are given in Table 5.

Calculated k and R^2 values from the kinetic models for the removal of PO_4^{3-} and NH_4^+ have been given in Table 5. According to the table, the most suitable model for PO_4^{3-} and NH_4^+ has been fit to the 1st-order kinetics with the R^2 values of 0.9634 for PO_4^{3-} and 0.8516 for NH_4^+, respectively. R^2 values close to 1 indicate the reliability and predictability of the nutrient removal of the process. In a study by Abirama et al. 2021, the biokinetic model of ammonium and phosphate removal from meat processing unit wastewater by Botrycoccus sp. microalgae was investigated. The kinetics of ammonium and phosphate removal of Botrycoccus sp. microalgae from wastewater were found to be 99.03 % and 99.93 %, respectively [42].

Table 5. R^2 and k values for the removal of PO_4^{3-} and NH_4^+ obtained from kinetic calculations

Kinetic model	PO_4^{3-}		NH_4^+	
	k	R^2	k	R^2
1st order	0.5383	0.9634	0.3911	0.8516
2nd order	54.853	0.5533	0.1693	0.5512

Moreover, k values have been figured out for kinetic models, and the first order model, which was the most appropriate, have been calculated as 0.5383 for PO_4^{3-} and 0.3911 for NH_4^+. Obtained experimental data and calculated theoretical results using the 1st order kinetic model for the removal of PO_4^{3-} and NH_4^+ have nearly overlapped, as seen in Figure 7 and Figure 8.

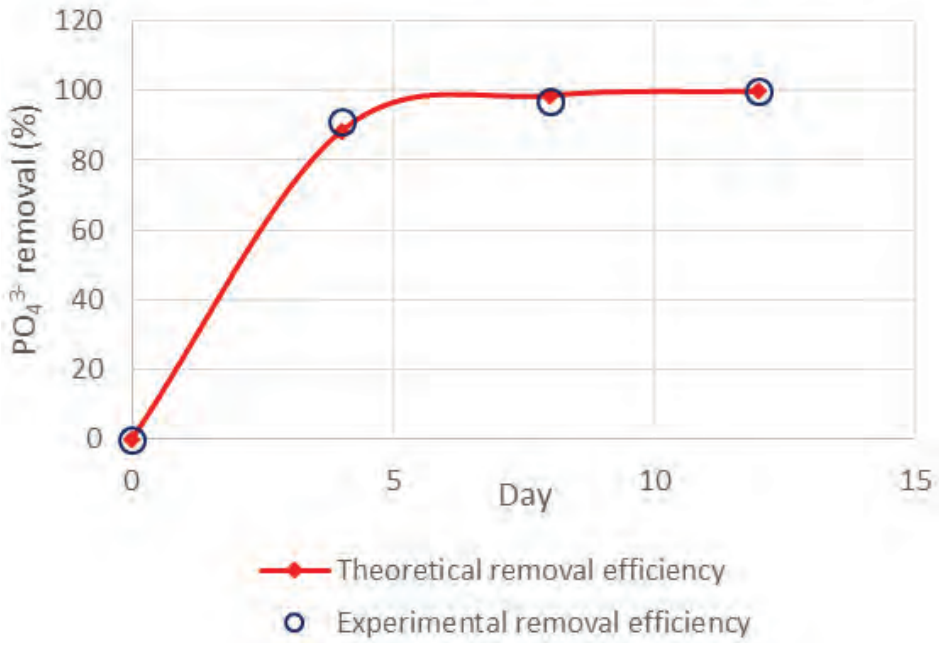
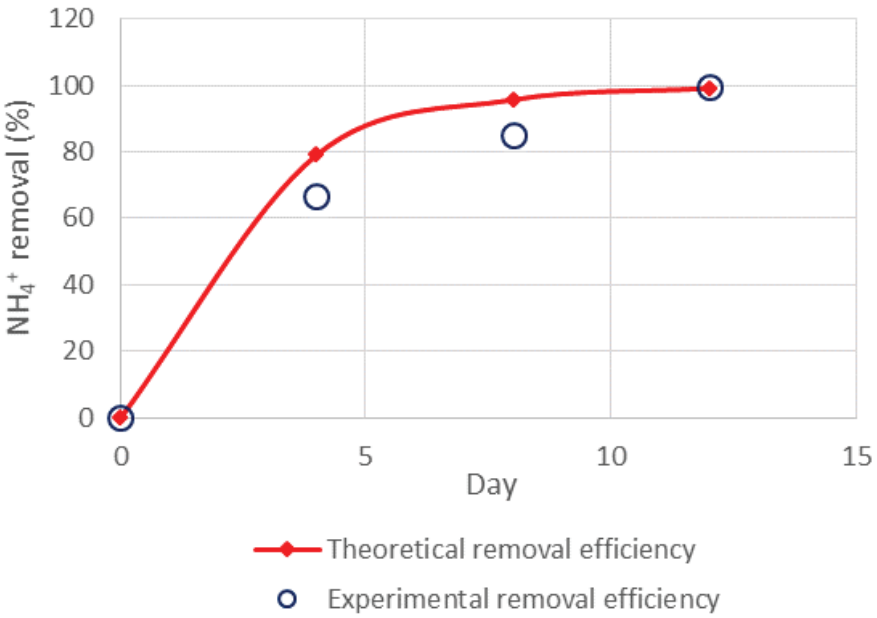


Figure 7. Experimental data and theoretical results using the 1st order kinetic model for the removal of PO_4^{3-}. Light intensity:100 μmolphoton/m^2s.



Figurel 8. Experimental data and theoretical results using the 1st order kinetic model for the removal of NH_4^+. Light intensity: 100 µmolphoton/m^2s.

CONCLUSION

This study aimed to investigate the advanced treatment of biologically treated poultry slaughterhouse wastewater and to find the optimum light intensity for developing microalgae culture. Although biologically treated wastewater contains suitable nutrients for microalgal growth, it is not possible due to rotifers and bacteria in wastewater. For this reason, suitability for microalgal growth and advanced treatment of the biologically treated poultry slaughterhouse wastewater were investigated by changing the pH of the nutrient medium with two different methods. The harmful organisms were destroyed by the alkaline pH method (pH 10.5), but fixing the pH at 10.5 restricted microalgal growth, as seen from the biomass, chlorophyll-a, total lipid and fatty acid methyl ester results. However, since the ammonium and phosphate removal rates from the wastewater are pretty high, it has been found suitable for using biologically treated wastewater with *Chlorella vulgaris* at pH 10.5 for advanced treatment of poultry slaughterhouse wastewater. The growth of *Chlorella vulgaris* in the biologically treated poultry slaughterhouse wastewater by pH shocking method gave similar results with the BG11 artificial medium in terms of biomass and lipid formation. High removal efficiencies (up to 100%) have been obtained at all light intensities for nutrient removal from the biologically treated poultry slaughterhouse wastewater with the C5.2 sample. Nutrient removal rates also increased with the light intensity. According to the study results, *Chlorella vulgaris* developed at 100 µmolphoton/m^2s light intensity is more suitable for biodiesel production in terms of fatty acid methyl ester formation and high lipid ratio in dry biomass. The most important contribution of this study to the literature is the development of *Chlorella vulgaris* culture with pH shocking method with wastewater taken after biological treatment at 100 µmolphoton/m^2s light intensity. Biomass grown in the biologically treated poultry slaughterhouse wastewater with pH shocking method, C5.2, has a higher total lipid amount and lipid ratio in dry biomass than biomass grown in BG11 artificial nutrient medium. The lipid ratio of dry biomass obtained at high light intensities was calculated as 18% for BG11 and 20% for C5.2. According to the results, higher lipid and fatty acid methyl ester were obtained with the establishment of stress conditions. According to European biodiesel quality standards, the obtained fatty acid methyl ester types are also suitable for biodiesel production (EN 14214). When approached in terms of advanced biological treatment and development of microalgae culture ideal for biodiesel production, large-scale application is possible in line with the information obtained in this study. The advantage of implementing this system on a large scale is that pH adjustments are low cost and easy to apply. Microalgae cultivation in large tanks has been done worldwide for many years. Similar designs can be made in addition to the existing biological treatment plants.

AUTHORSHIP CONTRIBUTIONS

Gamze KATIRCIOĞLU SINMAZ (PhD): Conceptualization, formal analysis, validation, investigation, writing - original draft preparation. İsmail Ayhan ŞENGİL (Prof. Dr.) Supervision, methodology, writing - review & editing.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Chisti Y. Biodiesel from microalgae. Biotechnol Adv 2007;25:294–306. [CrossRef]

[2] Knothe G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. Fuel Process Technol 2005;86:1059–1070. [CrossRef]

[3] Ananthi V, Brindhadevi K, Pugazhendhi A, Arun A. Impact of abiotic factors on biodiesel production by microalgae. Fuel 2021;284:118962. [CrossRef]

[4] Chang HX, Huang Y, Fu Q, Liao Q, Zhu X. Kinetic characteristics and modeling of microalgae Chlorella vulgaris growth and CO2 biofixation considering the coupled effects of light intensity and dissolved inorganic carbon. Bioresour Technol 2016;206:231–238. [CrossRef]

[5] Ogbonna JC, Tanaka H. Light requirement and photosynthetic cell cultivation - Development of processes for efficient light utilization in photobioreactors. J Appl Phycol 2000;12:207–218. [CrossRef]

[6] Seo SH, Ha JS, Yoo C, Srivastava A, Ahn CY, Cho DH, et al. Light intensity as major factor to maximize biomass and lipid productivity of Ettlia sp. in CO2-controlled photoautotrophic chemostat. Bioresour Technol 2017;244:621–628. [CrossRef]

[7] Nzayisenga JC, Farge X, Groll SL, Sellstedt A. Effects of light intensity on growth and lipid production in microalgae grown in wastewater. Biotechnol Biofuels 2020;13:1–8. [CrossRef]

[8] Cheirsilp B, Torpee S. Enhanced growth and lipid production of microalgae under mixotrophic culture condition: Effect of light intensity, glucose concentration and fed-batch cultivation. Bioresour Technol 2012;110:510–516. [CrossRef]

[9] George B, Pancha I, Desai C, Chokshi K, Paliwal C, Ghosh T, et al. Effects of different media composition, light intensity and photoperiod on morphology and physiology of freshwater microalgae Ankistrodesmus falcatus - A potential strain for biofuel production. Bioresour Technol 2014;171:367–374. [CrossRef]

[10] Pancha I, Chokshi K, Mishra S. Enhanced biofuel production potential with nutritional stress amelioration through optimization of carbon source and light intensity in Scenedesmus sp. CCNM 1077. Bioresour Technol 2015;179:565–572. [CrossRef]

[11] Kligerman DC, Bouwer EJ. Prospects for biodiesel production from algae-based wastewater treatment in Brazil: A review. Renew Sustain Energy Rev 2015;52:1834–1846. [CrossRef]

[12] Wang M, Keeley R, Zalivina N, Halfhide T, Scott K, Zhang Q, et al. Advances in algal-prokaryotic wastewater treatment: A review of nitrogen transformations, reactor configurations and molecular tools. J Environ Manage 2018;217:845–857. [CrossRef]

[13] Yan C, Luo X, Zheng Z. Effects of various LED light qualities and light intensity supply strategies on purification of slurry from anaerobic digestion process by Chlorella vulgaris. Int Biodeterior Biodegrad 2013;79:81–87. [CrossRef]

[14] Cheng T, Wei CH, Leiknes TO. Polishing of anaerobic secondary effluent by Chlorella vulgaris under low light intensity. Bioresour Technol 2017;241:360–368. [CrossRef]

[15] Yang J, Li X, Hu H, Zhang X, Yu Y, Chen Y. Growth and lipid accumulation properties of a freshwater microalga, Chlorella ellipsoidea YJ1, in domestic secondary effluents. Appl Energy 2011;88:3295–3299. [CrossRef]

[16] Cho S, Luong TT, Lee D, Oh Y-K, Lee T. Reuse of effluent water from a municipal wastewater treatment plant in microalgae cultivation for biofuel production. Bioresour Technol 2011;102:8639–8645. [CrossRef]

[17] Kim D, Kim EK, Koh HG, Kim K, Han JI, Chang YK. Selective removal of rotifers in microalgae cultivation using hydrodynamic cavitation. Algal Res 2017;28:24–29. [CrossRef]

[18] Katırcıoğlu Sınmaz G, Şengil İA. Microalgae culture development in biologically treated water of poultry slaughterhouse. In: Proc 16th Int Conf Clean Energy; 2018. p. 9–11.

[19] Katırcıoğlu Sınmaz G, Erden B, Şengil İA. Cultivation of Chlorella vulgaris in alkaline condition for biodiesel feedstock after biological treatment of poultry slaughterhouse wastewater. Int J Environ Sci Technol 2022;20:3237-3246. [CrossRef]

[20] Hernández D, Riaño B, Coca M, Solana M, Bertucco A, García-González MC. Microalgae cultivation in high rate algal ponds using slaughterhouse wastewater for biofuel applications. Chem Eng J 2016;285:449–458. [CrossRef]
[21] Azam R, Kothari R, Singh HM, Ahmad S, Ashokkumar V, Tyagi VV. Production of algal biomass for its biochemical profile using slaughterhouse wastewater for treatment under axenic conditions. Bioresour Technol 2020;306:123116. [CrossRef]
[22] Taşkan E. Performance of mixed algae for treatment of slaughterhouse wastewater and microbial community analysis. Environ Sci Pollut Res 2016;23:20474–20482. [CrossRef]
[23] Terán Hilares RA, Garcia Bustos KA, Sanchez Vera FP, Colina Andrade GJ, Pacheco Tanaka DA. Acid precipitation followed by microalgae (Chlorella vulgaris) cultivation as a new approach for poultry slaughterhouse wastewater treatment. Bioresour Technol 2021;335:125284. [CrossRef]
[24] Andersen RA. Algal culturing techniques. 1st ed. New York: Academic Press; 2005.
[25] APHA. Standard methods for the examination of water and wastewater. 19th ed. American Public Health Association Inc.; New York; 1995.
[26] EPA. ESS Method 150.1: Chlorophyll – Spectrophotometric. Environmental Sciences Section Inorganic Chemistry Unit; 1991.
[27] Lee I, Park JY, Choi SA, Oh YK, Han JI. Hydrothermal nitric acid treatment for effectual lipid extraction from wet microalgae biomass. Bioresour Technol 2014;172:138–142. [CrossRef]
[28] Bligh E, Dyer W. Rapid method of total lipid extraction. J Biochem Physiol 1959;37:911–917. [CrossRef]
[29] Soares AT, da Costa DC, Silva BF, Lopes RG, Derner RB, Antoniosi Filho NR. Comparative analysis of the fatty acid composition of microalgae obtained by different oil extraction methods and direct biomass transesterification. Bioenergy Res 2014;7:1035–1044. [CrossRef]
[30] Tanattı P, Şengil A, Özdemir A. Treatment of biodiesel wastewater by solvent extraction: Evaluation of kinetic and thermodynamic data. Environ Eng Manag J 2018;17:2657–2665. [CrossRef]
[31] Lam MK, Yusoff MI, Uemura Y, Lim JW, Khoo CG, Lee KT, Ong HC. Cultivation of Chlorella vulgaris using nutrients source from domestic wastewater for biodiesel production: Growth condition and kinetic studies. Renew Energy 2017;103:197–207. [CrossRef]
[32] Sydney EB, da Silva TE, Tokarski A, Novak AC, de Carvalho JC, Woiciecohwski AL, et al. Screening of microalgae with potential for biodiesel production and nutrient removal from treated domestic sewage. Appl Energy 2011;88:3291–3294. [CrossRef]
[33] Solmaz A, Işik M. Microalgae production with microalgal submerged membrane photo bioreactor (Msmpbr) and examining the nutrient removal yield. Sigma J Eng Nat Sci 2017;8:255–261.
[34] He Q, Yang H, Wu L, Hu C. Effect of light intensity on physiological changes, carbon allocation and neutral lipid accumulation in oleaginous microalgae. Bioresour Technol 2015;191:219–228. [CrossRef]
[35] Gim GH, Ryu J, Kim MJ, Kim II P, Kim SW. Effects of carbon source and light intensity on the growth and total lipid production of three microalgae under different culture conditions. J Ind Microbiol Biotechnol 2016;43:605–616. [CrossRef]
[36] Chinnasamy S, Bhatnagar A, Hunt RW, Das KC. Microalgae cultivation in a wastewater dominated by carpet mill effluents for biofuel applications. Bioresour Technol 2010;101:3097–3105. [CrossRef]
[37] Gao F, Yang HL, Li C, Peng YY, Lu MM, Jin WH, et al. Effect of organic carbon to nitrogen ratio in wastewater on growth, nutrient uptake and lipid accumulation of a mixotrophic microalgae Chlorella sp. Bioresour Technol 2019;282:118–124. [CrossRef]
[38] Deshmukh S, Kumar R, Bala K. Microalgae biodiesel: A review on oil extraction, fatty acid composition, properties and effect on engine performance and emissions. Fuel Process Technol 2019;191:232–247. [CrossRef]
[39] Mohd-Sahib AA, Lim JW, Lam MK, Uemura Y, Isa MH, Ho CD, et al. Lipid for biodiesel production from attached growth Chlorella vulgaris biomass cultivating in fluidized bed bioreactor packed with polyurethane foam material. Bioresour Technol 2017;239:127–136. [CrossRef]
[40] Komolafe O, Velasquez Orta SB, Monje-Ramirez IY, Noguez I, Harvey AP, Orta Ledesma MT. Biodiesel production from indigenous microalgae grown in wastewater. Bioresour Technol 2014;154:297–304. [CrossRef]
[41] Çirik S, Gökpınar Ş. Plankton bilgisi ve kültürü. 5th ed. İzmir: Ege Üniversitesi Yayınları; 2008. [Turkish]
[42] Abirama V, Mohamed RMSR, Al-Gheethi A, Abdul Malek M, Kassim AHM. Meat processing wastewater phycoremediation by Botryococcus sp.: A biokinetic study and a techno-economic analysis. Sep Sci Technol 2021;56:577–591. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 794–819, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2013

Research Article

Optimization of fibre laser parameters for micro-texturing in piston ring manufacturing

Atulkumar PATIL[1], Pankaj BELDAR[1,*], Prashant KUSHARE[1], Vishnu WAKCHAURE[2], Uddhav SHIRSAT[3]

[1]*Department of Mechanical Engineering, K.K.Wagh Institute of Engineering Education and Research, Savitribai Phule Pune University, 411007, India*
[2]*Department of Mechanical Engineering, Amrutvahini College of Engineering, Savitribai Phule Pune University, 411007, India*
[3]*Department of Mechanical Engineering, Bhivarabai Sawant College of Engineering and Research, Savitribai Phule Pune University, 411007, India*

ARTICLE INFO

Article history
Received: 04 November 2024
Revised: 13 January 2025
Accepted: 30 January 2025

Keywords:
Confocal Microscope; Fiber Laser; Taguchi Method; Texture Dimensions

ABSTRACT

Surface texturing with laser is a common method to enhance tribology of interacting surfaces, and, nevertheless, there are serious difficulties with controlling a perfect size of texture. This paper examines how fiber laser process parameters influence the depth of the texture, roughness, and diameter and how Taguchi methodology can be used to optimize this process. The test materials were commercial cast iron piston rings which had a surface roughness of 0.482 um, hardness of 410 HV, and a Youngs modulus of 250 Gpa. Surface texturing was performed on a Marko Laser fiber laser machine with power of 20 W, wavelength of 1062 nm, frequency of 200 KHz and a speed of 140 mm/s. We have used Taguchi method in conjunction with multi-objective optimization with the aid of the gray relational grade (GRG) rank method where an L9 orthogonal array was used to assess the influence of laser power, frequency, and scanning speed on the texture dimensions. A confocal microscope was used to determine surface topography. Findings have shown that laser power has the strongest effect on dimple attributes with 58.14 percent variance in diameter because of it, after which laser frequency (31.49 percent) and scan speed (0.35 percent) follow. The best set of parameters used gave an ideal dimple size and this was; scan speed of 140 mm/s, 25 KHz frequency, and laser power of 16 W. The originality of the study at hand is that, it uses a combination of Taguchi analysis and gray relational grade (GRG) rank to optimize the laser parameters including power, frequency, scan speed with dimple characteristics of cast iron piston rings. This method does not only enable one to have a clear picture on the influence of these parameters on the dimple sizes and the surface roughness but also helps to create more efficient and durable piston rings by using efficient laser texturing.

Cite this article as: Patil A, Beldar P, Kushare P, Wakchaure V, Shirsat U. Optimization of fibre laser parameters for micro-texturing in piston ring manufacturing. Sigma J Eng Nat Sci 2026;44(2):794–819.

***Corresponding author.**
*E-mail address: pankajrbell@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Surface texturing enhances the performance of moving mechanical parts due to their high load capacity, lesser wear and less friction. Scientists are making attempts to identify the optimal texture sizes under various circumstances. One can make these textures in a number of ways, such as using photochemical etching, abrasive jet machining, embossing, burnishing, and laser surface texturing. The most used is laser surface texturing that is capable of producing very finer textures. Yet, it is hard to make very small textures, particularly required in such applications as ring-liner systems. The textures are dependent on the laser settings such as power, scanning speed, frequency, and pulses. It is complex to find the appropriate environments. This research intends to come up with the influence of these laser settings on the texture size to simplify the process. Laser texture fabrication is a process that requires laser accuracy and the ability to create fine features on the surface, which is needed in most cases in industries. To this end, femtosecond (fs) and nanosecond (ns) pulse lasers are often employed, since they are very precise when creating textures. But in comparison with other fiber laser machines that run in the millisecond (ms) band, these systems are often prohibitively costly as a result of their advanced technology. Fiber lasers can be less accurate and less high quality of texture but still remain a rather cost-efficient solution that can address most of the industrial challenges where cost is a decisive factor.

Surface texturing with the help of laser is mainly aimed at the improvement of tribological properties of cast iron piston rings. This is done through the formation of accurate dimple designs on the face, which may contribute to better lubrication retention, and lessening friction, and wear hence making the complete functioning and life cycle of the piston rings in the mechanical systems more functional. The accuracy of texture dimensions is of great significance to tribological uses since surface texture influences directly the friction, wear resistance, retention of lubricants, and bearing capacity of load. The existence of uniform dimples are micro-reservoirs of lubricants that guarantee maximum lubrication, less friction, and uniform wear, which contribute to component longevity and better performance. The unreliability due to inconsistency of texture dimensions may result in uneven lubrication, localized wear, and load distribution. In the case of cast iron piston rings, tight texturing enhances the retention of oil, sealing, and minimization of friction loss as well as engine efficiency and is recommended in the quest to increase the service life and efficiency of vehicle engines. The key challenges in achieving precise dimensions in laser surface texturing include:

Laser Parameters Control

Laser power, frequency, scan speed and-pulse duration need to be accurately regulated to achieve uniform dimple size(diameter, depth, and roughness of the surface). Minor differences may lead to a mismatch of textures, which influences tribological performance [1].

Material Properties

Various materials respond differently to laser energy as in the case of cast iron. Their thermal characteristics (including melting point, thermal conduction and reflectivity) may affect the intensity with which the material may absorb laser energy and the precision with which texture may be controlled [2].

Heat Affected Zone (HAZ)

Laser surface texturing can induce localized heating, which may result in microstructural changes, residual stresses, or thermal deformation around the textured region. This may impact on the accuracy of the dimple formation adverse [3].

Debris and Surface Contamination

When performing the laser process, the molten material can harden differently resulting in debris or flaws that impact the consistency of the textured surface. These residues might need to be removed by post-processing[4].

Laser Spot Size and Focus

It is important to ensure that laser spot size and focus is constant in order to produce accurate dimples. Deviation in focus and alignment used in the process may result into variation in the dimensions of texture [1].

Surface Reflection and Absorption

Lasers are reflected by the surface and hence surfaces with a high amount of reflectivity cause the inefficiency of the process and erratic dimensions. Pre-treatments or surface coating can be required to alleviate this problem [5].

Reproducibility

It may be difficult to obtain the same and repeatable result using different samples because controlling all the factors at the same time can be difficult particularly in mass production settings.

To overcome those obstacles, careful optimization of laser parameters, strict control of the processes, and even post-processing procedures are necessary to achieve the required level of the surface texture. Laser processing has been experiencing remarkable progressions in an effort to enhance different technologies of laser and their uses. Recent investigations have discussed various issues of laser performance such as distribution of beam energy, optimization of parameters, and improvement in processing of materials. Moradi et al. [2] examined the consequences of energy distribution of laser beams on the surface transformation hardening by using Ndlasers and diode. Their research has showed that distribution of energy has a significant impact on the hardening process, which has an impact on the property of resulting material. This literature makes it possible to establish the parameters of laser, which may be customized to obtain the desired hardness characteristics in treated materials. The paper by Karamimoghadam et al. [1] gave an

overview of the developments in the area of laser surface transformation hardening of automotive metals. They have pointed out recent advances in technology that sought to enhance the life span and reliability of automobile parts by laser hardening. Their review explains that more research is necessary to streamline and perfect laser hardening processes and apply them to improved automotive interests. Mehrabi et al. [6] examined the use of direct laser metal deposition in the production of functionally graded materials. Their paper used experimental and response surface techniques to optimize additive manufacturing processes, and demonstrated that laser techniques can be successfully used to form complex graded structures with improved performance properties. Moradi et al. [7] discussed the processes of enhancing CO2 laser cutting on polycarbonate materials with a lot of focus on better cutting efficiency and quality. As evidenced in their study, process optimization had the potential of greatly improving the output of laser cutting in manufacturing process especially when dealing with plastic material. Baranowski et al. [8] have worked out and optimized a new process in laser-sintering to additively manufacture parts of carbon fiber reinforced polymer. Their experimental study aimed at the optimization of the process parameters that could yield high-quality continuous fiber-reinforced components to demonstrate the improvement of the laser-based manufacturing technologies. Ding et al. [4] used generalized regression neural networks and genetic algorithms to optimize the fiber laser cutting. Their method enabled a multi-objective optimization, cutting quality and efficiency. The paper is relevant to the research of the enhancement of laser cutting processes using highly developed optimization methods. The pump of a Tm-doped fiber laser was a compact design that Yao et al. [9] employed in environmental gas sensing. They were working on maximizing the design of laser to improve sensor operation, portending the ability of fiber lasers to monitor the environment. In a study conducted by Shehryar Khan et al. [10], the impacts of laser beam defocusing on welding process of automotive press- hardened steels were investigated. Their numerical and experimental confirmation gave them an insight into how to optimize the laser welding parameters to enhance the quality and productivity of the welds. Pan et al. [11] maximized the optical fiber laser linewidth with a self-injection technique of light. The study of their research led to the stability and performance of fiber lasers in thermal of the challenges to do with linewidth and operation of laser. Khozeymeh et al. [12] analyzed the applicability of the dysprosium-doped yellow fiber lasers in ophthalmology. They oriented their research to maximization of laser parameters to address a medical necessity, which brings the importance of fiber lasers to the medical practice. The in situ measurement of the temperature of fiber core laser by Lou et al. [13] was designed with consideration of the optical frequency domain reflectometry (OFDR). This development has made it possible to have a better temperature control of the high-power fiber lasers and this has enhanced overall performance. The approach taken by Lapre et al. [14] was to use genetic algorithms in optimizing the operation of noise-like pulse fiber lasers operating on broadband. Their work proved the usefulness of optimization methods in improving the work of laser and the working distance. Gopinath et al. [15] used Technique of Order of Preference by similarity to ideal solution (TOPSIS) to maximise the fibre laser microcutting parameters on duplex steel**.** Chen et al. [16] studied dissipative solitons in Er-doped fiber lasers by optimizing the tasks using machine-learning, using Gaussian processes. Their work shed light on how to optimize laser functioning with the help of the best machine-learning methods. Vora et al. [17] experimentally studied and Pareto optimized the processes of fiber laser cutting of Ti6AL4V. Their research was aimed at enhancing cutting quality and efficiency of titanium alloys, which prove useful in the practice of optimizing laser. In the study by Najjar et al. [18], neural networks and chimp optimization were utilized in forecasting the quality of the kerf when cutting basalt fiber-reinforced composites using lasers. Their output helped in the development of predictive models on enhancement of laser cutting in composite materials. Wu et al. [19] optimised the parameters of the fiber laser welding to achieve the trade-off between energy use and bead geometry. Their work helped to enhance efficiency and performance in welding with the emphasis of the multi-objective optimization. Zeng et al. [20] balanced stimulated Raman scattering (SRS) and thermal-mechanical instability (TMI) effects in direct diode-pumped high-power fiber lasers by optimizing them. Their study was to improve the work of lasers, which are high-power. Han et al. [21] studied the production, optimization, and use of ultrashort pulses of a few seconds in mode-locked fiber lasers. They worked on high precision of laser processing with ultrashort pulse. Kokhanovskiy et al. [22] used the mode-locked fiber laser inverse design based on the particle swarm optimization. Their efforts led to optimization methods in designing lasers and this has helped enhance the performance and efficiency of lasers. Yassin and Hojjati [23] conducted a review of the thermoplastic matrix composite processing in terms of automated fiber placement and tape laying. Their overview shows the developments in composite fabrication process, and the role played by automation in promoting the efficiency of the process and the characteristics of the materials. The review of artificial intelligence-based mode-locked fiber lasers was given in detail by Ma and Yu [24]. Their paper examined the possibility of optimizing the work of lasers using AI methods, with attention to the progress in mode-locking processes and their use in other applications in high-precision areas. Wu et al. [25] examined the intelligent soliton generation of breathing in ultra fast

fiber lasers. Their study showed the possibilities of employing the advanced control algorithms to improve the stability and performance of soliton generation, which is the part of the high performance ultrafast lasers development. Motard et al. [26] optimized the work of a monolithic Tm3+, Ho3+-codoped fiber laser through the control of the fiber gain and the reflected wavelength. They demonstrated that optimization of these parameters would be capable of improving the efficiency and output properties of fiber lasers. Kumar Mishra et al. [27] reviewed several laser beam drilling methods of fiber-reinforced composite that have compared Nd and CO 2 lasers. Their overview of the most common types of lasers in terms of their use in drilling composite materials showed the advantages and disadvantages of each type, which can be used to comprehend the most successful techniques in various cases. Mathew et al. [28] discovered how machine learning can be used to align the laser beams automatically using the Raspberry Pi auto-aligner. Their work reflected how machine learning could be used to automate the process of laser system alignment, making laser systems more precise and efficient. Kuprikov et al. [22] used the deep reinforcement learning to self tune the laser sources of dissipative solitons. Their study proposed a new method of optimizing the performance of lasers based on self-learning algorithms, which increases the stability and the performance of laser systems. Chengal Reddy et al. [29] used the grey relational analysis to optimize the laser drilling process on the AISI 303 material. Their research was aimed at the process parameter optimization to increase the efficiency and quality of drilling proving the usefulness of optimization techniques in the manufacturing process. The optimization of the production of polarization squeezed light in nonlinear optical fibers with femtosecond pulses was performed by Andrianov et al. [30]. Their efforts helped in the development of nonlinear optics, as they enhanced the methods of light generation so that they can perform better. Czelusniak and Amorim [31] used a selective laser sintering of carbon fiber-reinforced PA12 by using Gaussian process modeling, and stochastic optimization. Their experiment was based on optimization of process variables to enhance the quality and consistency of the 3D-printed parts.

The latest developments in laser machining have also played a significant role in the development of accurate surface texturing to be used in other engineering activities. Among such developments, the article by Chandan and Sahoo [32] is quite resourceful in the applications of low-power nanosecond lasers in scalable machining of surface on Ti-6Al-4V alloy, which is used broadly in aerospace and biomedical applications.

However, in spite of the tremendous improvement in the mode of laser processing and optimization procedures, the gap in total insight and optimization of laser parameter in specific purposes is still extremely significant, especially in the context of cast iron piston rings. Current research has mostly concentrated either on general laser processing information or on an individual material type, thus there is a gap in the optimization strategies applied to cast iron components, which play a significant role in most industrial applications. The present study fills this gap by concentrating on optimization of the machine settings of laser texturing of surfaces on cast iron piston rings, which is a major part of an automotive engine.

The originality of the current study is that it uses Taguchi analysis and gray relational grade (GRG) rank as the only means of optimization of laser parameters including power, frequency and scan speed in dimple characteristics of cast iron piston rings. Such a method is not only able to give a clear insight into how these parameters influence the nature of dimples and surface roughness but also leads to more efficient and robust piston rings that are formed by laser flints.

The advantage of this study is that it was focused on the optimization of laser texturing of the cast iron piston ring surface by applying a complex of new methodologies and detailed analysis. This study is not only a huge gap in the existing literature, but it also provides the means of improvement in the work of the piston rings and their service life using laser texturing that establishes a new standard of the same application in the industry. Finally, the use of ns and fs pulse lasers in metal texture fabrication is more precise, at the expense of cost-effectiveness so that the application is not very common in industries. Although fiber lasers are less precise, they offer a feasible economic option to many applications, which is sustained by the current research to optimize the parameters of fiber lasers to improve the texture quality and material properties in diverse metals and uses.

MATERIALS AND METHODS

The effects of laser processing parameters on texture characteristics were also evaluated through the composite method of assessment consisting of experimental and analytical methods. A fiber laser machine was employed to systematically change laser power, frequency and scan speed. Employing optical microscope and profilometer was employed to scan the texture characteristics like dimple diameter, depth and surface roughness. Taguchi designed experiments and gray relational grade (GRG) analysis were utilized to establish the optimal parameter settings while analysis of variance (ANOVA) was applied to find out the significance level of each parameter on the texture properties. This general approach allowed us to effectively determine how the parameters of the laser influence the surface texture. Figure 1 shows the Methodology of this Research work.

This was then followed by the integration of some response variables, such as dimple diameter, depth and surface roughness, into a single performance index using the gray relational grade (GRG) technique. This helped

facilitate the analysis for the overall quality of the parameter settings. Also, the Analysis of Variance (ANOVA) was performed to ascertain significance of every parameter towards the characteristics of texture and thus the best parameters of the processes of lasers that would provide the most desired outcome in terms of quality of the surface texture. Other traditional methods are response surface methodology (RSM) and genetic algorithms (GA) that are used in laser processing optimization. Although RSM is useful in the development of predictive models and parameter interaction, it needs more experiments to achieve reliable outcomes. Likewise, GA is good at global optimization that is computationally expensive and challenging to execute. Unlike that, our selected Taguchi-GRG method offers a more effective and realistic solution to multi-objective optimization that has a few experimental runs. The Taguchi method with its orthogonal array design reduces the amount of experimental effort necessary, whilst retaining accuracy and the Gray Relational Grade method has been found to offer a good way of transforming multiple responses into one optimization problem. This hybrid method is quite beneficial in the industrial sectors where the resources are limited, and speed is of the essence. Our work, concerned with cast iron piston rings, is the first attempt to apply this technique properly to obtain accurate surface texturing, which provides major gains in tribological performance.

Details of Specimen and Texturing

Some of the materials that were used in this study were commercially obtained piston rings made of cast iron. These piston ring segments were wire-cut electric discharge machining (EDM) to extract them out of the entire assembly to make them suitable samples to be used in laser texturing. Table 1 gives specifications of the test specimens in detail and the test specimens in this case are the cast iron piston rings that are to be texturized. The specimens have a roughness of the surface of 0.482 um, a Youngs modulus of 250Gpa and a hardness of 410 HV. The topography of the specimens following laser texturing was determined using a confocal microscope. The averaging values of the measurements of three different locations on each specimen were used to create depth of the textured areas and to determine the average diameter.

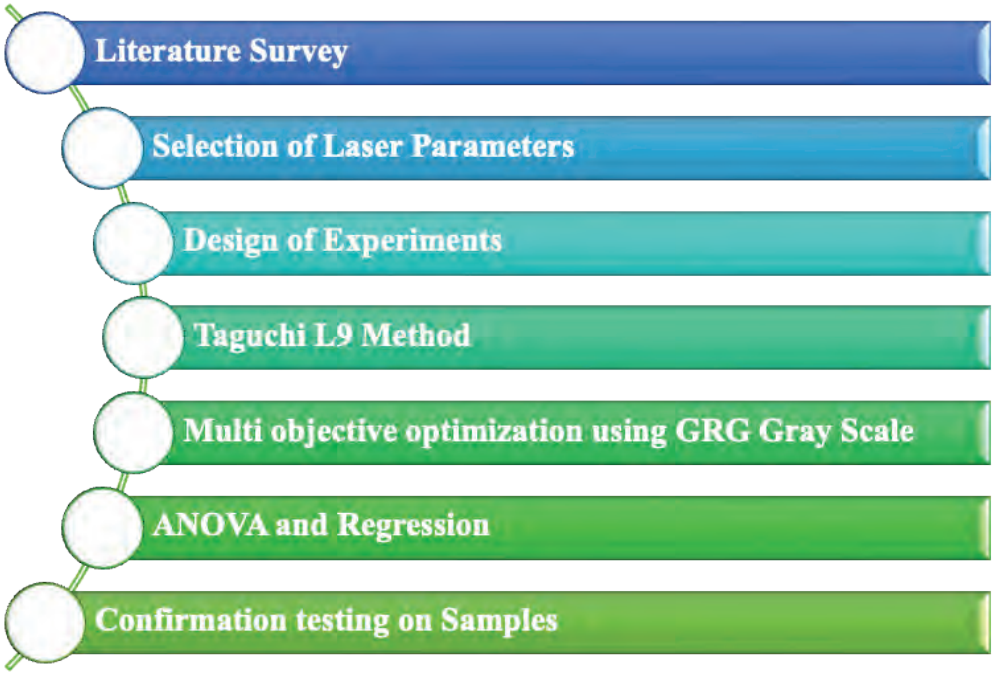


Figure 1. Methodology for Research work.

Table 1. Laser textured specimen details

Material	Chemical structure (%)	Hardness (HV)	Roughness of surface Ra (μm)
Piston Ring (Commercial) Samples	C-3.61, P-0.085, Mn-0.80, S-0.043, Si-2.80 and Fe- Remaining	410	0.482

Cast iron piston ring specimens are subjected to surface texturing using a fiber laser system with a maximum power of 20W, a maximum frequency of 500 KHz, and an operating wavelength of 1064 nm. In this study, a fiber laser was used for surface texturing. Below are the key specifications of the fiber laser machine employed in the experiments:

1. **Laser Type**: Fiber Laser
2. **Laser Wavelength**: 1064 nm
3. **Maximum Power Output**: 20W
4. **Pulse Frequency Range**: 10–50 kHz
5. **Scan Speed**: 100–300 mm/sec
6. **Beam Quality (M^2 factor)**: ≤ 1.3
7. **Spot Size**: Approximately 30–50 μm (depending on focusing lens)
8. **Pulse Duration**: Adjustable within the nanosecond range (ns)

These specifications were adjusted during the experiments to optimize the dimple dimensions (diameter, depth, and surface roughness) on the cast iron piston rings. The fiber laser was selected for its high precision, control, and ability to produce consistent surface textures. Figure 2-a depicts the schematic diagram of the fiber based laser system, While Figure 2b depicts the actual laser apparatus,

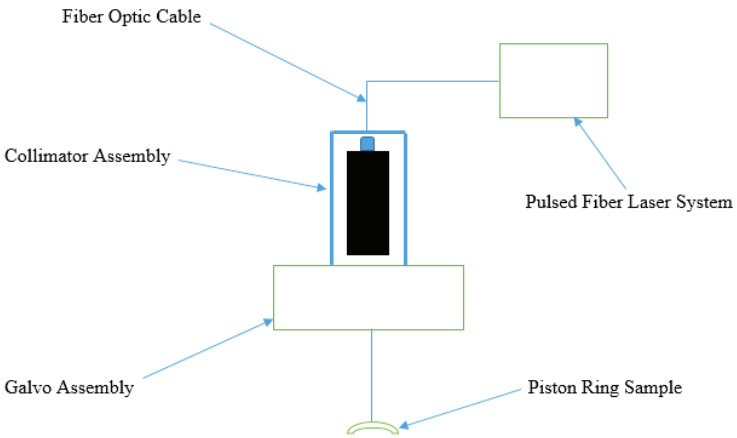


Figure. 2a Schematic of Fiber based laser system

Figure. 2b Actual laser apparatus.

The distance between the bottom of the galvo and the textured specimen is dependent on the focal length of the lens that has been absorbed into the galvo assembly. Table 2 presents the various input laser parameters that were altered when using a serial connection that was connected to a computer system. These parameters are frequency, scanning rate and power of laser. In order to determine the effect of every parameter on the texture dimensions of the piston ring material, the difference in values was broadly explored. The calibration and validation of the confocal microscopy has various major steps to be followed to ensure a good surface measurement. Reference standards with known step heights are used to perform vertical (Z-axis) calibration, whereas precision grids are in turn used to perform lateral (X and Y axes) calibration. Measures such as roughness are verified by using certified reference materials (CRMs) whose surface features are known. The resolution of the system is tested to sub-micron standards and its repeatability checked by scanning the system multiple times. Moreover, the data analysis algorithms in the software are verified to be accurate, and the surrounding conditions such as temperature and humidity are kept at the same level. The steps are used to obtain accurate and repeatable outcomes in measurements of surface topography and roughness.

Experimental Plan

Design of experiments (DOE) is a method which is systematically applied to plan, execute and analyze experiments in order to gain knowledge about the impact of different factors on a process or an outcome. Through a planned manipulation of the input factors, DOE assists in determining the factors that have a substantial effect on the outcomes and the interaction between these factors. This method permits effective experimentation of the test environment, as well as optimizing the processes in order to generate the desired results.

The Taguchi Method is a certain kind of DOE that is dedicated to the quality and performance improvement by minimizing variation and responsiveness to external factors. This technique was developed by a doctor by the name Genichi Taguchi to design experiments with a reduced number of runs, thus, a reduced amount of time and resources are saved. The Taguchi approach focuses on strong design by identifying favorable settings which maximize deviations at target values thereby improving process reliability and performance. The optimal parameters of laser processing to texturing the cast iron piston ring segments were determined with the help of Taguchi Method that considers three factors in an L9 orthogonal array and three levels [11-13]. This experimental design helped in evaluating the control factors, which were laser power, frequency and speed of scanning. According to a thorough review of literature, the ranges chosen in these factors are 8W-16W of laser power, 15 kHz-25 kHz of laser frequency, and 80mm/s-180mm/s of laser scanning speed [11-13]. Table 2 shows the particular components and their level used in the experiment.

Table 2. Variables and their intensities

Factors	Units	Levels		
		1	2	3
Laser power	Watt	8	12	16
Laser frequency	KHz	15	20	25
Laser scanning speed	mm/sec	80	140	180

Table 3. L_9 orthogonal experimental array with level details

Experiment no.	Laser power (Watt)	Frequency in (KHz)	Speed of the scanning (mm/sec)
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

The L9 experimental plan's specific levels for each of the nine experimental trials are outlined in Table 3, which is based on the factors and corresponding levels listed in Table 2. After conducting the laser texturing process according to this plan, the outcomes for each of the nine trials were evaluated in terms of texture diameter, texture depth, and texture roughness of the surface. The average values of these parameters were then calculated across all the textures created on the piston ring surface [7].

The Taguchi method can be used to reduce the need for numerous experiments while also saving money on them. The experiment's results were used to determine the (S/N) Signal to Noise ratio, which measures the quality features and makes it possible to comprehend how the results deviate from or approach the acquired results. The diameter of dimple, depth of dimple, and dimple roughness of the surface were the output parameters that were acquired from the designed experimentation. The ideal quality characteristics criteria for each of these output metrics are taken into consideration as they are covered below [15–17].

As dimple diameter increases it leads to increase in storage of lubrication for enhancing tribological applications and therefore "larger the better" optimum condition is considered in this case [2] shown by Equation 1.

$$\text{S/N ratio} = -10\log_{10}\left(1/n \sum_{i=1}^{n} \left(\frac{1}{y2}\right)\right) \quad (1)$$

Reducing dimple depth contributes to minimizing friction and wear on components, making the "smaller the better" criterion optimal in this context [2], as indicated by Equation 2. Moreover, optimizing dimple roughness of the surface, characterized by small peaks and valleys that serve as reservoirs for lubricating oils even under starvation conditions, enhances tribological performance. Hence, achieving a "smaller the better" condition for dimple roughness of the surface is crucial [1], also illustrated by Equation 2.

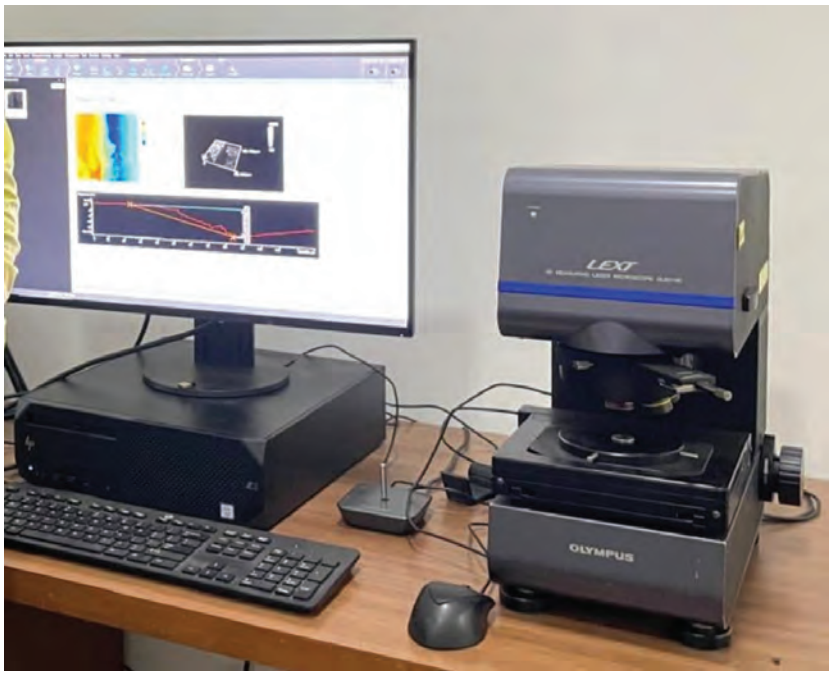
$$\text{S/N ratio} = -10\log_{10}\left(1/n \sum_{i=1}^{n} y2\right) \quad (2)$$

where y denotes the unique response of the experimental data and n the number of replications.

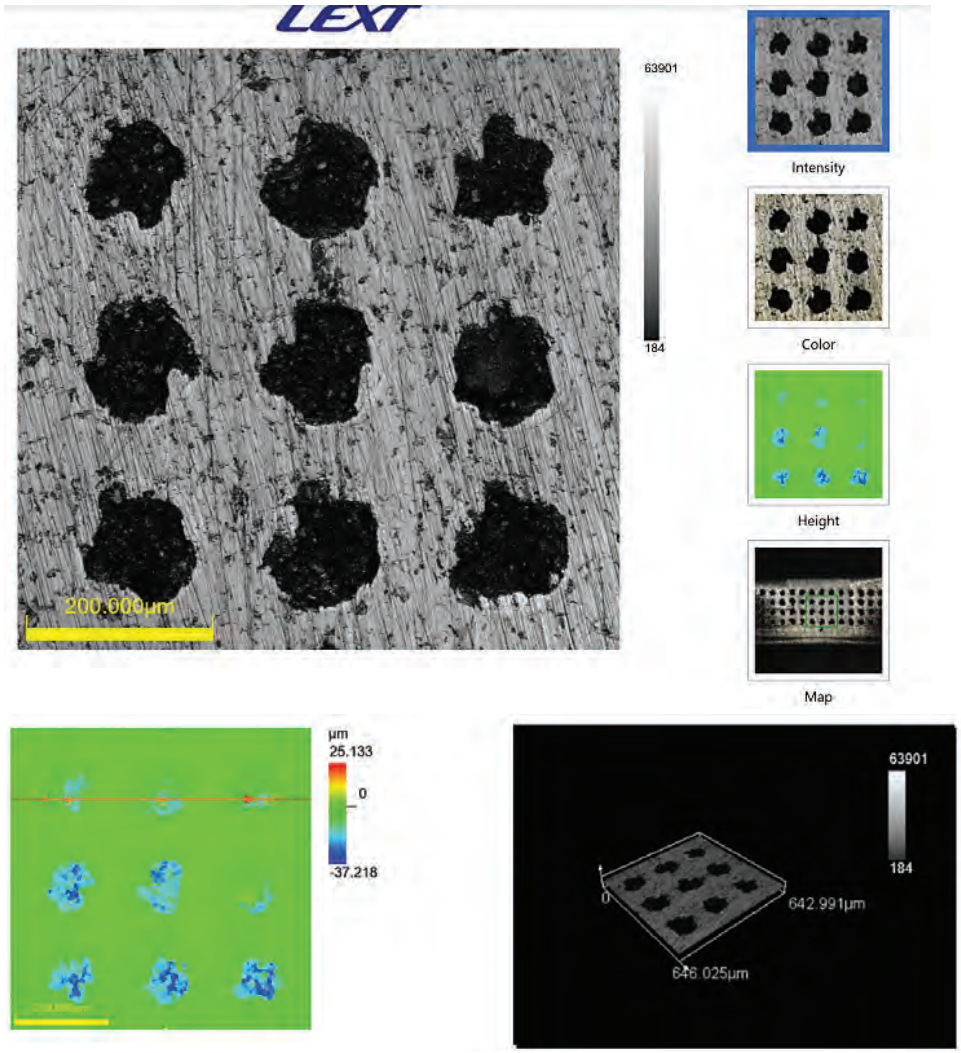
Characterization of the Specimen

An Olympus LEXT5000 confocal microscope was used to analyze the specimens of cast iron piston rings that have gone through laser texturing. The Olympus LEXT 3D OLS5000 is a 0.5 nm resolution high resolution confocal microscope and 3D profiler which has a 54x-17280x magnification and a 16-5120 micron field of view. With a reflection-type confocal laser, it can provide precise, non-contact surface profiling wherein it is suitable to analyze delicate surfaces without faulting them, as opposed to using a depth profiler. This microscope was used to measure the diameter, depth, and roughness of the dimples on the textured surfaces. The confocal microscope set up is illustrated in figure 3(a). We used confocal microscope to examine the topography of the surface of laser-textured cast iron piston rings in our study and this is essential to understand the efficiency of the laser processing parameters. Nevertheless, the original documentation of the topographical review was not detailed enough, which is why we decided to make our results more comprehensible. Confocal microscope was used to image the textured surfaces in 2D and 3D mode that enabled us to comprehensively examine the surface morphology including dimensional features of the dimple that were created by laser. We have created 3D images and this allowed us to visualize the spatial distribution and geometry of the textures, which are important in assessing their effect on the tribological performance of the piston rings. Also, we performed 2D roughness measurements, which quantifies surface properties and the parameters were Ra(average roughness), Rz (average maximum height of profile), and Rq(root mean square roughness). These measurements were important quantitative measurements of the surface finish, which allowed us to work out the correlations between roughness and possible performance under operational conditions. We made the collection of data representative by examining multiple scans on various parts of the surface, and this definition of the data collected any variability in the texture produced by the laser processing parameters. Such a combination of 2D and 3D analysis provided us with an overview of the surface properties and gave credence to the role of the laser processing parameters in our experiments. This elaboration allows us to enhance the methodological soundness of our research, as it helps us to make our work a good contribution to the existing body of knowledge and make the further study of the laser processing applications to different materials possible. Figures 3(b) to 3(d) present confocal microscope images of the piston ring sample textured under controlled conditions, in which the ranges for laser power, frequency, and scan speed were 8 W to 16 W, 15 KHz to 25 KHz, and 80 mm/s to 180 mm/s, respectively.

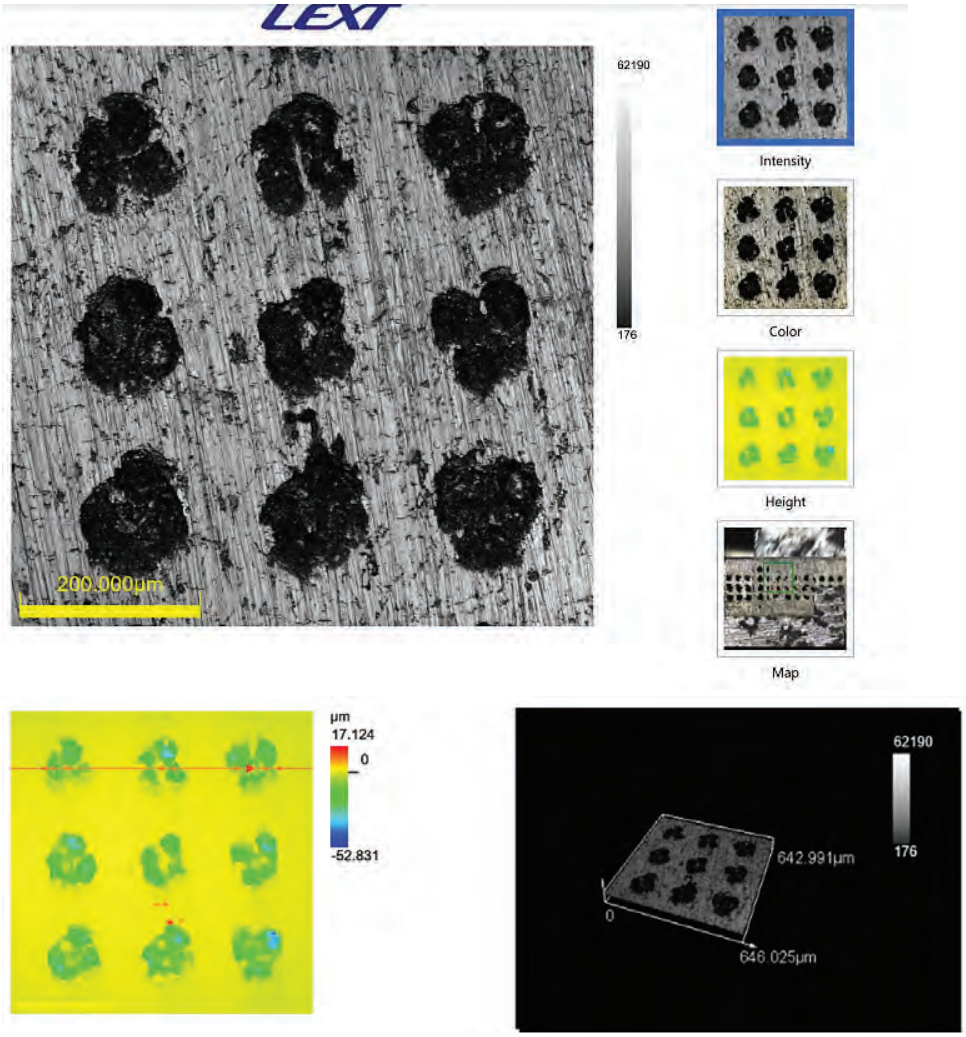
Detailed dimensional data for the specimen textured using these laser processing parameters are provided in Figures 4 to 7, sample readings of dimple diameter and depth for experimental trials 1, 4, and 8 as detailed in Table 3 of the Orthogonal array. The dimple roughness of the surface, averaged over a specific length of the experimental sample, was also measured using the confocal microscope.



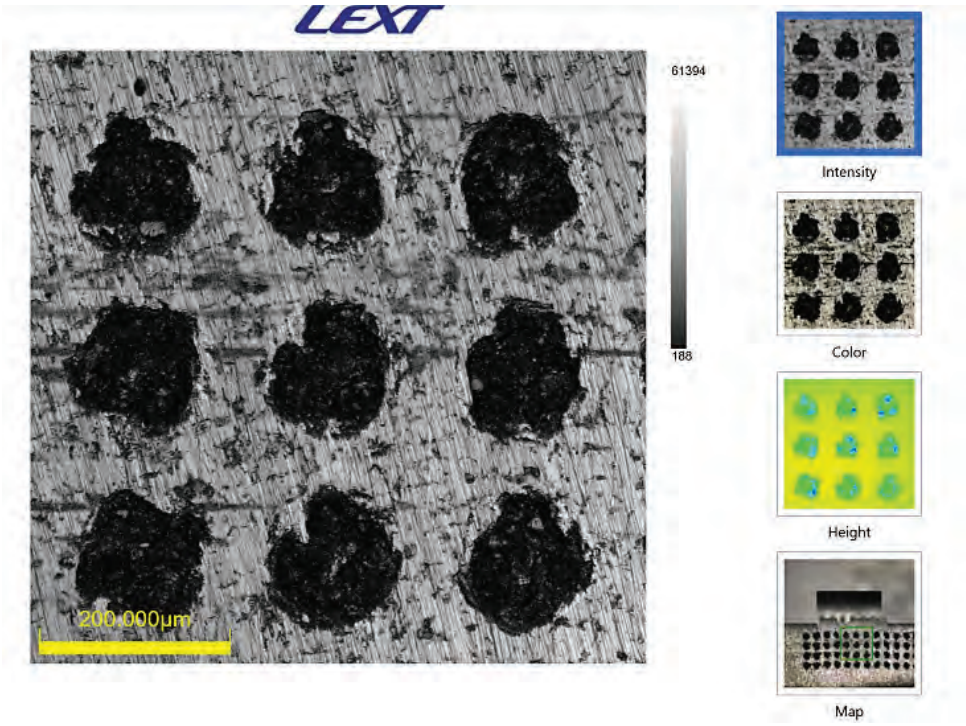
a)

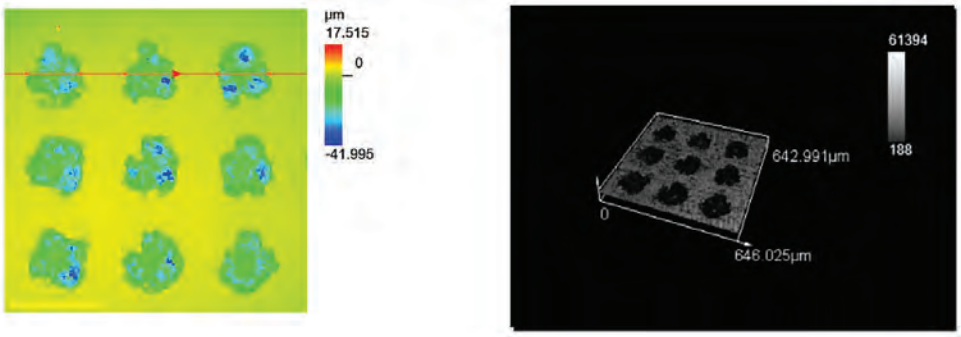


b) 8W, 15 KHz, 80 mm/sec



c) 12W, 15 KHz, 80 mm/sec





d) 16W, 20 KHz, 140 mm/sec

Figure 3. Laser Textured cast iron piston ring sample images for different laser process parameters.

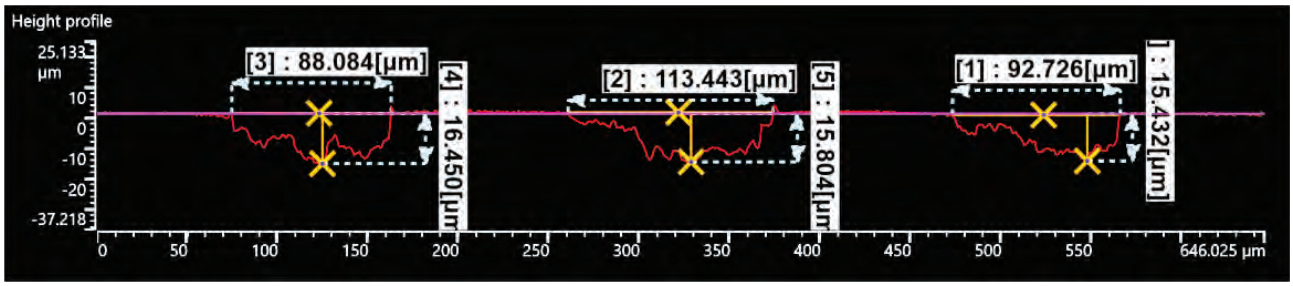


Figure 4. Dimple diameter and depth for textured sample with laser parameters 8W, 15 KHz, 80 mm/sec.

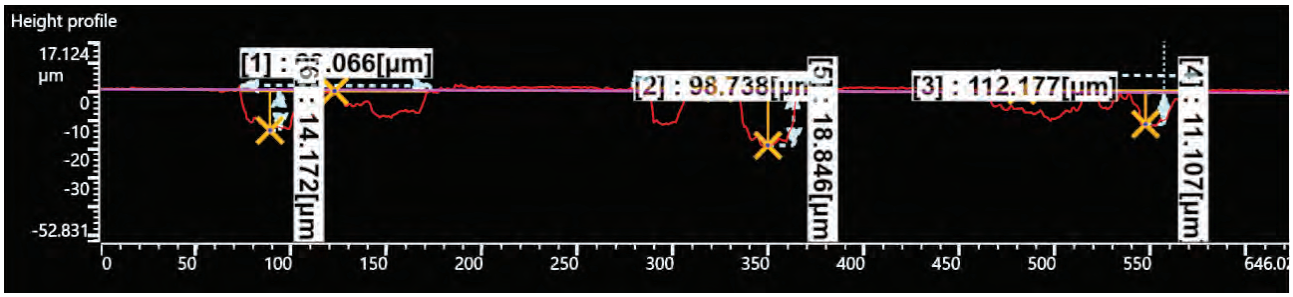


Figure 5. Dimple diameter and depth for textured sample with laser parameters 12W, 15 KHz, 140 mm/sec.

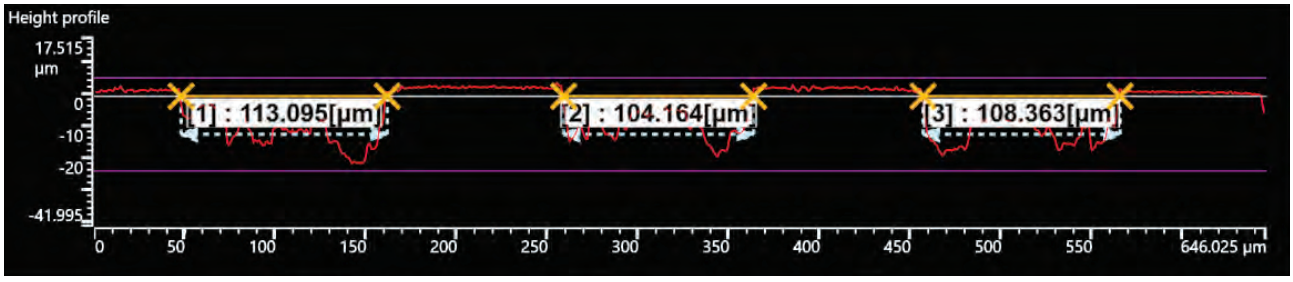


Figure 6. Dimple diameter for textured sample with laser parameters 16W, 20 KHz, 80 mm/sec.

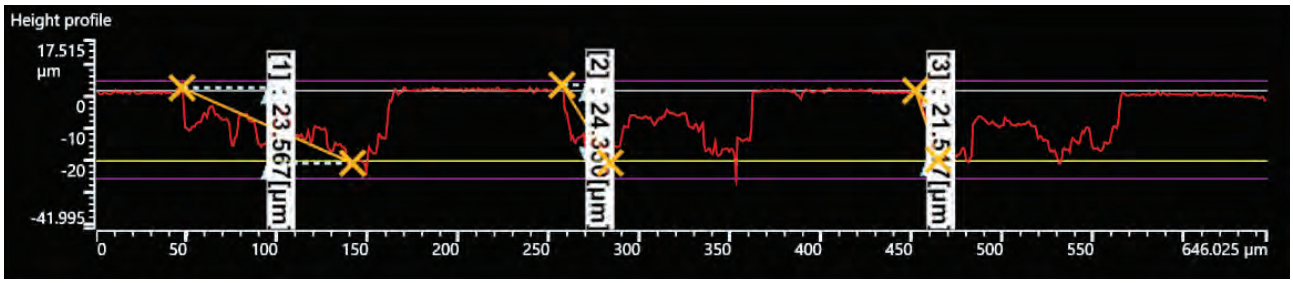


Figure 7. Dimple depth for textured sample with laser parameters 16W, 20 KHz, 80 mm/sec.

RESULTS AND DISCUSSION

Statistical analysis of the experimental data was conducted using Minitab 17. The study aimed to evaluate how laser processing parameters—specifically laser power, frequency, and scanning speed—affect the dimensions of the textured features (specifically the diameter of the dimples, their depth, and the surface roughness) created on the surface of cast iron piston rings. Table 4 provides a consolidated overview of the input parameters used in the 9 experimental runs and the corresponding output responses measured for each run.

The experiment design described in the table explores the relationship between the parameters of the laser processing and the cast iron piston ring dimple properties. The experiment involves three parameters viz. laser power, frequency and scanning speed and each parameter was varied in various experimental runs to observe their effect on the diameter of dimples, depth of the dimples and surface roughness.

Experimental Design Analysis

Laser power

As observed in the data, increased laser power between 8W and 16W is likely to give increased dimple diameter and depth. Indicatively, at power 16W, the dimple diameter and depth increase with the value being 116.782 um and 25.429 um, respectively, as compared to low powers. This is a pattern that indicates the greater power of the laser is involved in deeper and larger dimples.

Frequency

Frequencies that are high are likely to correspond with the more diameters and depth of dimple. Indicatively, the diameter and depth of the dimple at the 25 KHz frequency would be significantly elevated as is the case with the experimental runs 3, 5 and 6. However the effect of frequency with respect to roughness is not so definitive which implies that the frequency affects the diameter and depth but the effect on the roughness of the surface may be secondary or more complex.

Scanning speed

Scanning speed influences dimple sizes ambivalently. Indicatively, as the speed of scanning is slowed (e.g. at 140 mm/sec to 80 mm/sec), so does the depth and diameter of dimples. On the other hand, an increase in the scanning speed seems to reduce the dimple diameter and depth at the expense of increasing roughness. As an example, with

Table 4. Experimental input and output responses with L_9 Orthogonal Array

Expt. No.	Parameters (Input)			Responses (Output)		
	Laser power in watt	Frequency in KHz	Speed of the scanning in mm/sec	Dimple diameter in µm	Dimple depth in µm	Dimple Roughness of the surface in µm
1	8	15	80	98.084	15.895	2.451
2	8	20	140	102.993	16.555	2.525
3	8	25	180	104.03	17.344	2.765
4	12	15	140	102.993	14.708	3.545
5	12	20	180	104.03	17.344	3.434
6	12	25	80	107.507	21.901	4.23
7	16	15	180	107.13	23.069	4.3
8	16	20	80	108.541	23.144	4.23
9	16	25	140	116.782	25.429	4.442

a scanning speed of 80 mm/sec, deeper and larger dimples are created than with high speeds.

The experimental design has been effective in demonstrating the relationship between laser power, frequency and scanning speed with the achievement of intended dimple characteristics. Increasing laser power and frequency are more likely to encourage dimple diameter and depth rather than scanning speed whose measure varies irregularly. Knowledge of these types of behaviors helps in optimizing laser texturing conditions so as to achieve desired dimple profiles which may be essential in improving the performance of cast iron piston rings.

The response and parameters used in this study were informed by the aim of producing as many cast iron piston rings dimples as possible with perfection of the sharp dimples. The justification behind the following is as follows:

a. Parameters

Laser power

Justification: Laser power is an important parameter in laser processing because it directly influences the amount of energy that is applied on the material. Increase in the laser power usually augments the energy level of the laser beam which may cause bigger weldings and bigger dimples. This parameter was selected because it was to test the effect of that parameter on the dimensions of dimples so that the best power settings to apply to get the intended texture properties can be determined.

Frequency

Reasons: The frequency of the laser pulses determines the number of pulses to be applied on a unit time and this determines how the material reacts to the laser. Increased frequencies are able to put more thermal energy into the system, affecting dimple formation. The frequency aspect of the study would enable an evaluation of the degree of change in the rate of energy application on texture aspects.

Scanning speed

Justification Scanning speed determines the speed at which the laser moves across the surface and therefore the amount of energy deposited in each area. Faster scanning rate would mean low exposure time and less energy deposition, which may result in thinner and as well as shallow dimples. It is a parameter that was chosen to examine its effect on the formation of the desired dimple profile and surface finish.

b. Responses

Dimple diameter

Justification: Dimple diameter: This is a significant response as the dimension directly affects the functionality of the texture. Dimple diameter control is required in applications where there is a requirement to have a particular pattern of texture to make the substance more lubricated or less friction. This response can be measured and explicitly controlled in terms of texture profile.

Dimple depth

Rationale: Dimple depth has an impact on the behavior of the surface e.g.: can it retain lubricants or influence wear patterns. The dimension of the depth is essential to determine that the dimples are of the right size as well as possess the right depth to perform their purposes.

Dimple roughness

Justification: The general quality of the textured surface and its performance depend on the surface roughness. It determines friction, wear and retention ability of lubricants. Roughness measurement is done to verify that the surface finish is of acceptable quality concerning its intended application.

The parameters and responses adopted are in line with what the optimization of laser texturing entailed, in order to achieve some dimple characteristics. As the laser power, frequency and scanning speed are varied as well as the dimple diameter, depth and roughness are measured, the research ensures that a holistic examination is done as regards to how the contribution of these factors change to the final texture with the ultimate aim being to greatly customize the surface properties to improve its performance in the industrial processing.The selection to employ merely 9 samples in the project was based on the use of the Taguchi method experimental design, namely the L9 orthogonal array. Here is why this selection was done and why additional samples were not added:

The Taguchi approach aims at enhancing quality and performance with a small amount of experimentation through the utilization of orthogonal arrays. For this research, 3-level design was chosen for every one of the 3 variables (laser energy, frequency, and scanning velocity). For a full-factorial design with 3 variables and 3 levels, there would have been 3^3 = 27 experimental runs, which would involve a lot more time, resources, and expense.

But the L9 array of Taguchi method lets the same 3 factors and 3 levels be investigated with just 9 trials, reducing the number of tests by orders of magnitude while still allowing all factor interaction to be well represented. This reduces the process of experimentation to a more efficient scale without losing the scope for determining the best set of parameters.

More than 9 experiments would necessitate more time and cost in terms of resources. Laser texturing and measuring dimple size (diameter, depth, roughness) after that are complicated processes and equipment. The use of 9 experiments only minimizes the experimental load but not without valuable input into understanding how the input parameters influence the responses.

Practical constraints

Including more samples or trials would have increased the complexity of the study. Practical constraints such as

equipment availability, project deadlines, and resource limitations may also have played a role in limiting the number of samples.

Effect of Control Factors

The signal-to-noise (S/N) ratio response tables for the textured features of cast iron piston ring segments, specifically for dimple diameter, depth, and surface roughness, are shown in Tables 5 through 7. As explained in Tables 5-7 [10–13], these tables use delta or rank values to assess how control factors affect these dimensions. The results presented in Tables 5-7 indicate that the most significant factor influencing dimple diameter, depth, and surface roughness is laser power. Laser scan speed has the least bearing on these dimensions, whereas laser frequency comes in second. This analysis highlights that the size of the dimples formed on the piston ring surfaces is primarily determined by the laser power.

Figure 8, 9, and 10 illustrate the optimal laser processing parameters identified for creating textures on commercial piston ring segments composed of cast iron, summarized in Table 8. The results indicate that to achieve a precise dimple diameter in the circular textures on the ring samples, laser power should be set to 16 W, laser frequency to 25 KHz, and laser scan speed to 140 mm/sec. For achieving the desired dimple depth on the cast iron piston rings, the optimal parameters are 8 W of laser power, 15 KHz of laser frequency, and a laser scan speed of 140 mm/sec. Similarly, to attain the targeted dimple roughness of the surface on the studied cast iron piston ring sample, the recommended settings include 8 W of laser power, 20 KHz of laser frequency, and a laser scan speed of 140 mm/sec.

Figure 3-5 shows that while laser scanning speed has a distinct impact on each of these output variables, dimple diameter, depth and roughness of the surface all rise as laser power and frequency increase. Overall dimple roughness of the surface decreases with the increase in laser scanning speed.

ANOVA and Regression Analysis

ANOVA, a statistical technique grounded in the least squares method, is frequently employed to evaluate the effects of experimental parameters like laser power, frequency, and scan speed.

Table 5. S/N response table for dimple diameter

Level	Laser power in watt	Frequency in KHz	Speed of the scanning in mm/sec
1	40.14	40.23	40.39
2	40.41	40.44	40.62
3	40.89	40.77	40.43
Delta	0.74	0.54	0.23
Rank	1	2	3

Table 6. S/N response table for dimple depth

Level	Laser power in watt	Frequency in KHz	Speed of the scanning in mm/sec
1	-24.40	-24.88	-26.04
2	-24.98	-25.48	-25.28
3	-27.55	-26.57	-25.61
Delta	3.16	1.69	0.76
Rank	1	2	3

Table 7. S/N Response table for dimple roughness of the surface

Level	Laser power in watt	Frequency in KHz	Speed of the scanning in mm/sec
1	-8.222	-10.483	-10.947
2	-11.412	-10.429	-10.663
3	-12.716	-11.437	-10.740
Delta	4.494	1.008	0.284
Rank	1	2	3

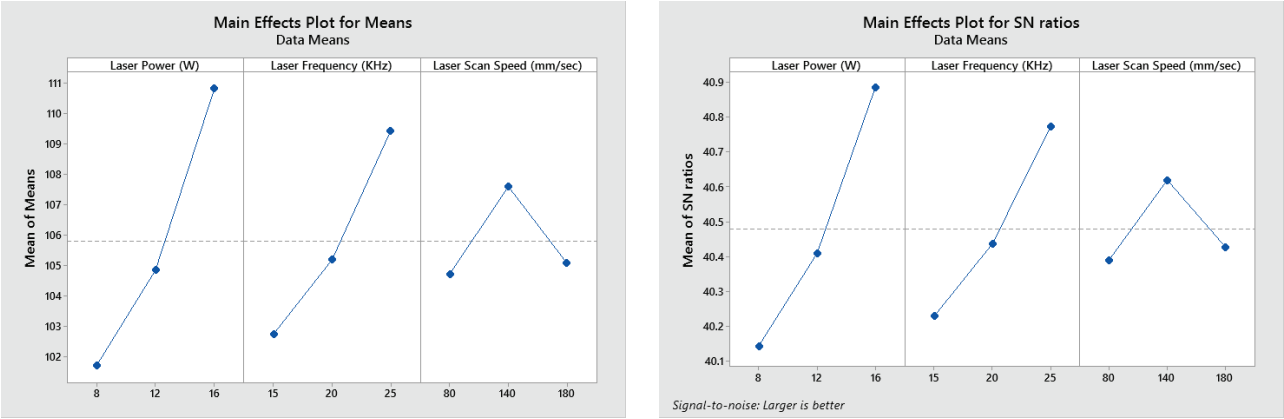


Figure 8. Plot of the main effects for the dimple diameter means and S/N ratios.

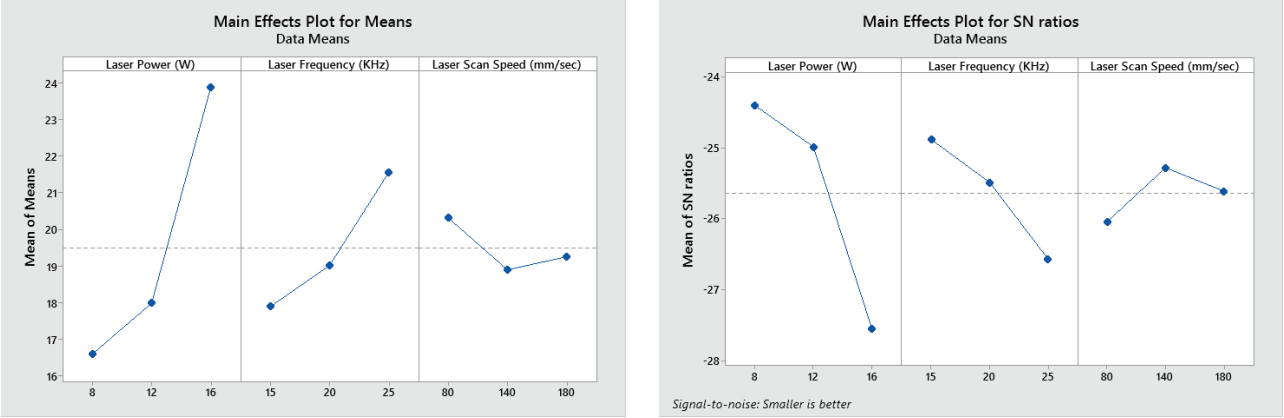


Figure 9. Plot of main effects for dimple depth means and S/N ratios.

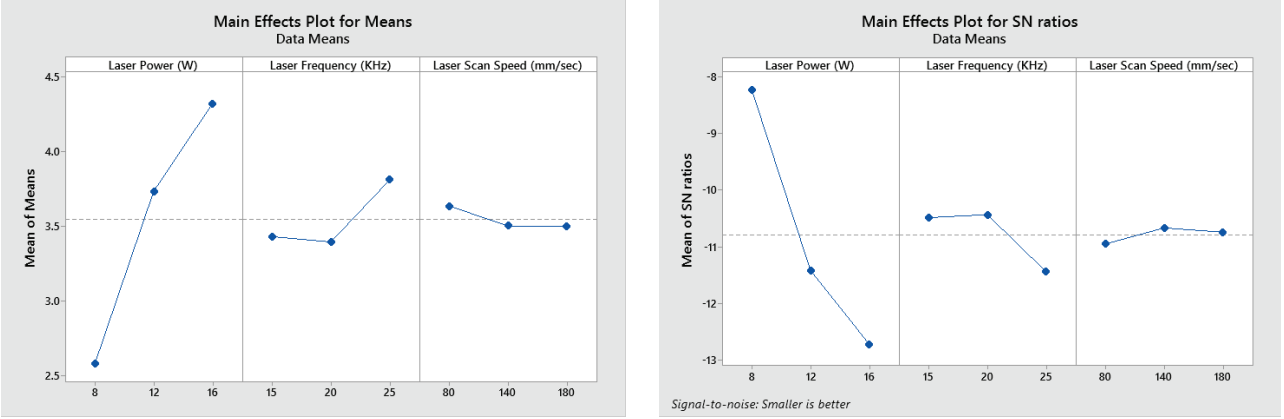


Figure 10. Principal effect plots for the surface dimple roughness's means and S/N ratios.

Table 8. Optimum laser processing parameter

Parameters	Laser power in watt	Frequency in kHz	Speed of the scanning in mm/sec
Dimple diameter	16	25	140
Dimple Depth	8	15	140
Dimple roughness of the surface	8	20	140

Tables 9, 10, and 11 present the ANOVA results for these factors, respectively. Each table includes the percentage contribution of these parameters to the measured responses, along with their significance levels indicated by p-values, where values below 0.05 are considered significant [10]. Table 9 reveals that laser power and laser frequency contributed significantly to the dimple diameter, with percentages of 58.14% and 31.49%, respectively. In contrast, laser scan speed had a minimal effect, contributing only 0.35%. For dimple depth, Table 10 shows that laser power had the highest influence at 65.82%, followed by laser frequency at 16.69%, and laser scan speed at 1.702%. Similarly, Table 11 indicates that laser power had a substantial impact on dimple roughness of the surface, contributing 88.19%. Laser frequency contributed 4.19%, while laser scan speed had the smallest influence at 0.60%. By examining these results, significant factors affecting the output dimensions can be identified based on their respective contributions and p-values.

Regression analysis serves as a tool to establish relationships between observed outcomes and controlled variables in experimental settings. In analyzing the current dataset, linear regression equations were utilized to link measured parameters—specifically dimple diameter, dimple depth, and dimple roughness of the surface—to controlled variables: laser power, laser frequency, and laser scan speed. Below is the regression equations derived from this statistical analysis for these measurable factors. The regression equation for dimple diameter in terms of the input control factors is presented below, where the coefficient of determination (R^2) assesses the goodness of fit of the model. An R^2 value closer to 1 indicates a better fit of the model [10]. The calculated R^2 of 89.94% suggests that the developed model explains approximately 89.94% of the total variance in the dimple diameter. This leaves less than 11% of the variance unexplained by the model. The adjusted R^2 value of 83.91% also supports the model's good fit.

Dimple Diameter (μm) =77.77 + 1.139 Laser Power (W) + 0.670 Laser Frequency (KHz) + 0.0070 Laser Scan Speed (mm/sec)

Similarly, the regression equation for the output parameter dimple depth, considering the input control parameters, is provided below. The coefficient of determination (R^2) for this model is 84.22%, indicating that approximately 84.22% of the total variance in dimple depth is explained by the developed model. This suggests that less than 16% of the variance remains unexplained. The adjusted R^2 value of 74.75% also indicates good agreement and robustness of the model.

Dimple Depth (μm) =2.78 + 0.910 Laser Power (W) + 0.367 Laser Frequency (KHz) - 0.0116 Laser Scan Speed (mm/sec)

Furthermore, here is the regression equation for the output parameter dimple roughness of the surface concerning the input control parameters. The coefficient of determination (R^2) for this model is 93.00%, signifying

Table 9. ANOVA table for dimple diameter

Source	DF (Degrees of Freedom)	Adj_SS (Adjusted Sum of Squares)	Adj_MS (Adjusted Mean Square)	F-Value	p	% Contribution
Laser power (Watt)	1	124.634	124.634	28.91	0.003	58.142
Laser frequency (KHz)	1	67.415	67.415	15.64	0.011	31.449
Laser scan speed (mm/sec)	1	0.752	0.752	0.17	0.693	0.350
Error	5	21.558	4.312			10.056
Total	8	214.360				

Note: S=2.07644 R^2=89.94%, R^2adj=83.91%, R^2pred=70.07%

Table 10. ANOVA table (dimple depth)

Source	DF (Degrees of Freedom)	Adj_SS (Adjusted Sum of Squares)	Adj_MS (Adjusted Mean Square)	F-Value	P-Value	% Contribution
Laser Power (Watt)	1	79.556	79.556	20.85	0.006	65.823
Laser Frequency (KHz)	1	20.174	20.174	5.29	0.070	16.691
Laser Scan Speed (mm/sec)	1	2.058	2.058	0.54	0.496	1.702
Error	5	19.074	3.815			15.781
Total	8	120.862				

Note: S=1.95316, R^2=84.22%, R^2adj=74.75%, R^2pred=44.14%

Table 11. ANOVA table (dimple roughness of the surface)

Source	DF (Degrees of Freedom)	Adj_SS (Adjusted Sum of Squares)	Adj_MS (Adjusted Mean Square)	F-Value	P-Value	% Contribution
Laser Power (Wall)	1	4.56056	4.56056	63.00	0.001	88.194
Laser Frequency (KHz)	1	0.21698	0.21698	3.00	0.144	4.196
Laser Scan Speed (mm/sec)	1	0.03151	0.03151	0.44	0.539	0.609
Error	5	0.36195	0.07239			6.999
Total	8	5.17101				

Note: S=0.26905,R^2=93.00%, R^2adj=88.80%, R^2pred=75.64%

that the constructed model explains approximately 93.00% of the overall variation in dimple roughness. This implies that about 7% of the variation remains unexplained by the model. The adjusted R^2 value of 88.80% also indicates strong agreement and reliability of the model.

Dimple Roughness of the surface (μm) =0.363 + 0.2180 Laser Power (W) + 0.0380 Laser Frequency (KHz) - 0.00144 Laser Scan Speed (mm/sec)

Figure 11 shows the all interaction plots obtained from ANOVA analysis which supports the results.

Figure 12 shows the pair plots of all parameters with other.

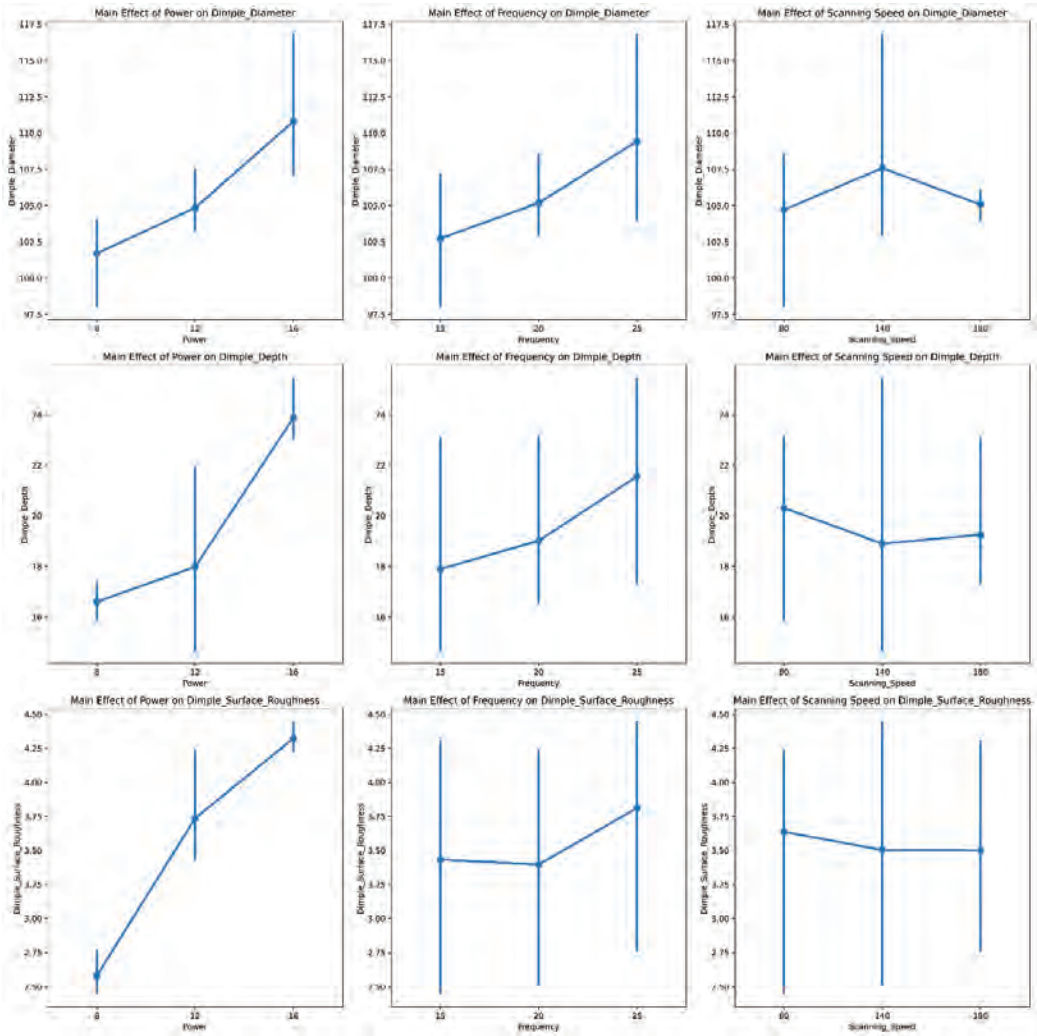
The interaction plots of the ANOVA (Fig. 11) indicate the effect of different parameters of the laser processing such as laser power, frequency and scan speed on the texture properties, such as dimple diameter, depth and roughness. The plots reveal that the largest effect on the size of dimples in a laser light is through the laser power, and the relationship between laser power and both laser frequency and scan speed gives interactions between dimples and laser power, and the relationship between dimples and laser frequency respectively. In the case of laser power variations, dimple diameter variations are mostly affected, although there are complex interactions between frequency and scan speed that affect the overall texture. These results are being supported by the pair plots (Fig. 12) which are the visual display of the relationships between the parameters in pairs. They show the effect of laser power, frequency and scan speed variation on a texture feature and illustrate such trends as increasing sizes of dimples as power and frequency increase. Those plots also imply that scan speed is compounded, it would tend to reduce the roughness of the surface, but with the various levels of power and frequencies.

The combination of these two numbers confirms that laser power is the most notable variable that can influence the size of dimples based on the contribution to the variance in the attributes of texture. ANOVA and pair plots provide a rough depiction of the combined effect of all the parameters and their interactions in creating the final texture to be able to finely-tune the process of laser texturing. The last step of the Taguchi analysis is the confirmation experiments for verifying the experimental data with the calculated outcomes obtained from regression analysis [15-17]. Table 12 shows the results of the confirmation studies that examine dimple diameter, dimple depth, and dimple roughness of the surface through regression equations to predict values based on the experimental data obtained. The verification experiments were performed under the optimal parameters found using Taguchi analysis: 16 W, 25 KHz, and 140 mm/sec for dimple diameter; 8 W, 15 KHz, and 140 mm/sec for dimple depth; and 8 W, 20 KHz, and 140 mm/sec for surface dimple roughness.

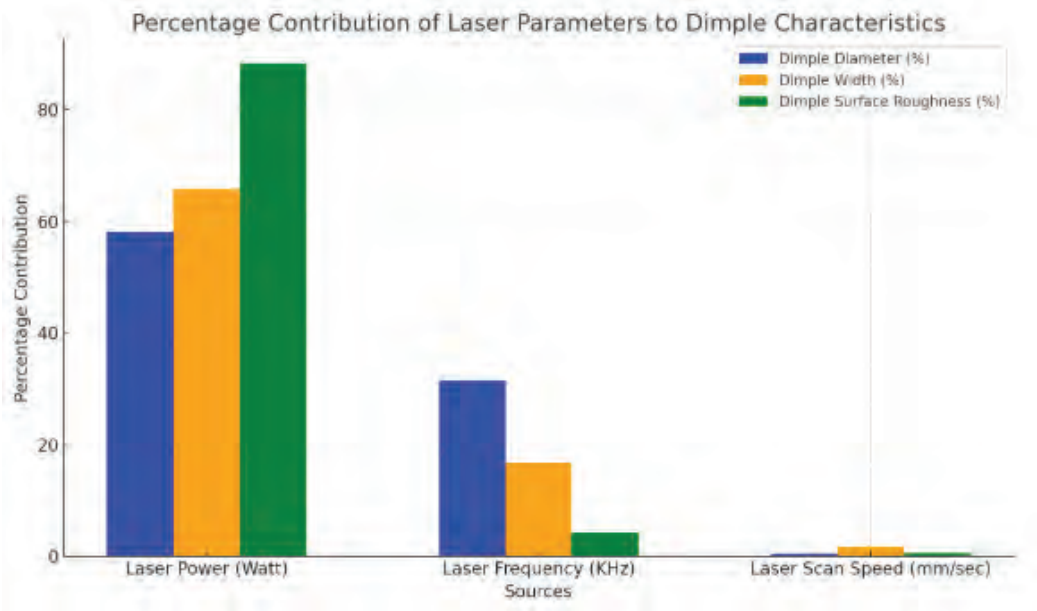
The outcome from the optimization of the laser surface texturing on cast iron piston rings, with regression equations analyzed, has a high degree of accuracy in the experiment and predicted values for dimple diameter, depth, and surface roughness. For the diameter of the dimple, with the optimum values of 16W, 25 KHz, and 140 mm/sec, the experimental value of 98.08 μm is almost the same as the calculated value of 98.57 μm, with an error of just 0.5%. This small error points to the accuracy of the regression equation to estimate the dimple diameter from the provided laser parameters. The depth of the dimple, with best parameters of 8W, 15 KHz, and 140 mm/sec, has an experimental result of 15.89 μm versus the predicted value of 16.24 μm, for a very small error of 2.2%. Although the deviation remains within acceptable limits, this implies that dimple depth might be less tolerant to variations in material properties or laser-material interaction during processing, which might cause the slight difference.

In the case of dimple surface roughness, the optimal values of 8W, 20 KHz, and 140 mm/sec gave an experimental result of 2.19 μm, whereas the calculated result was 2.24 μm, causing an error of 2.28%.

Surface roughness generally depends on smaller aspects of laser-material interaction and thus becomes prone to minute changes in experimental conditions. However, the low error rate shows that the regression model performs well in terms of predicting surface roughness with high accuracy.



A)



B)

Fig. 11A), Percentage Contribution of Laser parameters, B)ANOVA Interaction Plots of all parameters.

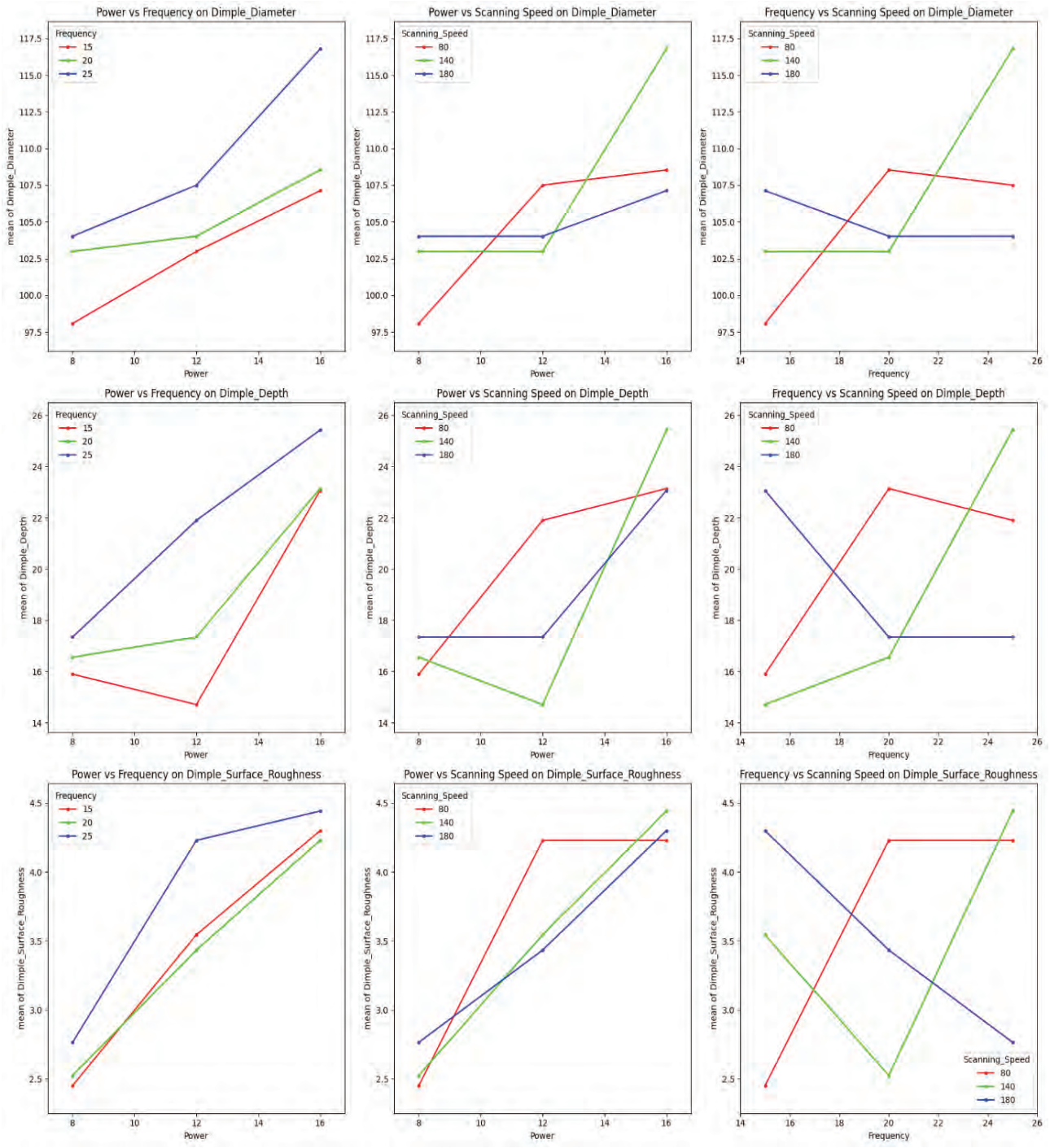


Figure 12. Pair plots of all parameters with each other.

Table 12. Confirmation Experiments for output responses

Responses	Optimum parameters	Experimental results	Predicted results by regression equations	Error (%)
Dimple diameter	16W, 25KHz and 140 mm/sec	98.08	98.57	0.5%
Dimple depth	8W, 15 KHz and 140 mm/sec	15.89	16.24	2.2%
Dimple roughness of the surface	8W, 20 KHz and 140mm/sec	2.19	2.24	2.28%

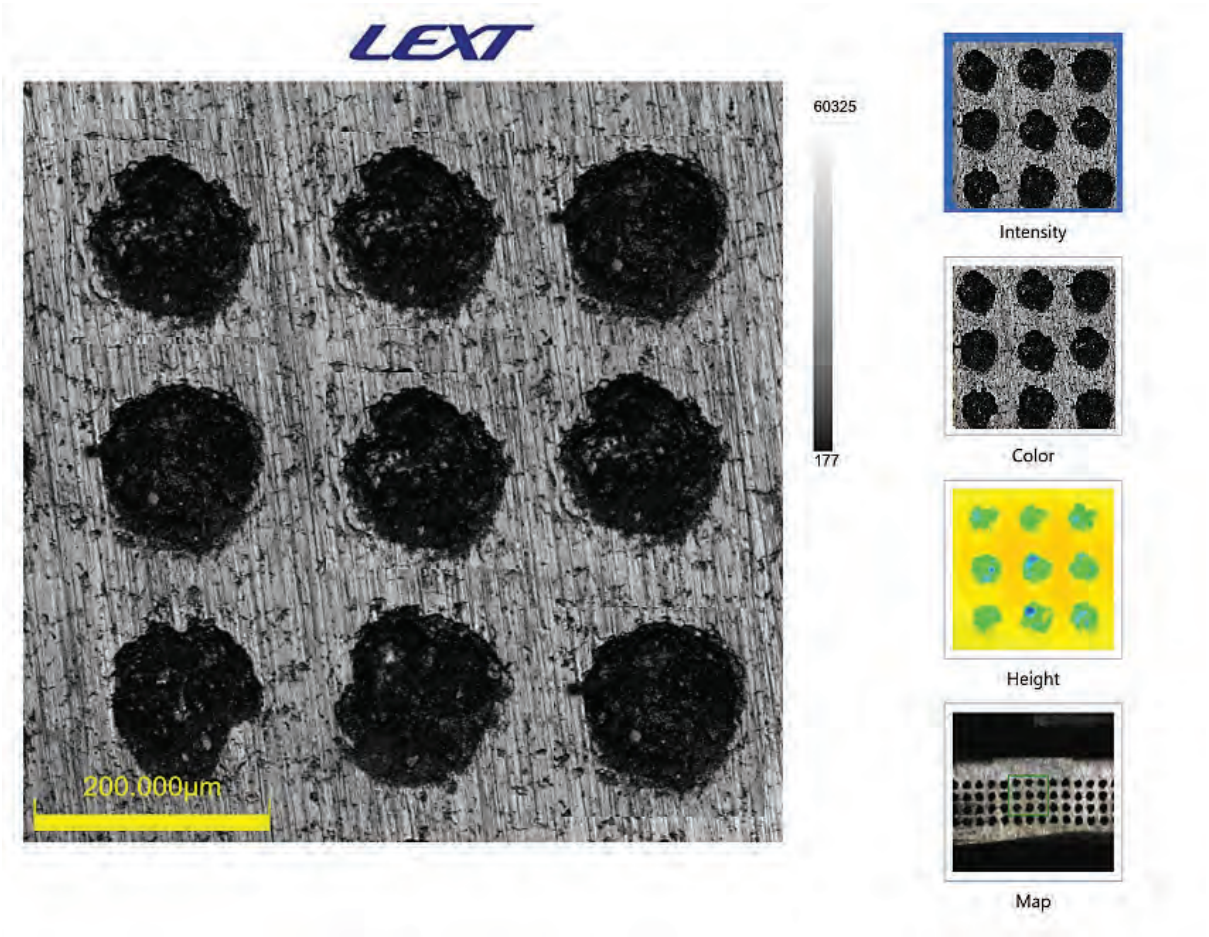


Figure 13. Laser textured cast iron piston ring sample images for 16W, 25 KHz and 140 mm/sec.

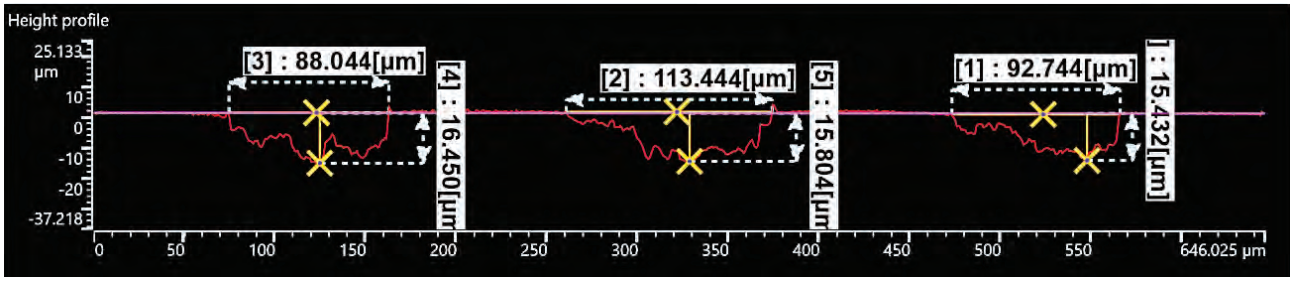


Figure 14. Dimple diameter and depth for textured sample with laser parameters 16W, 25 KHz and 140 mm/sec.

In addition to the quantitative analysis, Fig. 13 showcases the laser-textured cast iron piston ring sample images for the optimum parameters of 16W, 25KHz, and 140 mm/sec. The figure clearly illustrates the circular dimples, which are formed with high precision and accurate dimensions. The optimal parameters used ensure that the dimples are uniform, highlighting the effectiveness of the laser settings in achieving the desired texture.

Furthermore, Fig. 14 presents the dimple diameter and depth for the textured sample, using the same laser parameters (16W, 25 KHz, and 140 mm/sec). The figure confirms that these parameters result in consistent and well-defined dimple features, validating the accuracy of the regression model in predicting the dimple characteristics. Both figures support the conclusion that the optimized parameters produce high-quality surface texturing, with minimal error between experimental and predicted results.

Multi-objective Optimization Using the Gray Relational Grade (GRG) Rank Method

To perform multi-objective optimization using the Gray Relational Grade (GRG) rank method, we need to follow several steps:

1. **Normalization of data**: Normalize the data for each criterion to ensure that they are on a comparable scale.
2. **Calculation of gray relational coefficients (GRC)**: Determine the GRC for each experiment based on the normalized data.
3. **Calculation of gray relational grade (GRG)**: Compute the GRG by averaging the GRCs for each experiment.
4. **Ranking**: Rank the experiments based on the GRG.

Normalization of Data

The normalization of data can be done using the following formulas:

- For beneficial criteria (higher values are better):

$$Xi = \frac{Xi - Xmin}{Xmax - Xmin} \quad (3)$$

For non-beneficial criteria (lower values are better):

$$Xi = \frac{Xmin - Xi}{Xmin - Xmax} \quad (4)$$

Let's start by normalizing the data. For the given dataset, we will assume the following:

- Beneficial criteria: Dimple Diameter, Dimple Depth.
- Non-beneficial criteria: Dimple Roughness.

Calculation of Gray Relational Coefficients (GRC)

The GRC is calculated using the normalized data:

$$GRC = \frac{\Delta min + \xi \Delta max}{\Delta i + \xi \Delta max} \quad (5)$$

where Δi is the absolute difference between the reference sequence and the comparative sequence, Δmin and Δmax are the minimum and maximum values of Δi and ξ is the distinguishing coefficient (usually set to 0.5).

Calculation of Gray Relational Grade (GRG)

The GRG is the average of the GRCs for each experiment:

$$GRG = \frac{1}{n} \sum_{i=1}^{n} GRG_i \quad (6)$$

where n is the number of criteria.

Ranking

Rank the experiments based on the GRG.

The multi-objective optimization performed by the use of the Gray Relational Grade (GRG) process gives the overall ranking of the experiments in terms of the performance criteria of dimple diameter, dimple depth, and dimple roughness. The experiment with the highest GRG of 0.77777878, 16 watts of laser power, 25 KHz frequency and 140 mm/sec speed was ranked first in Experiment 9. The maximum dimple diameter (116.782 um) and the maximum dimple depth (25.429 um) were obtained in this experiment and the roughness was relatively high (4.442 um). Experiment 2 was second with a GRG of 0.570483 with the use of laser power of 8watts, frequency of 20 KHz and a speed of 140mm/sec which gave an average dimple size and minimum roughness of all the experiments. Experiment 1 has a GRG of 0.56441 close behind in third place, and with a laser power of 8 watts a frequency of 15 KHz and a speed of 80 mm/sec. The other experiments recorded different results with Experiment 4 recording the lowest GRG of 0.404606 and was positioned in position 9. The power of the laser used in this experiment was 12 watts, the frequency was 15 KHz, and the speed was 140 mm/sec, which yielded a moderate size of dimples but a comparatively high level of roughness. The rankings have shown that the greater the laser power and frequency combinations especially at moderate speed that the greater the overall performance in terms of the desired dimple characteristics.

Table 13. Results of Multi-objective optimization using the gray relational grade (GRG) rank method

Expt. No.	Laser power (Watt)	Frequency (KHz)	Speed (mm/sec)	Dimple diameter (μm)	Dimple depth (μm)	Dimple roughness (μm)	GRG	Rank
9	16	25	140	116.782	25.429	4.442	0.777778	1
2	8	20	140	102.993	16.555	2.525	0.570483	2
1	8	15	80	98.084	15.895	2.451	0.56441	3
8	16	20	80	108.541	23.144	4.23	0.530477	4
3	8	25	180	104.03	17.344	2.765	0.527303	5
7	16	15	180	107.13	23.069	4.3	0.512107	6
6	12	25	80	107.507	21.901	4.23	0.487958	7
5	12	20	180	104.03	17.344	3.434	0.441618	8
4	12	15	140	102.993	14.708	3.545	0.404606	9

Figure 15 shows the Experiment ranking by GRG method

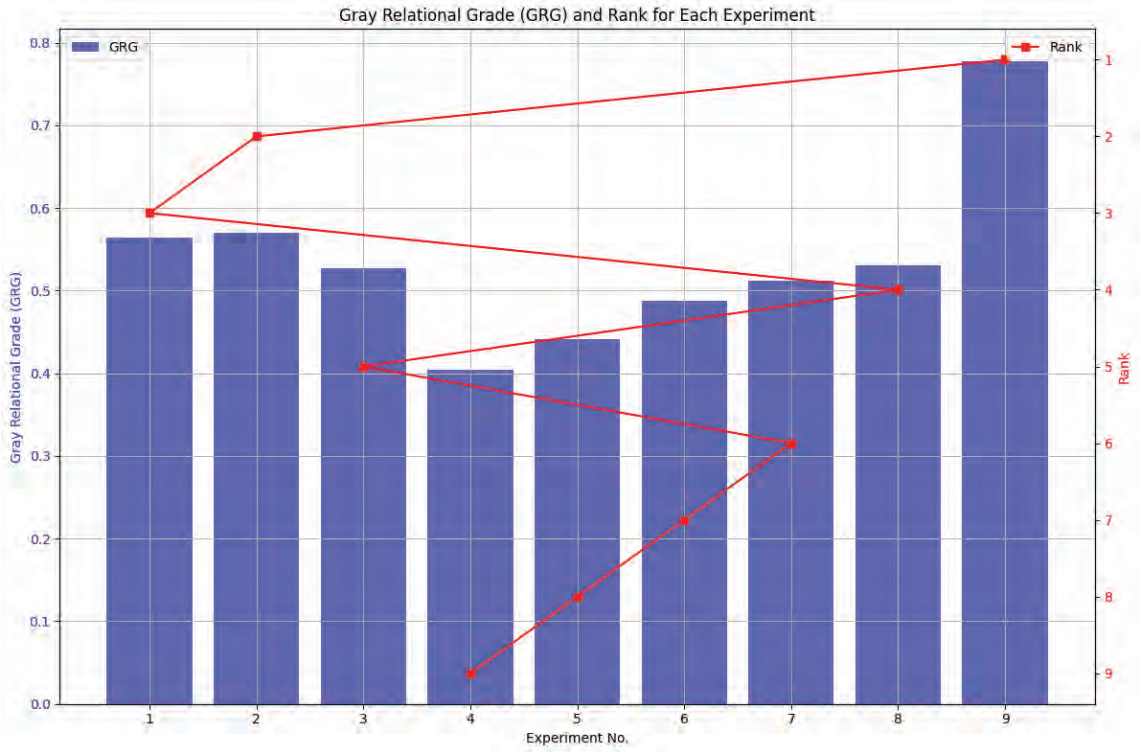


Figure 15. Experiment ranking by GRG method.

The results of the present research have the potential to be used in other materials other than cast iron. Here's how:

General rules of laser texturing

Laser surface texturing principles including laser power, frequency and scan speed optimization are very general to different materials. Parameters can be optimized in a way to adapt the study to other substrates by modifying the parameters to the material characteristics.

Material-specific adjustments

Although the experiment used cast iron, the process can be applied to other materials in that one should perform comparable experiments to find out the best laser parameters in the material. As an example, laser power or frequency may have to be varied to obtain the required dimple sizes and quality of the surface using different materials.

Test and validation

In an attempt to generalize the findings on other materials, one must conduct more tests and validation. This entails modification of the experimental apparatus to reflect the physical and thermal characteristics of the new material including thermal conductivity, melting point and absorption properties.

Scaling and transferability

Taguchi analysis and gray relational grade (GRG) are the optimization methods employed in the given study that can be applied to various materials. The systematic variation of the parameters and investigation of their influence on the characteristics of texture could be applied and modified to new materials.

Material-specific research

In order to be able to use it in practice, the material-specific research should be continued to know how the material reacts to the laser processing. The study would aid in narrowing down the parameters and seeing to it that the texturing of the surface is attaining the targeted results.

The laser surface texturing can be optimized and has a number of unlimited applications at the industrial level:

Improved tribological performance

Piston Rings in Internal combustion engines: Rough surfaces on the piston rings can have a huge impact on wear resistance and friction levels in the engine, which result in reduced fuel consumption, emissions, and engine life. Direct metal-to-metal contact can be reduced by laser surface texturing of piston rings and this increases their resistance to high pressures and temperatures. This application would have some advantages to the automotive industry, whereby it would enhance efficiency of the engines and minimize frequent maintenance.

Enhanced bonding and coating

Coatings in Automotive and Aerospace Components: Laser-textured piston rings or other engine parts can work

to enhance coating adhesion of other types of anti-corrosion or wear-resistance coating. These textured surfaces have superior mechanical interlocking of the coating to enhance their functionality and durability. This technique may be useful in the aerospace sector where parts have to be robust, such as high-speed friction and thermal loads, such that they are more likely to last longer, e.g., turbine blades and cylinder line.

Biomedical implants

Surface Texturing of medical devices: The principles of laser surface texturing that you are researching can be applied in biomedical implants, i.e., orthopedic joint replacement surgery or dental implantation. It is possible to develop micro or nano-scale textures on the surfaces of implants, which improves the mechanical connection of the implant with the bone, facilitating a faster and more successful process of osseointegration. This method has been especially beneficial in the medical sector where successful patient recovery and long-term success is reliant on the functionality of the implant.

Precision Engineering

High-precision components in electronics

Electronic and optical components Electronic and optical components can be made with textured surfaces at high precision, where the quality of the surface can make or break performance. The precision engineering of textured surfaces in precision optical lens design, semiconductor devices, or a micro-electromechanical system (MEMS) can be optimised in terms of functional performance, such as light reflection, flow control, and heat dissipation.

Functional surface design

Heat exchangers and fluid handling: In any industry where surface texture of parts such as heat exchanger fins or fluid handling systems is optimized this can improve heat dissipation or fluid friction. An example would be the texturing of the heat exchanger plate surfaces to improve thermal performance of the system by providing more area to the heat transfer resulting to more efficient cooling and use less energy to run the industrial processes.

Overall, the study offers a good methodology to texturize the surface of lasers, but its extension onto other materials needs to be experimented upon with a greater number of changes in parameters to suit the specific characteristics of materials.

To be able to compare trials and results, we undertook a comparative study with the available studies that have investigated such methodologies of laser surface texturing.

1. Wang et al. concentrated on the efficiency performance of axial piston motor with laser-textured valve plates. Their results suggested a 2.8-6.1 per cent all-around enhancement of efficiency in the presence of micro-dimples, with special focus on the impact of such a change on the fluid dynamics at pressure pulsations. Although our research focuses on cast iron piston rings, we find ourselves in agreement with Wang et al. in terms of the positive effects of micro-dimple scale in enhancing the mechanical performance, as laser texturing is capable of being a process of universal benefit to an array of applications [33].
2. In their study on the impact of lasers processing parameters on the texture dimensions in cast iron piston rings, Patil et al. met the target dimensions of 100 um in diameter, 20 um in depth, and 3 um in surface roughness in the target samples. Their findings showed that laser power and frequency played a significant role in determining texture characteristics, although our research agrees with this result by finding the best parameters of 116.782 um dimple diameter, 25.429 um depth, and 4.442 um roughness. The differences of the best results could be related to the differences in the experiment settings, and the parameter tuning in the special material use could be crucial [34].
3. Regarding the methodology, our study, as well as the research mentioned, had used the Taguchi method of optimizing the parameters of laser processing. This pattern of consistency justifies the usefulness of the Taguchi design to enhance the result of surface texturing of various materials and applications[7]. Moreover, our use of the Gray Relational Grade (GRG) technique gave us a useful ranking mechanism that helped us to make our analysis effective than conventional methods [35].
4. Our results also show that the laser power was the most impactful variable in all responses, and ANOVA analysis proves this argument. This observation is similar to the findings made by Patil et al. in the context of the importance of the laser parameters, and the idea that laser processing plays a critical role in the attainment of desirable surface properties.
5. The optimization of high-power fiber laser cutting was done by Lopez et al. [36] in the case of nuclear decommissioning. Their paper has discussed the efficiency and safety of the processes when using a high-power laser. Turkkan et al. [37] applied the multi-objective optimization of fiber laser cutting of stainless steel plates using Taguchi-based grey relational analysis. They enhanced the quality of cutting through optimization of several process parameters. Azaman et al. [38] did a test on the feasibility of harvesting oil palm with a pulse fiber laser system with the help of various lenses. Their study showed that there is a possibility of using the laser technology in the agricultural industry, which is harvesting. In the study, Kasman, S., Buyuker, B., and Ozan, S. (2023) are interested in machining of Ti-6Al-4V alloy through fiber laser technology and the impacts of various process parameters on the surface roughness. Their results show that, surface roughness tends to increase with a higher laser power, whereas surface quality with higher cutting speeds tends to be good. Optimal focal

position adjustment is also essential in getting required surface finishes. This paper contributes to the knowledge of laser machining parameters and can be effectively utilized in the optimization of surface quality to be applied in high performance. The contribution of the work is the development of laser machining technologies of Ti-6Al-4V alloys [39].

The comparative analysis highlights the applicability of using laser texturing of the surface as a material performance enhancement method. The observed differences and similarities between our results and those provided in the literature literature themselves indicate the need to conduct more research and optimization according to particular materials and applications [4]. The combination of the results of these investigations highlights the potential of laser texturing as a kind of transformation in many industrial applications, including auto-parts or even in a highly accurate engineering setting [11].

CONCLUSION

This research was able to examine the influence of the laser processing parameters in this case, power, frequency, and scan rate of the laser on the accuracy of texturing surfaces on cast iron piston rings. Optimal parameter combinations were determined using the Taguchi design method and multi-objective optimization based on gray relational grade (GRG) ranking to obtain the desired dimple dimensions, i.e. diameter, depth and surface roughness. Of the parameters, laser power was determined to have the most dramatic effect on the resultant dimple dimensions with frequency and scan speed also playing a role in the overall quality of the texture.

The results of the analysis demonstrated that the general tendency in the increases in laser power and frequency was larger dimple dimensions, whereas scan speed demonstrated a less general relationship with surface roughness. The ANOVA study offered a significant numerical measure of the significance of each parameter to the texture attributes, which can be useful in enhancing laser processing methods. Such conclusions demonstrate the usefulness of the selected techniques to attain accurate and reproducible surface textures, which play a significant role in promoting the tribological behavior of piston rings and other mechanical parts.

Although the research was done on cast iron, the methodologies are applicable to most materials to extrapolate the research to an industry setting. Future research such as the extended range of laser parameters, the combination of more sophisticated surface characterization methods, and the study of the long-term behavior under real-life conditions are also anticipated based on the research. In addition, laser texturing may be used together with other surface treatments to further increase the material characteristics and functionality.

Summarily, the findings of this research give a detailed methodology of maximizing the parameters of laser surface texturing, which have many implications in industry where high accuracy in surface features is necessary. This study will enhance the current research on the enhancement of advanced manufacturing processes to satisfy the requirements of the modern engineering applications by increasing the durability and efficiency of the elements, like piston rings.

ACKNOWLEDGEMENTS

The laser surface texturing facility provided by M/s. R. R. Engravers; Nashik is greatly appreciated by the authors. The authors further acknowledge the invaluable assistance that Mr. Gyanesh Singh and Mr. Vilas Bhangale from I. R. Technology Services Pvt. Ltd., Navi Mumbai, provided in the confocal microscopy characterization of the samples.

AUTHORSHIP CONTRIBUTIONS

AP: conceptualization, investigation, PB: resources, Methodology, investigation, visualization writing original and revised draft, US: revised draft, PK: supervision and project administration, VW: investigation, visualization writing original and revised draft. All authors have read and approved the manuscript.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Karamimoghadam M, Rezayat M, Moradi M, Mateo A, Casalino G. Laser Surface Transformation Hardening for Automotive Metals: Recent Progress. Metals 2024;14:339. [CrossRef]

[2] Moradi M, Moghadam MK, Shamsborhan M. How the laser beam energy distribution effect on laser surface transformation hardening process; Diode and Nd: YAG lasers. Optik 2020;204:163991. [CrossRef]
[3] Jiao J, Cheng X, Wang J, Sheng L, Zhang Y, Xu J, et al. A Review of Research Progress on Machining Carbon Fiber-Reinforced Composites with Lasers. Micromachines 2023;14:24. [CrossRef]
[4] Ding H, Wang Z, Guo Y. Multi-objective optimization of fiber laser cutting based on generalized regression neural network and non-dominated sorting genetic algorithm. Infrared Phys Technol 2020;108:103337. [CrossRef]
[5] Narayanan JA, Kaji F, Zimny M, Toyserkani E. Laser directed energy deposition-based additive manufacturing of Fe20Cr5.5AlY from single tracks to bulk structures: Statistical analysis, process optimization, and characterization. Machines 2023;11:58. [CrossRef]
[6] Mehrabi O, Seyedkashi SMH, Moradi M. Experimental and response surface study on additive manufacturing of functionally graded steel-inconel wall using direct laser metal deposition. Opt Laser Technol 2023;167:109707. [CrossRef]
[7] Moradi M, Mehrabi O, Azdast T, Benyounis KY. Enhancement of low power CO2 laser cutting process for injection molded polycarbonate. Opt Laser Technol 2017;96:208–218. [CrossRef]
[8] Baranowski M, Völger L, Friedmann M, Fleischer J. Experimental Analysis and Optimisation of a Novel Laser-Sintering Process for Additive Manufacturing of Continuous Carbon Fibre-Reinforced Polymer Parts. Appl Sci Switz 2023;13:5351. [CrossRef]
[9] Yao S, Chen SY, Pal A, Bremer K, Guan BO, Sun T, et al. Compact Tm-doped fibre laser pumped by a 1600 nm Er-doped fibre laser designed for environmental gas sensing. Sens Actuators Phys 2015;226:11-20. [CrossRef]
[10] Khan MS, Shahabad SI, Yavuz M, Duley WW, Biro E, Zhou Y. Numerical modelling and experimental validation of the effect of laser beam defocusing on process optimization during fiber laser welding of automotive press-hardened steels. J Manuf Process 2021;67:535–544. [CrossRef]
[11] Pan C, Liu W, Fan Z. Optical Fiber Laser Linewidth Optimization Based on Light Self-Injection Method. Laser Optoelectron Prog 2022;59:1914002. [CrossRef]
[12] Khozeymeh F, Poli F, Chowdhury W, Cucinotta A. Analysis and Optimization of Dysprosium-Doped Yellow Fiber Lasers for Ophthalmology Applications. IEEE Xplore 2024;16. [CrossRef]
[13] Lou C, Xie X. Multi-view intuitionistic fuzzy support vector machines with insensitive pinball loss for classification of noisy data. Neurocomputing 2023;549:126458. [CrossRef]
[14] Lapre C, Meng F, Hary M, Finot C, Genty G, Dudley JM. Genetic algorithm optimization of broadband operation in a noise-like pulse fiber laser. Sci Rep 2023;13. [CrossRef]
[15] Gopinath C, Lakshmanan P, Palani S. Fiber laser microcutting on duplex steel: parameter optimization by TOPSIS. Mater Manuf Process 2022;37:985–994. [CrossRef]
[16] Chen CT, Gu GX. Machine learning for composite materials. MRS Commun 2019;9:556–566. [CrossRef]
[17] Vora J, Chaudhari R, Patel C, Pimenov DY, Patel VK, Glasin K, et al. Experimental investigations and pareto optimization of fiber laser cutting process of Ti6AL4V. Metals 2021;11:1461. [CrossRef]
[18] Najjar IMR, Sadoun AM, Elaziz MA, Abdallah AW, Fathy A, Elsheikh AH. Predicting kerf quality characteristics in laser cutting of basalt fibers reinforced polymer composites using neural network and chimp optimization. Alex Eng J 2022;6:111005–11018. [CrossRef]
[19] Wu Q. Leafmap: A Python package for interactive mapping and geospatial analysis with minimal coding in a Jupyter environment. J Open Source Softw 2021;6. [CrossRef]
[20] Zeng C, Ghadimi H, Ding H, Nemati S, Garbie A, Raush J, et al. Microstructure Evolution of Al6061 Alloy Made by Additive Friction Stir Deposition. Materials 2022;15:3676. [CrossRef]
[21] Han Y, Guo Y, Gao B, Ma C, Zhang R, Zhang H. Generation, optimization, and application of ultrashort femtosecond pulse in mode-locked fiber lasers. Progress in Quantum Electronics 2020;71:100264. [CrossRef]
[22] Kokhanovskiy A, Kuprikov E, Bednyakova A, Popkov I, Smirnov S, Turitsyn S. Inverse design of mode-locked fiber laser by particle swarm optimization algorithm. Sci Rep 2021;11. [CrossRef]
[23] Yassin K, Hojjati M. Processing of thermoplastic matrix composites through automated fiber placement and tape laying methods: A review. J Thermoplas Compos Mater 2018;31. [CrossRef]
[24] Ma Q, Yu H. Artificial intelligence-enabled mode-locked fiber laser: A review. Nanomanuf Metrol 2023;6. [CrossRef]
[25] Wu S, Li H, Han C, Wu W. Effects of Full Chain Processes on the Performance of 316L Stainless Steel Composite by Fused Deposition Modeling and Sintering. Metals 2023;13. [CrossRef]
[26] Motard A, Louot C, Manek-Hönninger I, Dalloz N, Hildenbrand-Dhollande A. Optimizing the performance of a monolithic Tm 3+ , Ho 3+ -codoped fiber laser by FBG reflected wavelength and fiber gain matching. Opt Express 2023;31:18939–18948. [CrossRef]
[27] Kumar SA, Narayan YS. Tensile testing and evaluation of 3D-printed PLA specimens as per ASTM D638 type IV standard. Lect Notes Mech Eng 2019;79–95. [CrossRef]
[28] Mathew RS, O'Donnell R, Pizzey D, Hughes IG. The Raspberry Pi auto-aligner: Machine learning for automated alignment of laser beams. Rev Sci Instrum 2021;92:015117. [CrossRef]

[29] Reddy VC, Keerthi T, Nishkala T, Yadav GMP. Analysis and optimization of laser drilling process during machining of AISI 303 material using grey relational analysis approach. SN Appl Sci 2021;3. [CrossRef]

[30] Andrianov AV, Kalinin NA, Sorokin AA, Anashkina EA, Sanchez-Soto LL, Corney JF, et al. Optimizing the generation of polarization squeezed light in nonlinear optical fibers driven by femtosecond pulses. Opt Express 2023;31:765–773. [CrossRef]

[31] Czelusniak T, Amorim FL. Selective laser sintering of carbon fiber–reinforced PA12: Gaussian process modeling and stochastic optimization of process variables. Int J Adv Manuf Technol 2020;110:2049–2066. [CrossRef]

[32] Chandan GK, Sahoo CK. Development and analysis of scalable laser machined surface utilizing low power nanosecond laser on Ti-6Al-4V alloy. Mater Manuf Process 2022;38:589–597. [CrossRef]

[33] Wang Z, Gu L, Li L. Experimental studies on the overall efficiency performance of axial piston motor with a laser surface textured valve plate. Proc Inst Mech Eng B J Eng Manuf 2013;227. [CrossRef]

[34] Patil AS, Wakchaure VD, Shirsat UM. Effect of Fiber Laserprocess Parameters on Texture Dimensions Fabricated on Cast Iron Piston Ring Segments. Mater Sci Res India 2022;19. [CrossRef]

[35] Chen B, Zhao M, Liu X, Ye F, Fu HY, Li Q. Investigation of dissipative solitons in an Er-doped fiber laser through machine-learning online optimization based on the Gaussian process. J Opt Soc Am B 2022;39. [CrossRef]

[36] Lopez AB, Assunção E, Quintino L, Blackburn J, Khan A. High-power fiber laser cutting parameter optimization for nuclear Decommissioning. Nucl Eng Technol 2017;49:865–872. [CrossRef]

[37] Turkkan YA, Aslan M, Tarkan A, Aslan O, Yuce C, Yavuz N. Multi-Objective Optimization of Fiber Laser Cutting of Stainless-Steel Plates Using Taguchi-Based Grey Relational Analysis. Metals 2023;13:132. [CrossRef]

[38] Azaman MIH, Ramly AS, Mdradzi MKF, Ahmad MR, Khalid MRM, Bakri MAM, et al. Feasibility study of oil palm harvesting using pulse fibre laser system with different lenses. J Oil Palm Res 2022;34. [CrossRef]

[39] Kasman S, Buyuker B, Ozan S. Machining of Ti–6Al–4V alloy by fiber laser: Determining the effects of parameters on surface roughness. Sigma J Eng Nat Sci 2023;41. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2014

Research Article

Exploring fractional dynamics of fourth order parabolic partial differential equations with power law Kernel

Shelly ARORA[1], Wenxiu MA[2,3,4,5,*], Sukhjit SINGH DHALIWAL[6], Atul PASRIJA[1]

[1]*Department of Mathematics, Punjabi University, Patiala 147002, India*
[2]*Department of Mathematics, Zhejiang Normal University, Jinhua 321004, Zhejiang, China*
[3]*Research Center of Astrophysics and Cosmology, Khazar University, Baku 1096, Azerbaijan*
[4]*Department of Mathematics and Statistics, University of South Florida, Tampa, 33620-5700, USA*
[5]*Material Science Innovation and Modelling, Department of Mathematical Sciences, North-West University, Mmabatho, 2735, South Africa*
[6]*Department of Mathematics, SLIET, Longowal, Sangrur, Punjab, 148106, India*

ARTICLE INFO

Article history
Received: 20 November 2024
Revised: 02 January 2025
Accepted: 28 January 2025

Keywords:
Fourth-Order Parabolic Equations; GB Transform; Homotopy Perturbation Method; Liouville-Caputo Fractional Derivative

ABSTRACT

The aim of this study is to establish an efficient method for exploring the fractional dynamics of fourth-order parabolic partial differential equations. The generalized bivariate homotopy perturbation method is suggested to determine expansion solutions for the considered models. This scheme decomposes non-linearity using He's polynomials. It is applied within the non-local fractional framework of Liouville-Caputo sense. The impact of fractional phenomena is briefly described. A comparative analysis of the suggested scheme is presented through graphical and tabular illustrations. Error analysis, based on the evaluation of absolute error, demonstrates the high precision, authenticity, and superiority of the suggested scheme. The existence and uniqueness of continuous solutions are shown for the fractional variants of Cauchy problems. The semi-analytic results obtained from the suggested scheme effectively capture the wave dynamics of physical systems across various physical parameters. The proposed algorithm does not rely on the availability of an exact solution to demonstrate its efficacy.

Cite this article as: Arora S, Ma W, Singh Dhaliwal S, Pasrija A. Exploring fractional dynamics of fourth order parabolic partial differential equations with power law Kernel. Sigma J Eng Nat Sci 2026;44(2):820–842.

INTRODUCTION

The dynamical behavior of fourth-order parabolic equations has attracted global interest. These equations can be framed by non-linear functional equations defined by:

$$w = g + \mathrm{M}(w, w_{\eta}, w_{\xi}, w_{\xi\xi}, w_{\xi,\xi,\xi}, w_{\xi\xi\xi\xi})$$

where g is the known function and M is the non-linear operator.

*Corresponding author.
*E-mail address: mawx@cas.usf.edu, atulpasrija_rs@pbi.ac.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

The first-order derivative w_ξ reflects the advective transport. The term $w_{\xi\xi}$ accounts for instability at large scales. The inclusion of the third-order term $w_{\xi\xi\xi}$ yields dispersion effect. The fourth-order term $w_{\xi\xi\xi\xi}$ provides exponential damping out of very sharp spikes at small scales, and the non-linear term stabilizes the system dynamics by transferring energy between large and small scales.

In the past few decades, theoretical study and practical simulation have been addressed via fractional operators [1-4]. These operators have been found to be better predictors than integer-order models. Fractional operators have emerged as a non-local tool for providing an excellent description of a real evolution process with memory and hereditary properties. It naturally arises in multiple conducts of wave phenomena, exhibiting chaos in wave motions, solitons, wrinkled flame front propagation, phase turbulence in reaction-diffusion systems, and chaotic drifting waves induced by photon collision [5-10]. However, the inclusion of fractional derivatives with variable and non-variable order demands highly influential analytic tools. For numerical treatment, it requires a large number of variables to cope with reality. The implementation of fractional operators can be found in a wide range of fields such as electromagnetism, bioengineering, signal processing, optimal control theory, and finance [7-14].

There are innumerable alternative definitions of fractional operators available in the literature [11-15]. Although the Liouville-Caputo (L-C) derivative and the Riemann-Liouville (R-L) derivative are the most frequently used terminologies. With the definition of Caputo fractional derivative, a fractional derivative would be applicable for differentiable functions only. To deal with non-differentiable functions, R-L derivative is preferred, but it denies being zero for constant functions. To address non-local and non-singular kernel properties, the Caputo-Fabrizio (C-F) and Atangana Baleanu in Caputo sense (ABC) serve as extensions of the L-C operator. However, there are certain drawbacks when employing a non-singular fractional differential operator. The inability of a fractional differential operator with a non-singular kernel to satisfy the fundamental theorem of fractional calculus has been documented by Diethelm [16]. Another notion of local fractional derivatives extends the familiar limit definition of the classical derivative [17,18]. Because of their simplicity, the local fractional derivatives can meet a large number of well-known aspects of classical derivatives, such as the chain rule, quotient rule, integration by parts, Gronwall's inequality, fractional series representation of functions, and so on. However, the inclusion of the memory effect falls outside the scope of local fractional derivatives.

In present study, renowned models of fourth order partial differential equations featured with Caputo derivative are considered as:

- Fractional generalized Kuramoto-Sivashinsky (FGKS) equation:

$$C_\eta^\alpha w + ww_\xi + xw_{\xi\xi} + yw_{\xi\xi\xi} + zw_{\xi\xi\xi\xi} = 0. \quad (1)$$

- Fractional Swift-Hohenberg (FSH) equation without dispersion:

$$C_\eta^\alpha w + w_{\xi\xi\xi\xi} + 2w_{\xi\xi} + (1-x)w + w^3 = 0, 0 < x \leq 1. \quad (2)$$

- Fractional Swift-Hohenberg equation (FSH) with dispersion:

$$C_\eta^\alpha w + w_{\xi\xi\xi\xi} + 2w_{\xi\xi} - xw_{\xi\xi\xi} - yw - 2w^2 + w^3 = 0. \quad (3)$$

- Fractional Cahn-Hilliard (FCH) equation:

$$C_\eta^\alpha w = xw_\xi + \left(-w_{\xi\xi} - w + w^3\right)_{\xi\xi}, \quad (4)$$

or

$$C_\eta^\alpha w = xw_\xi + 6w\left(w_\xi\right)^2 + 3w^2w_{\xi\xi} - w_{\xi\xi} - w_{\xi\xi\xi\xi}, \quad (5)$$

where x and y are real constantss.

For α = 1, Equation 1 represents the classical generalized Kuramoto-Sivashinsky (GKS) equation [19]. The variants of GKS equation have been extensively explored with different initial and boundary data by numerous analytic and numerical tools, such as homotopy perturbation method [20] and homotopy analysis method, which have been projected to treat space fractional GKS equation [21], residual power series method has been implemented with the projection of Riemann-Liouville integral and Caputo's fractional framework [22], the amalgamation of the integral transform with variational iteration method [23] and q-HAM [24] are used to construct semi-analytic expansion solutions. The *G'/G* expansion method [25,26], exp-function method [27], modified Kudryashov method [28], tanh method and its extended version [29,30] are employed to interpret various soliton conductance. The non-integrability of K-S equation has been established via Painleve analysis [31]. The rich characterization of several dynamics of K-S model, such as multimodal, cellular, oscillatory, and chaotic solutions, have been studied so far [32-35].

In 1977, Jack Swift and Pierre Hohenberg have introduced and nurtured Swift-Hohenberg (S-H) equation as a universal model for the temperature and fluid velocity dynamics of convection [36]. S-H equation intrinsically occurs in nano-crystalline materials for expressing the average density of detachment during creation of shear microbands [37]. S-H equation addresses numerous issues such as patterns in Rayleigh-Bernard convections, patterns within thin vibrated granular layers, defect dynamics, spatial-temporal chaos and study of lasers [38-43]. S-H equation demonstrates the mechanism of amplitude of optical electrical field within the cavity embodying a non-linear medium [44]. Application of S-H equation rangs in the theory of pattern formation in fluid layers confined between horizontal well-conducting boundaries [45] Several researchers have conducted analytic and numerical experiments to unveil the dynamics of S-H equation. [46]

provides evidence for the existence of quasi-patterns for the S-H equation. With inclusion of dispersive effect, in [47] authors have reported the achievement of spatial reversibility. In [48], investigators have proved the existence of non-stationary meromorphic solutions of S-H equation. Consequently, they have derived the simple periodic and elliptic non-stationary travelling wave solutions of S-H equation. Semi-analytic techniques [49-52] are eminently aided to interpret the behavior of S-H equation.

The Cahn-Hilliard (C-H) equation is a non-linear, stiff, fourth-order parabolic equation. It shows how the conserved field has changed over time. In essence, the C-H equation has been created to simulate the spinodal breakdown of a binary system at a specific temperature. It records both a very gradual coarsening, which results in bulk phases with an interface separating them, and a quick phase separation, which creates a thin interface between two phases. The C-H equation has been solved by many researchers using various time discretization and spatial techniques. The C-H equation has been examined in 2D form by Grasselli et al. [53], who have established a number of findings about the well-posedness and large-time behavior of solutions. Analysis has been done in the presence of bounded absorbing sets in two distinct phase spaces. Additionally, the existence of exponential and global attractors has been proven. The proposed C-H equation has been solved by Lee et al. [54] utilizing the pseudo-spectral, finite element, and finite difference methods. Together with first- and second-order temporal algorithms, Chen et al. [55] have suggested a finite difference technique for the C-H system with a logarithmic Flory Huggins energy potential. For the C-H and Allen-Cahn equations, Zhang et al. [56] have suggested the optimal control issue. For optimum control issues, the optimality requirements have been applied. The C-H Biot model has been suggested by Storvik et al. [57] to describe flow through deformable porous media. Biot's theory has controlled the coupling between flow and deformation, resulting in a three-way coupled system. In [58], Manzenaro et al. have addressed a C-H model with wall boundary conditions. Using a curvilinear hexahedral model, the model equations have been discretized by the discontinuous Galerkin spectral element approach. The first-order implicit-explicit technique, which separated the system into two C-H equations, has been used to explore the discretization of time. For the space-time fractional order C-H equation, Khan et al. [59] have presented the traveling wave solution using the auxiliary approach. For complex dynamics, Zhou and Xie [60] have presented a multicomponent three-dimensional C-H equation. Using the projected gradient approach, lagrangian multiples of Gibbs *n* simplex phase constraints have been computed.

In its versatile framework, the present study incorporates potential models (FGKS, FSH and FCH) of thermal science, material science, biological science, hydrodynamics, study of lasers, oceanography and many more. The blend of these subjects forms the basis of vast studies in engineering and natural sciences. Therefore, the choice of FGKS, FSH and FCH equations will engage the interest of a large part of the scientific community. Despite the numerous semi-analytic studies for L-C fractional GKS, S-H and C-H equations, the establishment of reliability of solutions for $0 < \alpha < 1$ is extensively unexplored. While addressing this gap through proposing the first-time application of the generalized bivariate homotopy perturbation method (GB-HPM) for fourth-order equations, the present study demonstrates the competence of GB-HPM in terms of accuracy.

The manuscript is organized in the given manner. Section 1 is based on the introduction of the theory and applications of the considered problems. In Section 2, definitions, properties and terminologies of L-C derivative and generalized bivariate (GB) transform are presented. In Section 3, existence and uniqueness of a continuous solution are shown for time fractional K-S, S-H and C-H equations. In Section 4, solution procedure of the proposed algorithms is explained for general non-linear initial value problems. In Section 5, the application of theory developed in the above sections has been presented through several test problems. These are analyzed via comparative and error analysis to test efficacy and robustness of the proposed scheme. In the last section the concluding remarks are presented based on the above study.

Preliminary Results

This section covers the definitions, properties, and terminologies of L-C derivative and GB transform, which are discussed in further detail throughout the paper.

Definition 1. [15] The R-L fractional integral of order α is conceptualized as

$$I_{\eta}^{\alpha} w(\eta) = \begin{cases} \frac{1}{\Gamma(\alpha)} \int_0^{\eta} (\eta - \tau)^{\alpha - 1} w(\eta) d\tau; & \alpha > 0, \eta > 0, \\ w(\eta); & \alpha = 0. \end{cases} \quad (6)$$

Definition 2. [15] For a given function $w(\eta)$: $[0,\infty) \to R$, the L-C fractional derivative of w of order $n\text{-}1 < \alpha \leq n$ is defined by

$$C_{\eta}^{\alpha} w(\eta) = \frac{\partial^{\alpha} w}{\partial \eta^{\alpha}} = \begin{cases} \frac{1}{\Gamma(n-\alpha)} \int_0^{\eta} (\eta - \zeta)^{n-\alpha-1} \frac{\partial^n}{\partial \eta^n} w(\zeta) \mathrm{d}\zeta; & n-1 < \alpha < n, \\ \frac{\partial^n}{\partial \eta^n} w(\eta); & \alpha = n \in N. \end{cases} \quad (7)$$

Lemma 1. [15] For any $n\text{-}1 < \alpha \leq n$, $n \in N$. Then one hass

$$I_{\eta}^{\alpha} C_{\eta}^{\alpha} w(\eta) = w(\eta) - \sum_{k=0}^{n-1} w^{(k)}(0) \frac{\eta^k}{k!}. \quad (8)$$

In the subsequent development of fractional calculus, the operational features of integral transforms are extensively explored to construct solutions to fractional models [4,49,50,61]. However, each existing integral transform pursues some constraints, which stimulates interest in developing modified integral transforms.

In this series of developments, a large amount of research work has been devoted to formulate several novel integral

transforms [62-70]. In 2023, Arora et al. [70] proposed GB transform as the generalization of Laplace, Shehu, ARA and Formable transforms [63,64,68]. Mathematically, the GB transform of order m and its inverse can be defined as [70]:

$$\mathbb{A}_m[w(\eta)] = \mathcal{V}_m(s,\gamma) = \frac{s}{\gamma}\int_0^\infty \eta^{m-1} e^{\frac{-s}{\gamma}\eta} w(\eta)\, d\eta, \tag{9}$$

and

$$\mathbb{A}_m^{-1}[\mathcal{V}_m(s,\gamma)] = w(\eta) = \frac{(-1)^{2(m-1)}}{2\pi i}\int_{\delta-i\infty}^{\delta+i\infty} \frac{1}{\gamma} e^{\frac{s}{\gamma}\eta} \mathbb{S}(s,\gamma)\, ds, \tag{10}$$

over the set of functions

$$\mathcal{S} = \{w(\eta): \exists\, B, b_j > 0, |w(\eta)| < Be^{\frac{|\eta|}{b_j}},$$
$$i!\, f\ \eta \in (-1)^j \times [0,\infty)\ for\ j = 1,2\},$$

where $s > \gamma > 0$, δ is real constant and $\mathbb{S}$ is the Shehu transform [68].

Proposition 1. [70] GB transform holds the fundamental properties of integral transforms such as:

- Linearity Property

$$\mathbb{A}_m[\mu u(\eta) + \lambda w(\eta)] = \mu \mathbb{A}_m[u(\eta)] + \lambda \mathbb{A}_m[w(\eta)]. \tag{11}$$

- Convolution Property

$$\mathbb{A}_1[(u * w)(\eta)] = \frac{s}{\gamma}\mathbb{A}_1[u(\eta)].\mathbb{A}_1[w(\eta)]. \tag{12}$$

- Derivative Property

$$\mathbb{A}_m[w^n(\eta)] = (-1)^{m-1}\frac{s}{\gamma}\frac{\partial^{m-1}}{\partial s^{m-1}}\left[\left(\frac{s}{\gamma}\right)^{n-1}\mathbb{A}_1[w(\eta)] - \sum_{k=0}^{n-1}\left(\frac{s}{\gamma}\right)^{n-(k+1)} w^k(0)\right], \tag{13}$$

where μ and λ are arbitrary constants.

Theorem 1. [71] Formulation for application of GB transform over L-C derivative is given by

$$\mathbb{A}_m\left[{}_0^C C_\eta^\alpha(w(\eta))\right] = \frac{1}{\Gamma(n-\alpha)}\sum_{r=1}^m \binom{m-1}{r-1}\Gamma(n+m-\alpha-r)\frac{\gamma^{n-\alpha}}{s^{n+m-\alpha-r}}\mathbb{A}_r[w^n(\eta)], \tag{14}$$

where $n\text{-}1 < \alpha \le n$

Existence and Uniqueness

The present section illustrates the existence and uniqueness of continuous solutions for the considered fractional equations.

The FGKS reads as:

$$C_\eta^\alpha w = -ww_\xi - xw_{\xi\xi} - yw_{\xi\xi\xi} - zw_{\xi\xi\xi\xi\xi}. \tag{15}$$

Employing the fractional integral operator given in Equation 6 on Equation 15, gives

$$w(\xi,\eta) - w(\xi,0) = I_\eta^\alpha\left(-ww_\xi - xw_{\xi\xi} - yw_{\xi\xi\xi} - zw_{\xi\xi\xi\xi\xi}\right),$$

or explicitly

$$w(\xi,\eta) - w(\xi,0) = \frac{1}{\Gamma(\alpha)}\int_0^\eta (\eta-\tau)^{\alpha-1}\left(-ww_\xi - xw_{\xi\xi} - yw_{\xi\xi\xi} - zw_{\xi\xi\xi\xi\xi}\right)d\tau. \tag{16}$$

Let

$$N(\xi,\eta,w,x,y,z) := -ww_\xi - xw_{\xi\xi} - yw_{\xi\xi\xi} - zw_{\xi\xi\xi\xi\xi}.$$

In the same manner one can define

- For FSH equation:

$$N(\xi,\eta,w,x,y,z) := w_{\xi\xi\xi\xi} - 2w_{\xi\xi} + xw_{\xi\xi\xi} + yw + 2w^2 - w^3. \tag{17}$$

- For FCH equation:

$$N(\xi,\eta,w,x,y,z) := ww_\xi + \left(-w_{\xi\xi} - w + w^3\right)_{\xi\xi}. \tag{18}$$

Lemma 2. [72] The function N satisfies the Lipschitz condition.

Proof: Let w and w^* be two bounded functions. Using definition of N and triangular inequality gives

$$\begin{aligned} &\| N(\xi,\eta,w,x,y,z) - N(\xi,\eta,w^*,x,y,z)\\ &\|\le \| ww_\xi - w^*w_\xi^* \| + x \| (w-w^*)_{\xi\xi}\\ &\| + y \| (w-w^*)_{\xi\xi\xi} \| + z \| (w-w^*)_{\xi\xi\xi\xi},\\ &= \frac{1}{2}\| (w^2-w^{*2})_\xi \| + x \| (w-w^*)_{\xi\xi}\\ &\| + y \| (w-w^*)_{\xi\xi\xi} \| + z \| (w-w^*)_{\xi\xi\xi\xi} \|. \end{aligned}$$

Now suppose that $\exists$ constants $\kappa_1, \kappa_2 > 0$ such that $\forall(\xi,\eta)$ $\|w\| \le \kappa_1$ and $\|w^*\| \le \kappa_2$. $\kappa = \max\{\kappa_1, \kappa_2\}$.

Then their first derivative function $(.)_\xi$ fulfills Lipschitz condition and $\exists$ a non-negative number L_1 such that

$$\begin{aligned} &\le \frac{1}{2}L_1 \| (w^2-w^{*2}) \| + xL_1^2 \| (w-w^*) \| + yL_1^3\\ &\| (w-w^*)_{\xi\xi\xi} \| + zL_1^4 \| (w-w^*) \|,\\ &\backslash\\ &\le (\kappa L_1 + xL_1^2 + yL_1^3 + zL_1^4) \| w-w^* \|, \end{aligned} \tag{19}$$

where $\| w^2\text{-}w^{*2}\| < 2\kappa \| w\text{-}w^*\|$. Hence,

$$\| N(\xi,\eta,w,x,y,z) - N(\xi,\eta,w^*,x,y,z) \| \le L \| w\text{-}w^* \|,$$

where

$$L := \kappa L_1 + xL_1^2 + yL_1^3 + zL_1^4.$$

In the same manner one can deduce

- For FSH equation

$$L := L_1^4 + 2L_1^2 + xL_1^3 + y + 4\kappa + 4\kappa^2. \tag{20}$$

- For FCH equation

$$L := xL_1 + L_1^4 + L_1^2 + 4\kappa^2 L_1^2, \tag{21}$$

where $\| w^3\text{-}w^{*3} \| \le 4\kappa^2 \| w\text{-}w^* \|$.

Theorem 2. [72] The FGKS, FSH and FCH equations admit a unique continuous solution provided that $\frac{L}{\Gamma(\alpha+1)}\eta^{\alpha} < 1$.

Proof: The expressions 16, 17 and 18 can be expressed in a unified form as

$$w(\xi,\eta) - w(\xi,0) = \frac{1}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1} N(\xi,\eta,w,x,y,z)d\tau, \quad (22)$$

which recommends the consideration of recurrence relation as

- $w_0(\xi,\eta) = w(\xi,0)$,
- $w_{\mathrm{k}}(\xi,\eta) = \frac{1}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha}\,\mathrm{N}(\xi,\eta,w_{k-1},\mathrm{x},\mathrm{y},\mathrm{z})\mathrm{d}\tau$.

Let

$$\overline{w}(\xi,\eta) = \lim_{k\to\infty} w_k(\xi,\eta)\,. \quad (23)$$

Now the aim is to prove that $\overline{w}(\xi,\eta) = w(\xi,\eta)$ is a continuous solution. For that end, set the algebraic difference of successive terms as

$$\Theta_k(\xi,\eta) = w_k(\xi,\eta) - w_{k-1}(\xi,\eta).$$

Note that

$$w_k(\xi,\eta) = \sum_{i=0}^{k}\Theta_i(\xi,\eta).$$

It is easy to observe that

$$\|\Theta_k(\xi,\eta)\| = \|w_k(\xi,\eta) - w_{k-1}\| \le \frac{1}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1}$$
$$\|N(\xi,\eta,w_{k-1},x,y,z)N(\xi,\eta,w_{k-2},x,y,z)\|d\tau.$$

Lipschitz condition for N yields

$$\|\Theta_k(\xi,\eta)\| \le \frac{1}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1}\|w_{k-1}-w_{k-2}\|d\tau,$$
$$\le \frac{L}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1}\|\Theta_{k-1}(\xi,\eta)\|d\tau. \quad (24)$$

Following the recurrence procedure given in Equation 24 deduce

$$\|\Theta_k(\xi,\eta)\| \le \left(\frac{\eta^{\alpha}L}{\Gamma(\alpha+1)}\right)^{k}\|w(\xi,0)\|,$$

which implies the existence and continuity of the solution.

Now to prove that $w(\xi,\eta) = \lim_{k\to\infty} w_k(\xi,\eta)$ is the solution of the Cauchy problems defined as FKS, FSH and FCH equations. Let

$$\Psi_k(\xi,\eta) = \overline{w} - w_k(\xi,\eta),\ k \in N.$$

Thus,

$$\overline{w} - w_{k+1} = \frac{1}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1}[N(\xi,\eta,\overline{w},x,y,z)$$
$$- N(\xi,\eta,w_k,x,y,z)]d\tau.$$

Lipschitz condition for N yields

$$\|\overline{w}(\xi,\eta) - w_{k+1}\| \le \frac{L}{\Gamma(\alpha)}\int_0^{\eta}(\eta-\tau)^{\alpha-1}\|\overline{w}-w_k\|d\tau,$$
$$= \frac{L\eta^{\alpha}}{\Gamma(\alpha+1)}\|\Psi_k(\xi,\eta)\|. \quad (25)$$

Evaluating the limit $k \to \infty$ for Equation 25 yields $\Psi_k \to 0$ and left-hand side gives

$$\lim_{k\to\infty} w_k = \overline{w}.$$

The above result implies the certainty of taking $\overline{w}(\xi,\eta) = w(\xi,\eta)$ as a solution which is continuous. Finally for uniqueness consider w and $w^{\star}$ be two distinct solutions for FKS, FSH and FCH equations then Equation 24 and Lipschitz condition of N implies

$$\|w - w^{*}\| \le \frac{L\eta^{\alpha}}{\Gamma(\alpha+1)}\|w - w^{*}\|,$$

or equivalently

$$\|w - w^{*}\|\left(1 - \frac{L\eta^{\alpha}}{\Gamma(\alpha+1)}\right) \le 0.$$

Hence, $w = w^{\star}$ if $\left(\frac{L\eta^{\alpha}}{\Gamma(\alpha+1)}\right) < 1$. The theorem is proved.

Application of Generalized Bivariate Homotopy Perturbation Method

In this section, the fundamental idea behind the solution procedure of the proposed algorithms is elucidated. The general form of the fractional non-linear partial differential equation can be considered as:

$$C_{\eta}^{\alpha}w(\xi,\eta) = \mathcal{P}[w(\xi,\eta)] + \mathcal{Q}[w(\xi,\eta)]$$
$$+ \mathcal{N}(\xi,\eta); \xi \in \Omega \subset R^{n}, \eta \in [0,T], \quad (26)$$

subject to the initial condition

$$w(\xi,0) = h(\xi), \quad (27)$$

where w is the function of ξ and η. Ω is a bounded domain and T is finite time. $\mathcal{P}$ and $\mathcal{Q}$ are linear and non-linear operators, respectively, and $N(\xi,\eta)$ is the non-homogeneous term.

Apply A_1 on both sides of Equation 26 and use initial condition given in Equation 27, such that Equation 26 takes the following form:

$$\mathbb{A}_1[w(\xi,\eta)] = h(\xi) + \left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1[\mathcal{N}(\xi,\eta)]$$
$$+ \left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1\{\mathcal{P}[w(\xi,\eta)] + \mathcal{Q}[w(\xi,\eta)]\}. \quad (28)$$

Operating $\mathbb{A}_1^{-1}$ to Equation 28, yields

$$w(\xi,\eta) = h(\xi) + \mathbb{A}_1^{-1}\left\{\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1[\mathcal{N}(\xi,\eta)]\right\} + \mathbb{A}_1^{-1}\left\{\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1\{\mathcal{P}[w(\xi,\eta)] + \mathcal{Q}[w(\xi,\eta)]\}\right\}, \quad (29)$$
$$= \mathcal{G}(\xi,\eta) + \mathbb{A}_1^{-1}\left\{\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1\{\mathcal{P}[w(\xi,\eta)] + \mathcal{Q}[w(\xi,\eta)]\}\right\},$$

which is a recurrence relation for Equation 26, where $\mathcal{G}(\xi,\eta)$ denotes the term that arises from the non-homogeneous term and initial condition.

Here, it is worth noting that for existing integral transforms $\mathcal{I}[f.g] \neq \mathcal{I}[f].\mathcal{I}[g]$ where, $\mathcal{I}$ is an integral transform. Therefore, these integral transforms are insufficient for handling a large class of differential and integral equations due to the restrictions arising from non-linear terms. In the present scheme, GB transform is consolidated with homotopy perturbation method (HPM) [73] to overlap this difficulty.

According to the expansion methodology, which is followed by semi-analytic tools, let the solution be represented by the following infinite series:

$$w(\xi,\eta) = \sum_{k=0}^{\infty}\theta^k w_k(\xi,\eta). \quad (30)$$

Homotopy is a continuous process that associates a family of objects, which represent intermediatory stages of deformation of an object to another object. HPM is a semi-analytic strategy that relies on the principles of homotopy and perturbation methods. Application of HPM redistributes the existing non-linearities with the aid of homotopy parameter i.e., θ. The smooth tracing of homotopy parameter θ in [0,1] is equivalent to the variance in the strength of inherited non-linearity. This enhances to deform an accessible solution into the desired one.

To consolidate the limitations of GB transform for non-linear events, HPM offers the decomposition of non-linear terms such as:

$$\mathcal{Q}[w(\xi,\eta)] = \sum_{k=0}^{\infty}\theta^k\mathcal{H}_k = \mathcal{H}_0 + \theta\mathcal{H}_1 + \theta^2\mathcal{H}_2 + \theta^3\mathcal{H}_3 + \cdots, \quad (31)$$

where $\mathcal{H}_{\hbar}$ are considered to be the He's polynomials [74], given by

$$\mathcal{H}_k(w_0, w_1, w_2, \ldots, w_n) = \frac{1}{k!}\frac{\partial^k}{\partial\theta^k}\left[\mathcal{Q}\left(\sum_{k=0}^{\infty}\theta^k w_k\right)\right].$$

Using Equation 30 and Equation 31 in Equation 29 parametric homotopy equation with homotopy parameter θ can be constructed as

$$\sum_{k=0}^{\infty}\theta^k w_k(\xi,\eta) = \mathcal{G}(\xi,\eta) + \theta\left\{\mathbb{A}_1^{-1}\left[\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1\left(\mathcal{P}\left[\sum_{k=0}^{\infty}\theta^k w_k(\xi,\eta)\right] + \sum_{k=0}^{\infty}\theta^k\mathcal{H}_k\right)\right]\right\}.$$

Comparing the coefficients of like powers of homotopy parameter i.e., θ, derives

$$\theta^0: w_0(\xi,\eta) = \mathcal{G}(\xi,\eta),$$
$$\theta^1: w_1(\xi,\eta) = \mathbb{A}_1^{-1}\left[\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1(\mathcal{P}[w_0] + \mathcal{H}_0)\right],$$
$$\theta^2: w_2(\xi,\eta) = A_1^{-1}\left[\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1(\mathcal{P}[w_1] + \mathcal{H}_1)\right],$$
$$\vdots$$
$$\theta^k: w_k(\xi,\eta) = \mathbb{A}_1^{-1}\left[\left(\frac{\gamma}{s}\right)^{\alpha}\mathbb{A}_1(\mathcal{P}[w_{k-1}] + \mathcal{H}_{\hbar-1})\right].$$

The approximate solution of n^{th}-order for the Equation 26 can be obtained as:

$$W_n = \lim_{\theta\to 1}\sum_{k=0}^{n}\theta^k w_k = \sum_{k=0}^{n} w_k. \quad (32)$$

Test Problems

This section is devoted to extracting and discussing the semi-analytic results of the L-C fractional FGKS, FSH and FCH equations under various initial conditions. The comparative analysis of the proposed algorithms is developed using graphical and tabular representations of the results.

Example 1. Consider FGKS without dispersion effect at $x = -1$, $y = 0$ and $z = 1$ for which initial condition is extracted from the analytic solution:

$$w(\xi,\eta) = s + \frac{15tanh^3(\kappa(\xi - s\eta - \lambda)) - 45tanh(\kappa(\xi - s\eta - \lambda))}{19\sqrt{19}}, \quad (33)$$

where s, κ and λ are real constants.

The computational work carried out with choice of

$$\mathcal{P}[w(\xi,\eta)] = -\left(xw_{\xi\xi} + zw_{\xi\xi\xi\xi}\right),$$
$$\mathcal{Q}[w(\xi,\eta)] = -ww_{\xi}, \text{ and } \mathcal{N}(\xi,\eta) = 0,$$

produces components of the series solutions in terms of GB-HPM solution as

$$w_0(\xi,\eta) = s + \left(\sqrt{19}\left(45tanh(\kappa(\lambda-\xi)) - 15tanh(\kappa(\lambda-\xi))^3\right)\right)/361$$
$$w_1(\xi,\eta) = \left(45\sqrt{19}\kappa\eta^{\alpha}\left(tanh(\kappa(\lambda-\xi)) - 1\right)^2\left(tanh(\kappa(\lambda-\xi)) + 1\right)^2\left(361s - 15\sqrt{19}tanh(\kappa(\lambda-\xi))^3 20216\kappa^3 tanh(\kappa(\lambda\xi)) + 43320\kappa^3 tanh(\kappa(\lambda-\xi))^3 + 45\sqrt{19}tanh(\kappa(\lambda-\xi)) - 1444\kappa tanh(\kappa(\lambda-\xi))\right)\right)/\left(130321\Gamma(\alpha+1)\right)\ldots.$$

The solution profile of GB-HPM solution at $s = 5, \lambda = -25, \kappa = \frac{1}{2\sqrt{19}}$ is drawn in Figure 1 and Figure 2. These graphs represent the behavior for integer order K-S equation with $\alpha = 1$ and the variation arises for different values of α in 3D and contour surfaces. The sensitivity of GB-HPM solution profiles with respect to the time

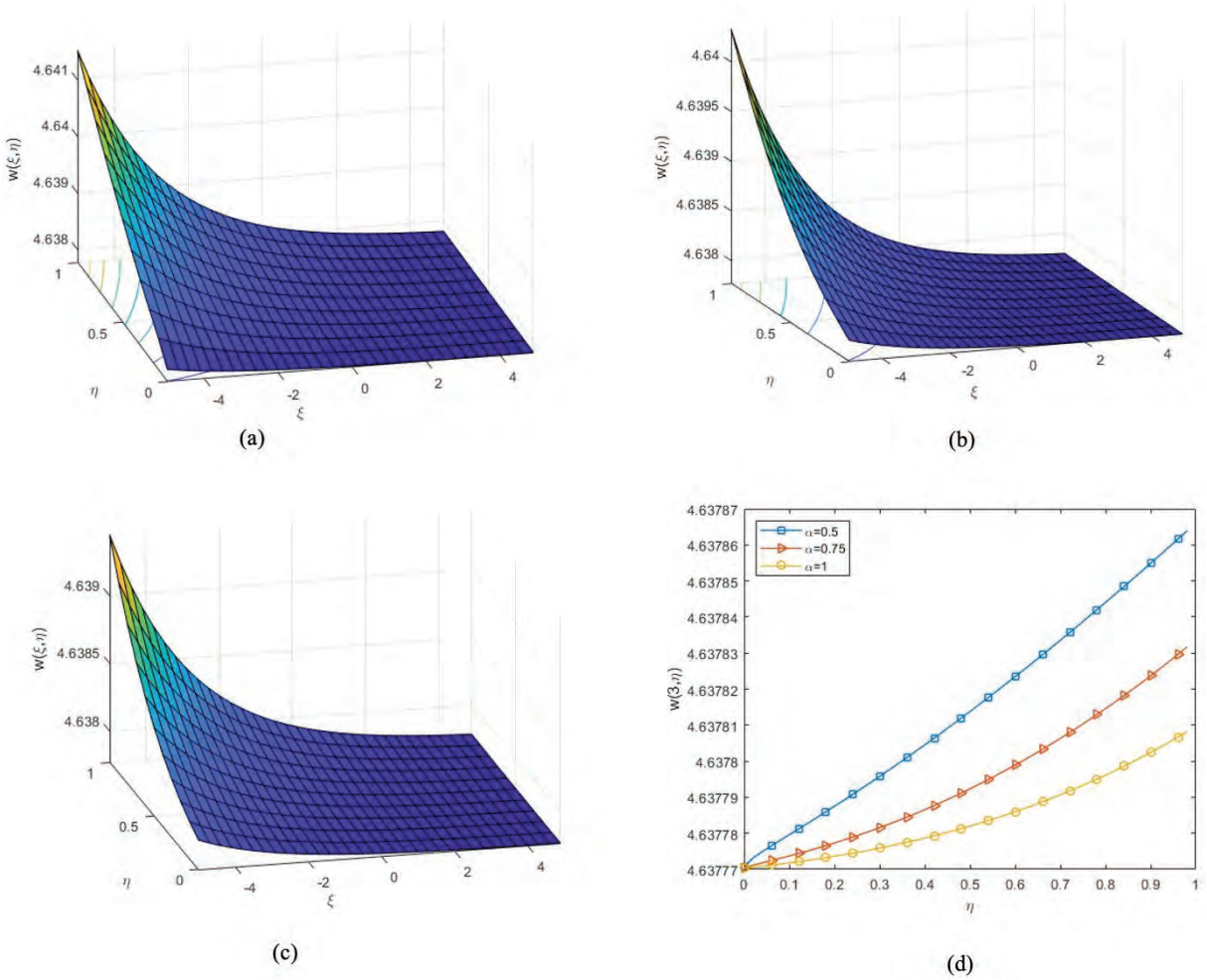


Figure 1. Solution profile of GB-HPM solution for Example 1 when (a) $\alpha = 0.5$ (b) $\alpha = 0.75$ (c) $\alpha = 1$ (d) $\xi = 3$.

fractional orders can be observed apparently through 2D simulations given in Figure 1(d) at $\xi = 3$ and contour surfaces in Figure 2. Fractional solution profile converges to the solution of integer order equation as $\alpha \to 1$. 3D surface plots of FGKS exhibit time evolution of flame-front. The contour surfaces effectively highlight the patterns in the data which reflects the heat distribution in the system. The fractional derivative reveals broader visualization of evolutionary process with its enhancing feature of controlling rate of change.

Table 1 shows the numerical findings for different time fractional orders. The absolute error is found competent and aligned to the error deduced for q-HATM scheme [24]. However, the present scheme is free from analysis of any auxiliary parameter. Table 2 demonstrates the convergence phenomenon of series solution for fractional derivative $\alpha = 0.75$. It provides the credibility of obtained solution for fractional phenomenon when exact solution is unknown. The convergence rate gets slower as time increases.

Example 2. Consider the FGKS equation with dispersion effect at $x = 1$, $y = 4$, and $z = 1$. The analytic solution is given by

$$w(\xi,\eta) = s + 9 - 15[tanh(\theta) + tanh^2(\theta) - tanh^3(\theta)], \quad (34)$$

where $\theta = \kappa(\xi - s\eta - s_0)$ with arbitrary constant s and s_0.

With choice of

$$\mathcal{P}[w(\xi,\eta)] = -\left(xw_{\xi\xi} + yw_{\xi\xi\xi} + zw_{\xi\xi\xi\xi}\right),$$
$$\mathcal{Q}[w(\xi,\eta)] = ww_{\xi}, \text{ and } \mathcal{N}(\xi,\eta) = 0,$$

and using the initial conditions from Equation 34 at $\eta = 0$, the suggested algorithm gives the explicit expressions of components of series solutions in terms of GB-HPM solution as

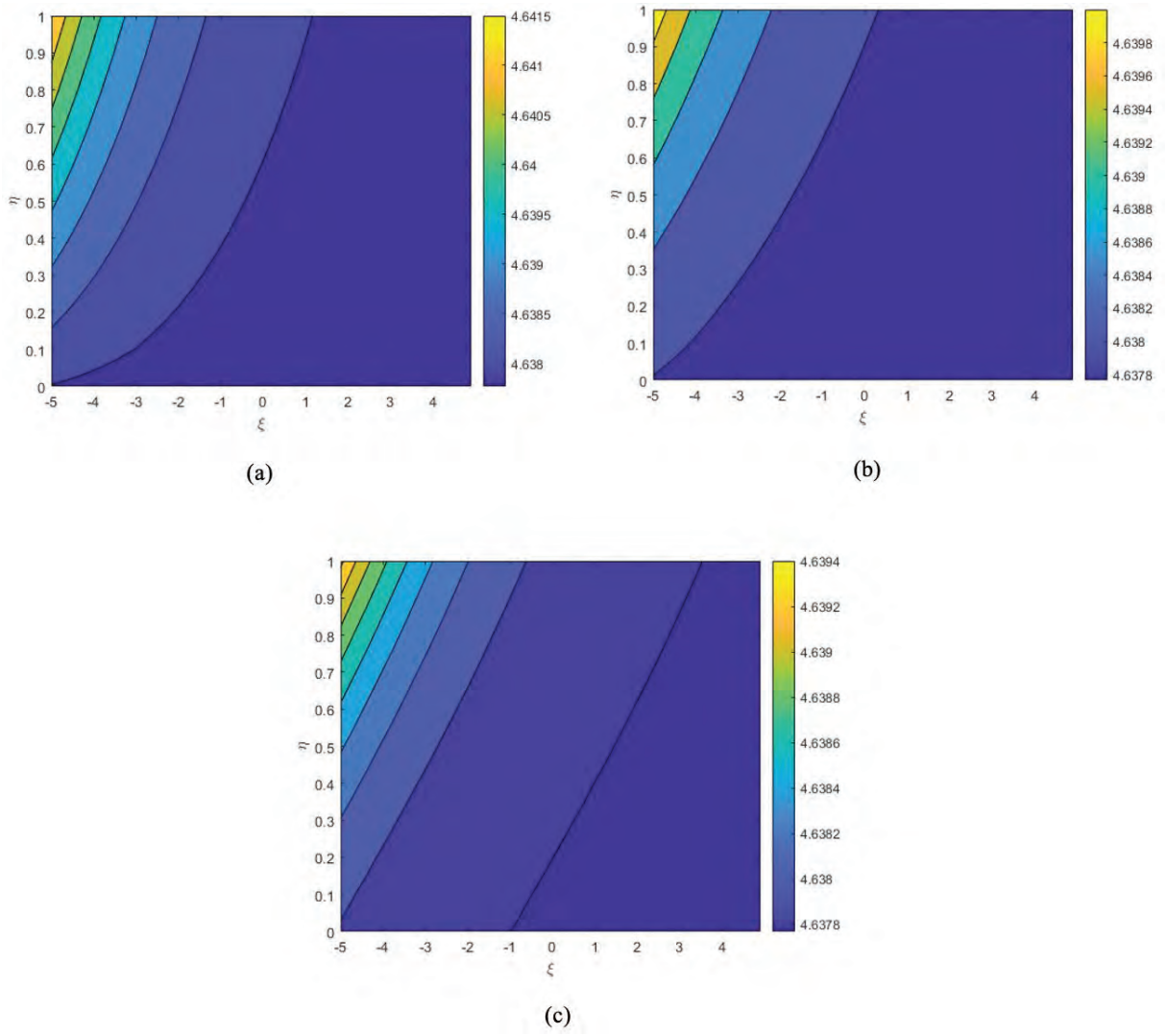


Figure 2. Contour surfaces of GB-HPM solution for Example 1when (a) $\alpha = 0.5$ (b) $\alpha = 0.75$ (c) $\alpha = 1$.

$$w_0(\xi,\eta) = s + 15tanh\big(\kappa(s_0 - \xi)\big) - 15tanh\big(\kappa(s_0 - \xi)\big)^2 - 15tanh\big(\kappa(s_0 - \xi)\big)^3 + 9,$$

$$w_1(\xi,\eta) = -\Big(15\kappa\eta^{\alpha}\big(tanh\big(\kappa(s_0 - \xi)\big) - 1\big)\big(tanh\big(\kappa(s_0 - \xi)\big) + 1\big)^2\Big(s + 2\kappa - 12tanh\big(\kappa(s_0 - \xi)\big) - 12\kappa tanh\big(\kappa(s_0 - \xi)\big)^2 + 96\kappa^2 tanh\big(\kappa(s_0 - \xi)\big) - 120\kappa^3 tanh\big(\kappa(s_0 - \xi)\big) - 60tanh\big(\kappa(s_0 - \xi)\big)^2 + 30tanh\big(\kappa(s_0 - \xi)\big)^3 + 45tanh\big(\kappa(s_0 - \xi)\big)^4 + 144\kappa^2 tanh\big(\kappa(s_0 - \xi)\big)^2 - 240\kappa^2 tanh\big(\kappa(s_0 - \xi)\big)^3 + 240\kappa^3 tanh\big(\kappa(s_0 - \xi)\big)^2 + 240\kappa^3 tanh\big(\kappa(s_0 - \xi)\big)^3 - 360\kappa^3 tanh\big(\kappa(s_0 - \xi)\big)^4 - 3stanh\big(\kappa(s_0 - \xi)\big) + 6\kappa tanh\big(\kappa(s_0 - \xi)\big) - 32\kappa^2 - 16\kappa^3 + 9\Big)\Big)/\Gamma(\alpha + 1),$$

and so on. For parameters $s = 3$, $\kappa = 0.5$ and $s_0 = -13$, Figure 3 and Figure 4 exhibit the dynamics that have been produced by the explicit expression of GB-HPM solution for Example 2. To visualize the variability in solution profile in temporal domain at $\xi = 3$, 2D solution profiles are drawn in Figure 3(d) for different values of α. Figure 4 simulates the projection of solution profile for distinct orders of fractional derivative. In this case of incorporating dispersion effect the flame-fronts exhibit some sharpness. The contour surfaces strikingly illustrate the regions of sharp transitions or abrupt changes which implies the occurring of steep thermal gradient.

Numerical values of the GB-HPM solution for severe fractional orders are tabulated in Table 3. In addition, error analysis in terms of absolute error of proposed solutions in comparison to q-HATM [24] reveals the supremacy of GB-HPM. This performance of accomplished results for FGKS equation with dispersion ensures the capability of the present method to capture every comprised effect of the system.

Table 1. Absolute error for approximate solution of Example 1 at $\kappa = 1/(2\sqrt{19})$, $s = 5$, $\lambda = -25$ for different α

ξ	η	$\alpha = 0.75$		$\alpha = 1$	
		GB-HPM	**q-HATM [24]**	**GB-HPM**	**q-HATM [24]**
1	0.2	9.11157×10^{-6}	9.11157×10^{-6}	2.80432×10^{-8}	2.80432×10^{-8}
	0.4	1.91218×10^{-5}	1.91218×10^{-5}	4.95160×10^{-7}	4.95160×10^{-7}
	0.6	2.93491×10^{-5}	2.93491×10^{-5}	2.78309×10^{-6}	2.78309×10^{-6}
	0.8	3.60085×10^{-5}	3.60085×10^{-5}	9.82802×10^{-6}	9.82802×10^{-6}
	1	3.27450×10^{-5}	3.27450×10^{-5}	2.69893×10^{-5}	2.69893×10^{-5}
2	0.2	5.77566×10^{-6}	5.77566×10^{-6}	1.78573×10^{-8}	1.78573×10^{-8}
	0.4	1.21255×10^{-5}	1.21255×10^{-5}	3.15448×10^{-7}	3.15448×10^{-7}
	0.6	1.86103×10^{-5}	1.86103×10^{-5}	1.77397×10^{-6}	1.77397×10^{-6}
	0.8	2.28157×10^{-5}	2.28157×10^{-5}	6.26876×10^{-6}	6.26876×10^{-6}
	1	2.06795×10^{-5}	2.06795×10^{-5}	1.72296×10^{-5}	1.72296×10^{-5}
3	0.2	3.65888×10^{-6}	3.65888×10^{-6}	1.13536×10^{-8}	1.13536×10^{-8}
	0.4	7.68382×10^{-6}	7.68382×10^{-6}	2.00632×10^{-7}	2.00632×10^{-7}
	0.6	1.17929×10^{-5}	1.17929×10^{-5}	1.12878×10^{-6}	1.12878×10^{-6}
	0.8	1.44489×10^{-5}	1.44489×10^{-5}	3.99096×10^{-6}	3.99096×10^{-6}
	1	1.30614×10^{-5}	1.30614×10^{-5}	1.09765×10^{-5}	1.09765×10^{-5}
4	0.2	2.31680×10^{-6}	2.31680×10^{-6}	7.20976×10^{-9}	7.20976×10^{-9}
	0.4	4.86653×10^{-6}	4.86653×10^{-6}	1.27441×10^{-7}	1.27441×10^{-7}
	0.6	7.46887×10^{-6}	7.46887×10^{-6}	7.17246×10^{-7}	7.17246×10^{-7}
	0.8	9.14654×10^{-6}	9.14654×10^{-6}	2.53702×10^{-6}	2.53702×10^{-6}
	1	8.25068×10^{-6}	8.25068×10^{-6}	6.98138×10^{-6}	6.98138×10^{-6}

Table 2. Absolute error in consecutive approximate solutions of Example 1 at $\kappa = 1/(2\sqrt{19})$, $s = 5$, $\lambda = -25$ for $\alpha = 0.7$

ξ	η	$\lvert W_2 - W_1 \rvert$	$\lvert W_3 - W_2 \rvert$	$\lvert W_4 - W_3 \rvert$
-5	0.01	8.37493×10^{-7}	3.07223×10^{-8}	9.03484×10^{-10}
	0.1	2.64839×10^{-5}	5.46328×10^{-6}	9.03484×10^{-7}
	0.5	2.96099×10^{-4}	2.04238×10^{-4}	1.12936×10^{-4}
-1	0.01	1.38596×10^{-7}	5.17944×10^{-9}	1.56727×10^{-10}
	0.1	4.38278×10^{-6}	9.21050×10^{-7}	1.56727×10^{-7}
	0.5	4.90010×10^{-5}	3.44323×10^{-5}	1.95909×10^{-5}
1	0.01	5.58599×10^{-8}	2.09697×10^{-9}	6.38892×10^{-11}
	0.1	1.76645×10^{-6}	3.72901×10^{-7}	6.38892×10^{-8}
	0.5	1.97495×10^{-5}	1.39404×10^{-5}	7.98615×10^{-6}
5	0.01	8.99584×10^{-9}	3.39273×10^{-10}	1.04094×10^{-11}
	0.1	2.84473×10^{-7}	6.03323×10^{-8}	1.04094×10^{-8}
	0.5	3.18051×10^{-6}	2.25545×10^{-6}	1.30118×10^{-6}

Example 3. Consider FSH equation without dispersion effect with initial data

$$w(\xi, 0) = \frac{1}{10} sin\left(\frac{\pi\xi}{l}\right), \tag{35}$$

where l is a real constant.

Following the computational procedure of GB-HPM with choice of

$$\mathcal{P}[w(\xi,\eta)] = -w_{\xi\xi\xi\xi} - 2w_{\xi\xi} - (1-x)w,$$
$$\mathcal{Q}[w(\xi,\eta)] = -w^3, \text{ and } \mathcal{N}(\xi,\eta) = 0,$$

yields the components of expansion solutions as

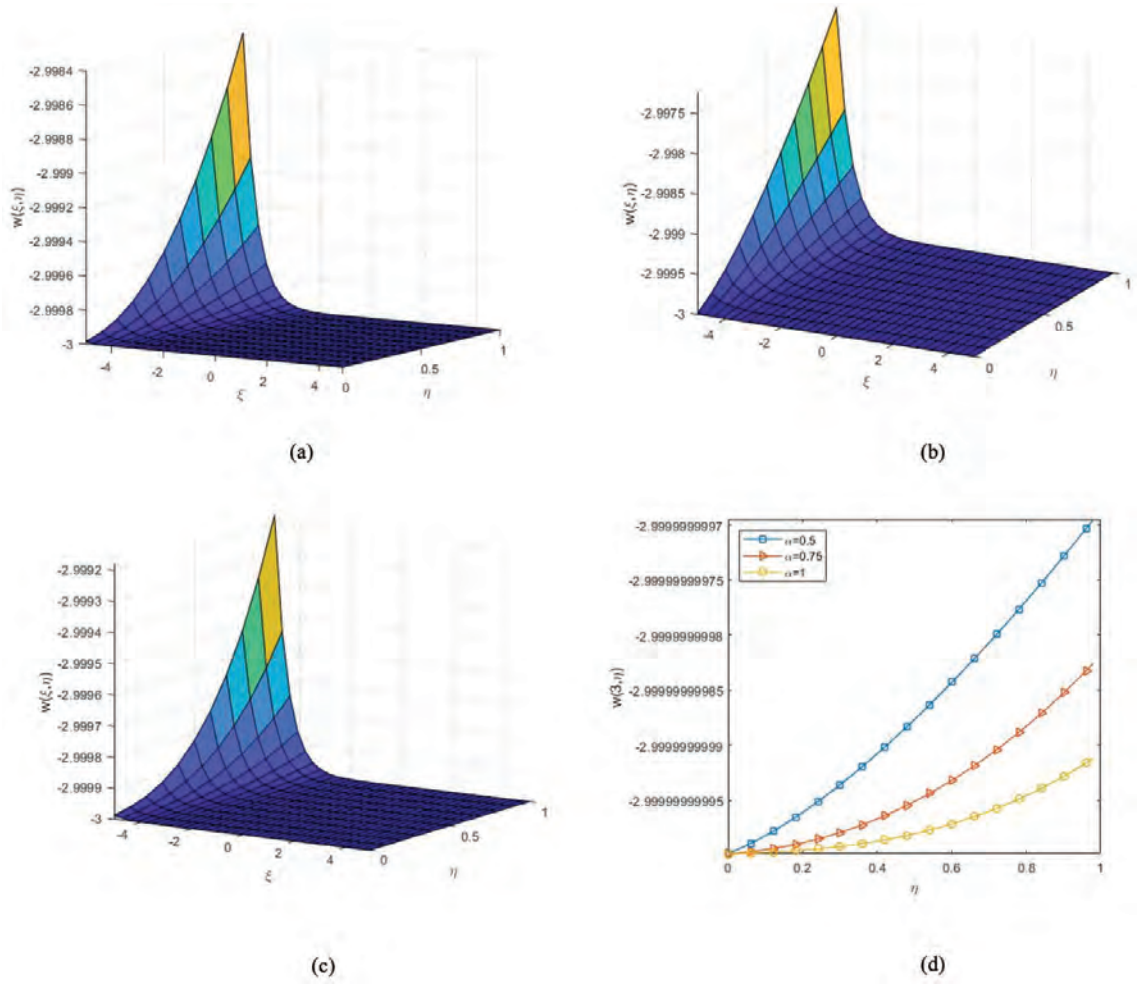


Figure 3. Solution profile of GB-HPM solution for Example 2 when (a) (a) $\alpha = 0.5$ (b) $\alpha = 0.75$ (c) $\alpha = 1$ (d) $\xi = 3$.

$$w_0(\xi,\eta) = sin\left((\pi\xi)/l\right)/10,$$

$$w_1(\xi,\eta) = \frac{-\left(\eta^{\alpha} sin\left((\pi\xi)/l\right)\left(100\pi^4 - 100\xi l^4 + 100l^4 - 200l^2\pi^2 + l^4 sin\left((\pi\xi)/l\right)^2\right)\right)}{\left(1000l^4\Gamma(\alpha+1)\right)},$$

$$\begin{aligned} w_2(\xi,\eta) = \frac{1}{\left(100000l^8\Gamma(2\alpha+1)\right)}\Big(&\eta^{2\alpha} sin\left((\pi\xi)/l\right)\Big(10000\pi^8 - 20000\xi l^8 \\ &+ 10000l^8 - 40000l^2\pi^6 + 54000l^4\pi^4 - 38800l^6\pi^2 + 10000\xi^2 l^8 \\ &+ 400l^8 sin\left((\pi\xi)/l\right)^2 + 3l^8 sin\left((\pi\xi)/l\right)^4 - 20000\xi l^4\pi^4 \\ &+ 40000\xi l^6\pi^2 - 400\xi l^8 sin\left((\pi\xi)/l\right)^2 + 8400l^4\pi^4 sin\left((\pi\xi)/l\right)^2 \\ &- 2400l^6\pi^2 sin\left((\pi\xi)/l\right)^2\Big)\Big), \end{aligned}$$

and so on. Solution profile of S-H equation signifies the physical effect of bifurcation and dispersive parameters on probability density function $w(\xi,\eta)$ for fractional Brownian and standard motions. For non-dispersive FSH equation, the graphical representation of GB-HPM solution in Figure 5, Figure 6 and Figure 7 simulate oscillatory and periodic solutions depending on the value of l. The behavior of obtained solutions varies harmonically in space and time. The variation of solution profiles at l = 3,6,8,10 for several fractional order displayed as 2D graphs is apparent in Figure 5, Figure 6 and Figure7. Behavior of obtained solution is more sensitive with respect to the α at $l = 3$, but at l = 6,8,10 there are no changes. For $\alpha = x = 0.5$ overshoots observed for probability density function $w(\xi,\eta)$ increase with increase in η for $l = 3$. However, overshoots of $w(\xi,\eta)$ show inverse relation with time η at l = 6,8,10.

Table 4 demonstrates the numerical results of 3^{rd}- order approximation for Example 3 at distinct fractional orders obtained by GB-HPM in comparison on of NTDM [49] and q-HATM [50] scheme. Obtained results are completely aligned with results of NTDM and agrees with q-HATM solution for atleast upto second significant digit. In comparison with NTDM scheme present method with alternative approach of He polynomials, overcomes the demerit of tedious calculations for Adomian polynomials [74]. In Table 5 the decreasing absolute error of consecutive approximate solutions ensures the convergence of the solution for $\alpha = 0.75$. Therefore, credibility GB-HPM solution is independent from the condition of availability of exact solution.

Example 4. Consider FSH equation without dispersion effect with initial data

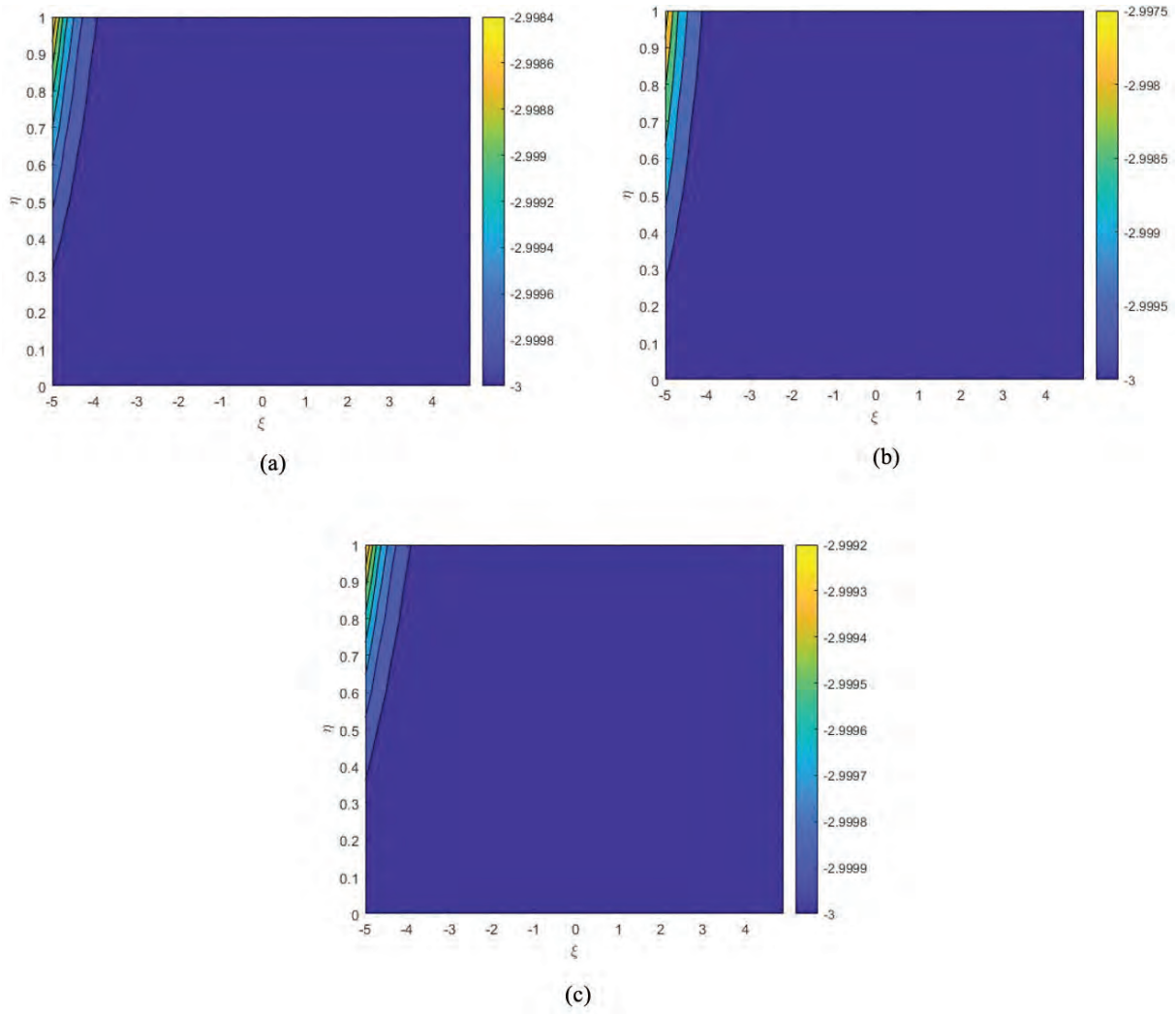


Figure 4. Contour surfaces of GB-HPM solution for Example 2 when (a) $\alpha = 0.5$ (b) $\alpha = 0.75$ (c) $\alpha = 1$.

$$w(\xi,0) = \frac{1}{10}sin\left(\frac{\pi\xi}{l}\right), \tag{36}$$

where l is a real constant.

Following the computational procedure of GB-HPM with choice of

$$\mathcal{P}[w(\xi,\eta)] = -w_{\xi\xi\xi\xi} - 2w_{\xi\xi} + xw_{\xi\xi\xi} + yw,$$
$$\mathcal{Q}[w(\xi,\eta)] = 2w^2 - w^3, \text{ and } \mathcal{N}(\xi,\eta) = 0,$$

yields the components of expansion solutions as

$$w_0(\xi,\eta) = sin\big((\pi\xi)/l\big)/10,$$

$$\begin{aligned} w_1(\xi,\eta) = -\Big(\eta^{\alpha}\Big(&100\pi^4 sin\big((\pi\xi)/l\big) - 20l^4 sin\big((\pi\xi)/l\big)^2 \\ &+ l^4 sin\big((\pi\xi)/l\big)^3 - 200l^2\pi^2 sin\big((\pi\xi)/l\big) \\ &- 100yl^4 sin\big((\pi\xi)/l\big) + 100xl\pi^3 cos\big((\pi\xi)/l\big)\Big)\Big) \\ &/\big(1000l^4\Gamma(\alpha+1)\big), \end{aligned}$$

$$\begin{aligned} w_2(\xi,\eta) = -\Big(\eta^{2\alpha}\Big(&200l^8 cos\big((4\pi\xi)/l\big) - 800l^8 cos\big((2\pi\xi)/l\big) \\ &- 160000\pi^8 sin\big((\pi\xi)/l\big) - 48000yl^8 \\ &- 9630l^8 sin\big((\pi\xi)/l\big) + 3215l^8 sin\big((3\pi\xi)/l\big) \\ &- 3l^8 sin\big((5\pi\xi)/l\big) + 600l^8 + 32000l^4\pi^4 - 64000l^6\pi^2 \\ &- 288000l^4\pi^4 cos\big((2\pi\xi)/l\big) + 192000l^6\pi^2 cos\big((2\pi\xi)/l\big) \\ &+ 640000l^2\pi^6 sin\big((\pi\xi)/l\big) - 644800l^4\pi^4 sin\big((\pi\xi)/l\big) \\ &+ 9600l^6\pi^2 sin\big((\pi\xi)/l\big) + 33600l^4\pi^4 sin\big((3\pi\xi)/l\big) \\ &- 9600l^6\pi^2 sin\big((3\pi\xi)/l\big) - 160000y^2l^8 sin\big((\pi\xi)/l\big) \\ &+ 48000yl^8 cos\big((2\pi\xi)/l\big) + 4800yl^8 sin\big((\pi\xi)/l\big) \\ &- 1600yl^8 sin\big((3\pi\xi)/l\big) + 160000x^2l^2\pi^6 sin\big((\pi\xi)/l\big) \\ &- 320000xl\pi^7 cos\big((\pi\xi)/l\big) + 640000xl^3\pi^5 cos\big((\pi\xi)/l\big) \\ &- 2400xl^5\pi^3 cos\big((\pi\xi)/l\big) + 12000xl^5\pi^3 cos\big((3\pi\xi)/l\big) \\ &+ 320000yl^4\pi^4 sin\big((\pi\xi)/l\big) - 640000yl^6\pi^2 sin\big((\pi\xi)/l\big) \\ &+ 160000xl^5\pi^3 sin\big((2\pi\xi)/l\big) + 320000yxl^5\pi^3 cos\big((\pi\xi)/l\big)\Big)\Big) \\ &/\big(1600000l^8\Gamma(2\alpha+1)\big), \end{aligned}$$

Table 3. Absolute error for approximate solution of Example 2 at $\kappa = 0.5$, s = 3, $s_0 = -13$ for different α

ξ	η	$\alpha = 0.75$		$\alpha = 1$	
		GB-HPM	**q-HATM [24]**	**GB-HPM**	**q-HATM [24]**
1	0.2	3.58527×10^{-10}	1.44623×10^{-7}	9.30287×10^{-12}	3.75185×10^{-9}
	0.4	9.03807×10^{-10}	9.87125×10^{-7}	2.02381×10^{-10}	8.16183×10^{-8}
	0.6	6.87414×10^{-10}	1.28593×10^{-6}	1.47209×10^{-9}	5.93605×10^{-7}
	0.8	3.67755×10^{-9}	1.12748×10^{-7}	7.11536×10^{-9}	2.86856×10^{-6}
	1	2.35700×10^{-8}	7.82679×10^{-6}	2.84106×10^{-8}	1.14490×10^{-5}
2	0.2	4.85206×10^{-11}	5.47368×10^{-8}	1.26051×10^{-12}	5.07855×10^{-10}
	0.4	1.22317×10^{-10}	1.33604×10^{-7}	2.73904×10^{-11}	1.10484×10^{-8}
	0.6	9.30283×10^{-11}	1.74032×10^{-7}	1.99230×10^{-10}	3.88396×10^{-7}
	0.8	4.97723×10^{-10}	1.53996×10^{-8}	9.62981×10^{-10}	1.55060×10^{-6}
	1	3.19000×10^{-9}	1.06033×10^{-6}	3.84510×10^{-9}	6.87344×10^{-11}
3	0.2	6.57005×10^{-12}	7.40806×10^{-9}	1.67069×10^{-13}	1.49536×10^{-9}
	0.4	1.65574×10^{-11}	1.80819×10^{-8}	3.70357×10^{-12}	1.08769×10^{-8}
	0.6	1.25933×10^{-11}	2.35527×10^{-8}	2.69595×10^{-11}	5.25725×10^{-8}
	0.8	6.73584×10^{-11}	2.09112×10^{-9}	1.30324×10^{-10}	2.09907×10^{-7}
	1	4.31727×10^{-10}	1.43554×10^{-7}	5.20385×10^{-10}	9.30101×10^{-12}
4	0.2	8.86264×10^{-13}	1.00258×10^{-6}	2.61299×10^{-14}	2.02384×10^{-10}
	0.4	2.23943×10^{-12}	2.44714×10^{-9}	5.03161×10^{-13}	1.47209×10^{-9}
	0.6	1.70319×10^{-12}	3.18751×10^{-9}	3.65031×10^{-12}	7.11536×10^{-9}
	0.8	9.11656×10^{-12}	2.83350×10^{-10}	1.76376×10^{-11}	2.84106×10^{-8}
	1	5.84287×10^{-11}	1.94306×10^{-8}	7.04272×10^{-11}	1.25588×10^{-12}

Table 4. Numerical values of approximate solution of Example 3 at $x = 0.9$ and $l = 10$ for different α.

ξ	η	$\alpha = 0.75$			$\alpha = 1$		
		GB-HPM	**NTDM [49]**	**q-HATM [50]**	**GB-HPM**	**NTDM [49]**	**q-HATM [50]**
1	0.2	0.031802	0.031802	0.031793	0.031446	0.031446	0.031446
	0.4	0.032452	0.032452	0.032452	0.032007	0.032008	0.032007
	0.6	0.033051	0.033051	0.033051	0.032587	0.032587	0.032587
	0.8	0.033627	0.033627	0.033627	0.033187	0.033187	0.033187
	1	0.034191	0.034191	0.034191	0.033810	0.033810	0.033810
2	0.2	0.060430	0.060430	0.060430	0.059780	0.059780	0.059780
	0.4	0.061608	0.061608	0.061608	0.060807	0.060807	0.060806
	0.6	0.062681	0.062681	0.062681	0.061860	0.061860	0.061860
	0.8	0.063701	0.063701	0.063701	0.062942	0.062942	0.062942
	1	0.064692	0.064692	0.064692	0.064055	0.064055	0.064055
3	0.2	0.083072	0.083072	0.083072	0.082224	0.082224	0.082224
	0.4	0.084594	0.084594	0.085442	0.083568	0.083568	0.083568
	0.6	0.085958	0.085958	0.087236	0.084933	0.084933	0.084933
	0.8	0.087236	0.087236	0.088458	0.086321	0.086321	0.086321
	1	0.088458	0.088459	0.097559	0.087732	0.087732	0.087731
4	0.2	0.097559	0.097559	0.099254	0.096607	0.096607	0.096607
	0.4	0.099254	0.099254	0.100751	0.098121	0.098121	0.098121
	0.6	0.100751	0.100751	0.102134	0.099646	0.099646	0.099646
	0.8	0.102134	0.102134	0.103439	0.101181	0.101181	0.101181
	1	0.103439	0.103439	0.102541	0.102725	0.102725	0.102725

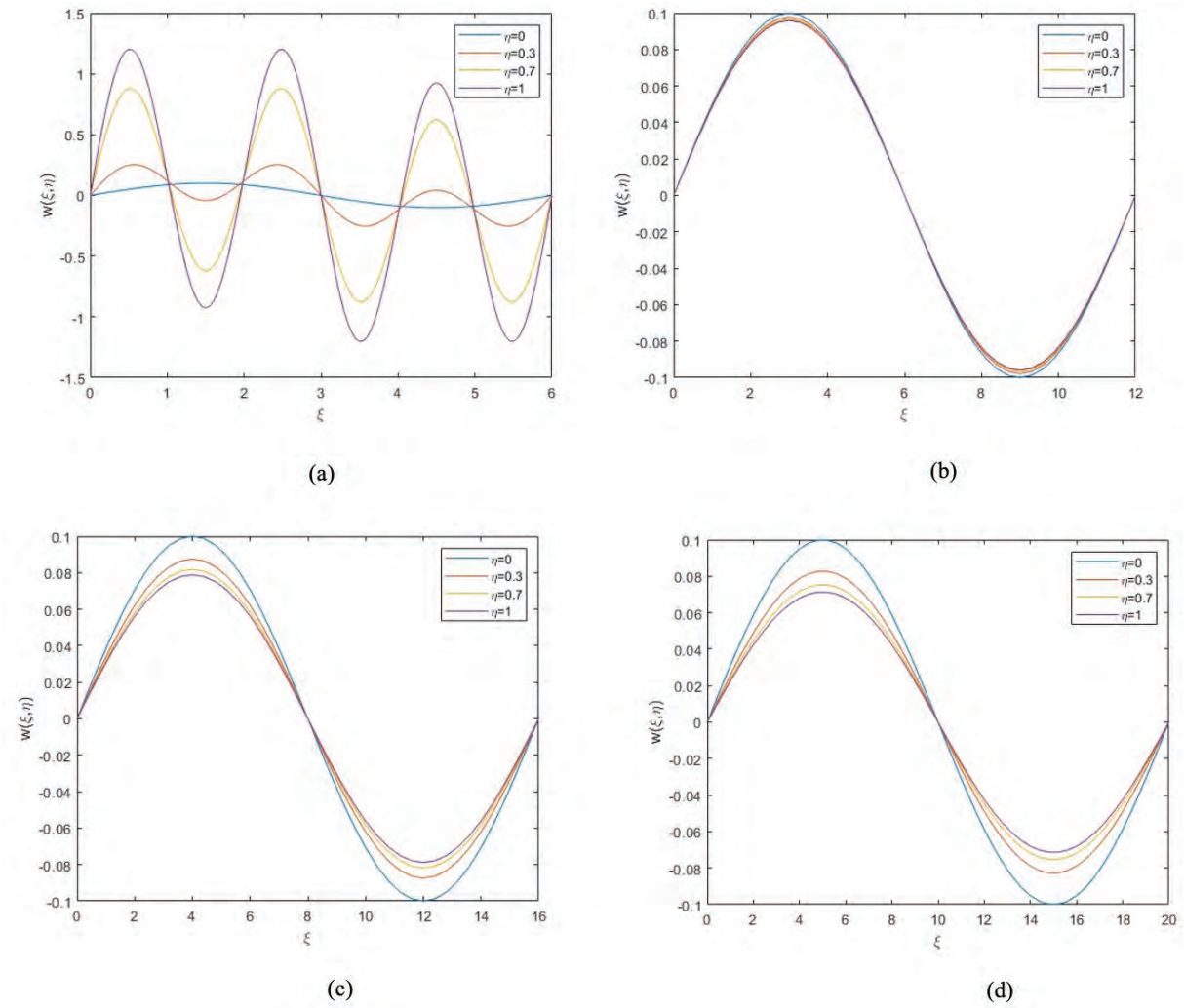


Figure 5. Solution profile of GB-HPM solution for Example 3 when $\alpha = 0.5$ and $x = 0.5$ at (a) $l = 3$ (b) $l = 6$ (c) $l = 8$ (d) $l = 10$.

Table 5. Absolute error in consecutive approximate solutions of Example 3 at $x = 0.9$ and $l = 10$ for $\alpha = 0.75$

ξ	η	$\lvert W_2 - W_1 \rvert$	$\lvert W_3 - W_2 \rvert$	$\lvert W_4 - W_3 \rvert$
1	0.01	2.92569×10^{-7}	2.59836×10^{-9}	3.01272×10^{-11}
	0.1	9.25185×10^{-6}	4.62061×10^{-7}	3.01272×10^{-8}
	0.5	1.03439×10^{-4}	1.72736×10^{-5}	3.76590×10^{-6}
2	0.01	4.30788×10^{-7}	3.09033×10^{-9}	3.49638×10^{-11}
	0.1	1.36227×10^{-5}	5.49547×10^{-7}	3.49638×10^{-8}
	0.5	1.52307×10^{-4}	2.05441×10^{-5}	4.37047×10^{-6}
3	0.01	3.82210×10^{-7}	1.10898×10^{-9}	1.07566×10^{-11}
	0.1	1.20866×10^{-5}	1.97208×10^{-7}	1.07566×10^{-8}
	0.5	1.35132×10^{-4}	7.37238×10^{-6}	1.34457×10^{-6}
4	0.01	2.51909×10^{-7}	1.68264×10^{-9}	2.19749×10^{-11}
	0.1	7.96607×10^{-6}	2.99221×10^{-7}	2.19749×10^{-8}
	0.5	8.90633×10^{-5}	1.11860×10^{-5}	2.74687×10^{-6}

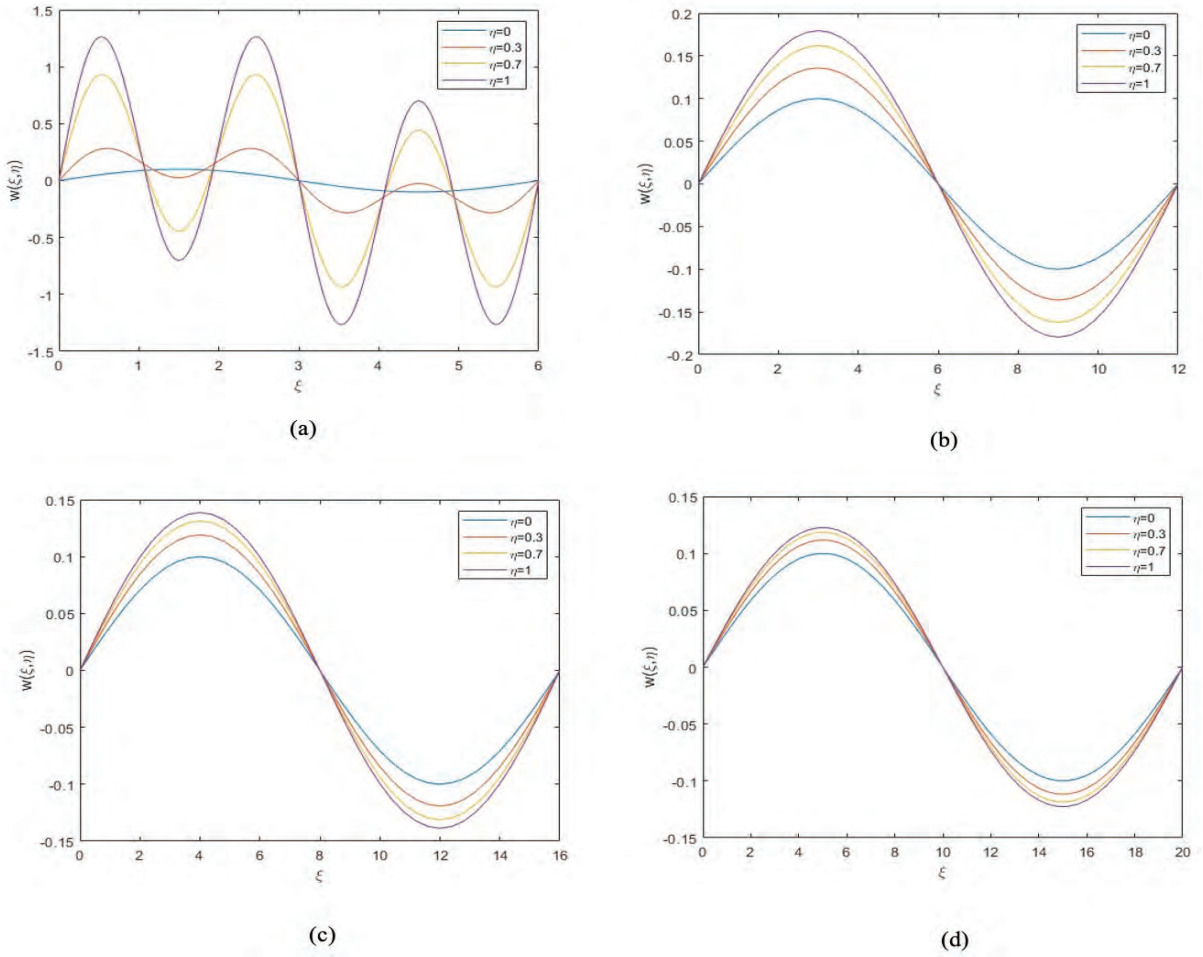


Figure 6. Solution profile of GB-HPM solution for Example 3 when $\alpha = 0.5$ and $x = 1$ at (a) $l = 3$ (b) $l = 6$ (c) $l = 8$ (d) $l = 10$.

Table 6. Numerical values of approximate solution of Example 4 at $x = 0.5$, $y = 0.7$ and $l = 10$ for different α

ξ	η	α = 0.75			α = 1		
		GB-HPM	NTDM [49]	q-HATM [50]	GB-HPM	NTDM [49]	q-HATM [50]
1	0.2	0.041383	0.041383	0.042774	0.036887	0.036887	0.037301
	0.4	0.050226	0.050226	0.054156	0.043729	0.043729	0.045381
	0.6	0.059170	0.059170	0.066390	0.051459	0.051459	0.055176
	0.8	0.068403	0.068403	0.079519	0.060108	0.060108	0.066716
	1	0.077989	0.077989	0.093524	0.069706	0.069706	0.080031
2	0.2	0.082115	0.082115	0.086069	0.071563	0.071563	0.072738
	0.4	0.104496	0.104496	0.115681	0.087199	0.087199	0.091901
	0.6	0.129531	0.129531	0.150079	0.106187	0.106187	0.116767
	0.8	0.157701	0.157701	0.189337	0.129029	0.129029	0.147837
	1	0.189231	0.189231	0.233444	0.156224	0.156224	0.185611
3	0.2	0.115950	0.115950	0.122485	0.099716	0.099716	0.101658
	0.4	0.151505	0.151505	0.169987	0.123489	0.123489	0.131258
	0.6	0.192814	0.192814	0.226768	0.153296	0.153296	0.170778
	0.8	0.240654	0.240654	0.292929	0.190213	0.190213	0.221290
	1	0.295424	0.295424	0.368481	0.235313	0.235313	0.283872
4	0.2	0.138479	0.138479	0.146962	0.118122	0.118122	0.120643
	0.4	0.183832	0.183832	0.207824	0.147741	0.147741	0.157827
	0.6	0.237549	0.237549	0.281626	0.185526	0.185526	0.208219
	0.8	0.300634	0.300634	0.368495	0.233036	0.233036	0.273379
	1	0.373628	0.373628	0.468467	0.291832	0.291832	0.354869

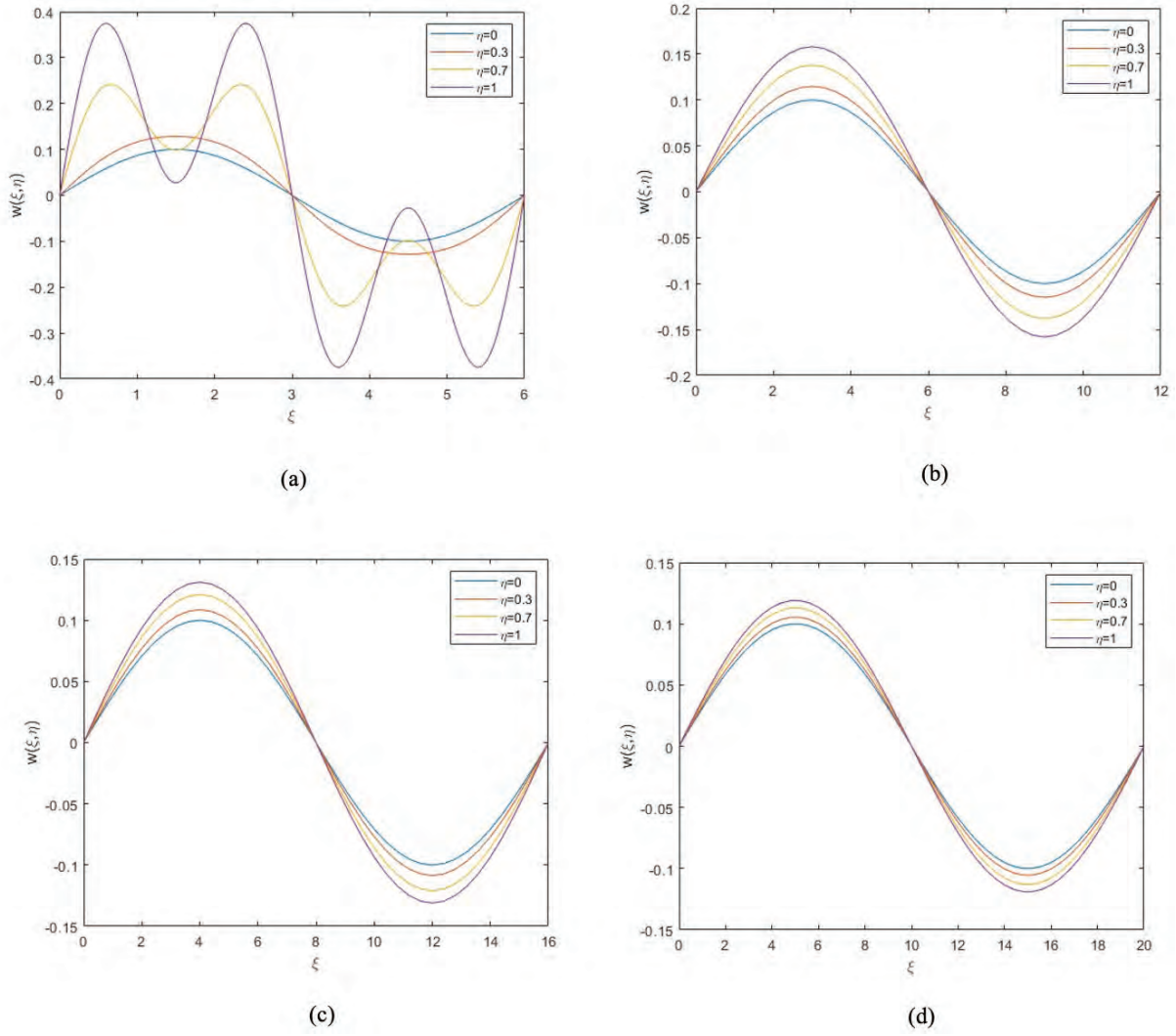


Figure 7. Solution profile of GB-HPM solution for Example 3 when $\alpha = 1$ and $x = 1$ at (a) $l = 3$ (b) $l = 6$ (c) $l = 8$ (d) $l = 10$.

Table 7. Absolute error for approximate solution of Example 5 at $x = 1$ and $\alpha = 1$

ξ	η	GB-HPM	Q-HAM [78]	
0.01	0	2.356974×10^{-12}	2.356975×10^{-12}	-
	1	2.823764×10^{-10}	2.823765×10^{-10}	-
	2	5.749510×10^{-11}	5.749512×10^{-11}	-
	3	3.757251×10^{-11}	3.757261×10^{-11}	-
0.1	0	2.352262×10^{-7}	2.352262×10^{-7}	-
	1	2.722916×10^{-6}	2.722916×10^{-6}	-
	2	5.848640×10^{-7}	5.848640×10^{-7}	-
	3	3.686350×10^{-7}	3.686350×10^{-7}	-
			RPS [76], HPM [79]	**HPM [77]**
0.2	0.1	2.55541×10^{-5}	2.55541×10^{-5}	4.68338×10^{-3}
	0.2	4.15291×10^{-5}	4.15291×10^{-5}	7.28902×10^{-3}
	0.3	5.42246×10^{-5}	5.42246×10^{-5}	9.61620×10^{-3}
	0.4	6.28898×10^{-5}	6.28898×10^{-5}	1.15931×10^{-2}
	0.5	6.72637×10^{-5}	6.72637×10^{-5}	1.31740×10^{-2}

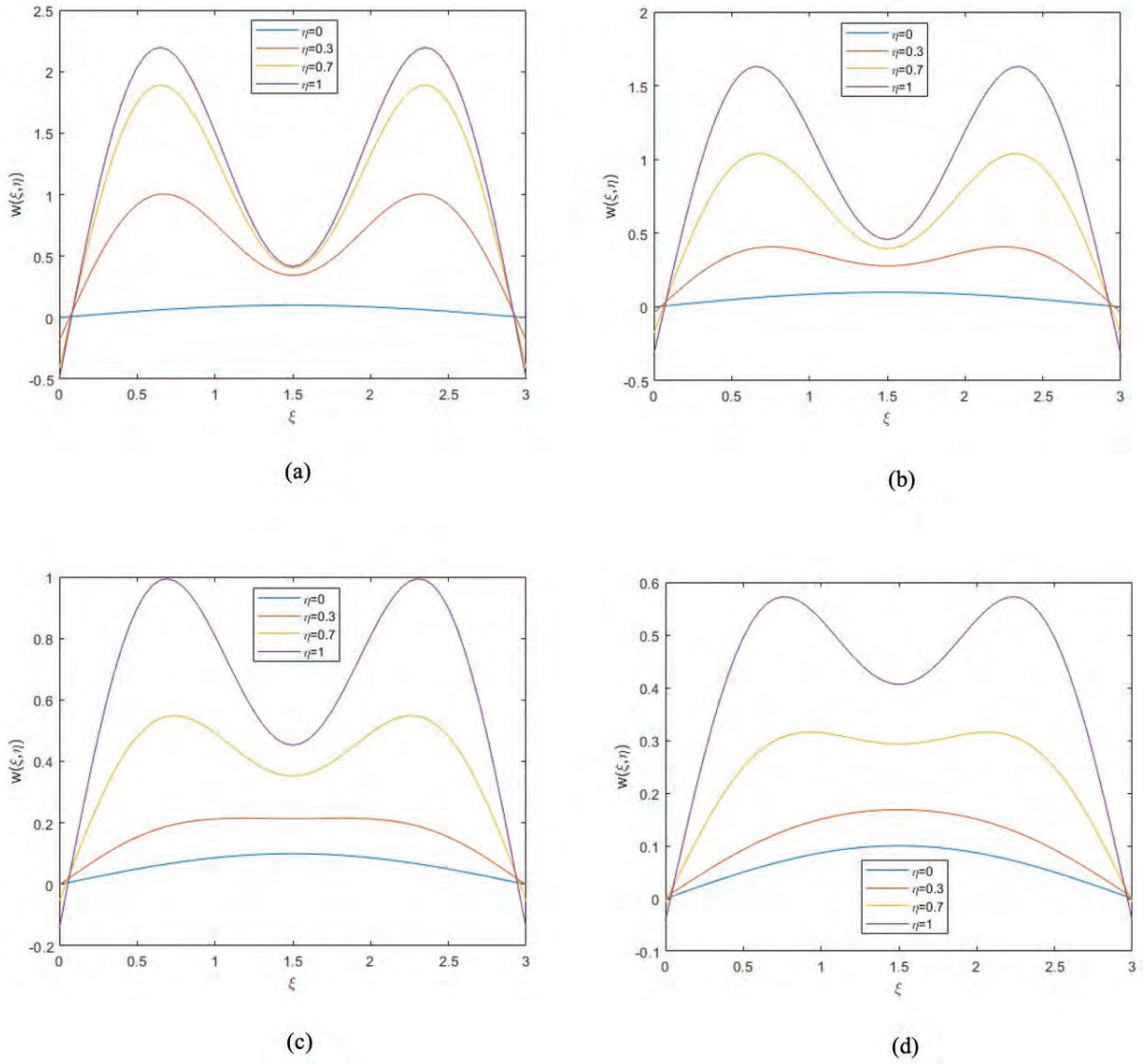


Figure 8. Solution profile of GB-HPM solution for Example 4 when $x = 0$, $y = 0.7$ and $l = 3$ at (a) $\alpha = 0.25$ (b) $\alpha = 0.5$ (c) $\alpha = 0.75$ (d) $\alpha = 1$.

Table 8. Absolute error in consecutive approximate solutions of Example 5 at x =1 and $\alpha = 0.75$

ξ	η	$\lvert W_2 - W_1 \rvert$	$\lvert W_3 - W_2 \rvert$	$\lvert W_4 - W_3 \rvert$
-5	0.01	2.54697×10^{-6}	5.99187×10^{-8}	7.61024×10^{-10}
	0.03	1.32344×10^{-5}	7.09717×10^{-7}	2.05477×10^{-8}
	0.05	2.84759×10^{-5}	2.23998×10^{-6}	9.51280×10^{-8}
-2	0.01	1.40856×10^{-4}	4.04291×10^{-6}	1.16966×10^{-6}
	0.03	7.31911×10^{-4}	4.78869×10^{-5}	3.15809×10^{-5}
	0.05	1.57482×10^{-3}	1.51139×10^{-4}	1.46208×10^{-4}
2	0.01	1.40856×10^{-4}	1.01417×10^{-6}	9.74371×10^{-7}
	0.03	7.31911×10^{-4}	1.20124×10^{-5}	2.63080×10^{-5}
	0.05	1.57482×10^{-3}	3.79133×10^{-5}	1.21796×10^{-4}
5	0.01	2.54697×10^{-6}	5.84709×10^{-8}	5.89081×10^{-10}
	0.03	1.32344×10^{-5}	6.92568×10^{-7}	1.59052×10^{-8}
	0.05	2.84759×10^{-5}	2.18586×10^{-6}	7.36351×10^{-8}

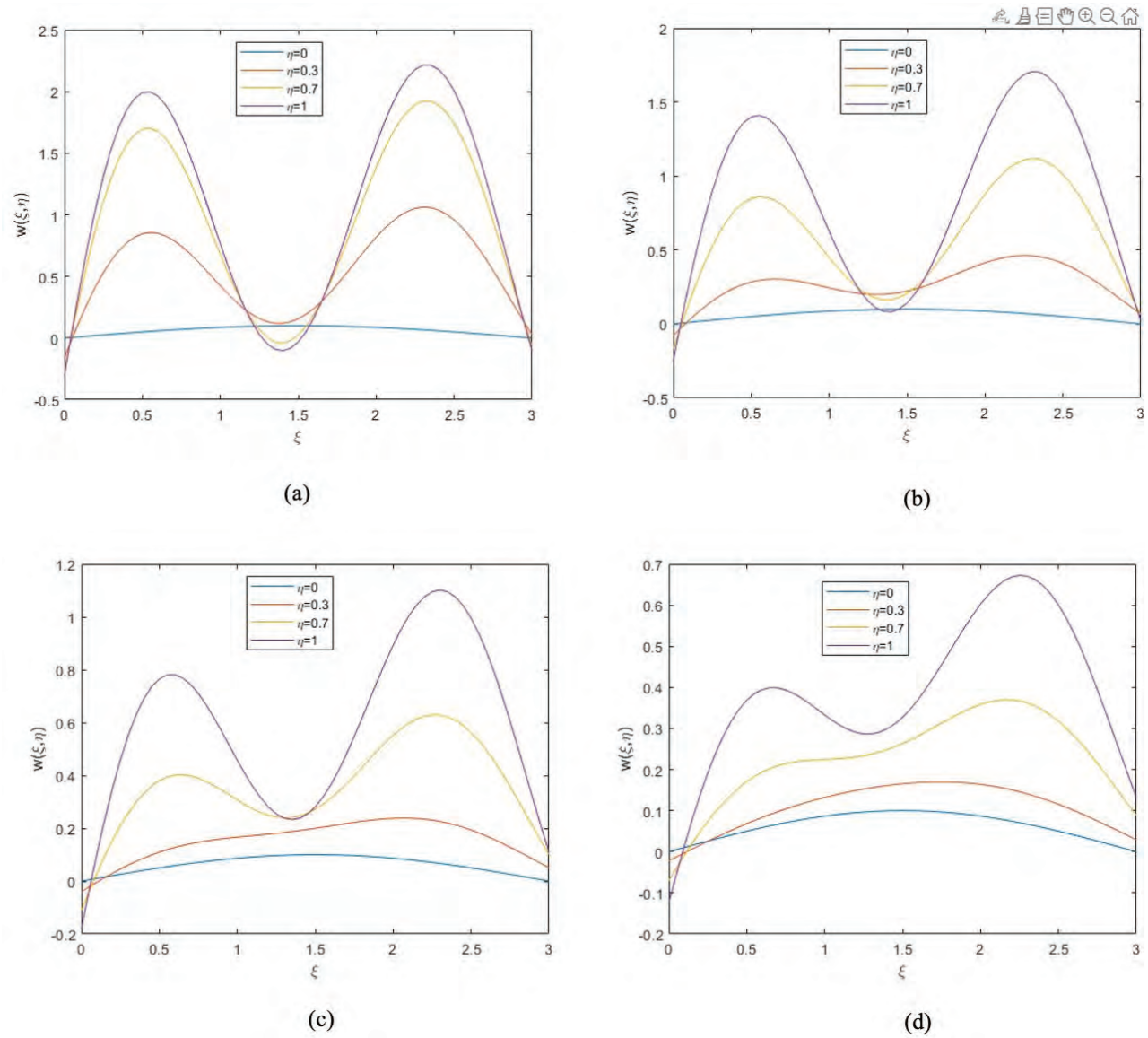


Figure 9. Solution profile of GB-HPM solution for Example 4 when $x = 0.5$, $y = 0.7$ and $l = 3$ at (a) $\alpha = 0.25$ (b) $\alpha = 0.5$ (c) $\alpha = 0.75$ (d) $\alpha = 1$.

and so on. The graphical illustrations for obtained GB-HPM solution to analyze the effect of dispersion and bifurcation parameters (i.e., x and y, respectively) on probability density function $w(\xi,\eta)$ have been presented in Figure 8, Figure 9, Figure 10 and Figure 11. From comparative visualization of presented figures for Example 4, it can be stated that $w(\xi,\eta)$ has a significant effect of included physical parameters for $l = 3$. However, for $l = 10$ solution profile does not depict much difference with respect to the variation in x and y. The presented figures also exhibit the evolution of solution profile with an increase in time for different values of fractional order α. Numerical values tabulated in Table 6 again demonstrate the agreement of GB-HPM solution with NTDM [49] and q-HATM [50] solutions.

Example 5. Consider FCH with the initial data $w(\xi, 0) = tanh\left(\frac{\sqrt{2}}{2}x\right)$ and for $x = 1$ on setting $\alpha = 1$ the corresponding exact solution is given by:

$$w(\xi,\eta) = tanh\left(\frac{\sqrt{2}}{2}(\xi+\eta)\right). \tag{37}$$

Following the criteria of proposed framework with choice of

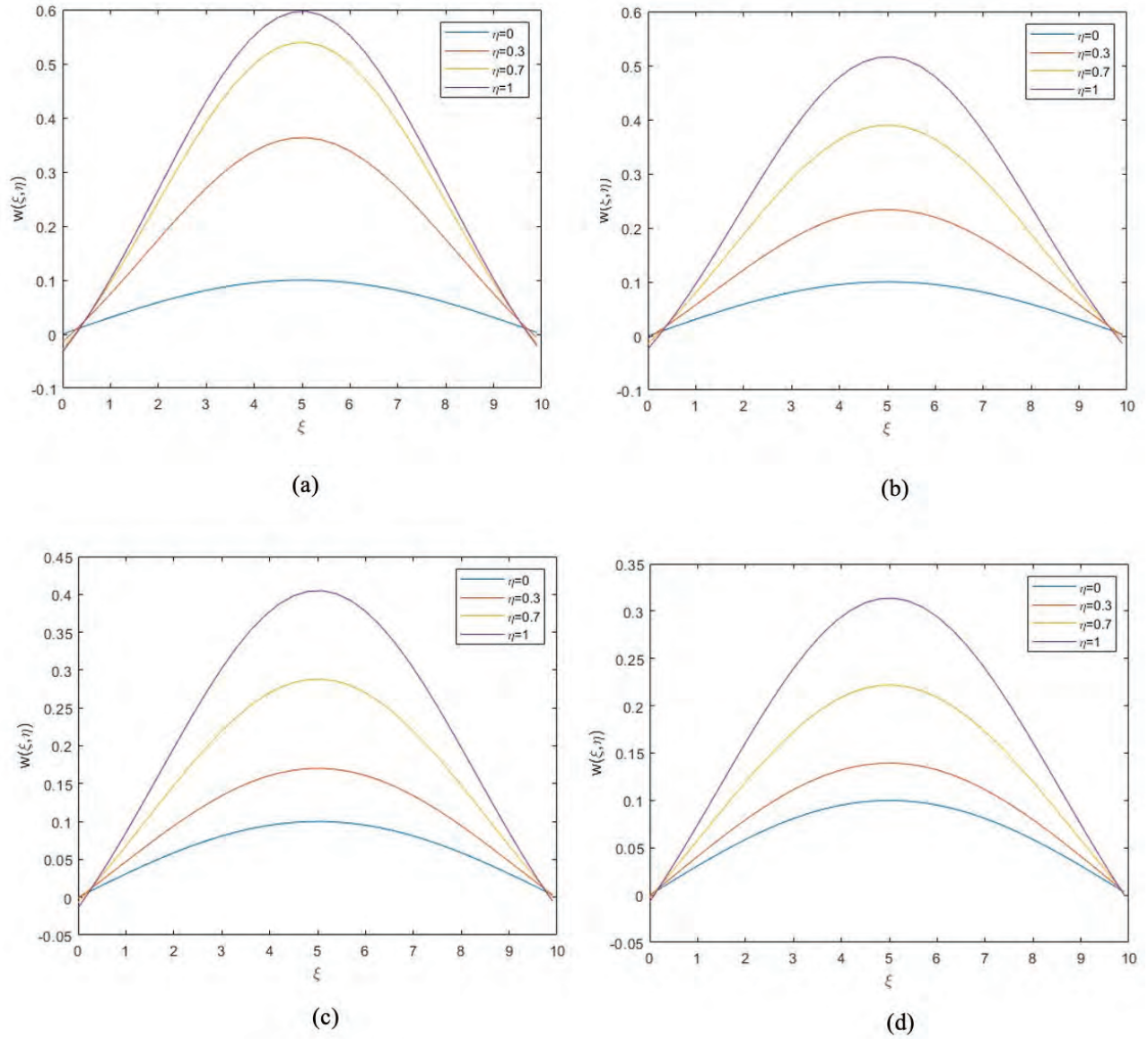


Figure 10. Solution profile of GB-HPM solution for Example 4 when $x = 0$, $y = 0.7$ and $l = 10$ at (a) $\alpha = 0.25$ (b) $\alpha = 0.5$ (c) $\alpha = 0.75$ (d) $\alpha = 1$.

$$\mathcal{P}[w(\xi,\eta)] = xw_{\xi} - w_{\xi\xi} - w_{\xi\xi\xi\xi},$$
$$\mathcal{Q}[w(\xi,\eta)] = 6ww_{\xi}^2 + 3w^2w_{\xi\xi} \; and \; \mathcal{N}(\xi,\eta) = 0,$$

the components of series solution can be obtained as:

$$w_0(\xi,\eta) = tanh\left((\sqrt{2}\xi)/2\right),$$

$$w_1(\xi,\eta) = (\sqrt{2}\eta^{\alpha})/\left(\Gamma(\alpha+1)(cosh(\sqrt{2}\xi)+1)\right),$$

$$w_2(\xi,\eta) = \left(\eta^{2\alpha}sinh\left((\sqrt{2}\xi)/2\right)\right)/\left(\Gamma(2\alpha+1)cosh\left((\sqrt{2}\xi)/2\right)^3\right)\ldots.$$

Continuing the same procedure, the remaining components of the series solution $W_n(\xi,\eta)$ $(n \geq 3)$ can be obtained.

Figure 12 depicts the geometrical configuration of the obtained semi-analytic solution at variant fractional orders. This Figure 12 suggests that the implementation of the proposed method is significantly efficient in finding a smooth wave solution to the specified problem. Surface plots of obtained solution describe the configuration of a kink wave. This wave represents a localized solution that rises or descends from one asymptotic state to another. In phase separation process, the solution profile of Cahn-Hilliard quantifies the transition of concentration of one of the two metallic components of the alloy. The present solution displays the stationary transition of phase in spinodal interval of $|w| < 1$, which directs to the unstable stationary state [75].

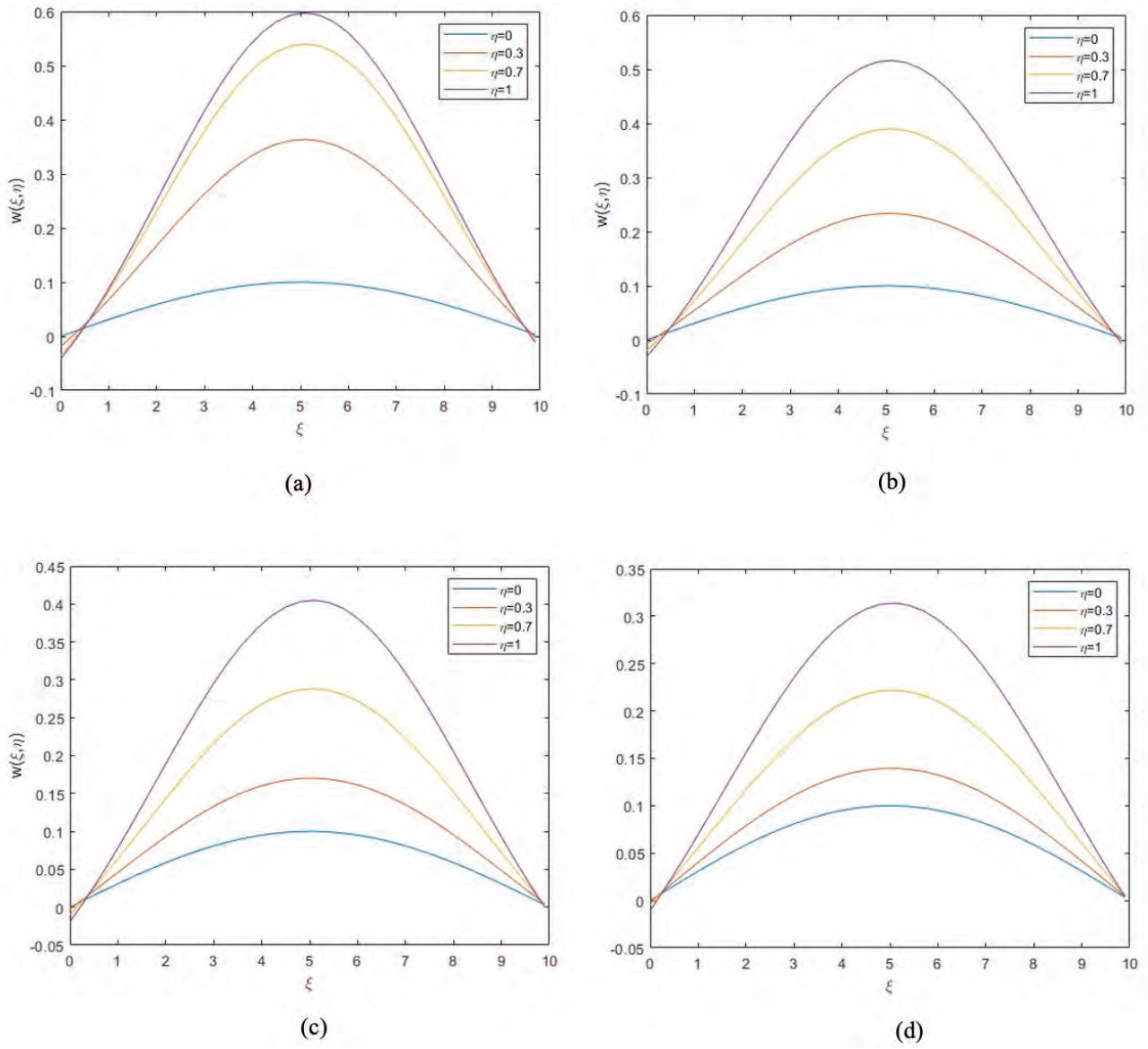


Figure 11. Solution profile of GB-HPM solution for Example 4 when $x = 0.5$, $y = 0.7$ and $l = 10$ at (a) $\alpha = 0.25$ (b) $\alpha = 0.5$ (c) $\alpha = 0.75$ (d) $\alpha = 1$.

In Table 7, a quantitative response of the achieved results in comparison of RPS [76], HPM [77] and q-HAM [78] is presented. For $x = 1$ obtained solution has acquired alike accuracy to RPS and q-HAM. In [76], RPS is found superior to HPM for solving fractional C-H equation. For the case $\alpha = 1$, the obtained solution converges to the solution obtained for the classical C-H equation, utilizing HPM and ADM [79]. Therefore, the GB transform can be considered a convergence booster of classical HPM for fractional equations. In addition, Table 8 reveals the convergence phenomenon with rapidly diminishing errors of consecutive approximations with increasing n for $\alpha = 0.75$. The consistency in terms of accuracy of the approximate solution in both cases of fractional and integer order equations confirms the efficiency of the proposed method.

CONCLUSION

In this article, the time-fractional generalized Kuramoto-Sivashinsky equation, Swift-Hohenberg equation, Cahn-Hilliard equation, and their special cases have been studied using the semi-analytic generalized bivariate homotopy perturbation method. Dynamics of several waveforms have been analyzed for different values of fractional order and physical parameters. Existence and uniqueness are presented for continuous solutions of fractional variants

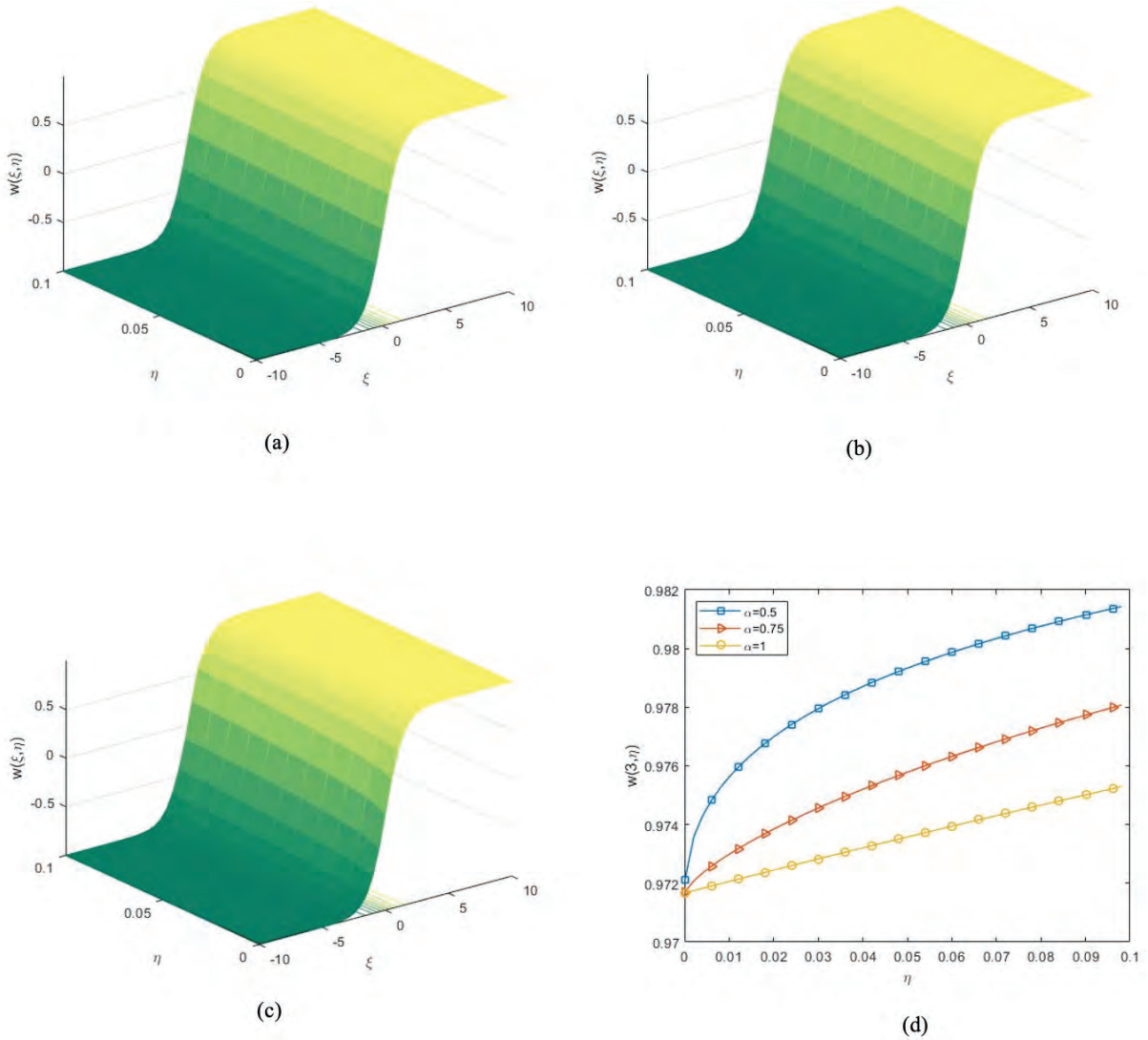


Figure 12. Solution profile of GB-HPM solution for Example 5 when x = 1 at (a) α = 0.5 (b) α = 0.75 (c) α = 1 (d) ξ = 3.

of Kuramoto-Sivashinsky, Swift-Hohenberg and Cahn-Hilliard equations. The suggested method is free from the assumption of requiring major or minor physical parameters in the problem. The graphical illustrations and tabulated absolute error of given results in contrast to existing analytic results establish the worthiness of the proposed scheme. While addressing non-linear problems by generalized bivariate homotopy perturbation method, the assessment of He's polynomials is much easier than the Adomain's polynomials. Generalized bivariate homotopy perturbation method is found highly competent in targeting fractional order challenges. Another advantage is that this method does not require any discretization of variables or linearization of non-linear terms, which preserves actual non-linear effects, large computer memory and extensive time. Beyond this, generalized bivariate homotopy perturbation method does not require exact solution to prove its credibility. For aspects of further advancements, the application of the proposed scheme to stochastic, hyperbolic and higher-dimensional systems can provide good supplements to the existing literature.

ACKNOWLEDGEMENT

Shelly Arora is thankful to SERB-POWER for grant support to complete this research via grant number SPG/2022/001269.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript

REFERENCES

[1] Çulha Ünal S. Approximate solutions of the fractional Harry Dym equation. Sigma J Eng Nat Sci 2024;42:1604–1611. [CrossRef]

[2] İskenderoğlu G, Kaya D. Lie symmetry analysis of Caputo time-fractional K(m,n) model equations with variable coefficients. Sigma J Eng Nat Sci 2024;42:885–899. [CrossRef]

[3] Merdan M, Öktem S. On solutions of time fractional order random HIV/AIDS modelling. Sigma J Eng Nat Sci 2024;42:1899–1906. [CrossRef]

[4] Günerhan H, Kaabar MK, Çelik E. Novel analytical and approximate-analytical methods for solving the nonlinear fractional smoking mathematical model. Sigma J Eng Nat Sci 2023;41:331–343.

[5] Yépez-Martínez H, Gómez-Aguilar F, Sosa IO, Reyes JM, Torres-Jiménez J. The Feng's first integral method applied to the nonlinear mKdV space-time fractional partial differential equation. Rev Mex Fis 2016;62:310–316. [CrossRef]

[6] Baleanu D, Machado JA, Luo AC. Fractional dynamics and control. New York: Springer; 2011. [CrossRef]

[7] Magin RL. Fractional calculus in bioengineering. Crit Rev Biomed Eng 2004;32:1–104. [CrossRef]

[8] Ellahi R, Alamri SZ, Basit A, Majeed A. Effects of MHD and slip on heat transfer boundary layer flow over a moving plate based on specific entropy generation. J Taibah Univ Sci 2018;12:476–482. [CrossRef]

[9] Li C, Qian D, Chen Y. On Riemann-Liouville and Caputo derivatives. Discrete Dyn Nat Soc 2011;2011:562494. [CrossRef]

[10] Tenreiro-Machado JA, Silva MF, Barbosa RS, Jesus IS, Reis CM, Marcos MG, Galhano AF. Some applications of fractional calculus in engineering. Math Prob Eng 2010;2010:639801. [CrossRef]

[11] Lotfi A, Dehghan M, Yousefi SA. A numerical technique for solving fractional optimal control problems. Comput Math Appl 2011;62:1055–1067. [CrossRef]

[12] Engheia N. On the role of fractional calculus in electromagnetic theory. IEEE Antennas Propag Mag 1997;39:35–46. [CrossRef]

[13] Cruz-Duarte JM, Rosales-Garcia J, Correa-Cely CR, Garcia-Perez A, Avina-Cervantes JG. A closed form expression for the Gaussian–based Caputo–Fabrizio fractional derivative for signal processing applications. Commun Nonlinear Sci Numer Simul 2018;61:138–148. [CrossRef]

[14] Sweilam NH, Abou Hasan MM, Baleanu D. New studies for general fractional financial models of awareness and trial advertising decisions. Chaos Solitons Fractals 2017;104:772–784. [CrossRef]

[15] Kilbas AA, Srivastava HM, Trujillo JJ. Theory and applications of fractional differential equations. Amsterdam: Elsevier; 2006.

[16] Diethelm K, Garrappa R, Giusti A, Stynes M. Why fractional derivatives with non-singular kernels should not be used. Fract Calc Appl Anal 2020;23:610–634. [CrossRef]

[17] Aydın M. An explicit solution of linear conformable systems with variable coefficients. Sigma J Eng Nat Sci 2024;42:1806–1812. [CrossRef]

[18] Ansar R, Abbas M, Mohammed PO, Al-Sarairah E, Gepreel KA, Soliman MS. Dynamical study of coupled Riemann wave equation involving conformable, beta, and M-truncated derivatives via two efficient analytical methods. Symmetry 2023;15:1293. [CrossRef]

[19] Lakestani M, Dehghan M. Numerical solutions of the generalized Kuramoto–Sivashinsky equation using B-spline functions. Appl Math Model 2012;36:605–617. [CrossRef]

[20] Zayed EM, Nofal TA, Gepreel KA. The travelling wave solutions for non-linear initial-value problems using the homotopy perturbation method. Int J Control 2009;88:617–634. [CrossRef]

[21] Song L, Zhang H. Application of homotopy analysis method to fractional KdV–Burgers–Kuramoto equation. Phys Lett A 2007;367:88–94. [CrossRef]

[22] Yin W, Xu F, Zhang W, Gao Y. Asymptotic expansion of the solutions to time-space fractional Kuramoto-Sivashinsky equations. Adv Math Phys 2016;2016:4632163. [CrossRef]

[23] Shah R, Khan H, Baleanu D, Kumam P, Arif M. A semi-analytical method to solve family of Kuramoto–Sivashinsky equations. J Taibah Univ Sci 2020;14:402–411. [CrossRef]

[24] Veeresha P, Prakasha D. Solution for fractional Kuramoto–Sivashinsky equation using novel computational technique. Int J Appl Comput Math 2021;7:33. [CrossRef]

[25] Ebadi G, Biswas A. Application of G'/G–expansion method to Kuramoto–Sivashinsky equation. Acta Math Appl Sin Engl Ser 2016;32:623–630. [CrossRef]

[26] Nazari-Golshan A. Fractional generalized Kuramoto-Sivashinsky equation: Formulation and solution. Eur Phys J Plus 2019;134:565. [CrossRef]

[27] Noor MA, Mohyud-Din ST, Waheed A. Exp-function method for solving Kuramoto–Sivashinsky and Boussinesq equations. J Appl Math Comput 2009;29:1–13. [CrossRef]

[28] Kilicman A, Silambarasan R. Modified Kudryashov method to solve generalized Kuramoto-Sivashinsky equation. Symmetry 2018;10:527. [CrossRef]
[29] Wazwaz AM. New solitary wave solutions to the Kuramoto-Sivashinsky and the Kawahara equations. Appl Math Comput 2006;82:1642–1650. [CrossRef]
[30] Sahoo S, Ray SS. New approach to find exact solutions of time fractional Kuramoto–Sivashinsky equation. Physica A 2015;434:240–245. [CrossRef]
[31] Conte R, Musette M. Painleve analysis and Backlund transformation in the Kuramoto-Sivashinsky equation. J Phys A Math Gen 1989;22:169. [CrossRef]
[32] Hyman JM, Nicolaenko B. The Kuramoto-Sivashinsky equation: A bridge between PDEs and dynamical systems. Physica D Nonlinear Phenomena 1986;18:113–126. [CrossRef]
[33] Hooper A, Grimshaw R. Travelling wave solutions of the Kuramoto-Sivashinsky equation. Wave Motion 1988;10:405–420. [CrossRef]
[34] Frisch U, She ZS, Thual O. Viscoelastic behaviour of cellular solutions to the Kuramoto-Sivashinsky model. J Fluid Mech 1986;168:221–240. [CrossRef]
[35] Abdel-Gawad HI, Abdusalam HA. Approximate solutions of the Kuramoto–Sivashinsky equation for periodic boundary value problems and chaos. Chaos Solitons Fractals 2001;12:2039–2050. [CrossRef]
[36] Swift J, Hohenberg PC. Hydrodynamic fluctuations at the convective instability. Phys Rev A 1977;15:319–328. [CrossRef]
[37] Malygin G. Mechanism of the formation of shear microbands under plastic deformation of nanocrystalline materials. Phys Solid State 2009;51:1814–1820. [CrossRef]
[38] Bodenschatz E, Pesch W, Ahlers G. Recent developments in Rayleigh-Bénard convection. Annu Rev Fluid Mech 2000;32:709–778. [CrossRef]
[39] Lega J, Moloney JV, Newell AC. Swift-Hohenberg equation for lasers. Phys Rev Lett 1994;73:2978. [CrossRef]
[40] Blair D, Aranson IS, Crabtree GW, Vinokur V, Tsimring LS, Josserand C. Patterns in thin vibrated granular layers: Interfaces, hexagons, and superoscillons. Phys Rev E 2000;61:5600. [CrossRef]
[41] Peletier LA, Rottschäfer V. Large time behaviour of solutions of the Swift–Hohenberg equation. C R Math 2003;336:225–230. [CrossRef]
[42] Fife PC. Pattern formation in gradient systems. In: Handbook of Dynamical Systems 2. Amsterdam: Elsevier Science; 2002.
[43] Hoyle RB. Pattern formation: An introduction to methods. Cambridge University Press; 2006. [CrossRef]
[44] Mandel P. Theoretical problems in cavity nonlinear optics. Cambridge University Press; 2005.
[45] Ryabov PN, Kudryashov NA. Non-linear waves described by the generalized Swift-Hohenberg equation. J Phys Conf Ser 2017;788:012032. [CrossRef]
[46] Braaksma B, Iooss G, Stolovitch L. Proof of quasipatterns for the Swift–Hohenberg equation. Commun Math Phys 2017;353:37–67. [CrossRef]
[47] Burke J, Houghton SM, Knobloch E. Swift-Hohenberg equation with broken reflection symmetry. Phys Rev E Stat Nonlinear Soft Matter Phys 2009;80:036202. [CrossRef]
[48] Kudryashov NA, Sinelshchikov DI. Exact solutions of the Swift–Hohenberg equation with dispersion. Commun Nonlinear Sci Numer Simul 2012;17:26–34. [CrossRef]
[49] Pavani K, Raghavendar K. Approximate solutions of time-fractional Swift–Hohenberg equation via natural transform decomposition method. Int J Appl Comput Math 2023;9:29. [CrossRef]
[50] Veeresha P, Prakasha DG, Baleanu D. Analysis of fractional Swift-Hohenberg equation using a novel computational technique. Math Methods Appl Sci 2020;43:1970–1987. [CrossRef]
[51] Almutlak SA, Shah R, Weera W, El-Tantawy SA, El-Sherif LS. Fractional view analysis of Swift-Hohenberg equations by an analytical method and some physical applications. Fractals 2022;6:524. [CrossRef]
[52] Li W, Pang Y. An iterative method for time-fractional Swift-Hohenberg equation. Adv Math Phys 2018;2018:2405432. [CrossRef]
[53] Grasselli M, Schimperna G, Zelik S. On the 2D Cahn–Hilliard equation with inertial term. Commun Partial Differ Equ 2009;34:137–170. [CrossRef]
[54] Lee D, Huh JY, Jeong D, Shin J, Yun A, Kim J. Physical, mathematical, and numerical derivations of the Cahn–Hilliard equation. Comput Mater Sci 2014;81:216–225. [CrossRef]
[55] Chen W, Wang C, Wang X, Wise SM. Positivity-preserving, energy stable numerical schemes for the Cahn-Hilliard equation with logarithmic potential. J Comput Phys X 2019;3:100031. [CrossRef]
[56] Zhang X, Li H, Liu C. Optimal control problem for the Cahn–Hilliard/Allen–Cahn equation with state constraint. Appl Math Optim 2020;82:721–754. [CrossRef]
[57] Storvik E, Both JW, Nordbotten JM, Radu FA. A Cahn–Hilliard–Biot system and its generalized gradient flow structure. Appl Math Lett 2022;126:107799. [CrossRef]
[58] Manzanero J, Redondo C, Rubio G, Ferrer E, Rivero–Jiménez Á. A discontinuous Galerkin approximation for a wall–bounded consistent three–component Cahn–Hilliard flow model. Comput Fluids 2022;225:104971. [CrossRef]
[59] Khan MA, Akbar MA, Binti-Abd-Hamid NN. Traveling wave solutions for space-time fractional Cahn-Hilliard equation and space-time fractional symmetric regularized long-wave equation. Alex Eng J 2021;60:1317–1324. [CrossRef]

[60] Zhou S, Xie YM. Numerical simulation of three-dimensional multicomponent Cahn–Hilliard systems. Int J Mech Sci 2021;198:106349. [CrossRef]
[61] Akgül A, Öztürk G. Application of the Sumudu transform to some equations with fractional derivatives. Sigma J Eng Nat Sci 2023;41:1132–1143. [CrossRef]
[62] Aboodh KS. The new integral transform 'Aboodh transform'. Glob J Pure Appl Math 2013;9:35–43.
[63] Saadeh R, Qazza A, Burqan A. A new integral transform: ARA transform and its properties and applications. Symmetry 2020;12:925. [CrossRef]
[64] Saadeh RZ, Ghazal BF. A new approach on transforms: Formable integral transform and its applications. Axioms 2021;10:332. [CrossRef]
[65] Elzaki TM. The new integral transform Elzaki transform. Glob J Pure Appl Math 2011;7:57–64.
[66] Maitama S, Zhao W. Beyond Sumudu transform and natural transform: J-transform properties and applications. J Appl Anal Comput 2020;10:1223–1241. [CrossRef]
[67] Khan ZH, Khan WA. N-transform properties and applications. Nust J Eng Sci 2008;1:127–133.
[68] Maitama S, Zhao W. New integral transform: Shehu transform a generalization of Sumudu and Laplace transform for solving differential equations. Int J Anal Appl 2019;17:167–190.
[69] Yang XJ. A new integral transform with an application in heat-transfer problem. Therm Sci 2016;20:677–681. [CrossRef]
[70] Arora S, Pasrija A. A novel integral transform operator and its applications. Iran J Numer Anal Optim 2023;13:553–575.
[71] Arora S, Dhaliwal SS, Ma WX, Pasrija A. Analysis of fractional order Schrödinger equation with singular and non-singular kernel derivatives via novel hybrid scheme. J Appl Anal Comput 2025;15:1039–1067. [CrossRef]
[72] Taneco-Hernández MA, Morales-Delgado VF, Gómez-Aguilar JF. Fractional Kuramoto–Sivashinsky equation with power law and stretched Mittag-Leffler kernel. Physica A Stat Mech Appl 2019;527:121085. [CrossRef]
[73] Khan Y, Wu Q. Homotopy perturbation transform method for nonlinear equations using He's polynomials. Comput Math Appl 2011;61:1963–1967. [CrossRef]
[74] Ghorbani A. Beyond Adomian polynomials: He polynomials. Chaos Solitons Fractals 2009;39:1486–1492. [CrossRef]
[75] Taneco-Hernández MA, Morales-Delgado VF, Gómez-Aguilar JF. The Willmore functional and instabilities in the Cahn-Hilliard equation. Commun Math Sci 2008;6:309–329. [CrossRef]
[76] Arafa A, Elmahdy G. Application of residual power series method to fractional coupled physical equations arising in fluids flow. Int J Differ Equ 2018;2018:7692849. [CrossRef]
[77] Bouhassoun A, Hamdi–Cherif M. Homotopy perturbation method for solving the fractional Cahn-Hilliard equation. J Interdiscip Math 2015;18:513–524. [CrossRef]
[78] Akinyemi L, Iyiola OS, Akpan U. Iterative methods for solving fourth and sixth-order time-fractional Cahn-Hillard equation. Math Methods Appl Sci 2020;43:4050–4074. [CrossRef]
[79] Ugurlu Y, Kaya D. Solutions of the Cahn–Hilliard equation. Comput Math Appl 2008;56:3038–3045. [CrossRef

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2015

Research Article

Finite element modeling and nonlinear analysis of tire deformation using hyperelastic material models

Ashok MACHE[1,*], Shivraj KOTHEKAR[2], Aniket SALVE[1], Nishant KULKARNI[1]

[1]*Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune, 411037, India*
[2]*Department of Mechanical Engineering, Vishwakarma Institute of Information Technology, Pune, 411048, India*

ARTICLE INFO

Article history
Received: 08 December 2024
Revised: 15 January 2025
Accepted: 03 March

Keywords:
Abaqus; Finite Element Modelling; Hyper Elastic Material; Mooney-Rivlin Model; Tire Deformation Simulation

ABSTRACT

Tires are critical components for automobile and aircraft applications from vehicle dynamics, safety and overall performance perspective. The work in this paper discusses finite lement analysis of tire deformation subjected to 3.5 bar pressure using ABAQUS tool. The hyperelastic material models viz Neo-Hookean and Mooney-Rivlin are used to capture the nonlinear behavior of tire. The simulation focuses on investigation of key tire parameters like tread width, tread pattern, and sidewall thickness in order to assess their impact on stress distribution and deformation. Stress concentrations (320 MPa) were observed in high-stress regions, such as the bead wire, belt edges, and sidewalls, which are imporatnt for maintaining structural integrity. Belt reinforcement were modelled using rebar layer function. This enhanced tread stability by distributing stress and reducing deformation in high stress zones. The results obtained from finite element simulation showed that the Mooney-Rivlin model gives more accuracy in predicting stress-strain behavior as compared to the Neo-Hookean model. Further contact pressure was observed varying across the tread region providing insights into tire-road interactions. The precise material modeling and reinforcement strategies found to be imporatnt for improving tire durability, safety, and performance.

Cite this article as: Mache A, Kothekar S, Salve A, Kulkarni N. Finite element modeling and nonlinear analysis of tire deformation using hyperelastic material models. Sigma J Eng Nat Sci 2026;44(2):843–855.

INTRODUCTION

Tires play a very important role in the automotive field as tires are the only components of a vehicle that are in direct contact with the road surface. Due to this, tires affects the vehicle dynamics, overall safety, and even fuel efficiency. The tire construction and their material properties decides their perfomance as these influnces important characteristics such as braking performance, traction, and also the vehicle handling. In recent developments, sustainability has become an important consideration in the design of tires. Researchers are focusing on developing effective and durable tire materials with reduced environmental impact caused by its production and disposal. Earlier, tyre design iterations and further

***Corresponding author.**
*E-mail address: ashok.mache@viit.ac.in

This paper was recommended for publication in revised form by Editor-in Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

improvements were validated mostly by doing prototype testing. However, physical testing for every alternative design is not only expensive but also time consuming. This does not fully align with the sustainability objective. As a result, numerical simulations based on the finite element method (FEM), have become practical approach for analyzing the tire behavior before carrying out any experimental testing. This approach is an efficient way to simulate and analyze complex mechanical behaviors under various loading conditions. This saves time and cost for prototype development and experimental testing.

Finite element modelling of the tire is carried out in different stages. Initially, in preprocessing stage, the tire geometry is discretised into finite elements called as meshing. Then material properties are assigned which are obtained from experimental testing, laboratory test result data or standard industry catalogues. Proper contact algorithm definitions are provided to represent the interaction between the tire and the road surface. Finite element model is then solved in FEA solver here ABAQUS. Once the analysis is completed, the results are interpreted during the post-processing stage. There parameters such as stresses, strains, deformations, and suitable failure criteria are studied to understand tire performance under different loading and operating conditions. One of the major difficulties in tire analysis is precisely modelling the tyre and road surface interaction. This interaction is basically three-dimensional in nature and involves force distributions at the contact area which is complex in nature. In real conditions, these forces continuously vary with load, road profile and also with inflation pressure. This makes the simulation more challenging. Further complexity arises due to the construction of the tire which has cross-section composite in nature. This is because tire consists of multiple components such as tread, sidewalls, bead cores, belts, and plies. Further they are made of different materials and serving specific structural functions. The overall mechanical behaviour of the tire depends on how these layers interact with each other, and therefore their coupling effects must be modelled and captured in the FE model. In addition to structural complexity, tires are subjected to various loading conditions which includes internal inflation pressure, vertical load due to vehicle weight, impact loads from curbs or potholes, and thermal effects generated by friction during rolling. All these factors together makes tyre modelling a highly detailed and computationally intensive task.

In commercial and non-commercial vehicle, tire mechanics is a very critical area to understand and requires special material models to simulate the deformation which occur in rubber components present in tire. To study this, hyperelastic material models have gaining the importance in simulation work. These material models allow researchers to report the nonlinear stress-strain behaviour of such materials. The previous two decades, the finite element method has become a popular tool for evaluating the performance of tire under different operating conditions. Korunovic et al. [1] conducted simulation work and reported various challenges under static loads on pneumatic tires and further presents a well-structured approach for modelling of tire using composite materials. Furthermore, study also examined how pressure and applied load affect the performance of tires and identified the critical stresses in bead as well as sidewall areas and validate with experimental observations. Top of the this Helnwein et al. [2], developed 3D model for tires with cord reinforcement. In their study, rebar elements were effectively utilized to represent the layers of cords in the structure. A significant advancement for later tyre models, their work used the Mooney material law to obtain accuracy while lowering computing costs.

In addition, Anoop et al. [3] has refined these techniques and compared the various tire balancing conditions by employing finite element analysis to study the effect on tire performance giving more importance on material nonlinearity in tread and sidewall areas. In another study, Nazari et al. [4] reported the direct relation between tread geometry, skid resistance with rubber properties. Similarly, Zhang et al. [5] carried out a study on heavy duty tire application and developed a Rivlin model for radial truck tire under heavy loading condition and accurately predicted deformation patterns along with stress distributions. Dring simulation work, selection of correct hyperelastic model is crucial. Models like Mooney-Rivlin, Neo-Hookean and Ogden are commonly utilized, and their effectiveness varies with end application. Kim et al. [6] confirmed that Mooney-Rivlin and Ogden models gives better results against large deformations while Neo-Hookean model may work for small-strain analyses. A similar kind of trend also reported by Jebur et al. [7] also reported a similar kind of trend in their research work carried out on tires made locally. Apart from rubber matrix, reinforcement layers in modelling are also important. In another study, using multiscale modelling of cord-rubber composites in intelligent tires, Behroozinia et al. [8] in their study demonstrated that reinforcing the layers reduces deformation and maximum stress. Behroozi et al. [9], carried out advanced finite element simulations to study failure of aeroplane tyre when it is inflated. Author demostrated the importance of accurate modelling of tyres to capture realistic behaviour in flated condition. Hernandez et al. [10] studied the effect of pavement stiffness on vertical stress patterns and rolling resistance. The results indicated that fuel efficiency is improved by optimizing tire–pavement interactions. Nowadays, researchers have explored non-pneumatic tires (NPTs), the reason behind this is to offer better safety and lower maintenance costs. Rugsaj and Suvanjumrat [11,12] studied and optimize the spoke design for NPT which reduces the stress concentrations and enhances the load carrying capacity. Similarly, in another study carried out by Kiran et al. [13] optimized tread patterns in non-pneumatic tires, which minimize energy losses. Further research was also carried out on recycled tire materials, such as tire-derived materials.

Montella et al. [14] tested tire-derived materials under compressive and shear loads. Author found that these materials can be useful for reducing vibrations. Hence waste rubber can be reused in construction. Kabe et al. [15] explored the tire corner area and suggested implicit methods for efficient steady-state simulation with detailed transient modelling. Reida et al. [16] prepared and studied a detailed LS-DYNA model for addressing complex impacts such as curb strikes and rock traversal. Sheshenin [17] studied behavior of pneumatic tires during rolling. He suggested using both shell theory and 3D elasticity models together to get more accurate results. The work focuses on complex properties of rubber, like anisotropy and viscoelastic behavior making tire analysis more realistic. Korunovic et al. [18] suggested Marlow model for defining the non-linear behaviour of cord reinforced composites. Additional research work has focused on the dynamic behaviour of the hybrid composite in terms of minimisation of vibration and analysing thermal effects in materials; all these contribute to improving the accuracy of tyre models [19-23] .

Overall, use of finite element modelling improves the tire behaviour prediction under various conditions. Studies underline the tire durability and hyperelastic tread performance and further validation with experimental results [24,25] Furthermore, both experimental and numerical methods improve the material modeling accuracy, making FEA-based tire models more reliable in terms of structural integrity and performance evaluation [26,27].

From above reviewed literature it is evident that, study of advanced simulation techniques like finite element analysis are important to accurately predict the tire performance in actual service condition. However, challenges remain in integrating dynamic loading conditions, thermal effects, and road-tire interactions into these models, suggesting areas for future research. The outcome of this research paper will help new researchers in this field, automotive tyre manufacturing companies and also belt manufacturing companies for virtual testing purposes since experimental testing is challenging and requires special equipment and instrumentation.

MATERIAL AND METHODS

Tire Nomenclature and Structural Components

Tire geometry and structural elements are imporatnt for developing accurate finite element model and carrying out nonlinear analysis using hyper-elastic materials. This section describes imporatnt aspects of tire design, tire construction and operating conditions. Tire size is generally printed on tire sidewalls as shown in Figure 1 [3].

The tire construction code gives information about the internal structure of the tyre. For example, letter "R" is for radial construction and "B" for bias-ply construction, which appear after the aspect ratio in the tyre specification. These letters indicate how the reinforcement cords are arranged inside the tire. In radial ply tires, the cords are placed radially from the centre to the outer edge, running perpendicular to the direction of travel, as shown in Figure 2. This type of construction is commonly preferred in passenger vehicles since it gives better traction during braking and acceleration. This in turn reduces rolling resistance resulting improved fuel efficiency, and ride comfort. Because of these advantages, radial ply tyres are widely preferred in modern automotive applications [28].

Figure 1. Standardized annotation of tire size.

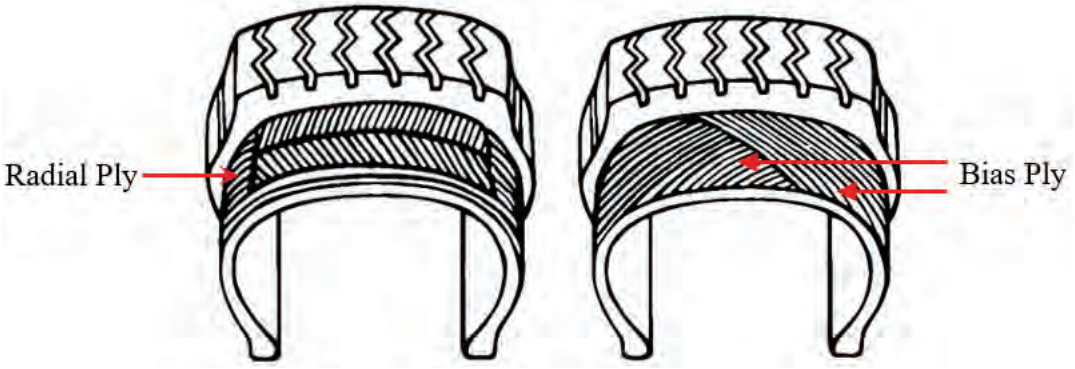


Figure 2. Radial tire configuration with core structural components.

In contrast, in bias-ply tires, cords are arranged diagonally in a crisscross pattern. This makes an angle between successive layers. This construction provides them distinct characteristics. Due to their higher rigidity and strength, bias-ply tires are well suited for heavy-duty and off-road applications. However, their comparatively stiff structure results in a harsher ride, and the crisscross cord arrangement may generate more heat, especially during prolonged high-speed or high-load operation.

Tire size information, typically marked on the sidewall, follows a standardized annotation format. The tire type prefix is an optional letter, such as "P" for passenger cars or "LT" for light trucks. Following this, the tire width in millimeters represents the width from sidewall to sidewall when mounted. The next figure, the aspect ratio, indicates the sidewall height as a percentage of the tire width. The wheel diameter, expressed in inches, specifies the intended wheel size, while an optional speed rating denotes the tire's maximum safe speed. Another Here it is considered a radial tire constituent with three sections.

- **Carcass:** The carcass represents the internal structure of tire. The carcass gives strength and flexibility to tire.
- **Belt:** Belt layer reinforces the the thread of tire. Belt is for improving stability and durability.
- **Bead wire:** Bead wire holds the tire firmly onto the wheel rim and ensure a secure fit. Figure 3 shows radial tire model and its major components.

Tire Materials

Tire materials are made from different components. ASTM D3492 sets standards for the rubber compounds used in tire manufacturing to ensure consistent quality. SAE J2047 provides guidelines for tire performance and durability. Table 1 shows the constituents of tires with percentage and examples.

Tire Manufacturing Process

Tire manufacturing is a multi-stage process that demands precision and expertise. It consist of follwing stages [30].

- Tire components are first assembled on a drum to form a green tire.
- The green tire is then placed inside a mould.
- A bladder inside the mould is inflated to give the tire its final shape.

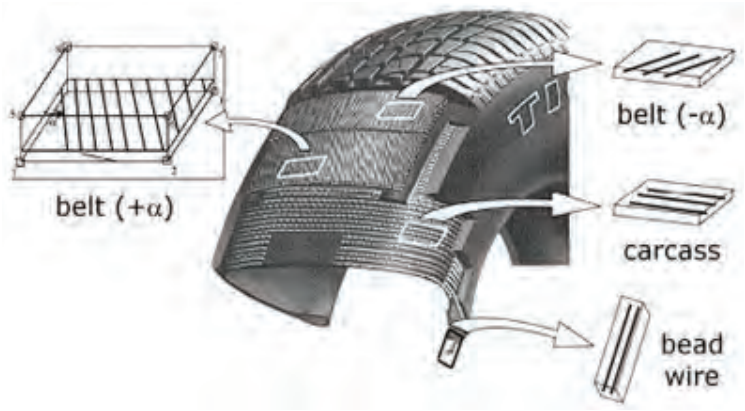


Figure 3. Radial tire and its major components [From Korunović et al. [1], with permission from Serbian Society for Computational Mechanics].

Table 1. Major constituents, ingredients, functions, and environmental aspects of rubber in vehicle tires [From Peterson [29] , with permission from ASME]

Constituent	Composition	Ingredients
Rubber polymer	40–60%	Natural rubber Isoprene rubber Styrene–butadiene rubber, Butadiene rubber, Chlornobyl rubber
Reinforcement aids	25–35%	Mainly carbon black
Softeners	15– 20%	Usually, HA oil but can be replaced with other oils
Activators	2–5%	Mainly Sulphur
Antidegradants	1–2%	Chemical compounds, e.g.: MBT, CBS, TBBS, MBS, HMT, MBTS, DCBS and TMTDb
Vulcanizers	1–2%	Zinc oxide+ fatty acid, or alternatively zinc stearate
Accelerators	0.5–2%	Antioxidants and antiozonants, e.g., TMQ, IPPD, 6PPD and 77PDc

- Heat is applied during curing to make the tire strong and durable.
- Finally, the mould is removed, and the finished tire is ready for use.

Tire Modelling

Rubber material reinforced in construction of tires is characterised by the phenomenon that it deforms under load application and regain its original shape and size when load is removed. This typical elastic behaviour of rubber makes it suitable choice for tire.

This typically follows Hooks law where force and deformation are proportional. In tire modelling, beyond elastic limits, model is not suitable to predict the rubber behaviour.

Elastic modulus is the measure of stiffness. Sidewalls use softer compounds to flex and absorb road bumps. Tread uses stiffer compounds to hold shape under cornering and braking forces. In tire models, this stiffness value directly controls contact patch size, load spread, and grip output.

Elastic limit

Elastic materials in tires flex within the elastic region only up to a defined elastic limit. Beyond elastic region, releasing the load does not restore the material to its original geometry, as the compound has entered the plastic region and the deformation becomes irreversible. The behavior of elastic materials such as mild steel and aluminum is shown below in Figure 4.

Elastic energy storage

When elastic material is subjected to deformation, it stores the energy. This energy is released when material returns to original shape. This is typical application used in springs and rubber band. Unlike this rubber is a hyper elastic material which certainly does not follow Hooke's law. Rubber is often described as a hyper elastic material due to its pronounced nonlinearity in response to deformation [31]. Behaviour of stress-strain graph for hyperelastic materials like rubber is as shown in the below Figure 5.

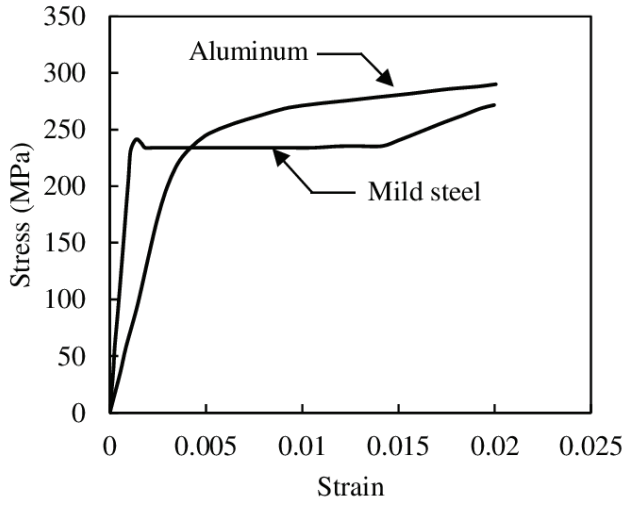


Figure 4. Stress versus strain curves for steel and aluminum.

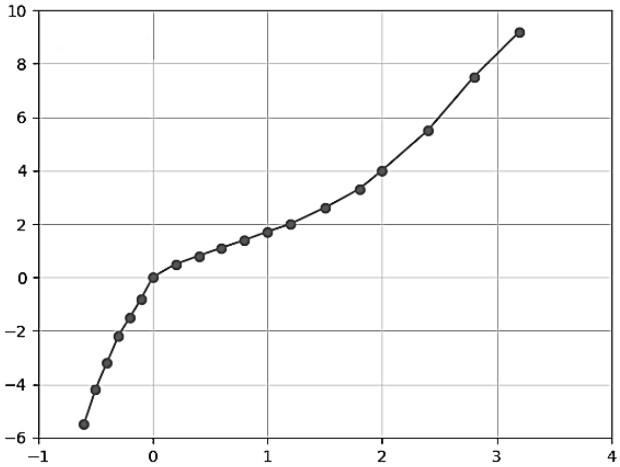


Figure 5. The stress-strain curve of the rubber material.

This behavior can certainly be seen in dynamic rolling of tires. While vehicle is stationary on ground, the tire contact patch is at centerline assuming side point of view. However, the front portion of the tire contacts the ground when the car drives, but the rear tire does not leave the ground as quickly as the front portion of tire encounters ground. This shifts the dynamic contact rearwards. Further this phenomenon causes pneumatic trail, and it has a major role in dynamic handling of vehicles.

1. Strain-Energy Function: Hyper elasticity in rubber is characterized by a strain-energy function, which quantifies the energy stored in the material as it deforms. The neo-Hookean model, for example, expresses the strain-energy function as a function of the deformation gradient.
2. Large Deformations: Rubber exhibits substantial deformations while remaining highly elastic. It can stretch many times its original length, and this behaviour is not described by linear elasticity theory.
3. Nonlinear Stress-Strain Behaviour: In rubber, the stress-strain relationship is nonlinear.
4. Incompressibility: Rubber materials are typically nearly incompressible. Rubber material has Poissons ratio near to 0.5 and they resist changes in volume during deformation.

Material Models for Tire Analysis

Hyperelastic material models viz. Neo-Hookean and Mooney-Rivlin models have been discussed in this section. These are widely used in tire modeling becasuse of their ability to represent the stress-strain behavior like of that rubber materials.

Neo-Hookean model

The A Neo-Hookean model is a hyperelastic material model that can be used for predicting the stress-strain behavior of materials, and the model is like Hooke's law but for hyper elastic materials. A Neo Hookean model is one of the simple models and the strain energy density function for an incompressible Neo-Hookean material is shown as follows [6]:

$$W = C_1(\bar{I}_1 - 3) \quad (1)$$

Where, C_1 is a material constant, and $\bar{I}_1$ is the first invariant of the left Cauchy-Green deformation tensor *C*. It is a tensor that describes how material elements within a deformable body change their relative distances and angles due to deformation.

The left Cauchy-Green deformation tensor *C* is defined as

$$C = F^T.F \quad (2)$$

Where *F* is the deformation gradient tensor, and F^T is its transpose. The deformation gradient tensor captures the local deformation of material points within the body, and $F^T.F$ describe strain energy without considering particle displacements directly. This approach is based on the thermodynamic behavior of cross-linked polymers. Neo-Hookean model effectively predicts behavior at low-strain. At high strains, it fails to provide accurate predictions because there is a rapid increase in the elastic modulus as covalent bonds approach their extension limits. This limitation makes it less accurate for high-stress applications in tires.

Mooney-Rivlin model

It is a progressed version of neo Hookean model, where strain energy is function of two invariants of left Cauchy-Green deformation tensor. The strain energy density function *W* for the Mooney-Rivlin model is expressed as [6]:

$$W = C_1(\bar{I}_1 - 3) + C_2(\bar{I}_2 - 3) \quad (3)$$

Where C_1 and C_1 are material constants determined empirically, and $\bar{I}_1$ and $\bar{I}_2$ represent the first and second invariants of the deviatoric component of *C* respectively. The inclusion of a second invariant $\bar{I}_2$ allows the model to better capture the material's response to multi-axial loading and large deformations, which are critical in tire performance under various operational stresses.

$$G = 2(C_1 + C_2) \quad (4)$$

This relationship explores need of quasi-static temperature dependency adjustments in the Mooney-Rivlin model to account for shear modulus variations as a function of temperature. The Mooney-Rivlin model is widely preferred in simulations of rubber-like materials because precise stress predictions across a range of strain levels and temperatures are crucial in the simulations of applications like tires.

These models provide the mathematical basis for precisely modeling tire materials in finite element analysis, enabling the accurate prediction of tire behavior under real-world conditions. These models possess this capability as they incorporate deformation invariants and strain energy density functions.

In ABAQUS, the parameters "C10," "C01," and "D1" are coefficients used for defining hyperelastic materials. Their roles are varying depending on the selection of specific model. Each parameter contributes in characterizing the material responses towards shear, volumetric, and compressibility effects. This is essential for accurate modeling of rubber-like materials. The breakdown of each term is given here.

- C10: C10 is a coefficient of material used in hyperelastic material models. The response of the material to shear deformation is represented by C10 in hyperelastic modeling. This coefficient meaning varies as per the hyperelastic model selected in ABAQUS.
- C01: C01 is also material coefficient used in some hyperelastic models. It is related to volumetric or dilatational deformation behavior of the material. Similar to C10, the use and exact meaning depend on the selection of specific hyperelastic model.
- D1: D1 is a parameter available under different hyperelastic material models in ABAQUS. It is associated with different material behavior parameters, such as the starting modulus or a scaling factor. Although the precise interpretation of D1 may differ based on the particular material model selected.

Values of the material coefficients required for tire modelling are given in Table 2.

Geometry Modeling and Simulation Setup

Tires are made by stacking different layers in a circular pattern Therefore, the finite element (FE) model of a tire can be developed using a 2D sectional model, which is then axially swept with the required modifications to represent the full geometry. This approach reduces computational time significantly without requiring a full 3D model. Figure 6 illustrates the axisymmetric model developed in the simulation software.

Table 2. Material property used for modelling materials [From Nazari et al. [4], with permission from ASME]

Element	C10 (MPa)	D1 (MPa)
Carcass	0.446	0.045
Sidewall	0.318	0.063
Rim strip	0.930	0.021
Belt	1.022	0.019
Tread	0.685	0.029
Apex	3.512	0.006

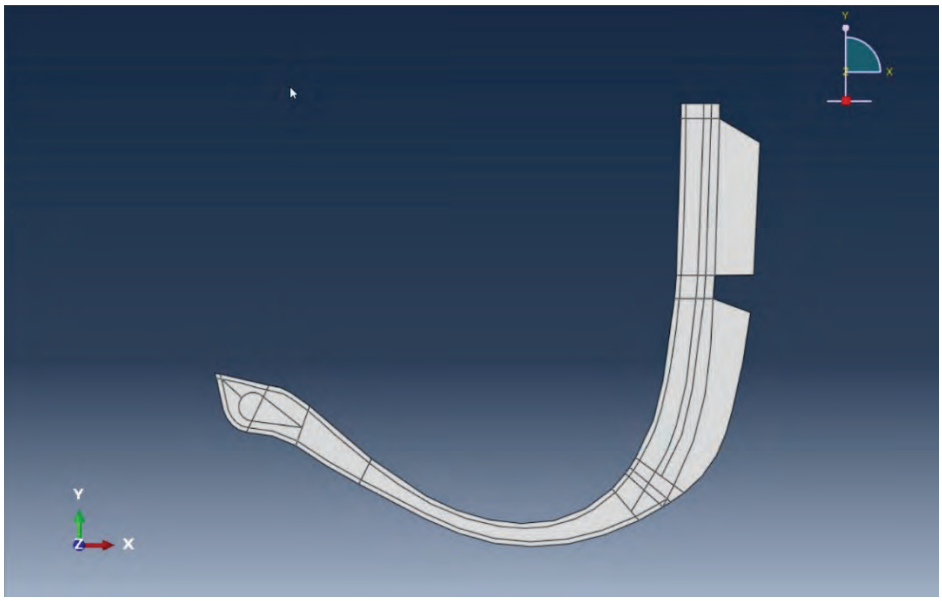


Figure 6. Axisymmetric model built in simulation software.

Figure 7 shows the assembly model of rim and tire components with assumed dimensions for creating the geometry.

After importing the 2D geometry, section creation and material assignment for each section are completed. The Rebar layer function is used to model rubber-reinforced belts [8]. Figure 8 shows this representing parameter definition like area per rebar and orientation angle.

The plies are embedded within the rubber and nylon material is assigned. The rim is modeled as a rigid body. In the post processing step, for rebar, output cards are defined to capture essential simulation results, including stress, rebar orientation angle changes, and reaction forces in the rebar. The contact defination of interference fit is applied for the rim-tire connection and surface smoothness parameter set to 0.2 with finite sliding. This ensures a realistic and secure connection. Four node quad elements are used for meshing.

The carcass, sidewall, and apex are meshed separately. The plies are modeled as wires with 1D element. The

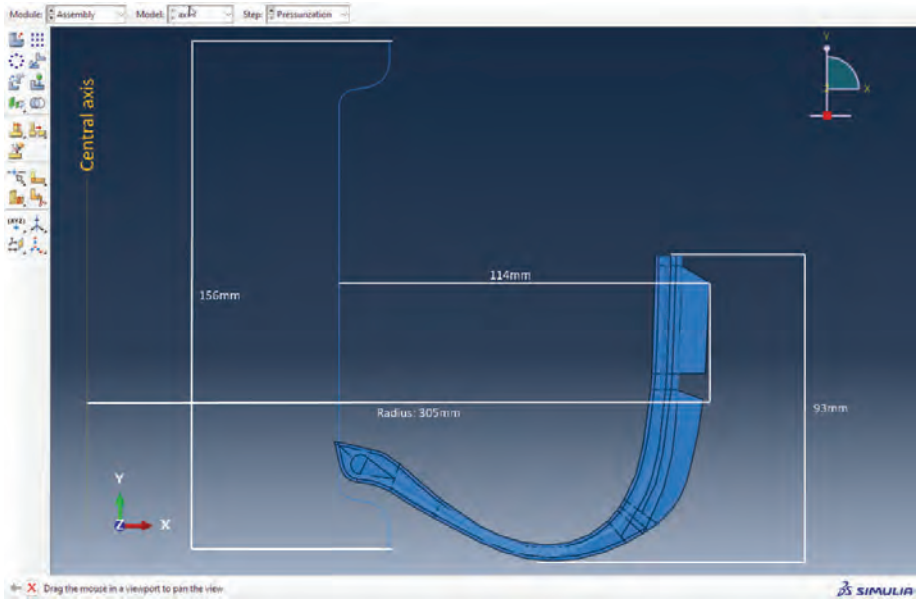


Figure 7. Axisymmetric model built in simulation software.

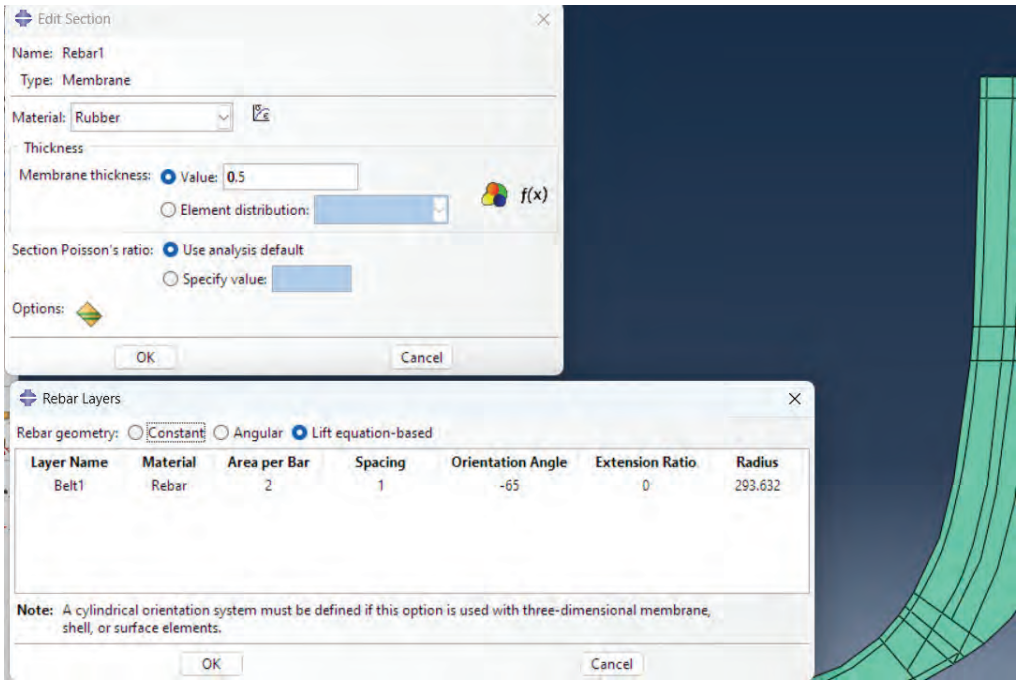


Figure 8. Rebar function for modelling reinforcement in plies.

element types CGAX4 and CGAX3 are used for modelling carcass-sidewall-apex assembly. This allows to capture in-plane stress distributions and structural deformations. For the belt section to have flexibility with twisting, they are meshed using axisymmetric membrane elements with twist (MGAX1), as shown in Figures 9 and 10. Element length of 8 mm is used for meshing carcass, whereas the apex and sidewall regions are meshed with more finer mesh size of 4 mm to capture the localized stress distributions more accurately. Similarly, the plies are also meshed with a 4 mm element size. The finite element model consists of a total of 390 nodes and 367 elements. Among these, 181 elements

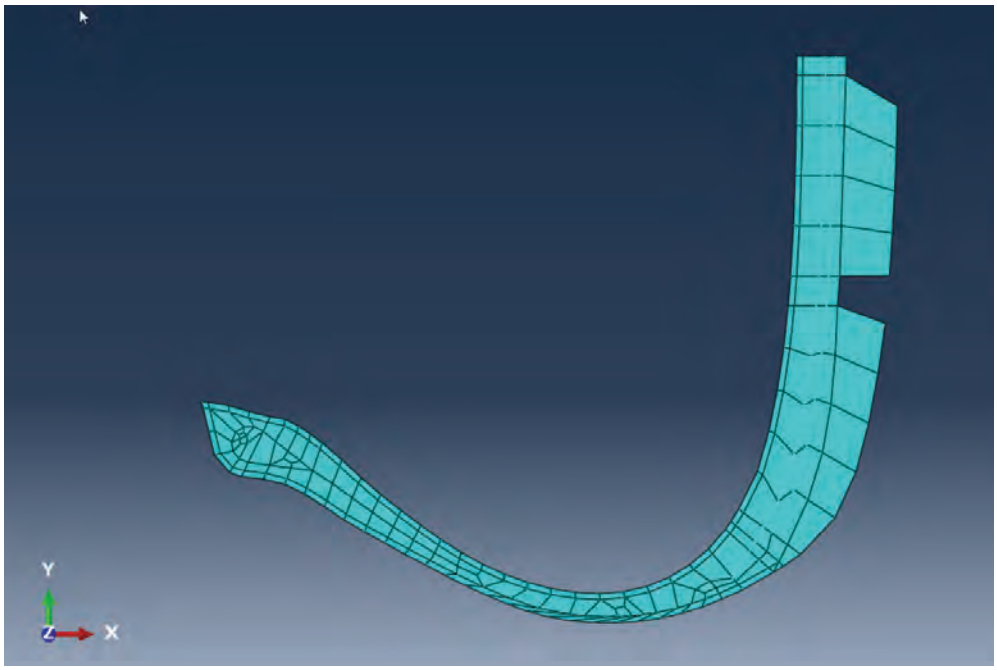


Figure 9. Carcass, sidewall, and apex meshing with suitable element types.

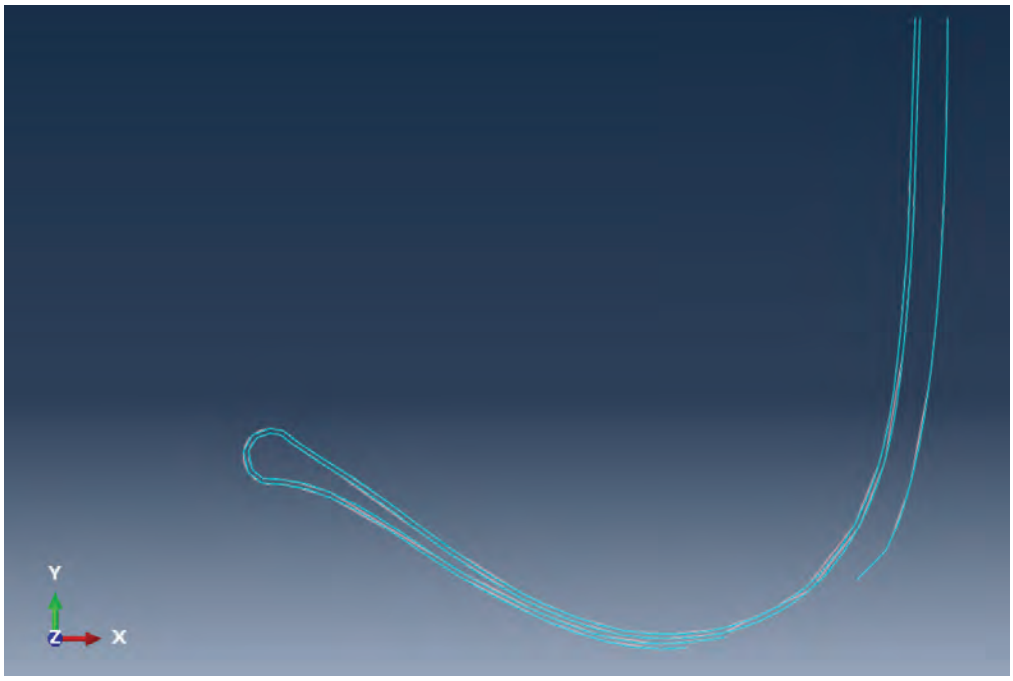


Figure 10. 1D Meshing of plies.

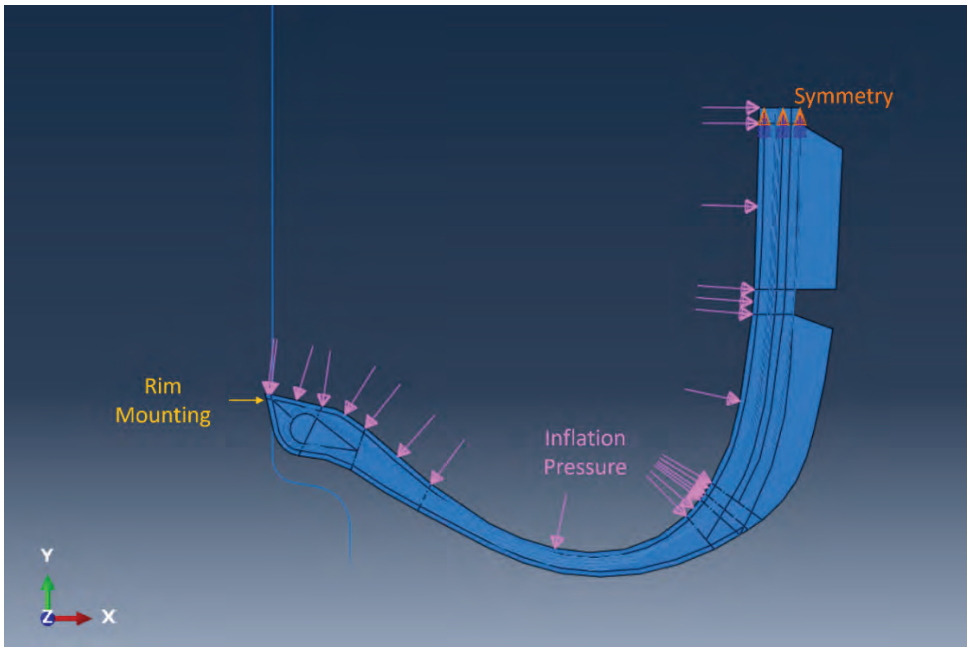


Figure 11. Visualization of boundary condition of an inflated tire.

are of type MGAX1, 184 elements are of type CGAX4, and 2 elements are of type CGAX3. The mesh quality check is performed before carrying out the simulation to ensure all elements pass the test.

Boundary conditions are applied as shown in Figure 11. The displacement boundary condition on the rim constrains all degrees of freedom except the axial direction. Inflation pressure is applied as an internal load of 3.5.

In the post-processing, specific field output requests such as deformation, max stress and reaction force at rebar are added based on the requirements of the simulation.

RESULTS AND DISCUSSION

The behavior of the tire under an inflation pressure of 3.5 Bar is obtained through simulations conducted in ABAQUS. The rubber components are represented using both Neo-Hookean and Mooney-Rivlin hyperelastic material models in the simulations. The stress analysis shows significant stress concentration around the belts edges and sidewalls. These are expected high strain aras in tire structure under inflation. The simulation results provided in Figure 12 show maximum stress reaching 319 MPa in the bead wire.

The stress level as shown is within acceptable limit. The material for bead wire is steel. This ensures its structural reliability for acting load. The significant expansion is observed in the sidewalls and contact patch areas. It is obvious as all degrees of freedom are constrained in simulation except the axial one to mimic real-world mounting conditions.

The energy stored in the material during deformation can be obtained using the strain energy density function, derived from the hyperelastic models. It will be maximum in the areas subjected to substantial strains. In this sense, the hyperelastic properties play a vital role in capturing tire

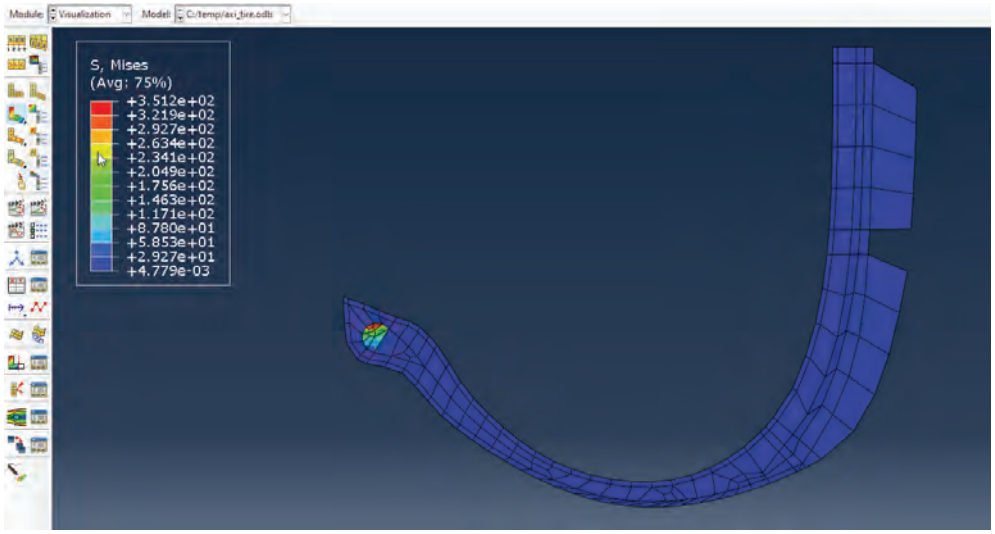


Figure 12. S-misses stress analysis.

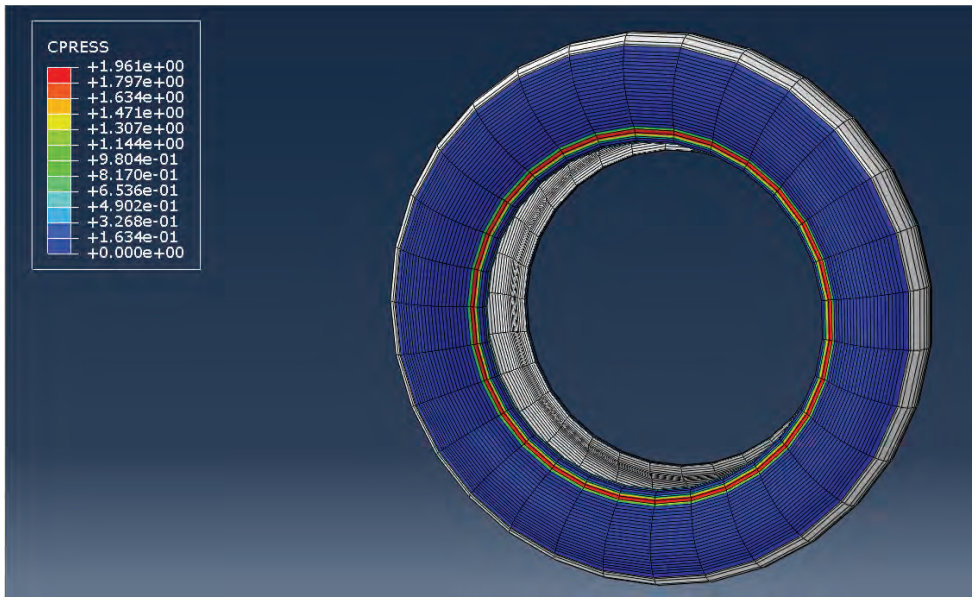


Figure 13. Contact pressure analysis.

material response under loading. The rebar layer function used to model the belt reinforcement mitigates stress and limiting deformation, particularly within the tread area. This reinforcing effect highlights the significance of reinforcement layers in tire performance and can increase durability and safety. Despite being computationally simple, the Neo-Hookean model exhibited limitations at higher strain levels when compred with the Mooney-Rivlin model. The Mooney-Rivlin model offers a more accurate depiction of the material behavior under considerable deformation as it incorporates an extra strain invariant. Therefore, it is concluded that the Mooney-Rivlin model is more suitable for forecasting tire performance in high-stress scenarios.

The details information about the forces acting on contacting surfaces can be derived through CPRESS analysis (contact pressure analysis). This analysis is also able to identify potential failure points and optimizing the designs for load distribution. It can also be used to ensure structural integrity, under situations where high contact pressures could lead to wear, deformation, or component failure. The results obtained by Contact Pressure analysis are provided in Figure 13.

The accuracy of the simulation depends on the factors like selected boundary conditions, the meshing techniques, etc. In the current analysis the meshing is carried out using quadrilateral surface elements and special attention is given to the aeas like sidewall and carcass regions in order to capture detailed deformation and stress gradients accuately. The boundary conditions and loadings applied during simulation study replicate real-world sinarios in order to get realistic insights into tire behavior under inflation. Additionally, the use of 2D elements used in the analysis helps in better visualization of the deformation patterns when swept applied as per need.

The importance of the non linear material properties and reinforcement structures in tire design is successfully underlined in the current research work by simulating the tire response subjected to inflation pressure. It is observed that the durability of the tire can be improved by optimizing the material properties and structural design specially in in high-stress regions. Future research in this area could focus on exploring advanced material models which can take into account thermal effects and investigating alternative reinforcement configurations to develop more safe and durable tires.

The behaviour of the tire rubber is investigated using Neo-Hookean and Mooney-Rivlin hyperelastic models in the current simulations. It is observed that the Mooney-Rivlin model is more accurate at higher strains. This observation is inline with the earlier researchers Gudsoorkar et al. [32] and Hernandez & Al-Qadi [10], who also considered Yeoh and Arruda-Boyce hyperelastic models additionally for tire rubber. The observations in the current study also highlight the limitations of the Neo-Hookean model at high strain levels as reported by Hernandez & Al-Qadi [10].

The stress analysis results show high stress concentration values around 319 MPa in the regions like the bead wire, belt edges, and sidewalls. Similar findings are reported by Zhang et al. [5], while Rugsaj & Suvanjumrat [11], emphasized on the role of reinforcement layers in stress mitigation. The tread stability can be improved using the rebar layer function for belt reinforcements which redirtibutes the stress. Rugsaj & Suvanjumrat [11] employed rebar elements in non-pneumatic tires (NPTs) for load distribution and stress reduction.

The realistic mounting conditions are achieved by constraining all degrees of freedom except axial one in the simulation study. The computational efficiency is ensured with the use of axisymmetric meshing with quadrilateral elements without sacrificing accuracy. Hernandez & Al-Qadi [10] validated boundary conditions using contact area and deflection measurements, while Zhang et al. [31] validated tire models subjected to inflation and normal loading. Similar approach is implemented in the current simulation study.

CONCLUSION

- The paper discusses FE analysis of tire with particular focus on structural configuration and modelling, material properties, and simulation methodology to evaluate its deformation behavior under different loading conditions. The FE simulation results demonstrated that accurate material modelling of the reinforcement layers plays an important role in predicting tyre deformation performance.
- The boundary conditions were applied to represent actual rim mounting conditions. The model was constrained such that movement was allowed only in the axial direction, while all other degrees of freedom (DOFs) were restricted. An axisymmetric mesh with quad elements was used. This reduces the overall computational time without compromising the accuracy of the results.
- The Neo-Hookean and Mooney–Rivlin hyperelastic material models were used to simulate the nonlinear stress–strain characteristics of rubber components and their results are compared. It was observed that the Neo-Hookean model provides good predictions at lower strain levels whereas the Mooney–Rivlin model gives better accuracy at higher strains. Strain energy density was found to be maximum in areas where there was large deformation. This confirms the significance of hyperelastic material modelling in tyre simulations. The comparison further shows the limitations of the Neo-Hookean approach at higher strain levels and can be concluded that the Mooney–Rivlin model stands out a more reliable option for analysing complex tyre behaviour.
- The FE simulations identified critical stress concentration zones, especially in regions such as the bead wire (approximately 320 MPa), belt edges, and sidewalls. These areas are structurally significant under an inflation pressure of 3.5 bar, as they directly contribute to

maintain tyre integrity and prevents the failure. The successful implemntation of rebar layer function to simulate belt reinforcements resulted in uniform distribution of stresses and enhanced tread stability. This is imporatnt for improving tire performance and safety.

- The overall study carried out here indicates the need for proper optimisation of structural design and material parameters to strengthen high-stress regions and improve tyre durability. Future work can extend this simulation by incorporating dynamic loading conditions, thermal effects due to rolling resistance, and alternative reinforcement configurations to develop a more comprehensive and realistic tyre model. The outcomes of this study provide a useful foundation for designing advanced tyres with improved safety, fuel efficiency, and longer service life.

ACKNOWLEDGEMENT

We extend our sincere appreciation to Dassault Systèmes for their support. Their generous support significantly elevated the quality and efficiency of our work.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Korunović N, Trajanović M, Stojković M. FEA of tyres subjected to static loading. J Serbian Soc Comput Mech 2007;1:87–98.

[2] Helnwein P, Liu CH, Meschke G, Mang HA. A new 3-D finite element model for cord-reinforced rubber composites—Application to analysis of automobile tires. Finite Elem Anal Des 1993;14:1–16. [CrossRef]

[3] Anoop PS, Sugumaran V. Comparative study on different balancing conditions of an air filled tyre using statistical features and classification via regression algorithm. IOP Conf Ser Mater Sci Eng 2021;1012:012031. [CrossRef]

[4] Nazari A, Chen L, Battaglia F, Ferris JB, Flintsch G, Taheri S. Prediction of hydroplaning potential using fully coupled finite element-computational fluid dynamics tire models. J Fluids Eng 2020;142:101012. [CrossRef]

[5] Zhang X, Rakheja S, Ganesan R. Nonlinear finite element modelling and incremental analysis of a truck tyre. Int J Heavy Veh Syst 2002;9:253. [CrossRef]

[6] Kim B, Lee SB, Cho S, Park H, Yeom S, Park SH. A comparison among Neo-Hookean model, Mooney-Rivlin model, and Ogden model for chloroprene rubber. Int J Precision Eng Manuf 2012;13:759–764. [CrossRef]

[7] Jebur QH, Hamzah MN, Jweeg MJ, Njim EK, Al-Waily M, Resan KK. Modeling hyperplastic elastomer materials used in tire compounds: numerical and experimental study. J Teknol 2024;86:77–87. [CrossRef]

[8] Behroozinia P, Taheri S, Mirzaeifar R. An investigation of intelligent tires using multiscale modeling of cord-rubber composites. Mech Based Design Struct Machines 2018;46:168–183. [CrossRef]

[9] Behroozi M, Olatunbosun OA, Ding W. Finite element analysis of aircraft tyre – Effect of model complexity on tyre performance characteristics. Mater Des 2012;35:810–819. [CrossRef]

[10] Hernandez JA, Al-Qadi IL. Tire–pavement interaction modelling: hyperelastic tire and elastic pavement. Road Mater Pavement Des 2017;18:1067–1083. [CrossRef]

[11] Rugsaj R, Suvanjumrat C. Study of geometric effects on nonpneumatic tire spoke structures using finite element method. Mech Based Design Struct Machines 2022;50:2379–2399. [CrossRef]

[12] Phromjan J, Suvanjumrat C. A suitable constitutive model for solid tire analysis under quasi-static loads using finite element method. Eng J 2018;22:141–155. [CrossRef]

[13] Kiran M, Aswath M, Shishir DA, Ponangi BR, Bhanumurthy R. Effect of material models on rolling resistance of non-pneumatic tires with hexagonal spokes. SAE Int J Commer Veh 2022;16:33–47. [CrossRef]

[14] Montella G, Calabrese A, Serino G. Mechanical characterization of a tire derived material: experiments, hyperelastic modeling and numerical validation. Constr Build Mater 2014;66:336–347. [CrossRef]

[15] Kabe K, Koishi M. Tire cornering simulation using finite element analysis. J Appl Polym Sci 2000;78:1566–1572. [CrossRef]
[16] Reida JD, Boesch DA, Bielenberg RW. Detailed tire modeling for crash applications. Int J Crashworthiness 2007;12:521–529. [CrossRef]
[17] Sheshenin SV. Three-dimensional modeling of tires. Mech Solids 2007;42:338–345. [CrossRef]
[18] Korunović N, Fragassa C, Marinković D, Vitković N, Trajanović M. Performance evaluation of cord material models applied to structural analysis of tires. Compos Struct 2019;224:111006. [CrossRef]
[19] Kumar KS, Muniamuthu S. Design and fabrication of nickel-chromium reinforced 2-stage energy efficient pyrolysis reactor for waste plastics applications. Int J Ambient Energy 2024;45:57–67. [CrossRef]
[20] Kumar Sunil et al. Experimental and theoretical analysis of exhaust manifold by uncoated and coated ceramics (Al2O3, TiO2 and ZrO2). Case Studies Therm Eng 2023;50:103465. [CrossRef]
[21] Muniamuthu S, Sunil Kumar K, Raja K, Rupesh PL. Dynamic characterization of hybrid composite based on flax/E-glass epoxy composite plates. Mater Today Proc 2022;59:1786–1791. [CrossRef]
[22] Sunil Kumar K, Arun S, Mohan A. Experimental analysis of noise and vibration reduction in windmill gear box for 5MW wind turbine. Int J Mech Eng Technol 2016;7:76–85.
[23] Muniamuthu S, Sunil Kumar K, Raja K, Rupesh PL. Dynamic characterization of hybrid composite based on flax/E-glass epoxy composite plates. Mater Today Proc 2022;59:1786–1791. [CrossRef]
[24] Fathi H, El-Sayegh Z, Ren J, El-Gindy M. Modeling and validation of a passenger car tire using finite element analysis. Vehicles 2024;6:384–402. [CrossRef]
[25] Gudsoorkar U, Bindu R. Computer simulation of hyper elastic re-treaded tire rubber with ABAQUS. Mater Today Proc 2021;43:1992–2001. [CrossRef]
[26] Arachchi NMLW, Abegunasekara CD, Premarathna WAAS, Jayasinghe JASC, Bandara CS, Ranathunga RRMSK. Finite element modeling and simulation of rubber based products: application to solid resilient tire. In: Dissanayake R, Mendis P, Weerasekera K, De Silva S, Fernando S (Eds). ICSECM 2019. Lecture Notes in Civil Engineering, vol 94. Singapore: Springer; 2021. p. 517–531. [CrossRef]
[27] Jebur QH, Hamzah MN, Jweeg MJ, Njim EK, Al-Waily M, Resan KK. Modeling hyperplastic elastomer materials used in tire compounds: numerical and experimental study. J Teknol 2024;86:77–87. [CrossRef]
[28] Thomas MV. Tire contact patch characterization through finite element modeling and experimental testing. Virginia: Virginia Tech 2016.
[29] Peterson MB. Wear control handbook. New York, USA: American Society of Mechanical Engineers; 2000.
[30] Stîngă F, Severin I, Mitrache IA, Lascu E. Redesign of the curing area of the tire manufacturing process. Sustainability 2020;12:6909. [CrossRef]
[31] Qin W, Qin H, Zheng H, Zhang Z. The coupled effect of bearing misalignment and friction on vibration characteristics of a propulsion shafting system. Proc Inst Mech Eng, Part M: J Eng Maritime Environ 2019;233:150–163. [CrossRef]
[32] Gudsoorkar U, Bindu R. Computer simulation of hyper elastic re-treaded tire rubber with ABAQUS. Mater Today Proc 2021;43:1992–2001. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 856–868, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00171

Research Article

Deep learning-based approach for automated detection of lumpy skin disease

Rohit SINGH[1], Ramraj DANGI[2], Lokesh MALVIYA[2,*], Akshay JADHAV[1]

[1]Manipal University Jaipur, Jaipur, 303007, India
[2]School of Computing Science and Engineering, VIT Bhopal University, Madhya Pradesh, 466114, India

ARTICLE INFO

Article history
Received: 10 October 2024
Revised: 07 December 2024
Accepted: 08 January 2025

Keywords:
Convolution Neural Network; Deep Learning; Lumpy Skin Disease; MobileNetv2; SMOTE

ABSTRACT

The developing viral illness of cattle known as lumpy skin disease (LSD) has terrible economic consequences. The virus responsible for lumpy skin disease is a member of the Poxviridae family and the Capripoxvirus genus. It is an economically significant transboundary disease that affects cattle, water buffalo, and camels. The symptoms of the disease include the development of lumps or nodules on the skin of the animal. This illness is widespread in Africa and the Middle East and has recently appeared in Asia. This article discusses the Lumpy skin disease outbreak and detection among cattle using Mobilenetv2 and Deep learning techniques. The lumpy skin disease dataset is utilized for this experiment and is balanced using the oversampling technique. MobileNetv2, a pre-trained neural network extracts the features from the images for image processing. Later deep learning model with the combination of a two-dimensional convolution neural network, max pooling, flatten and dense layer is utilized for classification purposes. The proposed model outperforms in terms of lumpy skin disease detection; the performance is compared using confusion metrics parameters with a classification accuracy of 99.88%.

Cite this article as: Singh R, Dangi R, Malviya L, Jadhav A. Deep learning-based approach for automated detection of lumpy skin disease. Sigma J Eng Nat Sci 2026;44(2):856–868.

INTRODUCTION

Lumpy disease, also known as lumpy skin disease (LSD), is a viral disease that affects cattle and is caused by the capripoxvirus. Early detection of Lumpy disease is essential to prevent its spread among cattle and reduce economic losses for farmers [1]. The most common detection method is through visual observation of the characteristic skin lesions. More reliable methods of detection include laboratory testing of blood or tissue samples using techniques such as PCR (polymerase chain reaction) and ELISA (enzyme-linked immunosorbent assay) [2].

Cattle, domestic water buffaloes across Africa and even camels [3] were more sus- ceptible to a viral illness known as lumpy skin disease. The migration of infected animals may be responsible for the long-distance spread of LSD. Still, the disease's contrasting seasonal patterns point to the possible involvement of arthropod-borne transmission in its rapid

***Corresponding author.**
*E-mail address: lokeshmalviya@vitbhopal.ac.in.

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

and aggressive short-distance expansion. Recent research suggests that the common house fly, Musca domestica, and other blood-sucking arthropods may have a role in LSDV transmission; however, this has not been shown in a clinical context [4].

Bangladesh reported the first case of the disease in South Asia in July 2019. In August 2019, India announced the first-ever case of Lumpy skin disease. Since May 2022, LSD has spread over various states in India [5]. Vaccination, bovine animal movement restriction, and quarantine are among of the current preventative measures being used, along with the implementation of biosecurity through vector management. In India, the standard method of treatment is vaccination with a goat pox vaccine that provides 60-70% protection against Lumpy skin disease. This vaccine takes 15-20 days to become fully effective. The Indian Council for Agriculture Research National Research Centre on Equines (ICAR-NRCE) in Hisar (Haryana) and the Indian Vet- erinary Research Institute (IVRI) in Uttar Pradesh have been working together to produce an indigenous vaccine called Lumpi-ProVacInd. There is no proof that it has any effect on people. There is a greater variation in sickness and mortality rates among cattle in the current epidemic in India compared to the previous outbreak [6]. LSD may also reduce the export of livestock and livestock goods [7].

The research mentioned in this article focuses on predicting LSD using machine and deep learning methodologies. Many academics have been working on LSD classi- fication during the last decade in order to develop very efficient machine learning and deep learning models. Because of the increasing accuracy and efficiency of prediction, the use of artificial intelligence and machine learning algorithms has gained promi- nence in recent years. The importance of research in this sector is in the capacity to construct and choose models with optimum accuracy and efficiency while minimising false negatives. The paper utilised image datasets of cattle infected and disinfected are two major classifications. The dataset is pre-processed and the later model is pre-trained using MobileNetv2, convolution neural network (CNN) models, utilised for feature extraction and classification respectively. Hyper-parameter tuning is per- formed for classifiers using a Keras tuner, and the performance of the proposed model is compared with the individual models and the existing work in the same domain over confusion metric parameters; Accuracy, Precision, F1-score, and Recall. The graphical comparison is also shown in the form of ROC and convergence curves. The overall architecture of the lumpy skin disease detection system adapted in this research is explained in Figure 1.

Key Contributions

- The dataset is pre-processed in order to remove unwanted noise and extract distorted images from the lumpy image dataset.
- The LSD dataset is imbalanced, in order to handle imbalanced data synthetic minority oversampling technique (SMOTE) is utilised.
- After successful preprocessing, the features are extracted from the images using MobileNetv2, which itself is a neural network-based classification algorithm for image datasets.
- The hyper-parameter tuning is performed for classifiers using a Keras tuner.
- The classification of the images as infected and disinfected is performed using a two-dimensional convolution neural network model.
- The performance of the models is evaluated using confusion metric parameters, ROC and convergence curves.

RELATED WORK

This sub-section describes the existing work and literature in the field of LSD by various researchers. The paper [8] used machine learning techniques to predict the prevalence of Lumpy skin illness using meteorological and geological factors. In order to predict the prevalence of a disease in unseen (test) data, the author first utilized the Extra Trees

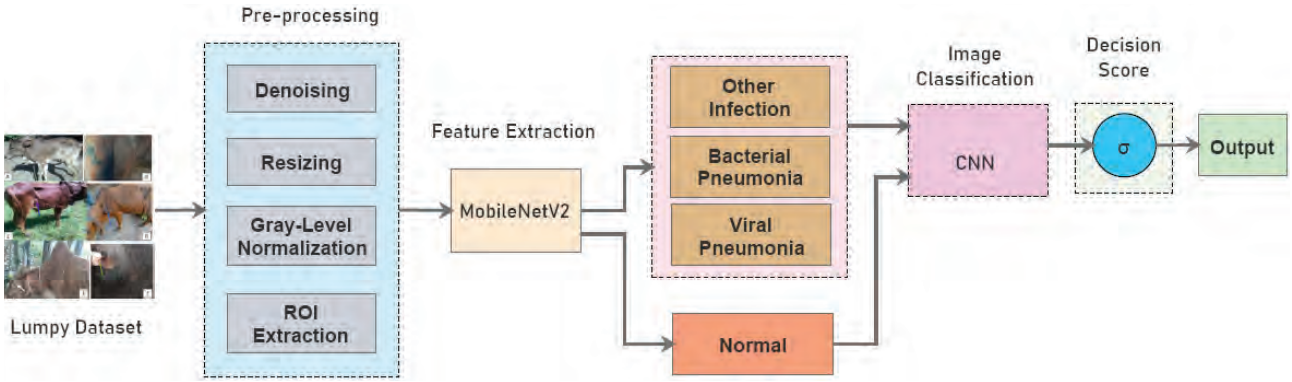


Figure 1. Basic architectural diagram of lumpy disease detection system.

Classifier method to pick out the most essential predictive factors, such as weather, animal population density, dominating land cover, and elevation. Test results showed that some machine learning methods were up to 97% accurate in predicting the development of the Lumpy skin illness. In predicting the existence of Lumpy skin infection in unseen data, the ANN model outperformed other machine learning algorithms, with an AUC and F1 performance measure score of 0.97 and 0.94, respectively.

This qualitative risk assessment was undertaken by the authors of [9] to learn more about the potential for the arrival of the LSD virus to Ukraine and the subsequent spread of the disease inside the country in the next year. The suggestions of the OIE import risk analysis for animals and animal products were used to inform a risk assessment that was carried out with the assistance of local experts assembled in an expert elicitation workshop. The Delphi approach was updated to get advice from specialists.

The research of [10] centered on the detection of illness in cattle. A mobile phone camera was used to record the outward manifestations of sickness in cattle. Information on palpated symptoms was gathered using conversational text. The picture analysis subsystem used a convolutional neural network (CNN) method to classify the symptoms into their respective categories. With 95% accuracy, the system categorized the given symptoms. The reasoner part of the expert system drew the final diagnostic conclusion by combining picture categorization results, location, and text input from the users. As proof of concept, we built a system that uses picture classification techniques and a reasoner. The created system's assessment results confirmed the new diagnostic system's capability of providing an efficient and timely diagnosis of cattle illnesses.

This article [11] presented a model to forecast if the cattle in a given geographic region are at risk of contracting Lumpy skin infection illness now or in the future. In this research, the authors use a dataset of instances of lumpy skin disorders to test the efficacy of multiple ML algorithms. Target columns in the dataset may take on values of 0 (showing the absence of lumpy sickness) or 1 (representing the presence of lumpy disease), and there are a total of 18603 cases and 16 characteristics. The authors found that of all the algorithms they examined, Random Forest performed the best.

Deep learning-based segmentation and classification using deep features are suggested for disease classification in this study [12]. Convolutional neural networks with 10 layers were used for this purpose. The framework is tested on a dataset built from gathered data. Due to the importance of the skin's color in determining the location of the damaged region during disease representation, we employed a color histogram to extract features from the input photos. A deep, pre-trained CNN is then utilized to extract features from the segmented region of impacted skin color. The output is then threshold-converted to a binary representation. Classification is performed using the Extreme learning machine (ELM) classifier. On CLSD, the suggested method's classification performance was able to reach an accuracy of 0.9012%.

The goal of this article [13] was to find the optimal settings for a re-sampling strategy for imbalanced datasets (SMOTE) and a hyper-parameter tuning strategy (Genetic Algorithm; GA) for the classification method (Random Forest; RF) while predicting LSD. Experiments indicated that combinations of SMOTE and GA in the RF algorithm improved performance measures such as Recall and F1-scores from 0.90 to 0.99 and AUC from 0.94 to 0.98. If the model has an accuracy of 0.99, it can reliably identify healthy cattle.

The study [14] is evaluated using a dataset that the authors themselves compiled, and its performance in feature extraction is compared to that of other, more cutting-edge methods including KNN, SVM, NB, ANN, and LR. Additionally, VGG-16, VGG- 19, and Inception-v3 were used to extract the features. Finally, Lumpy skin and normal skin were both successfully predicted using this deep convolutional neural network, with an accuracy of 92.5%.

The goal of this study [15] was to provide a web-based system for diagnosing skin diseases such as atopic dermatitis, acne vulgaris, and scabies using a convolutional neural network classifier built on top of the TensorFlow framework. The system was given the name medical-plus.

Red, Green, and Blue (RGB) colour characteristics and Grey Level Co-occurrence Matrix (GLCM) texture features were recovered from digital camera photos of chronic eczema, lichen planus, and plaque psoriasis gathered by the authors [16].

Comparison with Existing Surveys

In this section, Table 1 provides a comprehensive comparison of the discussed research articles, focusing on various datasets used for the analysis of Lumpy Skin Disease (LSD) and other skin diseases. The table highlights the application of machine learning and deep learning methods in these studies.

Challenges of existing approaches

The challenges in the field of skin disease analysis using machine learning and deep learning methods include:

- The creation of reliable models is hampered by the scarcity of complete outbreak data for training and validation.
- Future transmission patterns cannot be reliably predicted since qualitative judgements are used instead of solid facts.
- Accurate symptom categorization is challenging due to a lack of available annotated pictures for training CNN models.
- The generalizability of models that perform well on certain datasets might be compromised by overfitting.

The rest of the article is organized as follows; Section 1 contains introduction and literature related to LSD, along

Table 1. A state-of-the-art description on existing work

Ref	Proposed Model	Methodology	Dataset
[8]	ANN model for predicting Lumpy Skin Infection	The author demonstrated that the ANN performed better than other machine learning approaches in predicting the presence of Lumpy Skin Infection in unseen data. The model achieved an Area Under Curve (AUC) score of 0.97 and an F1 performance measure score of 0.94.	Lumpy skin disease outbreak data.
[9]	Qualitative Risk Assessment	The authors conducted a qualitative risk assessment to evaluate the likelihood of the LSD virus entering Ukraine and its subsequent transmission in the next year.	Live animals data.
[10]	CNN-based symptom classification	A CNN-based method was used for picture analysis to classify symptoms into their respective categories. The system achieved 95% accuracy in symptom categorization.	Debre Markos University, Addis Ababa University-School of Veterinary Medicine, the World Wide Web, and other secondary sources were used for photographic holdings.
[11]	Random Forest algorithm for Lumpy Skin Disease	The authors determined that the Random Forest method had the highest accuracy 97.7% among the tested algorithms.	Lumpy skin disease dataset with 18,603 instances and 16 features
[12]	CNN-based feature extraction and ELM classification	A pre-trained CNN was employed to extract features from segmented regions of impacted skin color. The ELM classifier was used for classification, achieving an accuracy of 0.9012%.	Dataset of different Cattel's Lumpy Skin Disease.
[13]	Random Forest algorithm with SMOTE and GA	Including SMOTE and GA to improved performance measure, such as Recall, F1 score, and AUC. The Recall and F1 score improved 0.99, and AUC improved 0.98.	Lumpy Skin Disease (LSD) dataset with 24,803 data.
[14]	CNN using VGG-16, VGG-19, and Inception-v3 for feature extraction	The developed framework successfully predicted both lumpy skin and normal skin using deep CNN, achieving an accuracy of 92.5%. The performance was compared with other advanced methodologies such as kNN, SVM, NB, ANN, and LR.	Dataset created by the authors.
[15]	Medilab-plus CNN classifier	The suggested method achieved high accuracies in classifying atopic dermatitis (88%), acne vulgaris (85%), and scabies (84.7%) in experimental settings.	Sample data (pictures) from four health care facilities in the Sunyani Municipality of Ghana.
[16]	ML methods for classification	Classifier performance was evaluated using ML methods such as ANN, LDA, Naive Bayes, and SVM for various feature combinations. LDA and SVM performed best in a test of four different classification methods.	Patient's images from the Department of Dermatology, Yenepoya Medical College, Mangaluru

with the comparison of existing surveys and challenges. Section 2 describes the methodologies used in the research. The proposed work is defined in Section 3, along with dataset description, data preprocessing, balancing techniques, and hyper- parameter tuning. Section 4 describes the performance metrics used for the evaluation of the proposed models. The evaluated results are presented and discussed in Section 5. Section 6 concludes the research findings and provides the future scope of further research in this field.

PRELIMINARY

This section includes a description of the methodologies used in this research. It includes the description of data balancing techniques, feature extraction techniques to extract the features from the image dataset and classification techniques for the identification of infected and disinfected cattle.

Synthetic Minority Oversampling Technique (SMOTE)

To overcome the issue of class imbalance, the classification problem of lumpy disease detection strongly depends on data balancing. A class imbalance might manifest itself, for instance, in a considerably lower proportion of lumpy image cases compared to non-lumpy image cases or pneumonia patients. To categorise lumpy diseases, we have acquired a lumpy dataset, which brings some degree of irregularity into the dataset. In the proposed system, we have used the SMOTE Algorithm [17, 18] to balance the data. To create a separate line between the samples in the feature space and create a unique example along the

line, the SMOTE algorithm first chooses examples that are close to the feature space. First, a sample from the under-represented group is randomly chosen. The nearest k neighbours to the sample are then found. K is often equal to 5. At some point in the feature space between the two samples, a neighbour is chosen at random, and a synthetic sample is created. The practice is repeated until the proportion of the minority group equals that of the majority group.

Feature Extraction Models

The paper utilises feature extraction techniques to extract the features of the image dataset. This section discusses various feature extraction techniques and their advantages [19, 20].

DenseNet-121

A popular deep-learning architecture for extracting features from images is DenseNet-121 [21]. In 2017, scientists from Facebook's AI Research Lab unveiled the concept of tightly linked convolutional layers [22]. Each layer in a DenseNet-121 architecture is fed feature maps from the layers above it and contributes its feature maps to the layers below it. This highly interconnected structure reduces the impact of disappearing gradients and promotes feature reuse [23]. DenseNet-121's dense connections may also be seen as a kind of feature aggregation, enabling the network to pick up more evocative elements. DenseNet-121, like ResNet-121, is a pre-trained model useful for extracting features from images. DenseNet-121's intermediate features may be utilized as input to a new classifier or for other downstream tasks like object recognition or picture segmentation by eliminating the final classification layer. DenseNet-121 is often more memory-efficient and has fewer parameters during training and inference than ResNet-121. However, owing to its high connection structure, it may be costlier to compute.

InceptionV3

Google's InceptionV3 [24] deep learning architecture was released in 2015. It is a kind of convolutional neural network (CNN) that is often used for feature extraction and image categorization. To provide a more expressive feature representation, InceptionV3 is notable for its novel design, which consists of numerous branches of convolutional layers with varying kernel sizes. This architecture was created with the idea that many pictures identification jobs benefit from capturing both fine-grained and coarse-grained characteristics at distinct sizes. Simply deleting the model's final classification layer will allow you to use the intermediate characteristics that InceptionV3 collected for downstream tasks like object identification or image segmentation. There are several ways in which InceptionV3 excels above other deep learning models. Parameter efficiency enables it to attain excellent accuracy with fewer parameters than other designs, while the use of numerous branches of convolutional layers allows it to collect information at various sizes. However, the utilization of many branches may cause its computational cost to be greater than that of other structures.

ResNet50

ResNet50 [25] is a deep learning architecture with similarities to ResNet-121 but with additional layers and enhanced feature extraction capabilities. Microsoft Research first introduced it in 2016, and since then, it has seen extensive use in the fields of feature extraction and image categorization. ResNet50 is a 50-layer deep convolutional neural network (CNN). It employs residual blocks, much like ResNet-121, to facilitate the learning of very deep neural networks. ResNet50's shortcut connections simplify the process of training deep neural networks and reduce the occurrence of vanishing gradients. ResNet50, like the other stated deep learning architectures, may be used as a pre-trained model for image feature extraction. ResNet50's intermediate features may be used as input to a new classifier or for other downstream tasks like object identification or picture segmentation by deleting the model's final classification layer. ResNet50 is more capable of feature extraction since it has more layers than ResNet-121. This has the potential to improve performance on increasingly difficult picture identification tasks, but it may need more computing power and training data.

VGG19

In 2014, researchers at Oxford University proposed a deep learning architecture called VGG19 [26]. It's a common tool for image categorization and feature extraction since it's a convolutional neural network (CNN). VGG19's intermediate features may be used as input to a new classifier or for other downstream tasks like object recognition or picture segmentation by deleting the model's final classification layer. There are a number of ways in which VGG19 excels over competing deep learning systems [27]. It uses modest filters to extract characteristics at various scales and has a unified design that makes it easy to learn and execute. The high number of model parameters, however, may make VGG19 costlier to compute than other designs.

MobileNetV2

MobileNetV2 [28] is a deep learning architecture that was introduced by Google in 2018. It is a convolutional neural network (CNN) that is designed for mobile and embedded devices, but it can also be used for image classification and feature extraction tasks on larger machines. MobileNetV2 has several advantages over other deep learning architectures. Its lightweight architecture makes it ideal for resource-constrained environments, and its use of depth-wise separable convolutions allows it to achieve high accuracy with fewer parameters and faster inference times. However, MobileNetV2 may not be as accurate as larger architectures such as ResNet50 or VGG19 for more complex image recognition tasks.

Classification using CNN

To classify the diseased and healthy cattle using images, the section discusses the detailed explanation of the layered deep learning classification models utilised in the experiment. Convolutional neural networks (CNNs) [29] [30] are a popular kind of deep neural network design for image categorization problem. Features are mapped to class probability in the fully connected layers at the network's output node. Adding additional layers, more filters, or more sophisticated methods like residual connections or attention processes may all make a CNN more complicated. These enhancements allow the network to pick up more complicated information and perform better on difficult picture categorization tasks.

A CNN typically consists of multiple layers [31], including:

Convolutional layers: Input data, which is often an image, is processed by these layers using a series of filters. When the filters pass over the picture, they calculate dot products between their components and the images.

Activation functions: These functions are used to incorporate non-linearity into the model and are applied element wise to the convolutional layer's output. In convolutional neural networks (CNNs), the rectified linear unit (ReLU) function is the most used activation function.

Pooling layers: To reduce the computational burden on the model and limit overfitting, these layers reduce the size of the feature maps.

Flatten layer: Flatten layer is typically used to transform the multi-dimensional feature maps into a one-dimensional vector.

Fully connected layers: The final classification or regression operation is carried out by these layers, which compute dot products between the feature maps and a set of weights before feeding the result into an activation function to provide a prediction.

Dropout layer: Overfitting may occur in deep neural networks, however regularisation methods like dropout can help. Figure 2 shows the pictorial representation of CNN architecture.

PROPOSED METHODOLOGY

The section describes the proposed methodology of the research paper, along with the description of the lumpy image dataset utilised. The section also narrates the flow of the proposed work in varying sub-sections.

Dataset Description

Secondary data was utilised for this research. The dataset, titled "Lumpy Skin Disease (LSD)", had 24,803 records and was downloaded from the Mendeley database [8]. There are only 3,039 records in this dataset about livestock with LSD (12.25%), whereas there are 21,764 records in this dataset about cattle without LSD (87.75%). It indicates an unbalanced dataset. The LSD dataset's properties are made up of geographical and

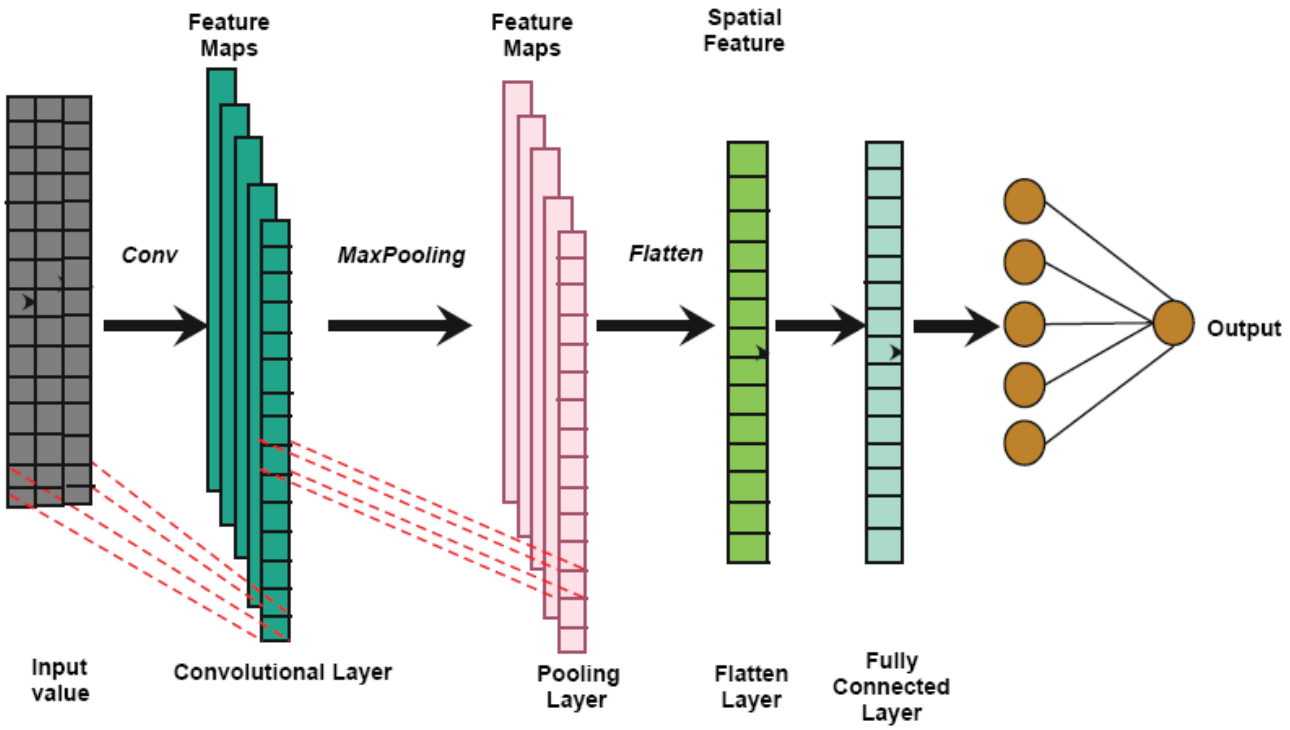


Figure 2. Convolution neural network.

meteorological features [8]. As regards the characteristics of meteorology, they include: Monthly Cloud Cover in Percent is represented by cld; The term "dtr" is for "diurnal temperature range"; "frs" stands for "frost day frequency"; "pet" stands for "potential evapotranspiration" in millimetres per day; and "pre" stands for "precipitation" in millimetres per month.

The following acronyms stand for temperature: tmn, which stands for daily mean, tmp, which stands for temperature, tmx, which stands for monthly average maximum and minimum, vap, which stands for vapour pressure, and wet, which stands for wet day frequency in days.

The geographical features are listed as follows: As with the other categories, "region" refers to the continent, "country" to the nation where the outbreak was first reported, "reporting date" to the day on which the outbreak was first reported, "elevation" to the height above sea level, and "dominant landcover" to the type of vegetation that predominated at the site in question. The X5 Ct 2010 Da file format provides a brief overview of asymmetric cattle GIS data, while the X5 Bf 2010 Da file format does the same for buffalo, and the lumpy file format classifies whether or not an animal is LSD-infected: 1, no code infection: 0.

Data Preprocessessing

There are always no mistakes in the dataset. Text, numbers, time series, continuity, and discontinuity are only a few of the data kinds that may be found in a dataset. Data quality issues such as noise, abnormalities, missing, erroneous, and duplicates may also be present. The data set is either too big or too little. For the model to be effective, the data must agree with it. To make the dataset suitable for the model, pre-processing is required. Data pre-processing includes actions like the elimination of singular features and the handling of missing information. Categorical variables were transformed into numeric values using the one-hot encoding approach so that they could be utilised by ML algorithms. Additionally, min-max scaling was used to standardise the values of several predictive features. In the end, the train test split class in the scikit-learn package was used to divide the dataset into test and training sets. Models were validated using the original training dataset and a new, independent test set.

Data Balancing

SMOTE is used to handle imbalanced data classes. SMOTE constructs synthetic examples by selecting minority class examples at random and employing interpolation techniques to build examples between the target and surrounding instances. The SMOTE method classifies data before sampling it either more or less thoroughly. These three operational elements are the neighbor count used to enter new cases, the up-sampling rate, and the down-sampling rate, respectively. The dataset is now clean and efficient for classification.

Hyperparameter tuning using Keras Tuner

The Keras Tuner library was developed as a hyperparameter tweaking tool for Keras- based deep neural networks. [33]:

- ✓ Define a model: Tuning a deep learning model requires first defining that model. A model constructed using either the sequential or functional Keras APIs will do.
- ✓ Define a hyperparameter search space: Establish a search space for the hyper- parameters you want to tune Next, you need to establish a search space for the hyperparameters you want to tune.
- ✓ Define a search algorithm: One need a search algorithm to sift through possible values for the hyperparameters and zero in on the optimal ones.
- ✓ Define a tuner: To encapsulate the model, the search space, and the search algorithm in order to execute the hyperparameter search.
- ✓ Run the hyperparameter search: Finally, one may start the tuner and look for opti- mal hyperparameter settings.

Once the hyperparameter search has been completed, the best hyperparameters may be used to train a final model using the whole training dataset [34], and then the model's performance can be assessed using the test dataset. Time and effort can be saved using Keras Tuner to automate the process of hyperparameter tweaking, which can lead to higher performance in your deep learning models [35].

Proposed Work

This section explains our proposed approach in order to classify lumpy disease in cat- tle and buffaloes. The lumpy image dataset in our research is preprocessed for the sake of removing distorted and blurred images from the dataset. Then SMOTE is used to handle imbalanced data classes. SMOTE constructs synthetic examples by selecting minority class examples at random and employing interpolation techniques to build examples between the target and surrounding instances. The SMOTE method classifies data before sampling it either more or less thoroughly. These three operational elements are the neighbor count used to enter new cases, the up-sampling rate, and the down-sampling rate, respectively. The dataset is now clean and efficient for classification.

Before classification, the various pre-trained models are utilized for extracting features from the images, these are pre-trained neural network architecture for large datasets such as ImageNet, where it learns to extract features that are useful for a wide variety of image classification tasks. The pre-trained models utilized are; MobileNetV2, VGG19, ResNet50, InceptionV3, and DenseNet-121.

The proposed work has taken the pre-trained network and removed the final classification layers. This leaves us with a network that can extract features from images but does not classify them. Then we used these features as input to a new classification network.

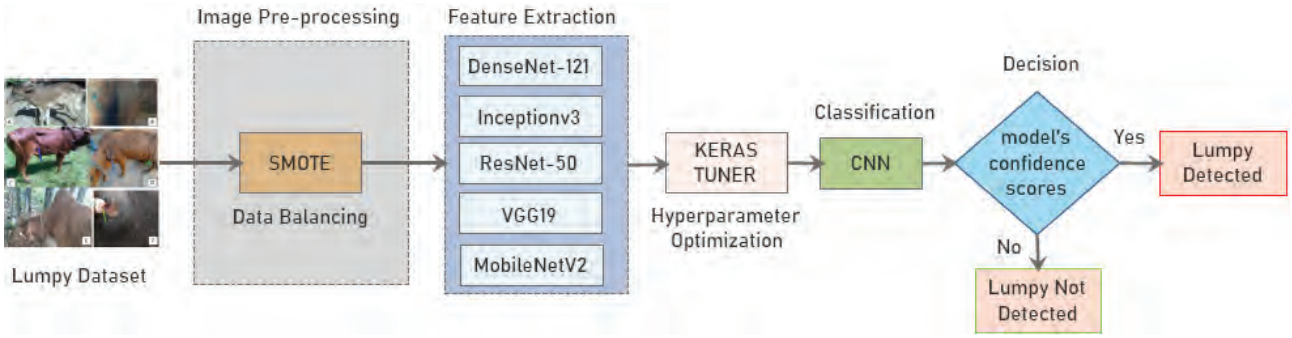


Figure 3. Proposed model.

The classification is performed using a two-dimensional Convolution Neural Net- work with the combination of max pooling layers, flatten layers, and dense layers, arranged in a fashion as visible in Figure 2. The parameters of the CNN are tuned using a Keras tuner which automates the process of hyper-parameter tweaking, which can lead to higher performance in the deep-learning classification models.

Finally, the dataset is trained with 70% of the lumpy image dataset and tested with the remaining 30% dataset. The experiment is performed using Python language, on jupyter notebook in Anaconda framework. The system with high processing with 32Gb RAM is used for the experiment purpose. In order to construct and train the CNN model, we must first load the required libraries, one of which is TensorFlow. We also import ImageDataGenerator, which is used to generate batches of augmented image data for training.

Next, we set up the directories that will hold our test and training images. Next, we create an image generator that uses data augmentation to have our training data ready to go. We imported keras and its keras tuner library to tune the parameter of the classification CNN model. Figure 3 provides the overall flow of the proposed model.

After successful execution (training and testing of model) of the models, the results are evaluated and analysed using confusion metric parameters; Accuracy, Precision, Recall, and F1-score. The graphical analysis is also performed using inbuilt Python libraries in order to generate ROC and Convergence curves, which are later discussed in the next section.

PERFORMANCE METRICS

The research employs the confusion matrix parameters (accuracy, recall, precision, and F1 score) to assess the effectiveness of the proposed model. The following details the confusion matrix [34] terms: To understand the evaluation criteria, one should be familiar with the following four terms:

- True Positive (TP): The number of occurrences in which the model correctly classifies them as positive.
- True Negative (TN): The number of occurrences in which the model correctly classifies them as negative.
- False Positive (FP): The number of negative occurrences in which the model incorrectly classifies them as positive.
- False Negative (FN): The number of positive occurrences that the model incorrectly classifies as negative.

Accuracy: It is the ratio of the number of correct predictions to the total num- ber of predictions made by the model. [36].

$$\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN \quad (1)$$

Precision: It is the ratio of true positives to the total number of positive predic- tions made by the model. [37].

$$\text{Precision} = TP/ (TP + FP) \quad (2)$$

Recall: It is the ratio of true positives to the total number of actual positive cases in the dataset. [16].

$$\text{Recall} = TP/(TP + FN) \quad (3)$$

F1-score: The F1-score balances accuracy and recall into a single number. It's helpful when there's an inequity in the dataset since it strikes a balance between accuracy and recall [38].

$$F\,1 = 2 * (\text{precision} * \text{recall})/(\text{precision} + \text{recall}) \quad (4)$$

RESULTS AND DISCUSSION

In this work, MobileNetV2 was used with a CNN classifier as a feature extractor to diagnose LSD. There are a total of 14 data characteristics and 24,803 data points included in the forecasts. This dataset had 3,039 infected records from set (1) and 21,764 from set (0). Resampling of the dataset is required since it is clear that the current data does not fairly represent cattles who are infected with lumpy compared to those who are not. Before resampling the data, the attributes such as, 'country', 'region', 'X5 Ct 2010 Da', 'X5 Bf 2010

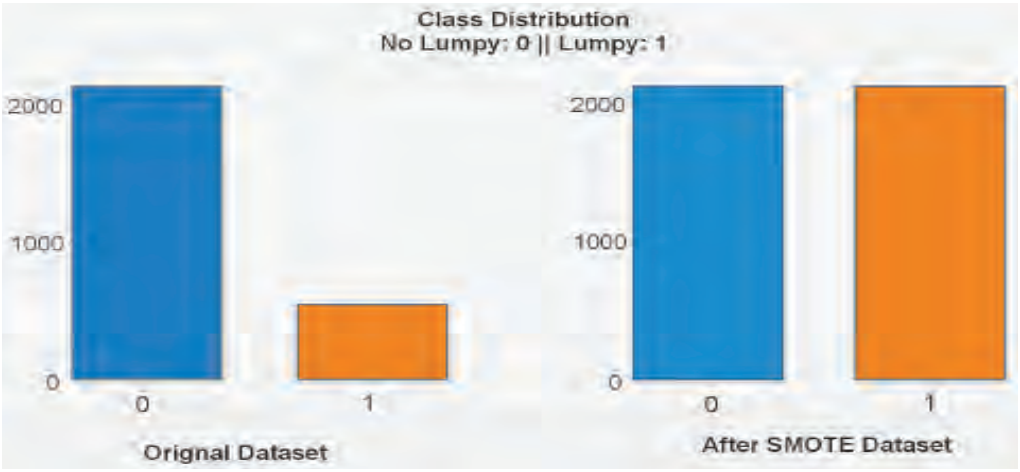


Figure 4. Dataset balancing using SMOTE.

Da', and 'reportingDate', will be eliminated, as they are not used in the classification process. The association between attributes was also analysed using a correlation heatmap in an effort to reduce the size of strongly linked features.

Some benefits of using MobileNetv2 for feature extraction are; first, since the network is pre-trained, we can take advantage of the knowledge learned from the original dataset to improve the accuracy of our new classification model. Second, by using depth-wise separable convolutions, MobileNetV2 is computationally efficient and requires fewer resources than other, more complex networks.

The original dataset should be used to run the training models, and then the LSD dataset should be resampled to evaluate the first model's efficacy. In order to put the model to the test, we must first partition the original data frame into a 70:30 split of training and validation data.

After employing the SMOTE fit resample oversampling strategy, the resampled dataset shape counter (1: 21.764, 0: 21.764), indicates that both the infected and uninfected datasets share the same total number of observations (21.764). Figure 4 depicts the balanced distribution of data that results from SMOTE re-sampling.

There are hyperparameters that need to be specified when using CNN for classification, including learning rate, batch size, filter size, pooling type, pool size, dropout rate, activation function, optimizer, and others. The optimal hyperparameters are optimized using Keras Tuner, which examines all possible combinations within the provided setup. learning rate: [0.0001, 0.001, 0.01, 0.1], Batch size: [16, 32, 64, 128], filter size: [3x3, 5x5,7x7], activation function: [ReLU, LeakyReLU, ELU, Sigmoid, Tanh], pooling type: [max pooling, average pooling], pooling size: [2x2, 3x3], dropout rate: [0.1, 0.2, 0.3, 0.4, 0.5], optimizer: [adam, RMSprop, SGD] are used in this study. The most ideal settings based on CNN are determined after hyperparameter tweaking, and they are as follows: learning rate: 0.001, batch size: 32, filter size: 3x3, pooling type: max pooling, pool size: 3x3, dropout rate: 0.5, activation function: ReLU, optimizer: adam.

The accuracy, precision, sensitivity, and F1-Score of the model are assessed using the Confusion Matrix. based on the test scenario utilizing MobileNetV2 and SMOTE with CNN classification on LSD datasets. The experiment's findings are shown in Table 2. The graphical comparison of the evaluation metrics is illustrated in Figure 5.

Receiver Operating Curve (ROC)

The confusion metric is used to assess the receiver operating characteristic (ROC) curve and to identify whether or not the identification is successful. The ROC curve establishes the ratio of true positives to false positives in

Table 2. Performance of various feature extraction models with CNN

Model	Accuracy(%)	Precision(%)	Sensitivity(%)	F1-Score(%)
DenseNet121-CNN	94.25	93.11	93.05	94.08
InceptionV3-CNN	93.29	91.85	91.17	92.88
ResNet50-CNN	94.74	93.66	92.87	93.21
VGG19-CNN	94.44	93.22	92.87	92.45
MobileNetV2-CNN	99.88	98.43	98.67	99.32

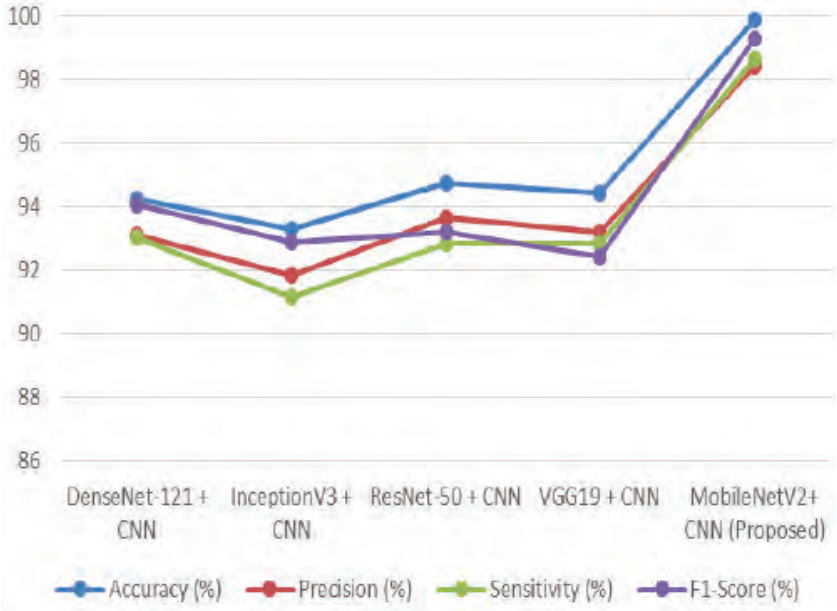


Figure 5. Graphical representation of model performance.

experimental data. True positive rate (Y axis) and false positive rate (X axis) are plotted on the ROC curve. An ideal ROC (receiver operating characteristic) curve climbs sharply toward the top-left corner, signifying that the model has a high true positive rate and a low false positive rate. The closer the curve gets to this corner, the more effectively the model differentiates between positive and negative classes.

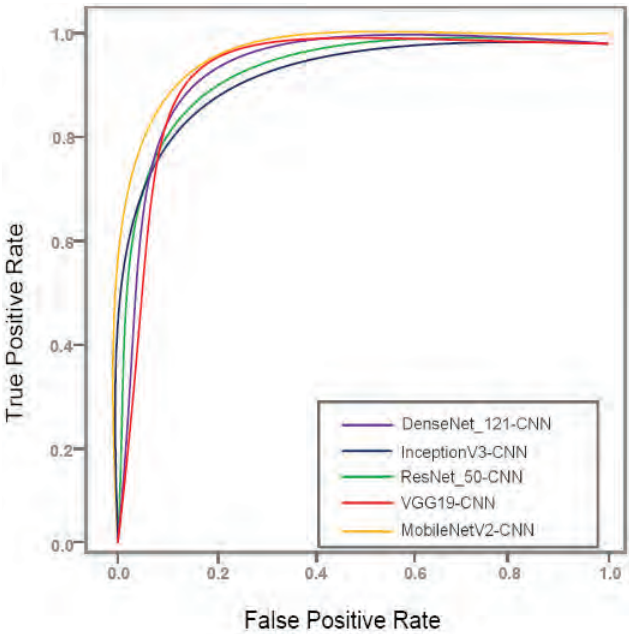


Figure 6. ROC of Various Feature Extraction techniques with CNN.

The Area Under the Curve (AUC) is a crucial indicator of performance, with values near 1 indicating excellent classification ability. In contrast, a random or weak model will produce a curve closer to the diagonal, where the AUC approaches 0.5. For optimal performance, the ROC curve should have an AUC near 1, reflecting strong predictive accuracy. Figure 6 depicts how the proposed model ROC stacks up against its competitors. The area under the ROC curve (AUC) is a two-dimensional measure of this connection that shows how well the suggested model outperforms alternative models.

Convergence Curve

While training and testing a model, the optimal value of the learning parameter is depicted by a convergence curve that plots the accuracy of the model against the loss function. The Figure 7 depicts the convergence curves of the proposed and alternative models. The sub-figures represent 7a DenseNet121, 7b InceptionV3, 7c ResNet50, 7d VGG19, and 7e proposed approach (MobileNetV2), all with the CNN model respectively. The 200-epoch training and validation accuracy curves for the proposed models are displayed. The proposed model outperforms the state-of-the-art with a training and validation accuracy of 99.88%, revealing a faster convergence rate.

Comparison with existing work

Table 3 gives a state-of-art analysis based on the performance of existing classification models and the proposed models on considering accuracy measures. The paper taken into account are using different classification techniques some of them are artificial neural networks, convolution

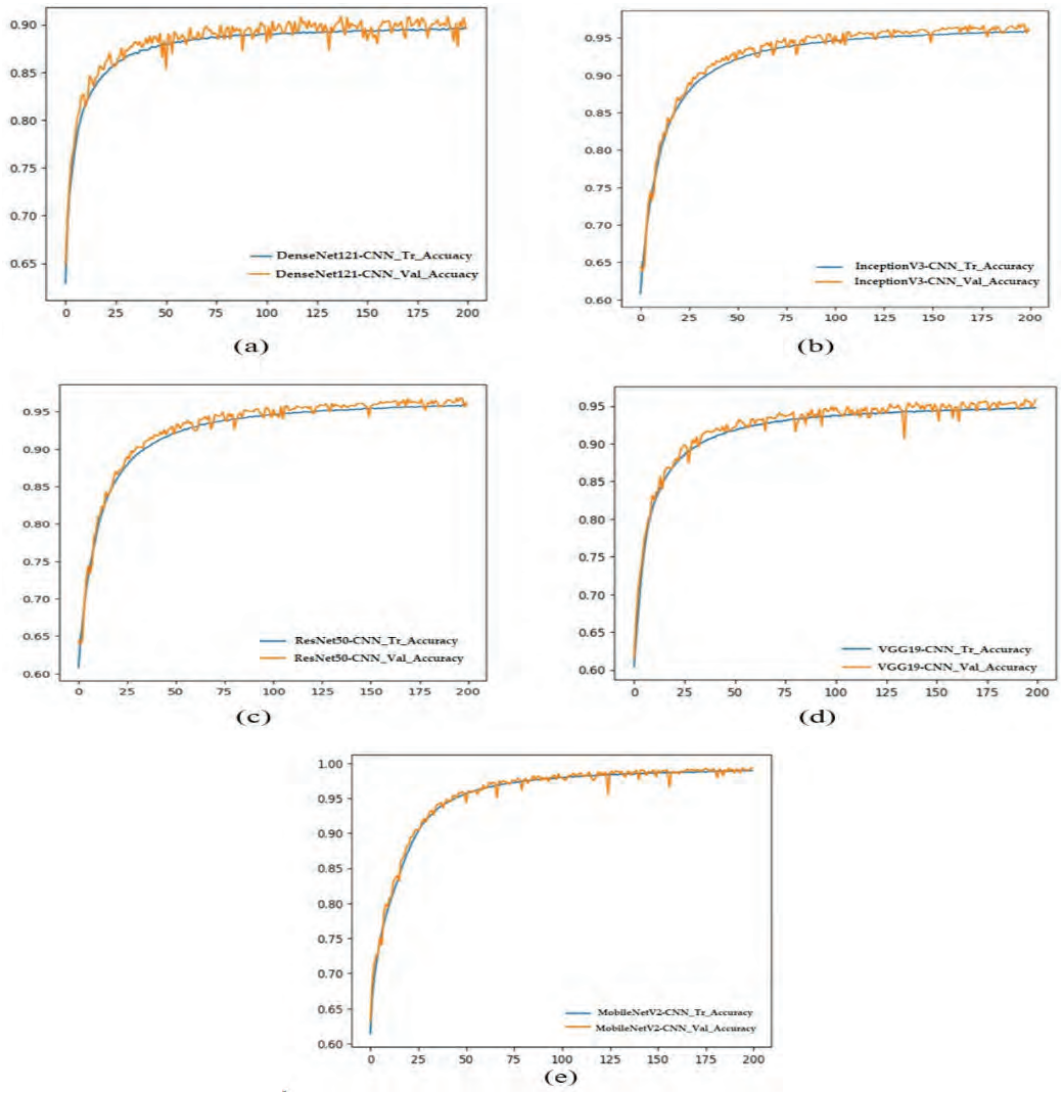


Figure 7. Convergence curve analysis: (a)DenseNet121, (b)InceptionV3 (c) ResNet50 (d)VGG19 (e) Proposed model.

Table 3. The state-of-art comparison of the proposed model with existing work

Ref	Dataset	Model	Accuracy (%)
[8]	Lumpy skin disease outbreak data	ANN	97
[10]	Debre Markos University, Addis Ababa University-School of Veterinary Medicine, the World Wide Web, and other secondary sources were used for photographic holdings.	CNN	95
[11]	LSD dataset with 18,603 instances and 16 features.	Random Forest	97.7
(12]	Dataset of LSD.	ELM	90.06
[14]	Dataset created by the authors.	Neural network model using Inception-v3 for feature extraction	92.5
[15]	Pictures from four health care facilities in the Sunyani Municipality of Ghana.	Medilab-plus CNN classifier	88 (atopic dermatitis), 85 (acne vulgaris), 84.7 (scabies)
Our Work	LSD Dataset	SMOTE+MobileNetV2+CNN	99.88

neural networks, ensemble learning models and individual models such as random forest.

The broader implications of our findings extend beyond the immediate scientific community, as the detection model can be implemented in various agricultural sectors, particularly in regions where LSD outbreaks are prevalent, such as Africa, the Middle East, and Asia. By incorporating this technology, veterinary services, livestock producers, and governmental agencies can utilize automated tools to rapidly detect and respond to disease outbreaks, improving animal welfare and reducing the impact of the disease on food security and trade. This work can also inspire further research into applying deep learning models for other transboundary animal diseases, enhancing overall disease control strategies.

CONCLUSION

Lumpy skin disease which is widespread in cattle, water buffalo and camels in South Asian countries, may be identified by the appearance of lumps or nodules on the animals' skin which can cause fever, anorexia, and reduced milk production. Lumpy skin disease has a considerable impact on economies in several countries. In order to detect and control the spread of the disease among animals, there are various measures taken by the government of various countries. The paper proposed a MobileNetv2- based Deep Learning model for the classification of lumpy skin disease (LSD) by training the machine using a lumpy dataset. The images in the dataset are preprocessed and then features are extracted using MobileNetv2, which is a pre-trained neural network model. The classification is performed using a 2D-layered convolutional neural network model with a combination of convolutional layer, flatten layers, max-pool layer and dense layer, which obtains the highest accuracy of 99.88%. The proposed model outperformed other individual models and existing work by different researchers in the domain.

In future, we intend to apply federation and transfer learning techniques in order to pre-process affected animal images and classify disease more accurately. We will also use the proposed technique on other image datasets in order to create a more generalized and robust model.

DECLARATIONS ETHICAL APPROVAL

Not Applicable

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest

FUNDING

Not applicable

DATA AVAILABILITY STATEMENT

The data used in this study will be made available upon reasonable request.

REFERENCES

[1] AAU. Lumpy skin disease (lsd): outbreak investigation, isolation and molec- ular detection of lumpy skin disease in selected areas of eastern shewa, ethiopia. Available at: https://etd.aau.edu.et/bitstreams/57dd46f6-bb31-48c2-a563-3760428b9fe3/download. Accessed on 4 Apr 2026.

[2] Al-Salihi K. Lumpy skin disease: Review of literature. Mirror Res Vet Sci Anim 2014;3:6–23.

[3] Kumar R, Godara B, Chander Y, Kachhawa JP, Dedar RK, Verma A, et al. Evidence of lumpy skin disease virus infection in camels. Acta Trop 2023;242:106922. [CrossRef]

[4] Sprygin A, Pestova Y, Wallace D, Tuppurainen E, Kononov A. Transmis- sion of lumpy skin disease virus: A short review. Virus Res 2019;269:197637. [CrossRef]

[5] Chaudhary N. A review on lumpy skin disease in india. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4351704. Accessed on 4 Apr 2026.

[6] Mathivanan E, Raju K, Rajaganapathy. Outbreak of lumpy skin disease in india 2022-an emerging threat to livestock & livelihoods. Global Biosecurity 2023.

[7] Gupta T, Patial V, Bali D, Angaria S, Sharma M, Chahota R. A review: Lumpy skin disease and its emergence in india. Vet Res Commun 2020;44:111–118. [CrossRef]

[8] Safavi EA. Assessing machine learning techniques in forecasting lumpy skin disease occurrence based on meteorological and geospatial features. Trop Anim Health Prod 2022;54:55. [CrossRef]

[9] Farra D, Nardi M, Lets V, Holopura S, Klymenok O, Stephan R, et al. Qualitative assessment of the probability of introduction and onward transmission of lumpy skin disease in ukraine. Microb Risk Anal 2022;20:100200. [CrossRef]

[10] Lake B, Getahun F, Teshome FT. Application of artificial intelligence algo- rithm in image processing for cattle disease diagnosis. J Intell Learn Syst Appl 2022;14:71–88. [CrossRef]

[11] Ujjwal N, Singh A, Jain AK, Tiwari RG. Exploiting machine learning for lumpy skin disease occurrence detection. IEEE Xplore 2022;1–6. [CrossRef]

[12] Genemo M. Detecting high-risk area for lumpy skin disease in cattle using deep learning feature. Adv Artif Intell Res 2023;3:27–35. [CrossRef]

[13] Utami E, Muhammad AH, Suparyati. Lumpy skin disease prediction based on meteorological and geospatial features using random forest algorithm with hyper- parameter tuning. IEEE Xplore 2022;99–104. [CrossRef]

[14] Rai G, Hussain A, Kumar A, Ansari A, Khanduja N. A deep learning approach to detect lumpy skin disease in cows. Springer 2021;369–377. [CrossRef]

[15] Akyeramfo-Sam S, Philip AA, Yeboah D, Nartey NC, Nti IK. A web- based skin disease diagnosis using convolutional neural networks. Int J Inf Technol Comput Sci 2019;11:54–60. [CrossRef]
[16] Hegde PR, Shenoy MM, Shekar B. Comparison of machine learning algo- rithms for skin disease classification using color and texture features. IEEE Xplore 2018;1825–1828. [CrossRef]
[17] Fern´andez A, Garcia S, Herrera F, Chawla NV. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res 2018;61:863–905. [CrossRef]
[18] Bhattacharya A. How to use smote for dealing with imbalanced image dataset for solving classification problems. Available at: https://medium.com/swlh/how-to-use-smote-for-dealing-with-imbalanced-image-dataset-for-solving-classification-problems-3aba7d2b9cad. Accessed on 4 Apr 2026.
[19] Krishna ST, Kalluri HK. Deep learning and transfer learning approaches for image classification. International J Recent Technol Eng 2019;7:427–432.
[20] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. IEEE Xplore 2020;109:43– 76. [CrossRef]
[21] Huang G, Liu Z, Weinberger KQ. Densely connected convolutional networks. arXiv 2016.
[22] Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: A powerful paradigm for scientific research. Innov 2021;2. [CrossRef]
[23] Pang S, Zhang Y, Ding M, Wang X, Xie X. A deep model for lung cancer type identification by densely connected convolutional networks and adaptive boosting. IEEE Xplore 2019;8:4799–4805. [CrossRef]
[24] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv 2015. [CrossRef]
[25] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv 2015.
[26] Darth M. Vgg 19 feature extraction. Available at: https://www.kaggle.com/code/omkarmodi/vgg-19-feature-extraction. Accessed on4 Apr 2026.
[27] Rajinikanth V, Raj ANJ, Thanaraj KP, Naik GR. A customized vgg19 network with concatenation of deep and handcrafted features for brain tumor detection. Appl Sci 2020;10:3429. [CrossRef]
[28] Sandler M, Howard AG, Zhu M, Zhmoginov A, Chen L. Inverted resid- uals and linear bottlenecks: Mobile networks for classification, detection and segmentation. arXiv 2018.
[29] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognit 2018;77:354–377. [CrossRef]
[30] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neu- ral networks: analysis, applications, and prospects. IEEE Xplore 2021. [CrossRef]
[31] O'Shea K, Nash R. An introduction to convolutional neural networks. arXiv 2015.
[32] Pon MZA, Prakash K. Hyperparameter tuning of deep learning models in keras. Available at: https://sparklinglightpublisher.com/index.php/staiqc/article/view/5/4. Accessed on 4 Apr 2026.
[33] Abdellaoui B, Moumen A, El Idrissi YEB, Remaida A. Training the fer2013 dataset with keras tuner. Available at: https://www.scitepress.org/Link.aspx?-doi=10.5220/0010735600003101. Accessed on 4 Apr 2026.
[34] Shawki N, Nunez RR, Obeid I, Picone J. On automating hyperparameter optimization for deep learning applications. IEEE Xplore 2021. [CrossRef]
[35] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–2830.
[36] Zhan J, Huang X, Qian Y, Ding W. A fuzzy C-means clustering-based hybrid multivariate time series prediction framework with feature selection. IEEE Xplore 2024. [CrossRef]
[37] Zhu C, Ma X, D'Urso P, Qian Y, Ding W, Zhan J. Long-term multivariate time series forecasting model based on Gaussian fuzzy information granules. IEEE Xplore 2024. [CrossRef]
[38] Zhu C, Ma X, Ding W, Zhan J. Long-term time series forecasting with multi-linear trend fuzzy information granules for LSTM in a periodic framework. IEEE Xplore 2023. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 869–879, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2016

Research Article

Bioconservative effect of *Cotula cinerea* Delile on Klila cheese

Rym Nouria BENAMARA[1,2], Fatima SAHNOUNI[1,3], Asmaa BELGHARBI[1,4], Kada IBRİ[1], Smail KHALED[5], Mirvette BENFRÉHA BENYELLES[1], Sihem ZITOUNI[1], Hadjer ZEMMOURI[1]

[1]*Faculty of Nature and Life Sciences, University of Mascara, 29000, Algeria*
[2]*Laboratory of Microbiology Applied to Agrifood and Environmental (LAMAABE), Facultyof SNV/STU, University of Tlemcen, 13000, Algeria*
[3]*Laboratory Research on Biological Systems and Geomatics, Department of Biology, University of Mustapha Stambouli, 29000, Algeria*
[4]*Laboratory of Bioconversion, Engineering Microbiology and Safety Health, Faculty of Natural and Life Sciences, University of Mustapha Stambouli of Mascara, 29000, Algeria*
[5]*Department of Computer Science, Faculty of Exact Sciences, Mustapha Stambouli University, 29000, Algeria*

ARTICLE INFO

Article history
Received: 10 April 2025
Revised: 30 July 2025
Accepted: 03 October 2025

Keywords:
Aqueous Extract; Antibacterial Activity; Antioxydant Activity; Bioconservation; Cotula Cinerea, Klila Cheese

ABSTRACT

The Klila cheese has been consumed by Algerians for many centuries, likely fromantiquity to the presentday. It is a traditional, artisanal North African cheese known for its high nutritional value. Its high dry matter content allowsit to bestored for extended periods without the risk of microbials poilage. This study aims to investigate the bioconservation potential of *Cotula cinerea* Delile for preserving Klilacheese. To this end, physical analyses of the aqueous extract of *Cotula cinerea* Delile were conducted, including studies on its antioxidant and antibacterial activities, as well as a phytochemical screening of the plant. Additionally, physico-chemical, microbiological, and sensory analyses of both the control and enhanced Klila cheese samples were performed over the storage period.

The results show that the physical characterization of *Cotula cinerea* Delile powder revealed certain properties, such as dry matter content (96.79%), ash content (3.57%), and a plant yield of 10.2%. Preliminary tests demonstrated the plant's richness in bioactive compounds. Flavonoid content was evaluated using the AlCl method, revealing a concentration of 0.04mg EQ/g of extract. The antibacterial activity of *Cotula cinerea* Delile extract was tested on *Staphylococcus aureus* and *Escherichia coli*, with an effective concentration of 0.3g.

In the physico-chemical analysis of Klila cheese, it was observed that the pH and acidity of the Klila with 0.3g of *Cotula cinerea* Delile powder were higher than those of the control Klila. Conversely, the control cheeses howeds lightly higher levels of protein, dry matter, and ash throughout the storage period. Microbiological analyses revealed that the control Klila had higher levels of total aerobic mesophilic bacteria, lactobacilli, and enterococci compared to the Klila sample with 0.3g of *Cotula cinerea* Delile. Additionally, no *Staphylococcus aureus*, sulfite-reducing Clostridium, yeast, or mold were detected in either type of Klila cheese. Sensory analysis suggests that the inclusion of *Cotula cinerea* Delile at a 0.3g

***Corresponding author.**
*E-mail address: asmaa.belgharbi@univ-mascara.dz, asmaabelgharbi@yahoo.fr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

concentration is highly acceptable as an additive for Klila cheese, as it does not introduce undesirable qualities or defects for the end consumer.

Cite this article as: Benamara RN, Sahnouni F, Belgharbi A, Ibri K, Khaled S, Benfréha Benyelles M, Zitouni S, Zemmouri H. Bioconservative effect of *Cotula cinerea* Delile on Klila cheese. Sigma J Eng Nat Sci 2026;44(2):870–879.

INTRODUCTION

Algeria has a long standing dairy tradition, with milk and dairy products being central to the Algerian diet and a primary source of animal protein [1]. Dairy products play a prominent role in Algerian culture, with this tradition being passed down through generations. A wide variety of fermented dairy products are traditionally prepared in Algeria for the purpose of milk bio-preservation [2]. As national consumption of milk and dairy, valued as a source of animal protein and calcium, grows, dairyremains essential in the Algerian diet [3]. Traditionally, milk was processed to extendits shelf life. Numerous scientific studies have shown that traditional dairy products made from fresh raw milk with distinct flavor and nutritional profiles are increasingly popular among consumers [4].

To prevent milk spoilage, given its perishable nature, most of the production must be processed [5]. The simplest preservation method is transforming milk into cheese, either through acidification or coagulation [6-8]. Dairy products are part of Algerian heritage and hold cultural, medicinal, and economic importance, [9] developed over time with rural women's expertise. Common fermented dairy products include Raib, Lben, Jben, and Klila [10].

Several traditional cheeses are produced in Algeria, withJben and Klila (alsoknown as Tiklilt) being the most popular [11]. Klila, traditionallymanufactured and primarily for householdconsumption, has been named by indigenous Algerian Berbers sinceancient times [12]. Itsnamelikelyoriginatesfrom the Berber word "Ikil," meaningcurdledmilk. Among the Chaoui people in the Aurès region of Algeria, when milk curdles, they say "T'kellel." In Kabylia, spontaneously curdled milkis called "Tiklitt," similar to traditional Klila cheese, whichis the curdrecovered after milk fermentation and coagulation [13].

Klila is one of Algeria's most popular traditional cheeses, produced using methods still in practice today (e.g., using the Chekoua for churning) [14,15]. This freshcheeseistypically made fromrawcow's, sheep's, or goat'smilk or a mixture, with spontaneous fermentation [16,17]. Klila can beconsumedfresh or dried. Its high dry matter content (over 80%) enables extended preservation without microbial spoilage [18].

The traditional production method of Klila, notably in the Naama Wilaya in southwestern Algeria [19]., contributes to itssensory and nutritional attributes, explaining its growing demand among consumers [20]. Research has confirmed that Klila production and consumption occuracross various Algerian regions, such as Guelma, Souk-Ahras, Oum El Bouaghi, Batna, and Khenchela [12] and furtherstudies have documenteditspopularity in Biskra, Tbessa, M'sila, Sétif, and other provinces.

Presently, chemical preservatives are the most common food preservation method, but they may have adverse health effects on consumers [21]. As a result, scientists are seeking natural bioconservatives, such as medicinal plants [22]. Biopreservation, which uses microorganisms or natural compounds, is an alternative to traditional chemical preservatives in food industries. Like other preservation methods, it aims to control pathogenic or spoilage flora growth while maintaining the product'sorganoleptic and nutritional qualities [23]. Thus, finding a natural and safe alternative to synthetic additives is essential for consumer health. In this context, this study focuses on *Cotula cinerea* Delile, also known as Guertoufa, widely used to treatailments like colic, cough, diarrhea, and digestive disorders [24].

This plant is commonly used in decoction, maceration, infusion, and inhalation. It is also used to filter go at butter due to its preservative properties [25]. This study is unique as itis the first to explore this topic, underscoring the importance of this research. The primary objective of this work is to study the bioconservation potential of *Cotula cinerea* Delile for traditional Klila cheese preservation.

MATERIALS AND METHODS

Biological Materials

Our work focused on studying the bioconservative properties of *Cotula cinerea* Delile on Klila cheese. For the purposes of our study, Klila cheese was prepared from traditional Lben obtained from raw milk. The Klila cheese samples studied were of cow's milk origin (Fig. 1A). The product used was the traditional Klila cheese made in the laboratory of the SNV faculty in Mascara. The manufacture of this type of cheese is based on the schemecited by [18]. The plant material used in this study corresponds to *Cotula cinerea* Delile, which was purchased in Ain Sefra (southwest Algeria) (Fig. 1B).

Cotula cinerea Delile

Determination of physical parameters of *cotula cinerea* delile

The dry matter content was determined following the NF B 51-004 standard (1985) [26]. Ash content was

Figure 1. Biological materials used (**A.** Klilacheese, **B.** *Cotula cinerea* Delile plant '*Guertoufa*').

measured using the method described by [27]. Within cineration temperature sranging between 500–600°C. The aqueous extract preparation was performed according to the method outlined by [28], where 100 g of powdered plant leaves were macerated at room temperature in 400 mL of distilled water with agitation for 24 hours, followed by decantation. After decanting the mixture, the aqueous extract was filtered through Wattman paper and then dried at 50°C until a constant weight was achieved. A yield determination was also conducted [29].

Qualitative analysis of polyphenolic compounds in *Cotula cinerea*Delile

A phytochemical screening using a colorre action method was performed for this analysis. The characterization tests relied on precipitation and complexation reactions, forming insoluble colored complexes. The observed coloration was induced by the appropriatere agent and generally results from conjugation formation within the molecule [30]. The tests performed included: Steroid Test [31], Terpenoid Test [32], Tannin Test [33], Flavonoid Test [34], Alkaloid Test [35], Saponin Test [36], Phenol Test [37], and Coumarin Test [38].

Determination of total flavonoid contain

Flavonoid content was measured based on the method described by [39]. One milliliter of an ALCl3 solution (2%) wasadded to 1 mL of the sample or standard solution (prepared in methanol) at various concentrations. After allowing the mixture to react for 10 minutes, absorbance was measured at 430 nm. Flavonoid concentration in the extracts was calculated from a calibration curve established with quercetin (0-14 µg/mL) and expressed as micrograms of quercetin equivalents per milligram of dry matter (µg EQ/mg DM).

Evaluation of antibacterial activity

The antimicrobial activity assessment determined the Minimum Inhibitory Concentration (MIC) and Minimum Bactericidal Concentration (MBC) of the aceticextract of *Cotula cinerea* Delile. The study used a macro dilution method in liquid media. *Escherichia coli* (Gram-) and *Staphylococcus aureus* (Gram+), strainsfrom the educational laboratory (Faculty of Natural and Life Sciences, University of Mustapha Stambouli, Mascara), wereused to evaluate the antibacterial activity of the aqueousextract of *Cotula cinerea* Delile. Following inoculum preparation and extract dilution, MIC, MBC, and the MBC/MIC ratio were determined.

Biopreservation of Traditional Klila Cheese

For the biopreservation of this traditional cheese, *Cotula cinerea* Delile (locallyknown as Guertoufa) was used. Preliminary tests led to a chosen concentration of 0.3% of the plant, homogenized with Klila cheese, while a negative control sample was prepared without the plant addition. Samples were stored at refrigeration temperature (5°C), and analyses were performed weekly (Days 0, 7, 14, 21, and 28).

Physicochemical analyses

To measure pH, Klila cheese samples were diluted 1:10 in distilled water, from which pH was measured [40]. Titrable acidity was determined by titration with 0.1 N NaOH. Protein content was measured via the formal dehyde titration method [41]. Total dry extract (TDE) was determined by drying at 103±2°C for 3 hours [42]. Ash content was determined by incinerating fresh Klila samples in a muffle furnace at 525±25°C until whitishash was obtained [43].

Detemination of microbiological quality

Aerobicmesophilic flora (FMAT) enumeration was conducted on nutrient agar (30°C, 48–72 hours, aerobic conditions, Biokar) according to NF EN ISO 4833-1:10-2013 [44]. Presumptiveme sophilic lactobacilli were tested on MRS agar, [45] anaerobic incubation at 30°C for 48–72 hours. Enterococci were numerated on BEA (Biokar) with aerobic incubation at 37°C for 24 hours. Total and fecal coliform counts were performed on bright green lactose broth (VBL) using three tubes per dilution with incubation at 37°C for 24–48 hours [46]. Using the most probable number method. Yeasts and molds were detected on oxytetra cycline glucose agar (OGA) [47]. *Staphylococcusspp*. Weretested on Baird Parker agar and incubated at 37°C for 24 hours [48]. Clostridium Sulfito-reductor spores were numerated on liver agar (VF) [49].

Sensory analysis

A triangular test was conducted for sensory analysis [50].

Statistical Study

All statistical analyses were performed using the R software (R Development Core Team, 2008). In this study, two types of statistical analyses were conducted: ANOVA and Tukey's test.

RESULTS AND DISCUSSION

Results of Cotula Cinerea Delile Analyses

Physical parameters

The results of somephysicalparameters of *Cotula cinerea* Delile powder showthat the ash content is 3.57%, the moisture content is 96.79% and the dry matter content is 3.21%. The aqueousextract of the *Cotula cinerea* Delile plant is brown in colour and viscous in appearance, with a yield of: 10.2%.

Table 1. Results of phytochemical screening of plant extracts

	Metabolitestested	Results
1	Steroid test	++
2	Terpenoid test	+++
3	Tannin test	+++
4	Flavonoid test	++
5	Alkaloid test	+
6	Saponin test	++
7	Phenol test	+
8	Coumarin Test	++

(+): Weakpresence; (++): Averagepresence; (+++): Strongpresence.

Qualitative analysis of polyphenolic compounds

Concerning the qualitative analysis of polyphenolic compounds and more specifically the phyto chemical analysis of *Cotula cinerea* Delile extract. Our phyto chemical study of *Cotula cinerea* Delile extractindicates a high presence of terpenoids and tannins, a moderate presence of steroids, coumarins, flavonoids and saponins, and a low presence of phenols and alkaloids. The results of this phytochemical screening are presented in Table 1 below.

Antioxydant activity

For antioxidant activity, flavonoids were measured using a colorimetric method based on the use of a standard, Quercetin. Our results indicate that our plant contains a low quantity of flavonoids of the order of 0.04 mg EQ/g extract.

Evaluation of antibacterialactivity

The antibacterial activity of *Cotula cinerea* Delile was studied in vitro using the macro-dilution method against two bacterial strains: *Staphylococcus aureus* and *Escherchia coli*. The results are shown in Table 2.

It can be seen that the MIC value was 0.312 g/ml for *Escherichia coli*, while *Staphylococcus aureus*was more sensitive with a MIC of 0.039 g/ml and the same BMC value = 0.312 for bothstrains.

In order to study the effect of the *Cotula cinerea* Delile plant on the preservation of traditional Klila, we compared

Table 2. Antimicrobialactivity of *Cotula cinerea* Delile

	MIC	BMC	CMB/CMI	Type of activity
Staphylococcus aureus	0.039	0.312	1.25	Bactericide
Escherchia coli	0.312	0.312	1	Bactericide

Table 3. Results of physico-chemical analyses

		pH	Acidity °D	Protein content (%)	Ash content (%)	TDE Total dry extract (%)	Moisture (%)
J0	**Control**	4.38±0.05	10.67±0.58	3.47±0.20	1.53±0.12	44.47±0.20	55.53±0.20
	0.30%	4.19±0.01	12.67±0.58	2.11±0.26	1.27±0.12	36.42±0.20	63.58±0.20
J7	**Control**	5.02±0.01	9.67±0.58	1.47±0.10	2.20±0.20	55.67±0.41	44.33±0.41
	0.30%	5.46±0.01	5.67±0.58	1.13±0.20	1.40±0.20	42.72±0.20	57.28±0.20
J14	**Control**	4.91±0.02	7.67±0.58	1.42±0.10	2.87±0.31	58.83±0.49	41.17±0.49
	0.30%	5.08±0.02	13.67±0.58	1.59±0.10	1.80±0.20	44.56±0.31	55.44±0.31
J21	**Control**	5.40±0.01	11±1.00	3.48±0.10	3.47±0.12	60.28±0.20	39.72±0.20
	0.30%	5.52±0.03	53.33±0.58	1.71±0.18	2.20±0.20	48.56±0.30	51.44±0.30
J28	**Control**	5.41±0.02	16±1.00	3.36±0.10	3.87±0.12	64.63±0.20	35.37±0.20
	0.30%	5.54±0.02	57±1.00	2.16±0.10	2.60±0.20	49.41±0.20	50.59±0.20

two Klila samples, control and Klila with 0.3g of *Cotula cinerea* Delileadded.

The effect of this plant was observed by monitoring the following parameters:

- The physico-chemicaleffect
- The microbiologicaleffect
- Sensoryeffect

Results of the Bioconservation of Traditional Klila Cheese

Physico-chemical analyses

The results of physico-chemical quality analyses of two types of Klila cheese analysed from D_0 to D_{28} (Klila control, Klila with 0.3% *Cotula cinerea* Delile) are shown in Table 3.

pH variation and titratable acidity in Klila cheese

The pH variation results for the two types of Klila cheese analyzed from Day 0 to Day 28 are shown in Figure 2A. As observed in Figure 2A, the average pH of the control Klila ranged from 4.38±0.05 to 5.41±0.02. For the Klila with the addition of 0.3 g/100 g *Cotula cinerea Delile*, the pH rangedfrom 4.19±0.01 to 5.54±0.02. These values are relatively similar, with a marked increase in pH levels over the preservation period. Initially, the pH of the enhanced Klila was lower than the control, but a slight increase in pH for the enhanced Klila compared to the control was observed during storage. According to the statistical analysis (ANOVA and Tukey tests), significant differences were noted between days and doses. For interaction results (Days x Doses), the highest pH was observed in Klila with 0.3% *Cotula cinerea Delile* on Day 28 at 5.54, while the lowest was in the enhanced Klila on Day 0 at 4.19.

The titratableacidity results for different Klila samples are presented in Figure 2B. Titratableacidity values for the control Klila varied from 7.6±0.58 °D to 16±1.00 °D, while for Klila enhanced with 0.3% Cotula cinerea Delile, acidity ranged from 5.67±0.58 °D to 57±1.00 °D. The results were similar from Day 0 to Day 7, with an increase in acidity for the enhanced Klila from Day 14 to Day 28. Statistical analysis (ANOVA and Tukey tests) showed significant differences between days and doses. For interaction (Days x Doses), the highest acidity (57 °D) was observed in the Klila with 0.3% Cotula cinerea Delile on Day 28, while the lowest (5.67 °D) was in the enhanced Klila on Day 7.

Protein and Ash Content in Klila Cheese

The protein content for different Klila samplesis shown in Figure 2C. In the control Klila, protein content ranged from 1.42±0.10% to 3.48±0.10%, while in Klila with 0.3 g of Cotula cinerea Delile, it ranged from 1.13±0.20% to 2.16±0.10% from Day 0 to Day 28. The control Klila exhibited higher protein content than the enhanced sample, except on Day 14. According to statistical analysis (ANOVA and Tukey tests), significant differences were noted between days and doses. For interaction (Days x Doses), the highest protein level was observed in the control Klila on Days 0, 21, and 28 (3.36 to 3.47%), while the lowest (1.13%) was recorded in the enhanced Klila on Day 7.

Ash content results are presented in Figure 2D. Ash content in the control Klila ranged from 1.53±0.12% to 3.87±0.12% from Day 0 to Day 28, while in Klila with 0.3% *Cotula cinerea* Delile, it ranged from 1.27±0.12% to 2.60±0.20%. The ash content was higher in the control Klila than in the enhanced sample, with a gradual increase over the storage period. Statistical analysis (ANOVA and Tukey tests) indicated significant differences between days and doses. For interaction (Days x Doses), the highestash content (3.47% to 3.87%) was observed in the control Klila on Days 21 and 28, while the lowest values (1.27%, 1.40%, and 1.53%) were found in the enhanced Klila on Days 0 and 7, and in the control on Day 0, respectively.

Total dry extract in Klila cheese

Total dry extractresults are presented in Figure 2. In the control Klila, the total dry matter ranged from 44.47±0.20% to 64.63±0.20% between Days 0 and 28. For Klilaenhancedwith 0.3 g Cotula cinerea Delile, itranged from 36.42±0.20% to 49.41±0.20%. The dry extract content showed an inverse relationship with moisture content. A notable increase in total dry extract was observed in the control Klila throughout the storage period, compared to the enhanced Klila. Statistical analysis (ANOVA and Tukey tests) for moisture content revealed significant differences between days and doses. For interaction (Days x Doses), the highest dry extract was found in the enhanced Klila with 0.3% *Cotula cinerea Delile* on Day 0 (63.58%), while the lowest was in the control Klila on Day 28 (35.37%).

Microbiological analysis results

Microbiological analysis results for both control and enhanced Klila samples with 0.3 g Cotula cinerea Delile are shown in Figures 3 and 4. For total aerobic mesophilic flora (FMAT), count sranged from 7.8 x Log UFC/g to 26.8 Log CFU/g for the control and from 7.4 Log CFU/g to 13 Log CFU/g for the enhanced Klila. Microbial loads increased in both samples over the storage period. For *Lactobacillus* counts, the control ranged between 2.78 Log CFU/g and 7.80 Log CFU/g, while the enhanced Klila with 0.3 g Cotula cinerea Delile ranged between 2.30 Log CFU/g and 6.59 Log CFU/g. Higher microbial loads were observed in the control Klila compared to the enhanced sample from Day 0 to Day 28. By Day 28, the control Klila sample showed uncountable bacterial levels, likely due to spoilage. Environmental factors, such as temperature, may influence bacterial growth. *Enterococci* counts for the control Klila varied from 3.0×10^2 CFU/g to 2.0×10^2 CFU/g, and for the enhanced sample from 1.1×10^2 CFU/g to 0.8×10^2 CFU/g. Lower counts were observed in the enhanced sample from Day 0 to Day 28.

Fecal coliforms were present only in the first week, with counts of 0.4×10^2 CFU/g for the control and 1.4×10^2 CFU/g for the enhanced Klila on Day 7. From Day 7 to Day 28, coliforms

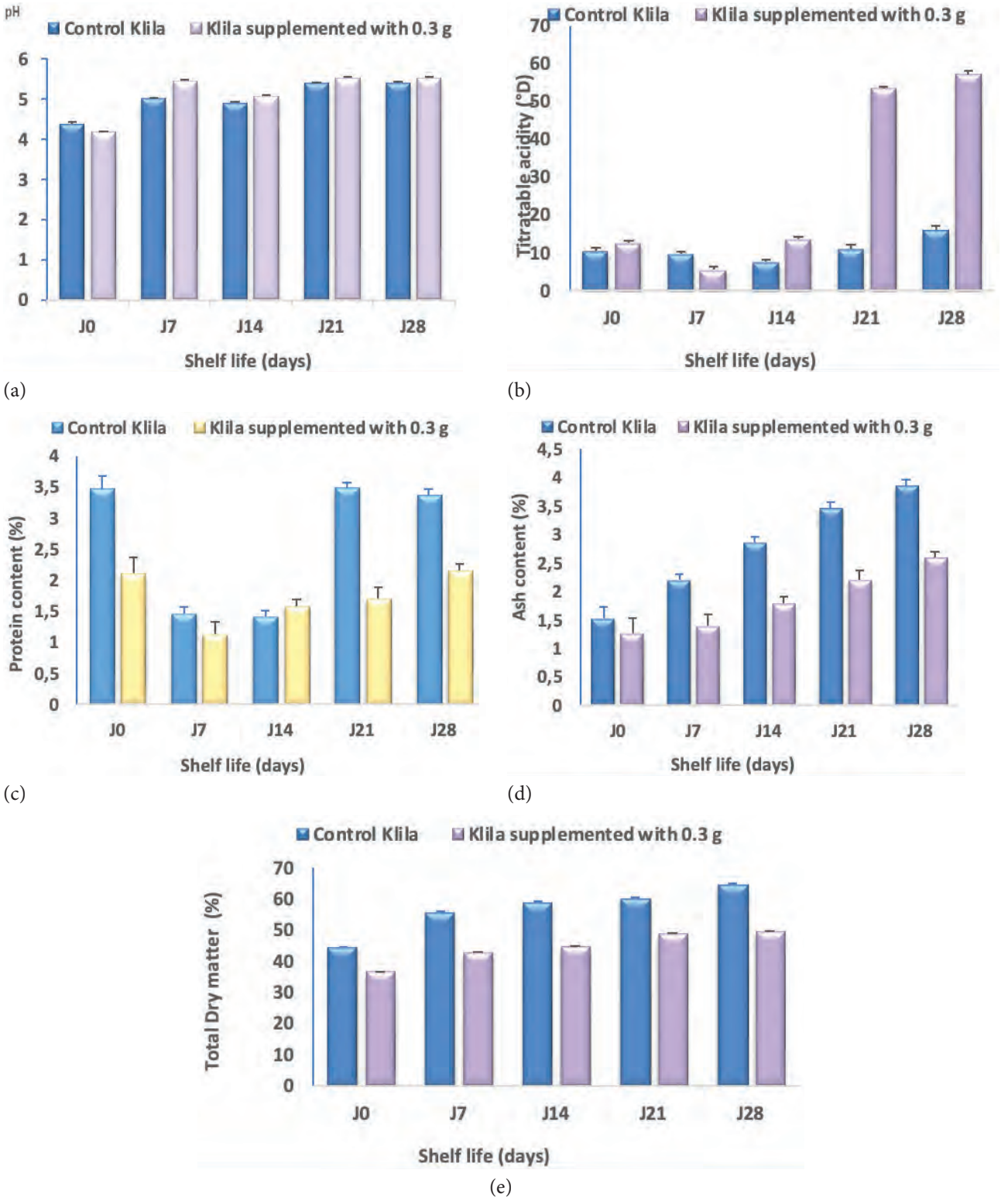


Figure 2. Variation of the different physicochemical parameters in the various Klila cheeses studied from Day 0 to Day 28 (a) pH, b) Titratable acidity, c) Proteins, d) Ash, e) Total Dry Matter).

were absent in both samples, indicating good hygiene practices during production. Yeasts, molds, *Staphylococcus aureus*, and Clostridium sulfite-reducerswere absent in both samples through out storage. Statistical analysis (ANOVA) showed no significant differences, indicating no variation between doses, days, and interactions (Dose x Day).

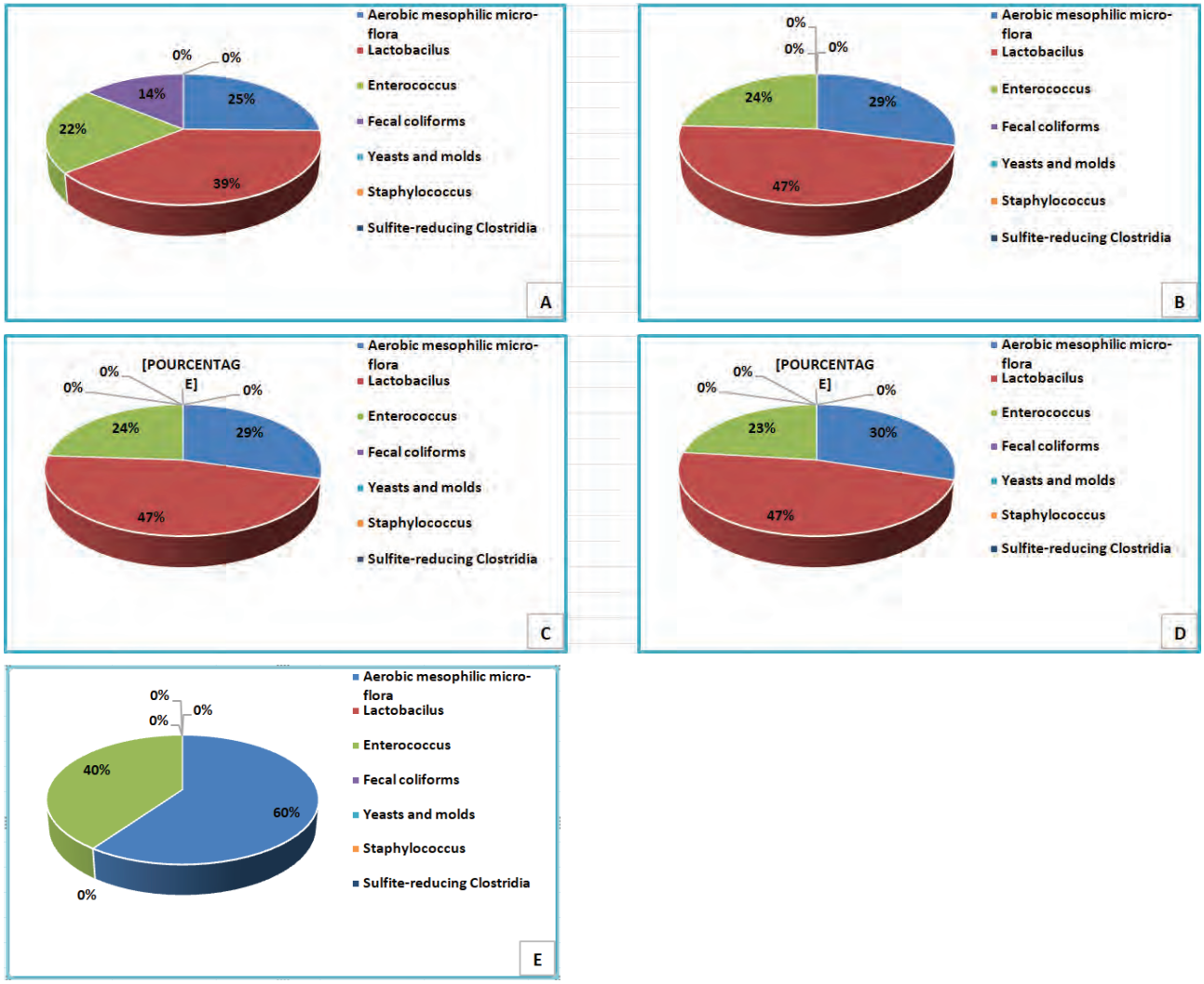


Figure 3. Distribution of the different microorganisms isolated from control Klila throughout the entire storage period.

Sensory analysis results

In this test, assessors were presented with three samples, two of which were identical. These samples were arranged in six possible combinations, with results summarized in the Table 4 below.

Half of the tasters found it easy to differentiate between the control Klila cheese and the enhanced Klila cheese. The reasons for the observed differences between the cheeses were not systematically recorded. Generally, color and odor are the primary factors for differentiation, while aroma and

Table 4. Results of the triangular test

Jury	Order of presentation	Sampleselection	Results
1	AAB	B	1
2	ABA	A	0
3	BAA	A	0
4	BBA	A	1
5	BAB	A	1
6	ABB	A	0

A: Klilawith the addition of *Cotula cinereaC* Delile; **B:** Control Klila (Witness).

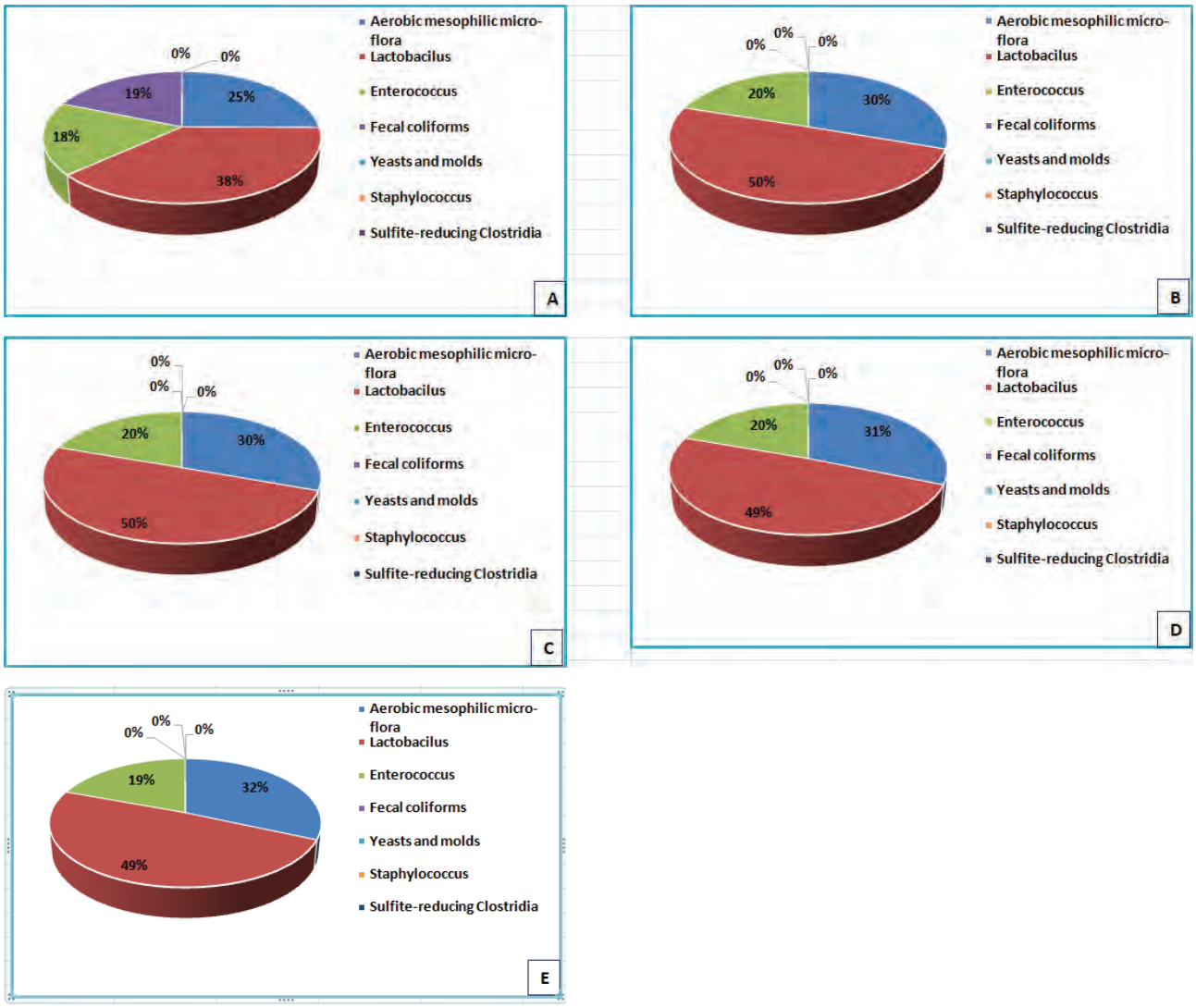


Figure 4. Distribution of the different microorganisms isolated from Klila containing 0.3% *Cotula cinerea* Delile throughout the entire storage period.

appearance are less frequently cited. In some rare instances, rancid sensations (for the enhanced Klila) and a bitter taste were also reported.

Our pH measurements on day J0 are close to those reported by [12] which indicate a pH of 4.6, as well as those of [16], who reported a pH ranging from 3.8 to 4.8, with an average of 4.2, in their studies on dried Klila. These values are comparable to those obtained by [51]; who report a pH of 4.71. Subsequently, pH values gradually increase during storage, from day J7 to day J28, varying from sample to sample due to several factors, including the method and duration of cheese preparation, the type of milk used, and the animals' diet [52].

The acidity levels measured in this study are lower than those reported by [53] at 27.71 °D, and by [51]; at 42.24 °D. Our results also fall below those of [16], who reported an average acidity of 79.4 °D for dried Klila, and those of [54], who found a value of 99 °D for fresh Klila. According to the study by [20], the titratable acidity of Klila varies between 38 °D and 42 °D. Boubekri and Ohta [55] reported titratableacidity values (39 and 42.25 °D) for two Klila samples, which are higher than our results.

The pH and acidity depend on the contents of casein, mineralsalts, and ions [56]; the hygiene conditions during milking, the total microbial flora, and its metabolic activity [57]. This increase in pH may also be attributed to the effect of the plant *Cotula cinerea* Delile.

Regarding protein content, the values obtained in our study are lower than those reported by [58], which indicated values for fresh Klila ranging from 7.11% to 13.48%. They are also lower than the values reported by [51]; who provided a rate of 53.86 g/100 g. Protein content varies based on factors such as breed, age, lactation stage of the

cow, milking frequency, diet, climate, season, and genetic criteria.

Our ash content results are slightly higher than those indicated by [59], who found an ash content of 1% in fresh Klila, and those reported by [58], which were 0.34% for dried Klila. It is observed that the control Klila is richer in minerals than the Klila supplemented with 0.3% of *Cotula cinerea* Delile. The variability in ash content may depend on several factors, including lactation stage, health status of the animal, climatic conditions, and feeding system.

Our findings regarding dry matter content are consistent with those of [53], who reported a dry matter content of 35.03% in dried Klila. However, this rate remain slower than the values reported by [51], who indicated a moisture content of 12.53% and consequently a total dry extract content of 87.47%. Moreover, in the study by [59], the dry matter content of freshcow Klila was found to be 36.6 g/100 g. The dry extract corresponds to the complement of water content up to 100%, influenced by the dry matter of milk and the degree of drainage, as the removal of whey leads to a significant increase in the dry matter content of the cheese [59]. This observation maybe explained by the addition of water to the milk, which reduces the total dry extract content, as well as by climatic, dietary factors and the lactation stage [60].

The count of total viable microorganisms (TVM) reflects the general microbiological quality of a natural product and allows for monitoring its evolution. The total germ count can indicate the state of freshness or decomposition (spoilage) of the product [61]. The values obtained during our study are similar to those found by [17], ranging from 1.43×10^5 CFU/g to 1.01×10^6 CFU/g. The TVM count by [53] yields an average value of 1.24×10^3 CFU/g, while [16] report an average value of 2.22 CFU/g. Furthermore, [62] noted an average TVM value of 1.2×10^6 CFU/g for traditional Klila. The failure to adhere to good hygiene practices during milking, collection of raw milk, or its processing into Klila is likely responsible for this observation.

*Lacticbacteria*are widely distributed in nature and are involved in numerous transformations, not only of milk but also of plant and meat products [17] noted an average Lactobacillus count for fresh Klila ranging from 2.6×10^2 to 4.0×10^3 CFU/g, while in the study by [18], a value of 6.15 log CFU/g isfound, corresponding to 1.4×10^4 CFU/g in dry Klila cheese. The obtained rate ishigherthan the indicated values.

Enterococci are lactic bacteriaused for centuries in food processing. These microorganisms play an essential role in preservation (extendingstorage time) and in the bacteriological quality of foods, while respecting their nutritional and organoleptic properties. However, they are also markers of fecal contamination (*Enterococcus faecalis and Enterococcus faecium*) and are involved in the emergence of nosocomial diseases [63]. Our results differ from those reported by [53], who indicated a value of 4×10^5CFU/g, and are also lower than the results of [10], who reported a value of 5.5×10^5 CFU/g in fresh Klila.

Coliforms are microorganisms associated with spoilage. Our results are consistent with those obtained by [17], which indicate the absence of this germ in this type of traditional cheese. According to [64], the presence of coliforms reflects a lack of hygiene in the animals' environment, during milking, and during the storage and preservation of milk. The absence of coliforms in the later days of storageis likely due to the hygienic quality of the cheese and may also depend on the intrinsic properties of *Cotula cinerea* Delile, which has antimicrobial effects that inhibit *E. coli*.

Our results show also the absence of yeasts and molds in both samples; on the other hand, [16] report yeast and mold counts ranging from 1.5×10^2 CFU/g to 2.2×10^2 CFU/g, while [18] found very high values for this type of count exceeding 5.7 log CFU/g in dry Klila cheese, likely due to poor drying conditions. The presence of yeasts and molds in the analyzed products could be attributed to external contamination, such as that from the stable or laboratory environment (walls, floor, air).

In the other hand; Our results show the absence of Clostridium sulfito-reductans in the two samples analyzed during their preservation. This finding aligns with the results for fresh Klila obtained by [53], who explained the absence of this bacterium in the analyzed samples by the lack of contamination.

Also the results obtained show a complete absence of this pathogenic germ in both types of Klila. Our findings are consistent with those obtained by [53], who indicated a total absence of this germ in all dairy product samples, revealing good hygiene practices from sampling through laboratory handling, as well as the good health of the animals (udder health).

CONCLUSION

The use of biopreservation techniques appears to be a promising alternative to methods involving "chemical" molecules and should meet consumer demands for "more natural" products. Consequently, the evaluation of such properties remains an interesting and useful task, particularly for identifying new sources of natural antioxidant and antimicrobial agents. In this context, we aimed to assess the bioconservation potential of *Cotulacinerea* Delile on the traditional cheese known as Klila. To this end, our study focused on analyzing the physical parameters, antioxidant and antibacterial activity of the aqueous extract of *Cotula cinerea* Delile, as well as the physico chemical, microbiological, and sensory analysis of this type of traditional cheese.

The results of our study allowed us to identify some properties of *Cotula cinerea* Delile powder, such as the dry matter content (3.21%), ash content (3.57%), and plant yield (10.2%). Furthermore, the phyto chemical study indicates a high presence of terpenoids, moderate presence of steroids, coumarins, flavonoids, and saponins, and low presence of

phenols and alkaloids. Additionally, regarding antibacterial activity, the aqueous extract obtained by maceration of *Cotula cinerea* Delile showed strong activity against the tested bacterial strains at a concentration of 0.3g. The physico chemical analysis revealed that the pH and acidity of Klila supplemented with 0.3 g of *Cotula cinerea* Delileare elevated compared to the control Klila. Moreover, for protein content, electrical conductivity, and ash content, the control Klila exhibited slightly higher values than the Klila supplemented with 0.3g during the entire preservation period. The results of the microbiological analyses indicate that the total viable microbial count (TVMC), lactobacilli, and enterococci were present at significant levels in the control Klila compared to the improved Klila with 0.3 g of *Cotula cinerea* Delile. Furthermore, there was a complete absence of *Staphylococcus aureus*, Clostridium sulfito-reductans, and yeasts and molds in both types of analyzed Klila cheese. Based on these results, we can deduce that the extract of *Cotula cinerea* Delile appears to be suitable as a natural agent for the preservation of food products.

ACKNOWLEDGEMENTS

The authors would like to Faculty of Natural and Life Sciences in University of Mustapha Stambouli of Mascara to their support.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Ghaoues S. Evaluation of the physicochemical and organoleptic quality of five brands of partially skimmed reconstituted milk sold in eastern Algeria. Available at: http://depot.umc.edu.dz/handle/123456789/10164 Accessed on Mar 26, 2026. [French]

[2] Benkerroum N, Tamime A. Technology transfer of some Moroccan traditional dairy products (lben, jben and smen) to small industrial scale. Food Microbiol 2004;21:399–413. [CrossRef]

[3] Achaglinkame MA, Dari L, Mörlein D. A review of dairy production and utilization in Ghana and Benin. Discover Food 2023;3:13. [CrossRef]

[4] Chammas GI, Saliba R, Béal C. Characterization of the fermented milk Laban with sensory analysis and instrumental measurements. J Food Sci 2006;71:156–162. [CrossRef]

[5] Bencharif A. Stratégies des acteurs de la filière lait en Algérie: état des lieux et problématiques. Options Méditerranéennes Ser B 2001;32:44. [French]

[6] Fox PF. Cheese | Overview. In: Fuquay JW, editor. Encyclopedia of Dairy Sciences, Second Edition. San Diego: Academic Press; 2011. p. 533–543. [CrossRef]

[7] Chamba JF. Applications des bactéries lactiques lors des fabrications fromagères. In: Corrieu G, Luquet FM, editors. Bactéries lactiques - De la génétique aux ferments. Paris: Lavoisier; 2008:787–815. [CrossRef]

[8] Fox PF, McSweeney PL, Cogan TM, Guinee TP, editors. Cheese: Chemistry, Physics and Microbiology, Volume 1: General Aspects. Elsevier; 2004.

[9] Papadopoulou OS, Argyri AA, Bikouli VC, Lambrinea E, Chorianopoulos N. Evaluating the quality of cheese slices packaged with Na-alginate edible films supplemented with functional lactic acid Bacteria cultures after high-pressure processing. Foods 2022;11:2855. [CrossRef]

[10] Braken A, Terki I. Analyse microbiologique et physicochimique du fromage Klila fabriqué à Ain Sefra wilaya Naama, mémoire de Master en sciences Biologiques. Université Mustapha Stambouli de Mascara, Algérie; 2016. [French]

[11] Bell V, Guina J, Fernandes TH. African fermented foods and beverages. Potential impact on health. In: Microbial fermentations in nature and as designed processes. In: Hurst CF, editor. Newyork: Wiley; 2023. p. 293–322. [CrossRef]

[12] Leksir C, Chemmam M. Contribution à la caractérisation du klida, un fromage traditionnel de l'est de l'Algérie. Livest Res Rural Dev 2015;27:5. [French]

[13] Leksir C, Boudalia S, Moujahed N, Chemmam M. Traditional dairy products in Algeria: Case of Klila cheese. J Ethnic Foods 2019;6:1–14. [CrossRef]

[14] Leksir C. Caractérisation, fabrication et consommation du dérivé laitier traditionnel "Klila" dans l'Est algérien. Doctoral dissertation, Université 8 Mai 1945 de Guelma; 2018. [French]

[15] Krim T, Rekhou L. Etude du procédé artisanal de fabrication du fromage traditionnel «Tiklilt» dans la région de Tizi-Ouzou [Doctoral dissertation]. Université Mouloud Mammeri; 2021.

[16] Guetouache M, Guessas B. Characterization and identification of lactic acid bacteria isolated from traditional cheese (Klila) prepared from cows milk. Afr J Microbiol Res 2015;9:71–77. [CrossRef]

[17] Mennane Z, Faid M, Lagzouli M, Ouhssine M, Elyachioui M, Berny E, Ennouali M, Khedid K. Physico-chemical, microbial and sensory characterisation of Moroccan Klila. Middle East J Sci Res 2007;2:93–97.

[18] Benamara R, Gemelas L, Ibri K, Moussa-Boudjemaa B, Demarigny Y. Sensory, microbiological and physico-chemical characterization of Klila, a traditional cheese made in the south-west of Algeria. Afr J Microbiol Res 2016;10:1728–1738. [CrossRef]

[19] Sahraoui F. Essai d'amélioration du fromage traditionnel "KLILA" face au bacille de la tuberculose bovine. Master's thesis in Agronomy, Biotechnologie Alimentaire. Université Algeria:Abdelhamid Ibn Badis-Mostaganem; 2016. [French]

[20] Lahsaoui S. Etude de procédé de fabrication d'un fromage traditionnel (Klila). Master's thesis in Engineering, Agronomy Department. Batna, Algeria: Université El Hadj Lakhdar; 2009. [French]

[21] Leyva Salas M, Mounier J, Valence F, Coton M, Thierry A, Coton E. Antifungal microbial agents for food biopreservation—a review. Microorganisms 2017;5:37. [CrossRef]

[22] Shaltout FA. Impact of preservatives on food preservation and their effect on the bacteria. Univ Lib Innov Res Stud 2024;1:1. [CrossRef]

[23] Garry P, Christieans S, Cartier P. Procédés de bio-préservation. Viandes 2008;203.

[24] Guaouguaou FE, Es-Safi NE. Products with potential biological activities. Drug Discov Design Using Nat Prod 2023;465. [CrossRef]

[25] Bouziane M, Badjah-Hadj-Ahmed Y, Hadj-Mahammed M. Chemical composition of the essential oil of Brocchia cinerea grown in South Eastern of Algeria. Asian J Chem 2013;25:3917. [CrossRef]

[26] NF B 51-004 standard. Wood Determination of Moisture. AFNOR, French standard; 1985. Available at: https://www.boutique.afnor.org/en-gb/standard/nf-b51004/wood-determination-of-moisture-content/fa032831/3900 Accessed on Apr 02, 2026.

[27] Pinta M, Bourdon B, Rousselet F. Spectrophotometrie d'absorption atomique. Paris; Masson et Arston; 1980. [French]

[28] Moroh JLA, Bahi C, Dje K, Loukou YG, Guede-Guina F. Étude de l'activité antibactérienne de l'extrait acétatique (EAC) de Morinda morindoides (Baker) Milne-Redheat (Rubiaceae) sur la croissance in-vitro des souches d'Escherichia coli. Bull Soc Roy Sci Liège 2008;77:44–61. [French]

[29] Bssaibis F, Gmira N, Meziane M. Activité antibactérienne de Dittrichia viscosa (L.) W. Greuter. Rev Microbial Ind San Environn 2009;3:44–55.

[30] Bekro YA, Mamyrbekova J, Boua B, Tra Bi F, Ehile E. Étude ethnobotanique et screening phytochimique de Caesalpinia benthamiana (Baill.) Herend. Et Zarucchi (Caesalpiniaceae). Sci Nature 2007;4:217–225. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 880–893, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00051

Research Article

Multi-zone commercial market building HVAC control strategy based on reinforcement learning algorithm models

Ganesh MURADE[1,*], Bhanu Pratap SONI[2], Ankit Kumar SHARMA[1]

[1]University of Engineering & Management, Jaipur, Rajasthan, 700160, India
[2]SEEE, Fiji National University, 1544, Fiji

ARTICLE INFO

Article history
Received: 16 July 2024
Revised: 04 October 2024
Accepted: 05 March 2025

Keywords:
Actor-critic Learning; Deep Deterministic Policy Gradient (DDPG); Deep Reinforcement Learning (Deep RL); Multi-Zone Commercial HVAC; Linear RL; Reinforcement Learning

ABSTRACT

The optimization of residential heating, ventilation, and air conditioning (HVAC) systems is crucial for reducing energy consumption and maintaining user comfort. As urbanization increases, there is a demand for smart buildings with energy-consuming appliances. Researchers are developing HVAC control strategies, especially for commercial buildings with complex load patterns. This paper presents a novel model-free deep reinforcement learning (RL) approach. RL allows systems to learn and adapt to individual occupant preferences, creating customized comfort using the deep deterministic policy gradient (DDPG) algorithm to design an optimal control strategy for multi-zone residential HVAC systems. DDPG aims to reduce energy costs while maintaining occupant comfort. The DDPG reinforcement learning technique is applied to control a multi-zone commercial HVAC system, aiming to minimize energy costs while maintaining occupant comfort. The DDPG method achieves a 56% faster convergence time compared to linear HVAC control methods and a 15% improvement over linear reinforcement learning models. The mean steps required for DDPG and linear RL models are 9.9 and 115.3, respectively.

Cite this article as: Murade G, Soni BP, Sharma AK. Multi-zone commercial market building HVAC control strategy based on reinforcement learning algorithm models. Sigma J Eng Nat Sci 2026;44(2):880–893.

INTRODUCTION

It is estimated that 41% of India's annual energy consumption [1] is attributed to the building sector. Consequently, building temperature regulation has garnered significant attention [2,3] recently. The primary aim is to strike a balance between minimizing energy usage and ensuring high comfort levels for office workers. However, uncertainty arises due to occupants' behavior and insufficient data on their comfort preferences, impacting decision-making regarding optimal control strategies [4,5]. Integrating these considerations into the design process poses a pressing challenge. Occupant behavior adds complexity to building systems, characterized by continuous and discrete states, further complicated by variable sensor requirements. Hence, developing a scalable control framework is imperative. Drawing inspiration from recent

*Corresponding author.
*E-mail address: ganesh.murade26@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

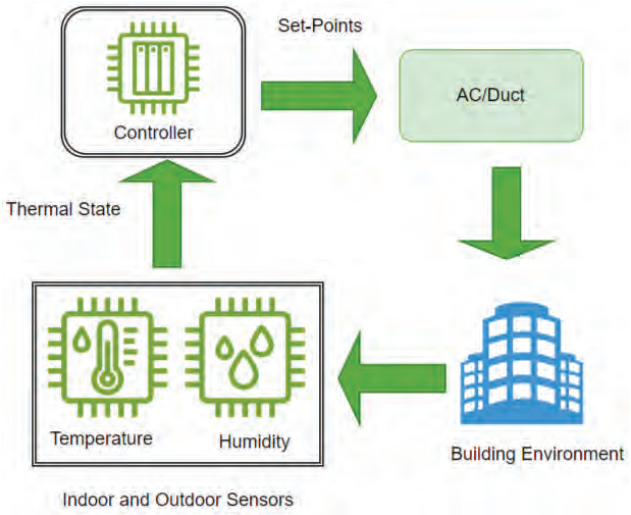


Figure 1. The design of the thermal control system and Controller set Points (sensors).

advancements in RL [6,7], this study aims to evaluate the efficacy of RL in addressing intelligent temperature control in building environments.

The air conditioning system, depicted in Figure 1, currently stands as the most sophisticated device for maintaining a comfortable temperature within a structure. Additionally, efficient demand-side energy management measure, aid in reducing peak loads and stabilizing overall system performance [8]. Several studies have been conducted to determine the optimal methods for optimizing HVAC control systems to enhance efficiency.

A primal-dual algorithm is used in [9] to determine the energy supplier's pricing strategy and the HVAC operating states that are most beneficial to the user, within the context of a load prediction error model for HVAC energy management. For responsive commercial HVAC demand scheduling a day in advance, another study uses a regression method [10]. According to [11], a hierarchical control method might allow the HVAC system to serve as the principal frequency regulator for the bulk system. In [12], an optimization method based on the work of Yuri Lyapunov is presented for managing HVAC loads without requiring estimates of price or temperature fluctuations. A distributed Tran's active control market approach for HVAC systems in commercial buildings is presented in [13] to demonstrate the efficiency of HVAC systems during peak shaving and load shifting (Fig. 2). Using an adaptive control system, Andres-Chicote et al. [11], managed the building's HVAC system to maximize both occupant thermal comfort and the building's energy efficiency. Giuseppe et al. [12], proposed a model prescient control and hereditary calculation-based streamlining structure to diminish warming energy utilization and related inconvenience. Wei and others introduced an information-driven technique for controlling air conditioning frameworks with changing air volume giving profound support to learning. Deng et al. [14], proposed a unique HVAC control system that actively detects changes in the building's environment by using DQN, resulting in significant energy savings.

All of the previously mentioned methodologies fall under the umbrella term of "model-based techniques," which need logical arrangement tool compartments for pragmatic runtime of the executives as well as displaying the particular warm elements of the air conditioning while at the same time considering the impacts of the encompassing climate. Since the structure and hardware models should be acclimated to a given structure to deliver the right

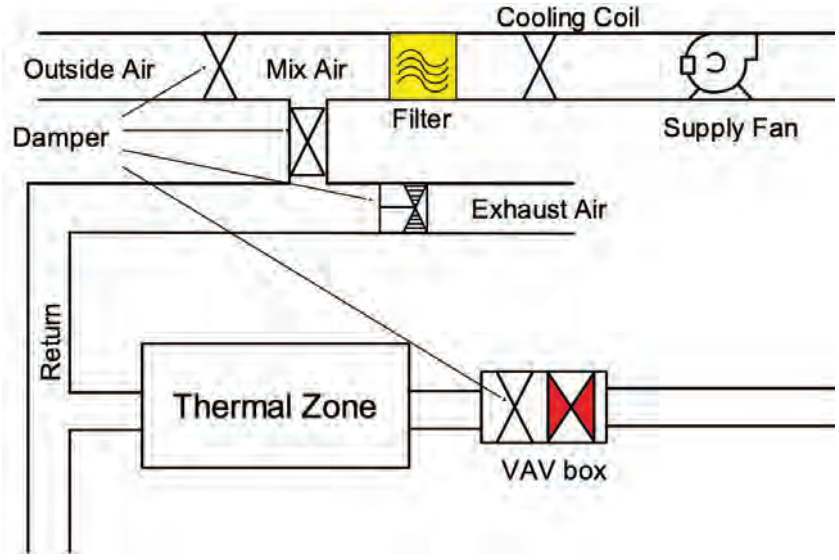


Figure 2. Schematic of the HVAC air distribution system.

discoveries, the model-based approaches might experience the ill effects of estimation mistakes (for instance, building model incorrectness) and figuring failures. This presents a significant obstacle to the widespread application of model-based methods.

In the meantime, the outcome of Alpha Go [14] shows the headway that has been made in AI advancements like profound learning and support learning. In [15], the creator presents a wide outline of the possible utilizations of AI in the field of force framework research. In [16], an illustration of how AI-based technology is beginning to be implemented in industrial applications in actual control centers, the primary known control-room utilization of man-made intelligence-driven circulated includes choice for an enormous, certifiable power matrix is shown.

Specifically, profound Reinforcement learning (RL), which consolidates a profound brain organization (DNN) with RL, has collected critical interest as of late for its capability to successfully address exceptionally complex control and improvement issues across many aspects. Two unique ways to deal with upgrading the energy the board strategies for HEVs are twofold Q learning [17] and ceaseless profound deterministic approach inclination (DDPG). Reinforcement learning (RL), a machine learning method that has emerged in recent years, features self-learning and online learning capabilities. RL is an information driven control approach since, utilizing the "activities and prizes" component, it can achieve the versatile improvement of regulators in any event, when no control framework models are free. In [18], the no concurrent advantage entertainer pundit is utilized to decide the most practical times for the activity of an organization of decentralized power plants. In [19], a profound Q learning approach is created to help mass power framework upkeep direction. In [20], a protected profound RL technique is investigated to get the ideal control plan of the dynamic dissemination network considering voltage level cutoff points. To do this, we add a secure layer to the standard actor-network, which helps to ensure that voltage limitations are not breached.

A portion of the primary endeavors on the air conditioning framework control issue zeroed in on the most proficient method to best utilize the strong profound RL technique to boost both energy and monetary effectiveness. A convolutional neural network (CNN) approximates the state-activity esteem capability in [21] to more readily catch the spatial and worldly connections in the info state information. In [22], scientists investigate the utilization of a profound strategy slope (DPG) way to deal with dealing with a wide assortment of responsive necessities. Utilizing an actor-critic approach, [23] focuses on optimizing HVAC's thermal comfort and energy consumption. In [24], an advantageous actor-critic-based HVAC control framework is created for a comprehensive building energy model. To accomplish appropriate control balance across various HVAC systems, linear reinforcement is used [25].

All of these studies show that using deep RL techniques to optimize the HVAC temperature management strategy is more successful than using custom-built benchmarks. To reduce the search space, previous studies frequently discretize continuous HVAC control activities like HVAC set point or air flow rate. At low granularities or without combining action spaces, discretization may achieve satisfactory performance. However, when the action space is high-dimensional, as when controlling HVAC in a building with multiple zones, it encounters the exponential explosion issue. The algorithm's performance suffers as a result, and deep RL technique training necessitates more simulations. When the action space is high-dimensional, like when controlling HVAC in a building with several zones, however, it runs into the problem of exponential growth. As a consequence, the algorithm performance degrades and more simulations are required for training deep RL techniques.

As a result of the above worries, this study additionally employs the DDPG technique to fine-tune the continuous thermal management strategy of commercial HVAC. Here is a quick rundown of what this study adds to the body of knowledge:

- Designed benchmark scenarios without RL to illustrate that the implemented DDPG may give larger monetary gains while keeping consumer comfort.
- To demonstrate the superiority of the DDPG technique in managing the never-ending activity space, which is more typical in a plethora of verifiable scenarios, we lead a comprehensive correlation between it and the widely used direct assistance strategy.
- The optimum HVAC control technique is a well-trained deep RL approach with good generalizability and adaptability that can handle a wide range of pricing signals and physical constraints.

Research paper insights in-depth exploration of the air conditioning control problem. A brief overview of two prominent deep reinforcement learning approaches. The paper also emphasizes the linear reinforcement method and simulation results of the DDPG method, along with a comparative analysis against the linear reinforcement method. The paper underscores the importance of the linear reinforcement method for its simplicity and ease of implementation.

PROBLEM FORMULATION FOR MULTI-ZONE CLIMATE CONTROL IN COMMERCIAL MARKET SETTINGS

Commercial buildings, such as malls, offices, and supermarkets, contain distinct climate zones with different heating and cooling requirements. Controlling multi-zone efficiently is a challenge due to factors like occupancy fluctuations, weather variations, building thermal properties, and energy costs. Multi-zone climate control systems aim to enhance energy efficiency while maintaining comfort and regulatory standards. Multi-zone HVAC (Heating,

Ventilation, and Air Conditioning) systems are designed to regulate the temperature and air quality across different zones within a building. Section 2.1. gives the brief information about the various issues related to implementation of HVAC in multizone.

Introduction to the HVAC Multi-Zone Control Issue

In this analysis, we focus on a multi-floor commercial/market structure. The HVAC system has a setpoint that may be changed to regulate the interior temperature in each zone individually. In addition to the "Cooling" and "Heating" settings, an "Auto" setting is also available for the HVAC system. In view of the room's ongoing temperature and the client's chosen temperature edge, the air conditioning framework will consequently switch among cooling and warming when the indoor regulator is in "Auto" mode. To keep individuals agreeable, the warming, ventilation, and cooling framework will turn on when within temperature decreases beneath the setpoint. We will zero in on the situation in which all zones require warming without forfeiting over-simplification. The motivation behind central air framework control is to give an agreeable inside climate at the most reduced conceivable energy consumption.

Multi-Zone HVAC System

A multi-zone HVAC system divides your house into smaller regions or zones, allowing you to independently manage the temperature in each. You could simply establish a distinct zone for each level or even separate zones for each room in the home, depending on the arrangement of your home and your individual wants and needs.

No matter how many zones you have, each one will have its own thermostat that solely measures and regulates the temperature in that zone. This allows you to control the temperature for that zone while leaving the rest of the building alone. Because all of the air still flows via the same air handler and ductwork, the only true constraint is that you cannot have one zone set to heating and another set to cooling.

Dampers, which are simply metal plates situated inside the ductwork that may open and close to control the airflow to each zone, are used in a multi-zone system. This is analogous to blocking the supply vents to a specific room or area of the house. The distinction is that sealing the vents simply prevents air from entering that room, implying that air will continue to flow via that branch of the ductwork.

This is normally not suggested since it might cause a pressure imbalance inside the HVAC system, affecting how well and efficiently the system warms or cools. It also puts more strain on your furnace or air conditioner, which can contribute to premature wear and tear on its components. Multi-zone HVAC systems are not affected by pressure imbalance difficulties since the damper cuts off the whole branch or section of ductwork rather than merely turning it off at the end of the branch as happens when the vents are closed.

Despite the fact that each zone has its own thermostat, the system is still managed by a single central control panel. When a zone requires hot or cold air, the control panel opens the damper for that zone, allowing air to flow to the zone. When the zone achieves the desired temperature, the thermostat in that zone sends a signal to the central control panel, which closes the zone's damper. Each zone may have a single damper or many dampers that govern the airflow to that zone, depending on the size and structure of the zones.

Application of Markov Decision Process (MDP) in HVAC Control Issue

In this subsection, we present a Markov Choice Cycle (MDP) detailing the multi-zone Business/Market air conditioning the board issue that will be tended to in Area 3 utilizing a without model profound RL-based approach. The current interior temperature, according to the enhanced warm elements model of air conditioning in [26], is only associated with previous state boundaries, such as the indoor temperature at the previous time span, and is independent of the indoor temperature at certain other time spans. Since climate control may be represented as a restricted Markov process, the RL method might be used to find optimal configurations.

A controlled stochastic process with the Markov property with costs assigned to state changes is referred to as a Markov decision process. A Markov decision problem consists of a Markov decision process and a performance criterion. A policy, mapping states to actions, that (possibly stochastically) determines state transitions to reduce the cost according to the performance criterion is a solution to a Markov decision problem. Markov decision problems (MDPs) provide the theoretical underpinnings for decision-theoretic planning, reinforcement learning, and other sequential decision-making tasks of interest to artificial intelligence and operations research academics and practitioners. MDPs use dynamical models based on well-understood stochastic processes and performance criteria based on known theory in operations research, economics, combinatorial optimized performance, and the social sciences. It appears that MDPs have a unique structure that can be used to accelerate their resolution. In investment planning, for example, the beginning state is frequently known with certainty (the present price of a stock or commodity), limiting the range of likely achievable states (future values) and possible investment plans in the near-term future.

In general, MDPs have inherent properties such as time, action, and reachability in state space that can be used to create efficient algorithms for solving them. It is critical that we grasp the computational challenges involved in these sources of structure in order to understand the potential for efficient sequential and parallel algorithms that compute both exact and approximate answers. The dynamics of an agent interacting with a stochastic environment are described by a Markov decision process. The Markov decision process describes the following development of the

system state over a (potentially infinite) sequence of times referred to as the stages of the process, given a starting state or distribution of states and a sequence of actions. This study focuses on the infinite-horizon scenario, when the stage sequence is unlimited.

The four fundamental parts of a MDP are the state (s), the activity (a), the probability of changing to another state (p), and the award (r). These four elements are at play when a multi-zone HVAC control issue arises in a commercial or industrial setting:

The user's minimum acceptable temperature, $T_{lower}(t)$, should be reported together with the current outside temperature, Tout(t), and the current inside temperature, T_{in},z(t), for each zone z. Fourth, the retail price is the retail(t) in this time step.

Keep in mind that the user's minimum acceptable degree of discomfort is a time-dependent component of the condition. This is because we expect that HVAC occupants have varying comfort needs throughout the day. This makes sense, since the interior temperature's comfort range may be adjusted to save money on utility bills when no one is home throughout the day. When the home is used in the evenings and on weekends, the temperature may be returned to a more comfortable setting.

In order to achieve the preheating impact of HVAC, the current retail price is also included in the state parameters. Setting the air conditioning framework's setpoint to a moderately large number while the retail cost of energy is low takes into consideration early warming of within climate, saving money on energy costs that would otherwise be incurred as the temperature outside drops.

$$\text{HVAC} = \begin{cases} 1, if\ Tin\ (t) < setpoint - deadband \\ 0, if\ Tin\ (t) > setpoint \\ remain\ at\ the\ current\ status, elsewise \end{cases} \quad (1)$$

In this review, we exclusively center on the warming utilization of the central air model. The dead band in Eq. (1) is the temperature range outside of which the thermostat will not switch between the on and off positions, hence preventing rapid cycling. In Eq. (1), when the interior temperature is above the user-specified comfort level, we can see that the HVAC system is turned on when the temperature drops below that level.

- Reward: During the control period, the cost of energy usage is added to the cost of comfort deterioration, which is defined as follows:

$$r(t) = -\omega_c \sum_{t=t-\Delta t}^{t} \lambda^{retail}(t)\, E_{HVAC}(t) - \sum_{t=t-\Delta t}^{t} c^{penalty}(t) \quad (2)$$

The HVAC system's energy cost is represented by the first component in Eq. (2), where $\lambda^{retail}(t)$, $E_{HAVC}(t)$, and t, respectively, denote the Δt control interval, power consumption, and retail price; the subsequent term addresses the punishment for client solace infringement, still up in the air as follows.

$$c^{penalty}(t) = \begin{cases} 1, for\ T_{in}\ (t) < T_{lower}\ (t) - T_{in} \\ 0, elsewise \end{cases} \quad (3)$$

T_{th} has a low value, which is a threshold used in Eq. (3). If the magnitude of the temperature deviation is less than T_{th}, it is disregarded. The dead band in the HVAC system makes it difficult to maintain the desired interior temperature at all times. There's wiggle room for the thermostat inside the house because of the threshold.

Since the prize incorporates the energy cost and the punishment, the multi-objective limit is provoked by relegating various loads to the two goals, proposed by ω_c and ω_p in Eq. (2). The reduction of r(t) is a clear goal of air conditioning warm control, which is the cost of energy consumed in addition to the penalty incurred throughout the control cycle, to the bare minimum: r(t): ΣNT t=1r(t). As a result, a multi-step decision-making dilemma arises from the need for a long-term control approach to forestall the effects of unknowable future conditions.

Considering the abovementioned, this study utilizes the sans model profound RL way to deal with managing the imperceptibility inborn in the multi-zone Business/ Market air conditioning control issue. To be effective, the sans model RL approach requires no earlier information on the climate or state changes. It learns from the results of its decisions and exchanges information with its surroundings over time to improve its decision-making process. In this manner, mistakes made in forecasting due to unknown variables, as well as those made in measuring the thermal mass of a structure, may be avoided. In the following paragraphs, we will go into further depth about the deep RL approach.

DDPG-BASED MULTI-ZONE HVAC CONTROL STRATEGY

Deep deterministic policy gradient (DDPG) is an RL algorithm that integrates deep learning with both policy-based and value-based approaches. It is particularly effective in multi-zone HVAC management, where it improves energy performance, maintains ideal comfort conditions, and adapts to environmental variations. Section 3.1. insights the various methods in deep reinforcement learning.

Methods in Deep Reinforcement Learning

The RL approach is an AI procedure for further developing MDP dynamic systems. The MDP-specified reward serves as an input for the RL algorithm's evolutionary process. If the reward for continuing in the present direction is high enough, the algorithm will continue to look there, and vice versa. When there are time limitations or a hidden state space involved in a decision, the RL approach excels.

The two most prevalent types are policy-based and value-based RL techniques. The approaches to activity assessment are where the two methodologies diverge. Esteem

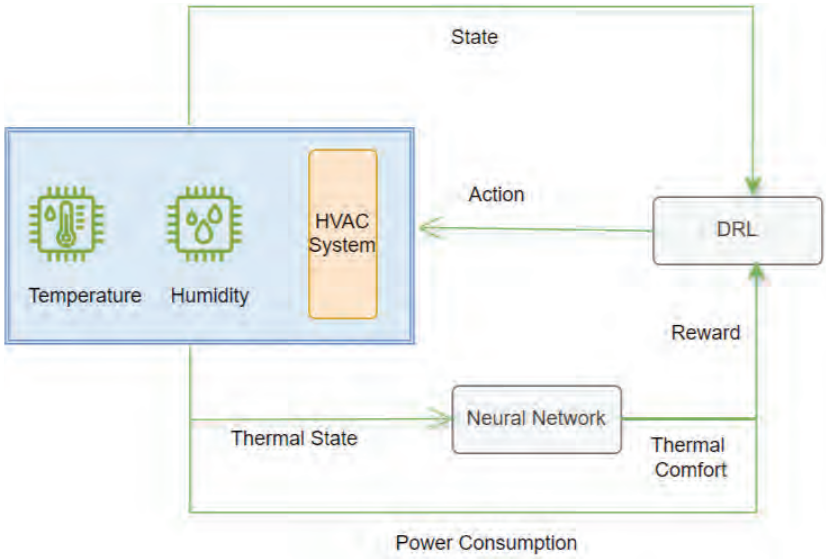


Figure 3. The flow of the building control system using DNN for Thermal comfort and DRL for thermal control.

based RL assesses the Q worth of a state-movement pair (s, a), which is the all-out compelled reward starting from performing activity at state s, though procedure-based RL produces probabilities of all potential activities at the ongoing status and chooses the activity with the most noteworthy probability.

The main RL strategy utilizes a DNN combined with RL. Figure 3 depicts a flow of the building control system using a deep neural network for thermal comfort and Deep Reinforcement learning for thermal control. In significant RL, the DNN is utilized as a backtracking device for foreseeing the Q regard (in accordance with the worth-based RL strategy) or the development likelihood (in accordance with the method-based RL approach). Figure 4 depicts a typical DNN layout for RL regression.

To perform significant level control for extremely confounded circumstances, for example, those with persistent state space or activity space, without the plain limits is the major advantage of the profound RL approach over the customary RL technique. In contrast to traditional Q realizing, which uses an actual Q table to record all possible activity levels, advanced RL creates a more rounded relapse model. On account of constant control, this summed up relapse model gives more strong and versatile strategies against obscure circumstances. Straight Support will act as a substitute for esteemed based profound RL strategies, and the DPG approach will act as a representation of strategy based profound RL techniques. Then, at that point, the DPG strategy, a ceaseless control technique that consolidates the previously mentioned approaches, will be described in depth as a means of determining an ideal multi-zone Commercial/Market HVAC management strategy.

DDPG for Constant Temperature and Humidity Regulation in a Building

The DDPG (Deep Deterministic Policy Gradient) technique was specifically designed to address challenges involving continuous variables. Unlike the Straight Assistance or DPG methods, where the Deep Neural Network (DNN) generates Q-values or probabilities for all possible actions

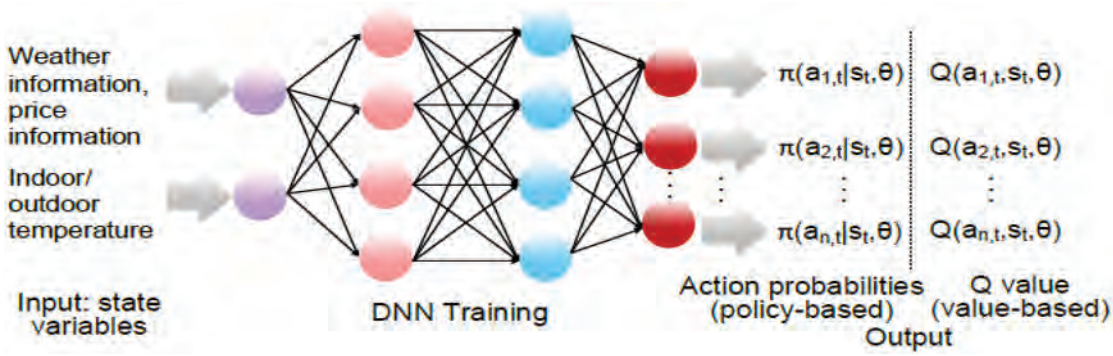


Figure 4. Approximating functions in RL using a DNN framework.

for the agent to choose from, the term "deterministic" in DDPG highlights that the DNN produces a single, definitive output rather than a set of options. This unique characteristic allows the action space to be defined continuously, facilitating effective handling of continuous control tasks. A key advantage of DDPG is its integration of Linear Reinforcement with DPG. The Deep Deterministic Policy Gradient (DDPG) algorithm is a reinforcement learning method that employs two neural networks working together to solve continuous action space problems. The functions they serve are described below.

The entertainer network gets the present status and plays out a deterministic activity; the current status and the action made by the performer network are dealt with into the savant association, in which the value of the state-action pair, Q, is sent. Utilizing this new Q esteem, the entertainer organization's settings will be changed. The entertainer organization's misfortune capability is the greatest squared blunder (MSE) of the Q esteem, as per DPG rationale, while the pundit organization's misfortune capability is the mean squared mistake (MSE) of the Q esteem. Taking everything into account, the entertainer network is answerable for choosing acts, while the pundit network assesses the picked activity [26-30].

As illustrated in Figure 5, the algorithm initially acquires state information for the external environment, such as temperature and retail price value. The program also receives a task ID from the external environment, which is a 0-1 binary variable that indicates whether the situation is cooling or heating. The task ID is an important indicator of the current task on which the actor is working. The state parameters will then be normalized. Normalization is required because the state characteristics of the two jobs can differ greatly [31-32]. For example, in the cooling scenario, the outdoor temperature is substantially higher than in the heating scenario. Data that is not normalized might cause algorithm divergence [33]. The concatenation of the normalized state parameters with the task ID is then delivered to the deep neural networks. The DDPG method employs two types of neural networks: the actor-network and the critic network. The actor-network generates HVAC control actions, while the critic network computes the Q value as an evaluation of the chosen action. There is also a behaviour network and a target network for both the actor-network and the critic network. The control action is produced by the behavior network, while the target network generates a target value for the behavior network to learn, which is analogous to labelled data in supervised learning. The target network aids in the stabilization of the training process [34,35]. The DDPG algorithm has four neural networks in total. Actor Network, Critic Network, Target Actor Network, and Target Critic Network. To improve action selection, the actor network updates itself using policy gradients based on state inputs. The critic network calculates

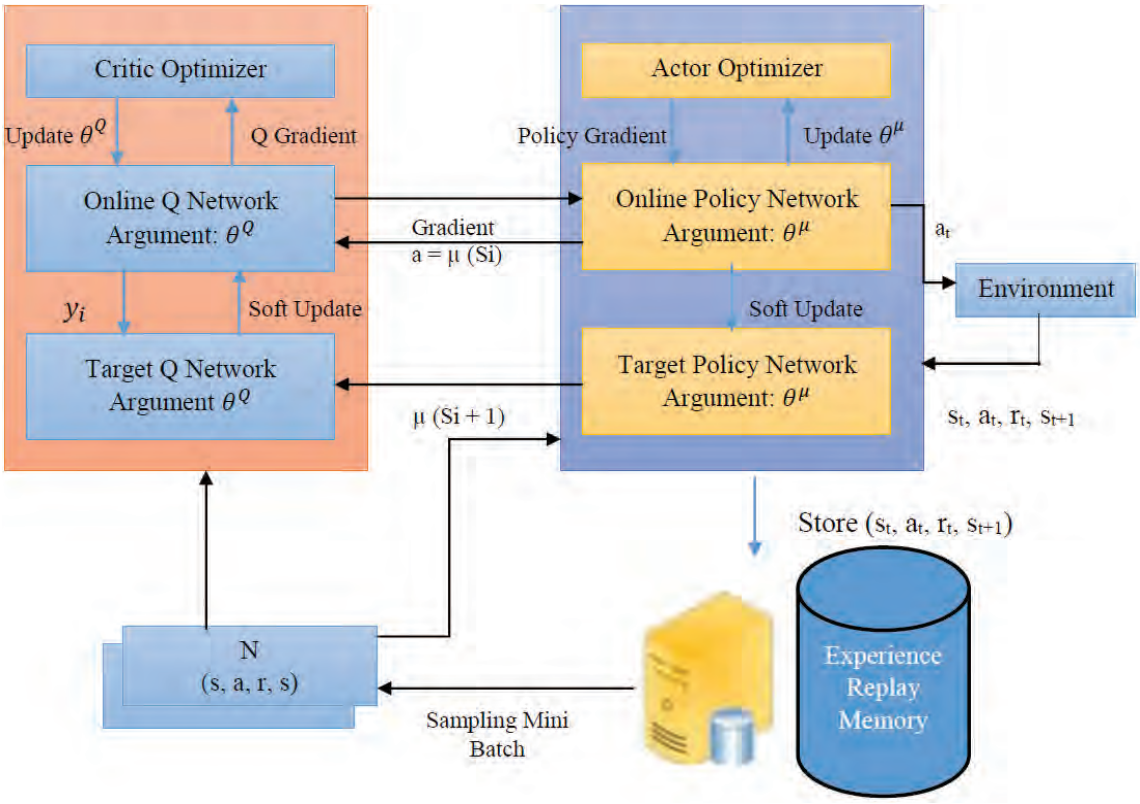


Figure 5. Multi-task DDPG for multi-zone HVAC control.

the Q-value, representing the expected reward for a specific state-action pair [36-38]. Designed for stability, the target actor network is a delayed version of the main actor network with slow updates. The target critic network acts as a backup of the primary critic, updating gradually to prevent fluctuations in Q-values.

Similar to the Straight Support approach, the Deep Deterministic Policy Gradient (DDPG) algorithm utilizes two types of neural networks: the actor network (policy network) and the critic network (Q-value network). Each of these networks is paired with a corresponding target network, resulting in a total of four neural networks. The inclusion of target networks helps stabilize the training process by providing more consistent and reliable updates, which improves the convergence of the algorithm.

The DDPG calculation is depicted in additional profundity in the following segment

Algorithm DDPG for Optimal HVAC Control Design

Similar to Linear Reinforcement, an actor-network is constructed using the DDPG method to choose a deterministic activity. The DDPG calculation utilized is made sense of more meticulously underneath. Before initiating their respective target networks, we randomly initialize the actor network and the critic network, both of which are neural networks [39]. The system state is initialized at the start of each cycle, and the current actor-network is used as the setpoint for selecting an HVAC control action [40]. In order to encourage the algorithm to further investigate the chosen action, some noise is included [41]. During the full t period, the specified action is carried out in the environment while the outcome states and rewards are monitored. To train an algorithm, the state transition (s(t), Setptz(t), r(t), s(t + t)) is recorded in a replay buffer. Once enough transitions have been gathered, a random subset of transitions is used to update the parameters of the actor-network and the behavior network. The assumption of independence and uniform distribution in the learning model may be preserved by the use of random selection, which can sever temporal connections between transitions. The efficiency with which the transitions may be used is further improved by the fact that they can be sampled several times [42].

In response to the loss functions, adjustments are made to the Q and of the neural network. The mean squared mistake (MSE) between the objective Q respect and the genuine Q respect is utilized to portray the blunder-making ability of the predominant scholarly union. The objective Q esteem is the sum of the continuing prize and a limited Q esteem from the objective pundit organization "Q," where is the contrast error, for the new control stretch "t + t," where is the discount variable [40]. Subsequent to deciding the misfortune capability, the Q conduct pundit organization's boundaries are changed utilizing the angle. The pace of learning is denoted by ηQ.

The actor network's loss function is designed to optimize the quality factor (Q):

$$\max \frac{1}{M} \sum_{i=1}^{M} Q\left(s^{(i)}(t); a^{(i)}(t), \theta^{Q}\right) \mid a^{(i)}(t) = \pi s^{(i)}(t); \theta^{\pi}) \quad (5)$$

The actor-network Q(s;θ^{Q}) is used to provide the solution an (i)(t) in Eq. The chain rule is utilized to work out the Q worth's slope according to. While the conducting network is continually being refreshed, the objective pundit organization and target entertainer organization's boundaries, θ^{Q} and θ^{π}, are changed at a slower pace. This more gradual update serves to strengthen the reliability of the learning process. The actor network is updated using policy gradients, influenced by the critic's Q-values. The critic network is updated using the TD error, learning to predict accurate Q-values. Both networks gradually improve through these updates, with target networks stabilizing the process. Through this iterative process, both the actor and critic networks become better at selecting actions and evaluating them, leading to improved decision-making and long-term rewards [43,44].

PROBLEM FORMULATION

Energy management in smart homes has become increasingly important due to rising energy costs and the need for sustainable practices. This research focuses on two primary objectives: cost minimization and efficiency in energy management. By leveraging time-of-day tariffs and various optimization algorithms, this study aims to develop strategies that reduce energy costs and enhance efficiency without compromising the residents' comfort.

Rating Performance

Here, we prove the benefits of the DDPG method by simulating it with real-world data and comparing it to the linear reinforcement-based discrete control method and the benchmark cases. This is done so as to show that the DDPG method is the superior choice for multizone Commercial/ Market HVAC systems.

Modelling And Simulation Setting

Real-world weather data is used for training and testing in a two-zone house HVAC model from 1/1/2021 to 31/12/2021 for 8 different zones (TOUT, ITUZ1, ITUZ2, TCWZ2, TCH2, TDICZ1, TIECZ1, and TDXZ1). The goal of using such a dynamic pricing sequence is to test the deep RL agent's ability to decode market signals and adapt its control tactics accordingly. It is also believed that the user's minimum acceptable degree of comfort fluctuates 144 times every day.

The RL agent's control interval is set at 10 minutes Δt = 10.

Since we only care about how the HVAC system affects heating, we chose the November weather data for training. Each training session is considered one episode. Each episode will provide 24 transitions of the form:

$$(s^{(i)}(t), \text{Setpt (i) z (t)}, r^{(i)}(t), s^{(i)}(t + \Delta t)) \quad (6)$$

Table 1. The DDPG and linear RL algorithms both make use of a DNN-based framework

Algorithm	DDPG		Linear RL
	Critic network	Actor network	
	Air Conditioner	1500	6
Input Size	(1,7)	(1,5)	(1,5)
Number of Secret Levels	2	2	2
the thickness of each covert layer	(7,20), (20,10)	(5,20), (20,10)	(5,20), (20,10)
Quantity of results	(1)	(2)	(25)
The secret layer's activation function	ReLU	ReLU	ReLU
Rate of learning (η)	0.001	0.01	0.01
Payoff significance	$\omega_{c:}$ 10, ω_p: 1		

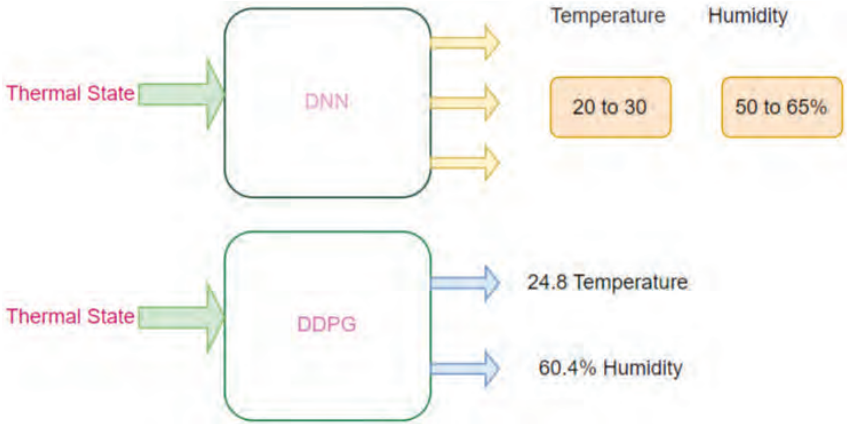


Figure 6. A comparison of the thermal control strategies, DDPG and DNN.

The RL agent learns from 10,000 simulated events. After the RL agent has been trained, it will be used in further tests under varying environmental circumstances.

Deep RL's DNN Architecture Design

The DDPG's actor and critic network architecture is laid out in full in Table 1. Linear reinforcement learning's architecture is included for reference, as well. The present configurations of the DDPG model and the linear reinforcement learning model are the best of many potential outcomes gained via trial and error.

The scalar estimated Q value is output by the critic network in the DDPG method, whereas the actor network outputs the setpoint for each zone after receiving the vector of state variables and the vector of action variables as inputs.

Temperature and humidity are continuous variables, hence DDPG can be used for continuous action control directly, whereas DNN requires action space discretization. Although the setpoint is a consistent variable, there is dependably a scope for it to keep clients agreeable. Accordingly, the entertainer organization's result layer utilizes tanh as the actuation capability, restricting the result to the stretch [-1,1].

$$Setpt_z = T_{lower} + \Delta T \cdot (y_{out} + 1) \quad (7)$$

Yields the genuine setpoint, where y_{out} is the entertainer organization's result and ΔT is the set point's upper reach. ΔT is set to 2oC in the reenactment. As a result, the DDPG RL model chose a value for the setpoint that is between $[T_{lower}, T_{lower} + 2]$.

The outputs of the Linear RL algorithm are identical to the inputs. With a stage size of 0.5 ^{0}C, we discretize the setpoint space to meet the requirements of Linear RL's discrete action space. This means the 2-zone HVAC system has 25 possible configurations, with 5 actions each zone. Figure 6 shows a comparison of the thermal control strategies, DDPG and DNN. Temperature and humidity are continuous variables, hence DDPG can be used for continuous action control directly, whereas DNN requires action space discretization. A vector of 25 Q values, each of which represents a unique set of actions, is what you get when you run Linear RL.

Effectiveness of Real-Time Climate Control

Real-time climate control refers to the dynamic adjustment of heating, ventilation, and air conditioning (HVAC) systems based on live data from environmental sensors, occupancy patterns, and external weather conditions.

THE DDPG AND RL LINEAR ALGORITHMS HAVE REACHED CONVERGENCE

The average performance improvements achieved after each training session were evaluated for both the deep deterministic policy gradient (DDPG) algorithm and the linear reinforcement learning (linear RL) approach.

```
def train(env, ENV_NAME, training_steps):
nb_actions = env.action_space.shape[0]
agent = build_agent(env.action_space.shape[0], env.observation_space
agent.compile(Adam(lr=.001, clipnorm=1.), metrics=['mae''])
agent.fit(env, nb_steps=training_steps, visualize=False, verbose=1, |
agent.save_weights('results/weights/ddpg_{} weights.h5f".format(ENV_|
agent.test(env, nb_episodes=1, visualize=False, nb_max_episode_steps
def test(env, ENV_NAME , num_episodesf):
nb_actions = env.action_space.shape[0]
agent = build_agent(envjaction_space.shape[0], env.observation_space
agent.compile(Adam(lr=.901, clipnorm=1.), metrics=['mae"])
agent.load_weights('results/weights/ddpg_{}_weights.h5f".format(ENV_I
```

Linear-based approaches and linear reinforcement models generally refer to traditional control methods or simpler reinforcement learning (RL) techniques where policies, value functions, or both are modeled using linear functions. These approaches are commonly used when the system dynamics are relatively simple. For DDPG, the gains demonstrated a non-linear progression, characterized by significant jumps in performance as the model fine-tuned its policy through exploration and exploitation of the action space. In contrast, the linear RL approach exhibited a more steady and incremental improvement pattern, reflecting the constraints of its simpler policy structure and limited adaptability. This contrast underscores the differences in learning dynamics and capability between the two methodologies. In the first few episodes, returns seem to be on average higher than in the last few. This happens because an arbitrary training day is chosen for each episode. On days with moderate exterior temperatures, the energy cost and penalty of exercise may be relatively low, and vice versa. As training progresses, however, more episodes are added, decreasing the average return. However, the DDPG RL approach often yields greater returns than the Linear RL approach. The linear RL model's output is higher than the DDPG RL models, and after 10,000 episodes, the combination of actions has not been thoroughly explored, resulting in a lower average return. The algorithm used for the present research work is as follows:

Figure 7 shows the convergence of Linear RL and DDPG RL model for all episodes of different zones. From Figure 7 it is observed that the convergence of the Linear RL model is much less as compared to the DDPG model whose converge is 100 % of all episodes.

Computational per unit efficiency

Using the real-world data the DDPG RL agent is trained and then used for 20 test days in January 2021 to provide the best possible HVAC management plan. The whole examination process takes no more than eight minutes. The software utilizes the free and open-source TensorFlow framework for deep learning and is developed in Python 3.6. The platform consists of a laptop equipped with 16.00 GB of RAM and a 2.8 GHz Intel®CoreTM i7- 7600U CPU.

Figure 8 shows the Per Unit Efficiency of the DDPG RL model for Different zones such as TDECZ1, TIECZ1, and TDXZ1. The light green color in the above figure shows the higher Per-Unit Efficiency for most of the episodes which lies between 1.3 to 1.4.

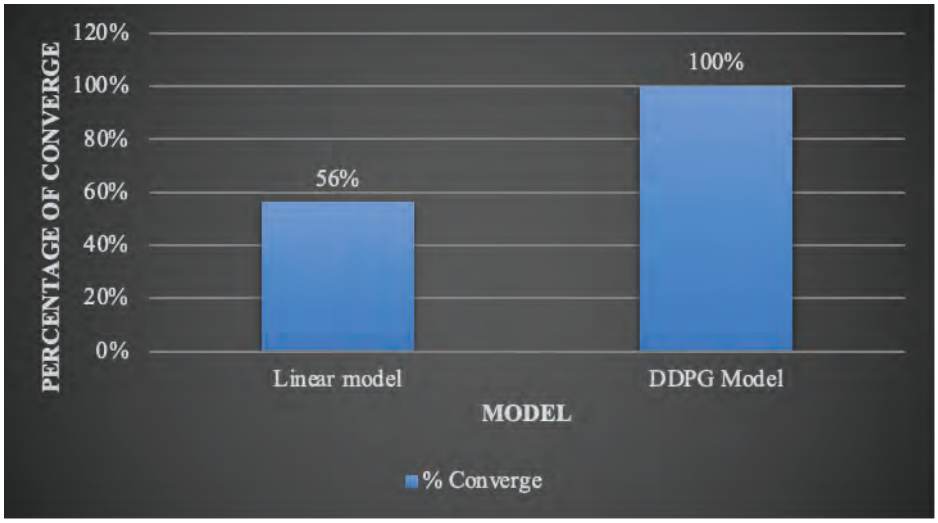


Figure 7. Converge of linear RL and DDPG RL model.

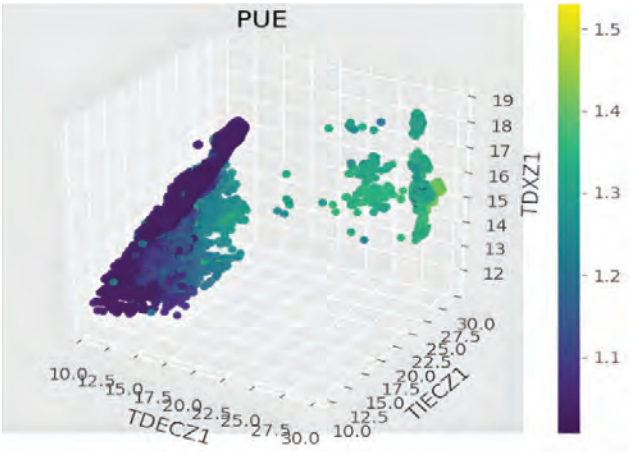


Figure 8. Per unit efficiency of DDPG for different zones.

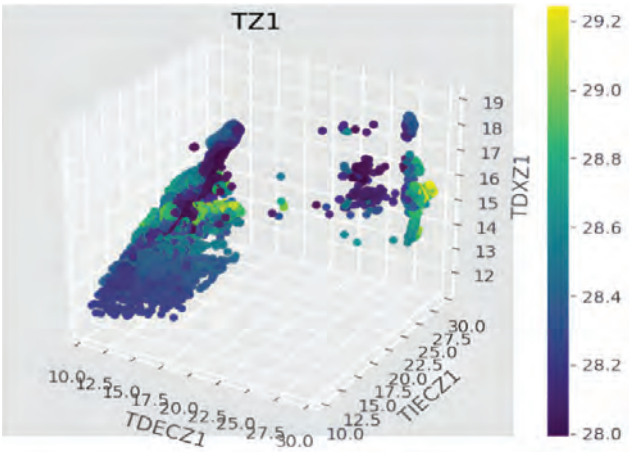


Figure 10. Temperature of TZ1.

Comparison of the DDPG with the linear RL model

After step of convergence, we count the mean number of steps required for converge of the episode and rewards. The mean number of steps taken by linear is 115.4 which is very higher that steps taken by DDPG model and it's also time consuming too. Mean steps taken by DDPG model is only 9.9. The mean and maximum step taken by both the model also mentioned in above Table 2. Figure 9 shows that the proper difference between numbers of steps taken by each selected model.

Temperature of selected zone

The temperature range in Zone 1 and Zone 2 after the application of a designed model for controlling the temperature of Commercial/Market buildings is shown in Figure 10 and Figure 11. In zone one control temperature lies between 28.4 °C to 28.8 °C, in zone 2 this temperature lies between 20 °C to 30 °C. This minimization of temperature is possible by the DDPG RL algorithm model.

Table 2. Comparison of the DDPG model with linear RL model

Type of model	Total samples	% Converge	Mean (Steps)	Min (Steps)	Max (Steps)
Linear model	10,000	56%	115.4	5.0	200.0
DDPG model	10,000	100%	9.9	3.0	20.0

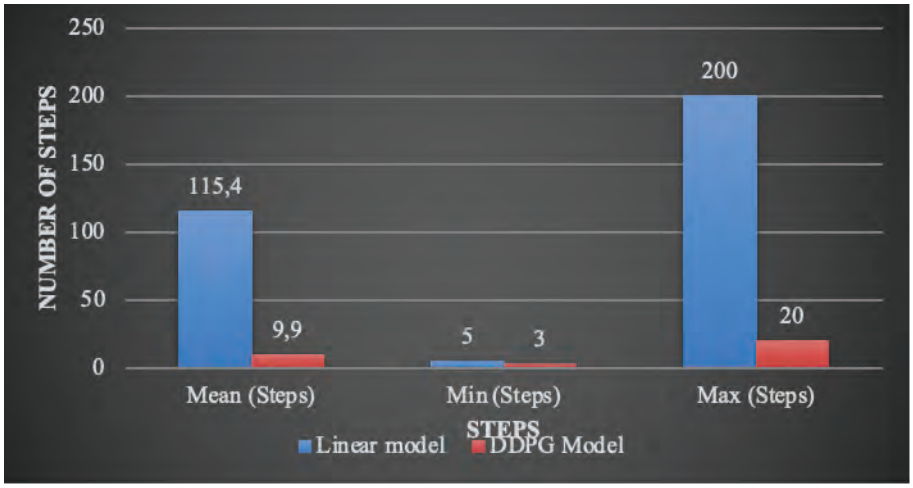


Figure 9. Mean step required for convergence of sample for both the models.

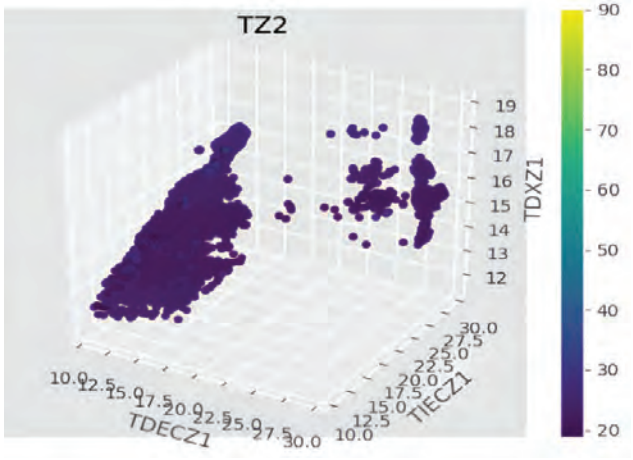


Figure 11. Temperature of TZ2.

CONCLUSION

In this study, the DDPG RL technique is used to regulate a commercial/market HVAC system with many zones to reduce energy costs without sacrificing comfort. Using DNNs, the DDPG can keep the temperature and humidity in the building at a constant level. The simulation results demonstrate that a well-trained DDPG RL expert can act wisely to adjust the various improvement targets and that it can also acquire speculation and flexibility to concealed climate, recommending its true capacity for future web-based applications in tackling MDP issues with stowed away data or with consistent pursuit space. The RL-based control method's stability will be further enhanced by research on two main areas in the future:

- The deep RL agent must first master the ability to adjust to a variety of seasonal conditions by automatically switching between cooling and heating modes of operation in order to provide HVAC customers with cost-effective management techniques over a year-long period;
- The deep RL specialist should have the option to learn a set point plan that takes client preferences into account more in order to provide HVAC management strategies that are more flexible. We can increase the deep RL agent's adaptability and resistance to the uncertainties that arise in real-world applications by exploring these two directions.
- The study highlights the potential of DRL, particularly the DDPG approach, in revolutionizing air conditioning control systems. While the linear reinforcement method serves as a valuable baseline, the superior performance of DDPG underscores the importance of advanced learning techniques in tackling complex, dynamic control problems. This research contributes to the growing body of knowledge aimed at developing energy-efficient and user-centric AC control solutions.

REFERENCES

[1] U.S. Dept. Energy. 2011 Buildings Energy Data Book. Available at: https://ieer.org/wp/wp-content/uploads/2012/03/DOE-2011-Buildings-Energy-DataBook-BEDB.pdf. Accessed on 17 Mar 2026.

[2] Erdinc O, Tascıkaraoglu A, Paterakis NG, Eren Y, Catalao JPS. End-User Comfort-Oriented Day Ahead Planning for Responsive Residential Hvac Demand Aggregation Considering Weather Forecasts. IEEE Xplore 2016;8:362–372. [CrossRef]

[3] Hao H, Corbin CD, Kalsi K, Pratt RG. Transactive Control of Commercial Buildings for Demand Response. IEEE Xplore 2016;32:774–783. [CrossRef]

[4] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Lanctot M, et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. Nature 2016;529:484–489. [CrossRef]

[5] Huang T, Guo Q, Sun H. Distributed Computing Platform Supporting Power System Security Knowledge Discovery Based on Online Simulation. IEEE Xplore 2017;8:1513–1524. [CrossRef]

[6] Hua H, Qin Y, Hao C, Cao J. Optimal energy management strategies for energy Internet via deep reinforcement learning approach. Appl Energy 2019;239:598–609. [CrossRef]

[7] Rocchetta R, Bellani L, Compare M, Zio EE, Patelli E. A reinforcement learning framework for optimal operation and maintenance of power grids. Appl Energy 219;241:291–301. [CrossRef]

[8] Ma Y, Kelman A, Daly A, Borelli F. Predictive Control for Energy Efficient Buildings with Thermal Storage: Modeling, Stimulation, and Experiments. IEEE Xplore 2012;32:44–64. [CrossRef]

[9] Dounis AI, Caraiscos C. Advanced Control Systems Engineering for Energy and Comfort Management in a Building Environment–A Review. Renew Sustain Energy Rev 2009;13:1246–1261. [CrossRef]

[10] Aswani A, Master N, Taneja J, Culler D, Tomlin C. Reducing Transient and Steady State Electricity Consumption in Hvac Using Learning-Based Model-Predictive Control. IEEE Xplore 2012;100:240–253. [CrossRef]

[11] Manuel A-C, Tejero-Gonzalez A, Illera-Riesgo J, Salins SS. Integrated evaluation of energy and water use in air-cooled and adiabatic condensers – a case study in an office building. Energy Nexus 2026;22:100706. [CrossRef]

[12] Razzano G, Brandi S, Piscitelli MS, Capozzoli A. Rule extraction from deep reinforcement learning controller and comparative analysis with ASHRAE control sequences for the optimal management of Heating, Ventilation, and Air Conditioning (HVAC) systems in multizone buildings. Appl Energy 2025;381:125046. [CrossRef]

[13] Tianshu W, Wang Y, Zhu Q. Deep Reinforcement Learning for Building HVAC Control. Proceedings of the 54th Annual Design Automation Conference. New York: Association for Computing Machinery; 2017. p. 1–6. [CrossRef]
[14] Deng X, Zhang Y, He Q. Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning. Build Environ 2022;211:108680.
[15] Dobbs JR, Hencey BM. Model Predictive Hvac Control with Online Occupancy Model. Energy Build 2014;82:675–684. [CrossRef]
[16] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Graves A, et al. Human-Level Control Through Deep Reinforcement Learning. Nature 2015;518:529–533. [CrossRef]
[17] Kuruganti T, Kou X, Li F, Dong J, Starke M, Munk J. A distributed energy management approach for residential demand response. IEEE Xplore 2019. [CrossRef]
[18] Lin Y, Barooah P, Mathieu JL. Ancillary services through demand scheduling and control of commercial buildings, IEEE Xplore 2016;32:186–197. [CrossRef]
[19] Li F, Du Y. From AlphaGo to Power System AI: What Engineers Can Learn from Solving the Most Complex Board Game. IEEE Xplore 2018;16:76–84. [CrossRef]
[20] Wu Y, Tan H, Peng J, Zhang H, He H. Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. Appl Energy 2019;247:454–66. [CrossRef]
[21] Kou P, Liang D, Wang C, Wu Z, Gao L. Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks. Appl Energy 2020;264:1–12. [CrossRef]
[22] Mocanu E, Mocanu DC, Nguyen PH, Liotta A, Webber ME, Gibescu M. On-line building energy optimization using deep reinforcement learning. IEEE Xplore 2018;10:3698–3708. [CrossRef]
[23] Ahn KU, Park CS. Application of deep Q-networks for model-free optimal control balancing between different HVAC systems. Sci Technol Built Environ 2019;26:61–74. [CrossRef]
[24] Zhang Z, Chong A, Pan Y, Zhang C, Lam KP. Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning. Energy Build 2019;199:472–490. [CrossRef]
[25] Vrancx P, Ruelens F, Claessens BJ. Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control. IEEE Xplore 2016;9:3259–3269. [CrossRef]
[26] Wang Y, Velswamy K, Huang B. A long-short term memory recurrent neural network-based reinforcement learning controller for office heating ventilation and air conditioning systems. Processes 2017;5:46–63. [CrossRef]
[27] Jiang T, Zou Y, Yu L. Online energy management for a sustainable smart home with an HVAC load and random occupancy. IEEE Xplore 2017;10:1646–1659. [CrossRef]
[28] Hu G, Spanos CJ, Ma K. Energy management considering load operations and forecast errors with application to HVAC systems. IEEE Xplore 2016;9:605–614. [CrossRef]
[29] Lu N. An evaluation of the HVAC load potential for providing load balancing service. IEEE Xplore 2012;3:1263–1270. [CrossRef]
[30] Sharma VK. Secret image scrambling and dwt-based image steganography using smoothing operation and convolution neural networks. J Discret Math Sci Cryptogr 2023;26:695–705. [CrossRef]
[31] Markad K, Lal A. Thermal post buckling analysis of smart SMA hybrid sandwich composite plate. Polym and Polym Compos 2021. [CrossRef]
[32] Cui B, Munk J, Jackson RK, Fugate DL, Starke M. Building thermal model development of typical house in US for virtual storage control of aggregated building loads based on limited available information. In: 30th International Conference on Efficiency, Cost, Optimisation, Simulation and Environmental Impact of Energy Systems; California, USA. 2017.
[33] Bulut M. The impact of the establishment of the cross-border balancing market on the integration of RES into the regional electricity grid. Int J Energy Stud 2024;309–329. [CrossRef]
[34] Lal A, Markad K. Static and dynamic nonlinear stability analysis of hybrid sandwich composite beam under variable inplane loads. J Mech Sci Technol 2021;35:3895–3908. [CrossRef]
[35] Pogaku N, Prodanovic M, Green TC. Modeling, analysis and testing of autonomous operation of an inverter-based microgrid, IEEE Xplore 2007;22:613–625. [CrossRef]
[36] Wu H, Liu X, Ding M. Dynamic economic dispatch of a micro-grid: Mathematical models and solution algorithm. Int J Elect Power Energy Sys 2014;63:336–346. [CrossRef]
[37] Choi J, Shin Y, Choi M, Park W, Lee W. Robust control of a micro-grid energy storage system using various approaches, IEEE Xplore 2018;2702–2712. [CrossRef]
[38] Kanif M, Vivek P, Kishor K. Experimental and Numerical Investigation of Nano-material based Structural Composite. Curved and Layered Structures 2024;11. [CrossRef]
[39] Gu S, Lillicrap T, Sutskever I, Levine S. Continuous deep Q-learning with model-based acceleration. Proceedings of the 33rd International Conference on Machine Learning; New York, USA. 2016.

[40] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv 2015;1–15.

[41] Kuznetsova E, Li Y, Ruiz C, Zio E, Ault G, Bell K. Reinforcement learning for microgrid energy management. Energy 2013;59:133–146. [CrossRef]

[42] Lal A, Markad K. Influence of Dynamic Temperature Variation and Inplane Varying Loads over Post-Buckling and Free Vibration Analysis of Sandwich Composite Beam. Int J Comput Mater Sci Eng 2020;S2047684120500128. [CrossRef]

[43] Salpakari J, Lund P. Optimal and rule-based control strategies for energy flexibility in buildings with pv. Appl Energy 2016;161:425–436. [CrossRef]

[44] Rajabi H, Hu Z, Ding X, Pan S, Du W, Cerpa A. Modes: Multisensor occupancy data-driven estimation system for smart buildings. Proceedings of the Thirteenth ACM International Conference on Future Energy Systems; New York, USA. 2022;228–239. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.1904

Research Article

Narrowband and highly linear LNA design for 4.5 GHZ on 180NM CMOS

Serdar KAYA[1,*], Alparslan Çağrı YAPICI[1], Hayrullah YILDIZ[2], Mübarek KAYA[1], Mustafa D. CERCI[1]

[1]*Department of Research and Development Teknokar Defence and Aerospace Inc. Ankara, 06800, Türkiye*
[2]*Department of Electrical and Electronics Engineering, Başkent University, Ankara, 06790, Türkiye*

ARTICLE INFO

Article history
Received: 25 November 2024
Revised: 26 February 2025
Accepted: 30 July 2025

Keywords:
4.5 GHz; 180nm; CMOS; LNA; Narrowband; Linear

ABSTRACT

In this paper, a narrowband and highly linear Low Noise Amplifier (LNA) is designed and implemented on 180 nm CMOS technology, operating at a center frequency of 4.5 GHz. This LNA design is particularly suitable for modern communication systems and emerging 5G applications. For simulation and analysis, the 180 nm Generic Process Design Kit (GPDK) in Keysight's Advanced Design System (ADS) was utilized. To improve stability and gain, resistive feedback was chosen, while inductive source degeneration was employed to achieve better noise performance and linearity. The proposed design aims for a simple architecture, and its narrowband specifications are achieved through optimized input and output matching networks. The analysis showed that the designed LNA provides a gain of 14.1 dB, a minimum noise figure of 0.928 dB, an input 1-dB compression point (P1dB) of 0 dBm, and an input third-order intercept point (IIP3) of 4.942 dBm across a 360 MHz band. These results indicate excellent linearity and signal integrity, and show the design is competitive with relative studies. The novelty of this work lies in achieving high linearity and low noise with a notably simplified design, overcoming common trade-offs. Compared to recent state-of-the-art work, the proposed LNA offers superior linearity and a lower noise figure while maintaining a simple circuit topology.

Cite this article as: Kaya S, Yapıcı AÇ, Yıldız H, Kaya M, Cerci MD. Narrowband and highly linear LNA design for 4.5 GHZ on 180NM CMOS. Sigma J Eng Nat Sci 2026;44(2):894–907.

INTRODUCTION

Wireless communication systems have become an essential part of our lives for a very long time. Over time, the needs and specifications expected from these systems became more and more advanced. This trend led the subsystems and components to evolve to provide the performance requirements. The receivers used in these systems employ many components. Low noise amplifiers (LNA) are one of the most important parts of wireless receivers [1]. Their main role is to amplify the received signal while adding minimal noise to it. This enables them to improve the quality of the signal in terms of the signal-to-noise ratio (SNR) [2]. It also offers non-linearity to the circuit since it is an active device. The LNA's non-linearity is undesired since it leads to signal

***Corresponding author.**
*E-mail address: serdar.kaya@teknokar.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

distortion and the generation of undesired products like harmonic and inter-modulation distortion [3]. In order to overcome the non-linearity, a filter prior to the LNA is placed to suppress the unwanted bands. These filters are also crucial for noise figure calculations since cascaded noise figure calculations show that the first element of the network provides the most significant contribution to noise [4]. Therefore, it is important to have a low noise figure (NF) in the first element of the receiver, which is typically a passive band-select filter that introduces a noise equal to its insertion loss [5]. This condition leads to a trade-off between selectivity and the total NF of the receiver. To overcome the high NF contribution issue of band select filters, various different types of band selective or narrowband LNA designs were offered [6-9]. These studies offer varied solutions like frequency translational resistive feedback architectures, current conveyors, nested reactive feedback loops, and integrated frequency selective networks. These designs were mainly focused on complex topologies such as multistage matching circuits, baluns, and feedback networks to provide better selective characteristics. Since the filter is not used in these designs to lower the unwanted frequency components, the linearity of the designed LNA becomes more decisive in terms of performance. Yet, the linearity of the designs mostly was not prioritized as a design goal. On the other hand, regular LNA studies focusing on these bands are focused on either the noise figure or the linearity improvements by making trade-offs on the non-focused side. This approach led to bad affection for their overall performance.

In this paper, a narrowband and highly linear LNA design was implemented to overcome the trade-off between noise figure and selectivity. The design is made for a rather less popular band which is reserved for military communications and future 5G communication systems, 4.5 GHz [10,11]. To minimize complexity, size, and cost, maintaining circuit simplicity was a key design objective. To provide better stability and matching characteristics, resistive feedback and source inductive degeneration techniques are employed within the design [12,13]. Input and output matching circuits, on the other hand, are designed to obtain a narrowband response by applying a simple network design with two elements. This design method provides simpler circuits in terms of components as well as offers a good band selectivity performance. The nonlinear behavior of the device is also taken into account within the design by determining the parameters of the components used. GPDK 180 nm process is used to implement the design in the Advanced Design System (ADS) of Pathwave Design to run S-parameter and nonlinear simulations.

THEORY

Design Objectives

Design objectives of the LNA design can be seen in Table 1.

Table 1. Design objectives of the Lna design

Specification	Unit	Value
Center frequency	MHz	4500
Bandwidth	MHz	350
Maximum noise figure	dB	<1.3
IP1dB	dBm	>-5
IIP3	dBm	>4
Minimum gain	dB	>11

The main objective is to have a frequency-selective response both on input and output matching circuits so the amplifier can also act close to a filter over the band. It is also aimed to have NF lower than 1.3 dB over the passband of the design while having a gain of more than 11 dB. Since the LNA is designed to act as a filter and therefore as the first element of a receiver structure, linearity objectives are also considered to be important. In order to obtain a linear characteristic, IIP3 values bigger than 7 dB and IP1dB values bigger than -5 dB are aimed to be achieved.

Transistor Sizing

RF field effect transistors (FET's) have lots of parameters which has a direct effect on the performance. In this study, the physical parameters of transistors which are length, width, and number of fingers, are selected to have the transistor optimized for the design objectives. The length of the transistor has an inversed relationship with the cutoff frequency (f_T) [14]. Formula for the f_T can be seen (1).

$$f_T = \frac{V_a}{2*\pi*C_{total}} \quad (1)$$

f_T represents the transition frequency or cutoff frequency, V_a is the Early voltage and C_{total} refers to the total capacitance of the transistor.

Unlike the f_T, the length of transistor has a direct relationship with the transconductance (g_m) of the transistor. In this design, the length of the transistor is picked as the minimum length the process can offer, 180nm. This selection was made to ensure the maximum high-frequency performance of the transistor.

The channel width of the transistor, on the other hand, is mainly effective on the current flow capability and physical coverage of the component on the layout [15]. This parameter was mainly considered to have sufficient power to obtain the nonlinear specifications of the amplifier.

The number of fingers to be used on the transistor is also an important parameter since it can provide additional improvement in terms of linearity and g_m [16,17] Yet, it should be noted that it can lead to performance degradation in terms of f_T. Depending on optimization results, 200 µm is chosen to be the length with 4 fingers in total to have a 50 µm finger width.

Topology

There are various topologies to form an LNA using a single transistor element. The most common of these topologies are common source and common gate amplifiers [18]. The main advantage of the common source is the minimized NF whilst the common gate topology offers more bandwidth due to the limited miller effect [19]. Since the objective is to have a narrowband response and low NF as possible, common source topology is used as the reference. The common source design topology can be seen in Figure 1. In this topology, the input signal is applied to the gate (G) and the amplified and inverted output signal is taken from the drain (D), with the source (S) terminal being common to both the input and output.

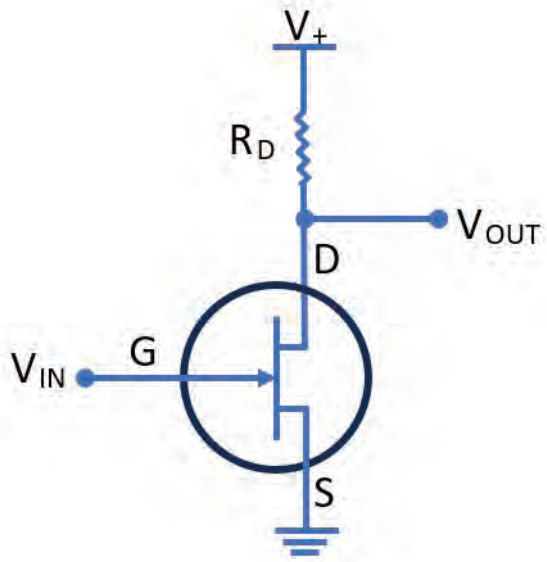


Figure 1. Common source topology.

After deciding the main topology, various techniques to be added to improve performance on different aspects such as stability, gain and bandwidth are also chosen. In order to obtain the unconditional stability of the transistor over the operating bandwidth, resistive feedback is added to the main topology. This addition also had its impacts on other parameters such as matching. Also to improve stability and noise characteristics over the operating band, the source inductive degeneration technique is chosen to be added in the topology. Matching circuits to be designed are mainly focused on obtaining selectivity in terms of the operating band while using a minimum number of components to avoid complexity and therefore large area occupation within the die. With the same concern regarding circuit size, lumped comments offered within the process which are MIM capacitors [20] and spiral inductors [21], will be used as elements of the matching network.

DESIGN

The design procedure of the LNA consists of analysis and considerations covering DC biasing, stability, S parameter analysis and matching network design.

DC Analysis

The chosen 50x4 µm width transistor's I/V curve analysis can be seen in Figure 2. The optimum biasing point is chosen to be at 0.7 V VGS and 2 V VDS, which stands as a valid point regarding the class A amplifier operating conditions [22].

The transconductance(gm) analysis at the selected bias point can be seen in Figure 3. The transconductance value is chosen to be 0.002 S for the design.

In order to calculate the DC power consumption of the proposed LNA, voltage and current values determined from the DC analysis can be used. The main reason behind

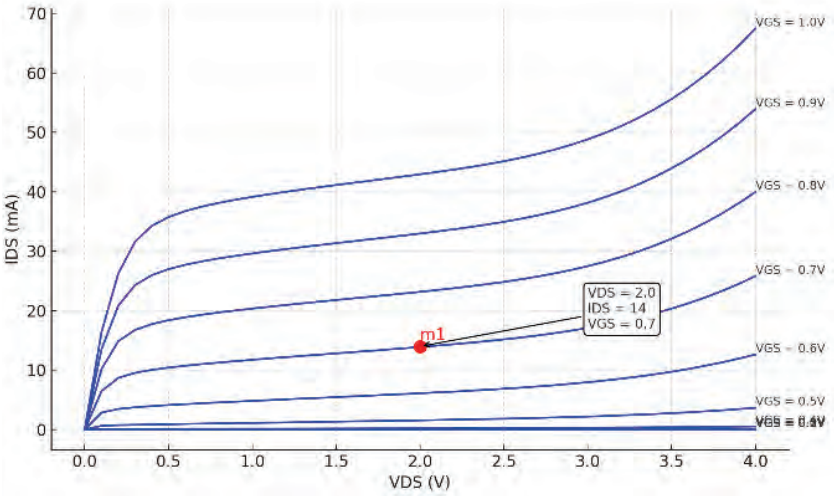


Figure 2. 50X4 µm transistor I/V curve analysis.

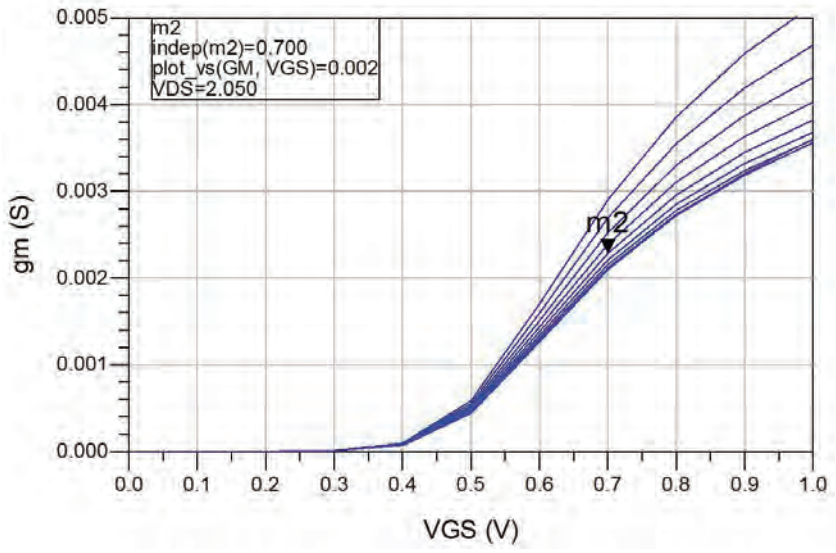


Figure 3. Transconductance analysis of the 50X4 μm transistor.

it is that the LNA is assumed to work with small signals which does not affect the bias point and operating region of the transistor.

DC power consumption can be calculated with formula (2).

$$P_{del} = I_D * V_{DS} \quad (2)$$

P_{del} represents the total DC power consumption, I_D is the drain current, and V_{DS} is the voltage applied to the drain. By using formula (2), DC power consumption can be calculated as 28 mW.

RF Analysis

After DC analysis, a biasing network is formed to get the design ready to stability and s parameter analysis. The design schematic can be seen in Figure 4. Due to the determinations made on design methodology, a source degeneration inductor and a feedback resistor are also added.

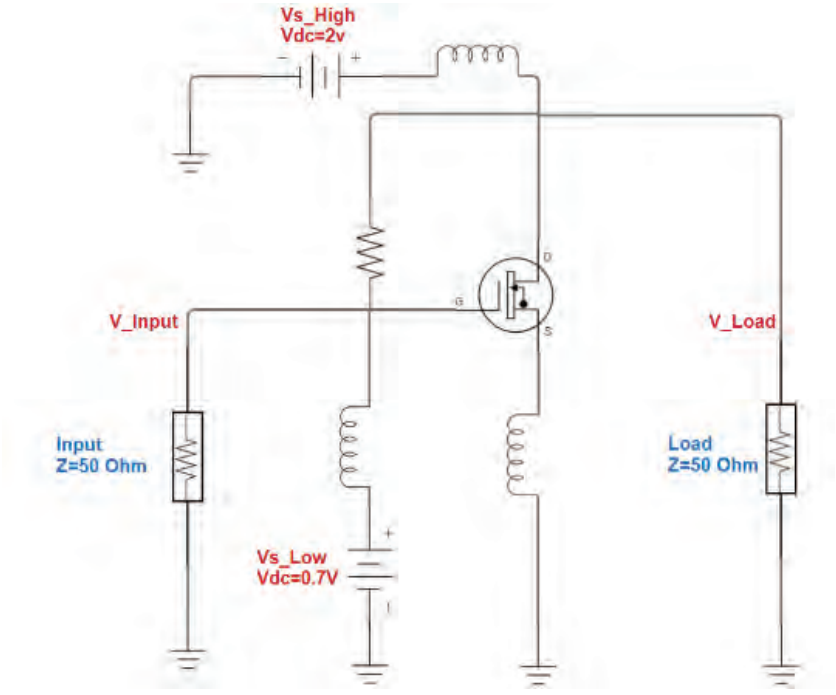


Figure 4. Design circuit schematic.

The circuit's stability analysis made for a wideband frequency range (2-6 GHz) can be seen in Figure 5. It is clear that the circuit is unconditionally stable within the operating bands (4-5 GHz).

S parameter analysis can be seen in Figure 6. Reverse isolation (S12) is low enough for unilateral consideration [23], and forward transmission parameters are within the needs of design objectives in 4.5 GHz.

Maximum gain analysis can be seen in Figure 7. It is clearly seen that this biasing condition provides enough gain for design objectives.

Matching

For matching network design, a narrowband frequency response is targeted by using a minimum number of components. Impedance points for optimum source, load and noise figure matching can be seen in a normalized smith chart for frequencies between 2-6 GHz in Figure 8.

For a target center frequency of 4.5 GHz with a narrow bandwidth (less than 10%), the matching networks must be highly frequency selective. On the Smith chart, this means that while the impedance should be perfectly matched at 4.5 GHz, it should move away from the center point at adjacent frequencies. Also, the source matching network must

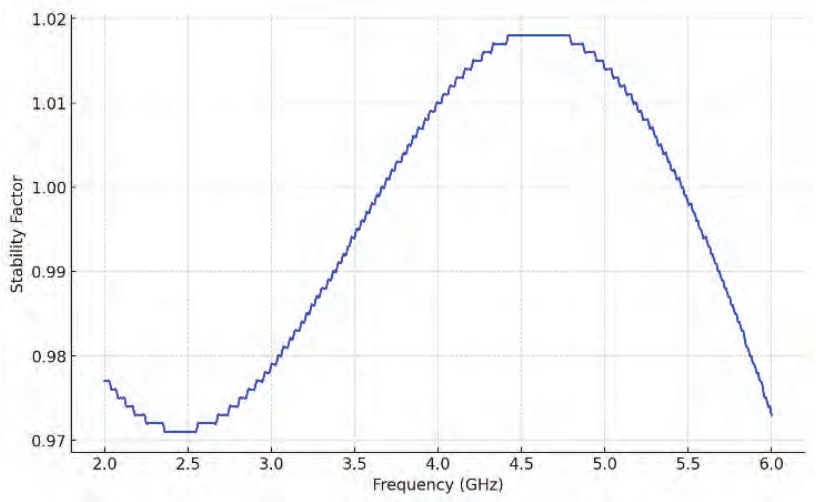


Figure 5. Stability analysis of the circuit.

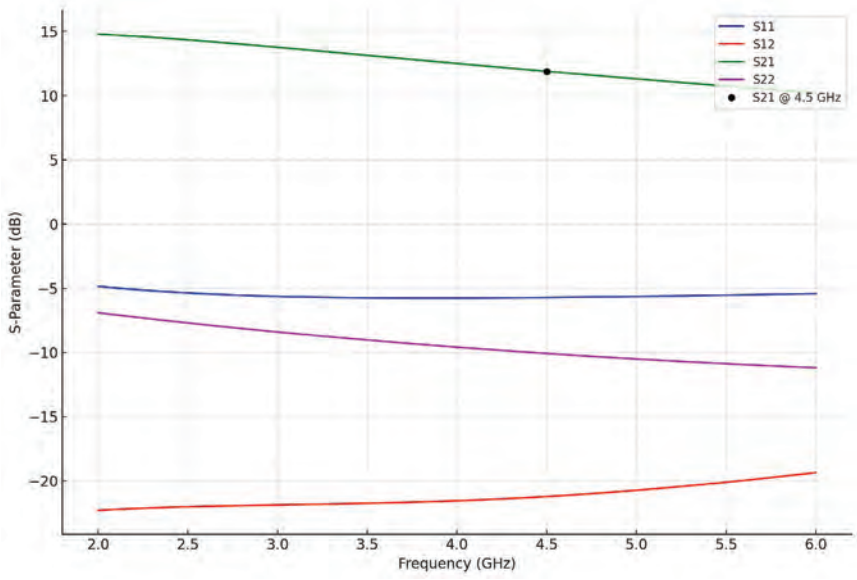


Figure 6. S parameter analysis of the circuit.

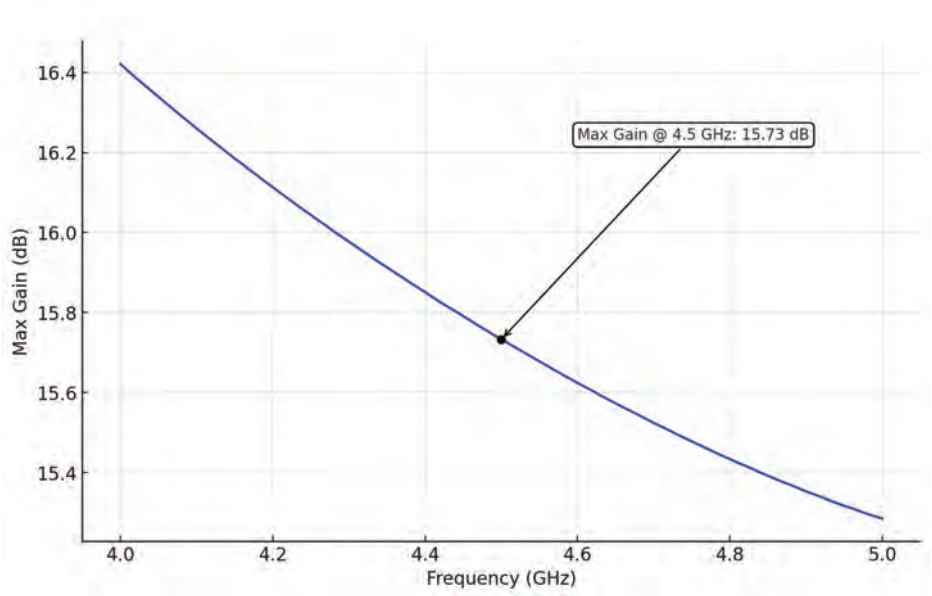


Figure 7. Maximum gain analysis of the circuit.

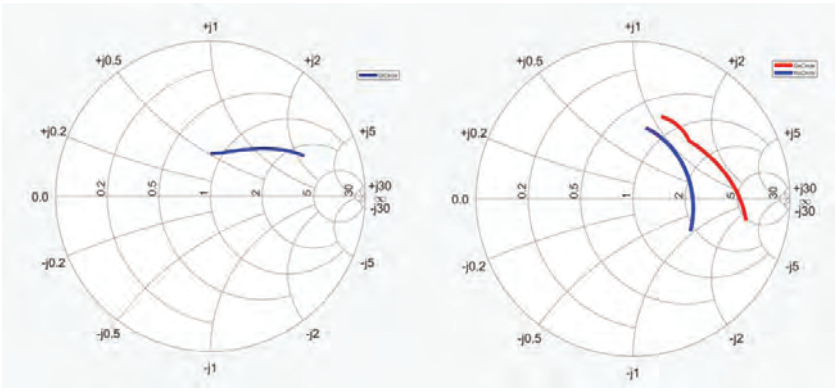


Figure 8. Impedance matching points of the circuit.

be designed to achieve an optimal trade-off between gain and noise figure, with a clear priority given to minimizing the noise figure.

To have all this, a simple series LC network was offered for both source and load matching. Series capacitance was needed to adjust the impedance points to R=1 circle and also to act as a DC block, and since the aimed matching points to obtain the objectives for 4.5 GHz is placed close enough to the R=1 circle, there are no additional parallel components used. A series inductor is good enough to match the reactance for both sides of the transistor. The chosen matching network can be seen in Figure 9. A spiral inductor and a MIM capacitor are chosen from the GPDK library to be used within the network. The parasitic of these devices were not dominant on the operating band and therefore, ignored within this design.

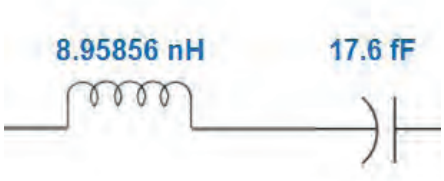


Figure 9. Matching network proposed for the LNA.

Reactance value of the selected capacitor value (17.6 fF) can be calculated with formula (3).

$$X_C = \frac{1}{2*\pi*f*C} \quad (3)$$

X_C represents the reactance value of capacitor, f is the frequency, and C is the capacitance value. For the capacitor on the source and load matching circuits (17.5 fF), reactance value can be calculated as 2012 Ω. High reactance value allows the impedance values to be adjusted on R=1 circle.

Reactance value of the selected inductor values (8.269 nH and 8.958 nH) can be calculated with formula (4).

$$X_L = 2 * \pi * f * L \tag{4}$$

X_L represents the reactance value of inductor, f is the frequency, and L is the inductance value. For the inductor on the source matching circuits (8.958 nH), reactance value can be calculated as 253.1 Ω. On the load matching circuit, the inductor's (8.269 nH) reactance value is equal to 233.77 Ω. Both inductors have reactance values close to each other, which indicates that the impedances of source and load branches are close to each other.

Final Circuit

In the final design stage, all passive and active components were integrated to form the complete LNA circuit. The finalized schematic, shown in Figure 10, incorporates the biasing network, stabilization elements, and matching circuits. The spiral inductors and MIM capacitors used are selected from the GPDK library due to their compact footprint and acceptable Q factor in the 4–5 GHz range. The values for all components used in the LNA are summarized below. These values are derived based on impedance transformation requirements, stability goals, and noise figure optimization. Calculations were carried out using standard L-network synthesis equations and Smith chart mappings. The source and load matching networks are tuned specifically to target the 4.5 GHz center frequency with a fractional bandwidth of less than 10%. After the implementation of matching networks, the final circuit design was complete and can be seen in Figure 10. The values of the components used are summarized in Table 2.

Table 2. Component values used in Lna design

Name/Indicator	Unit	Value
Feedback resistor	Ω	1050
Source inductor	pH	215.49
Drain inductor (RF choke)	nH	42.13
Gate inductor (RF choke)	nH	8.367
Source matching inductor	nH	8.958
Source matching capacitor	fF	17.6
Load matching inductor	nH	8.269
Load matching capacitor	fF	17.6

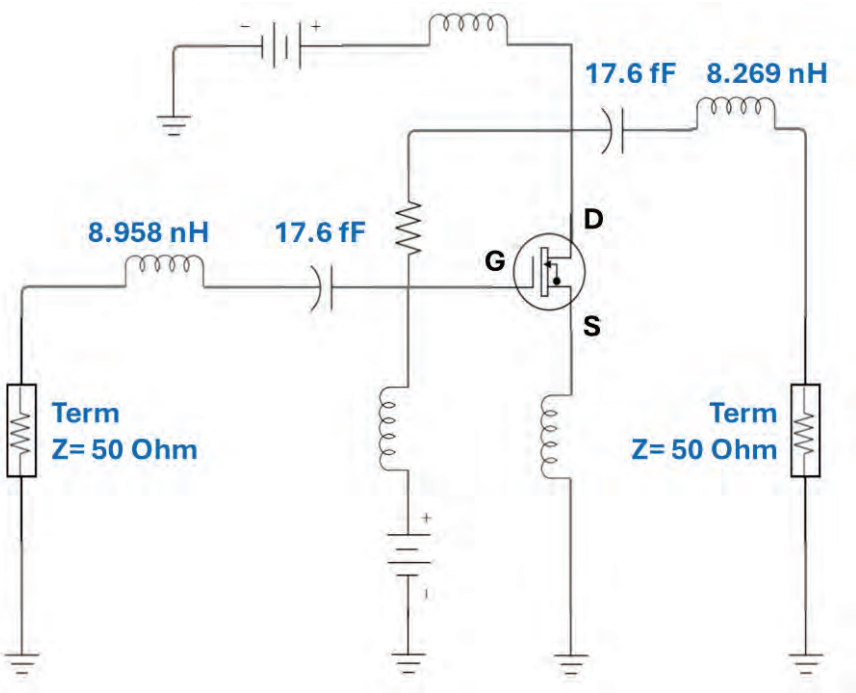
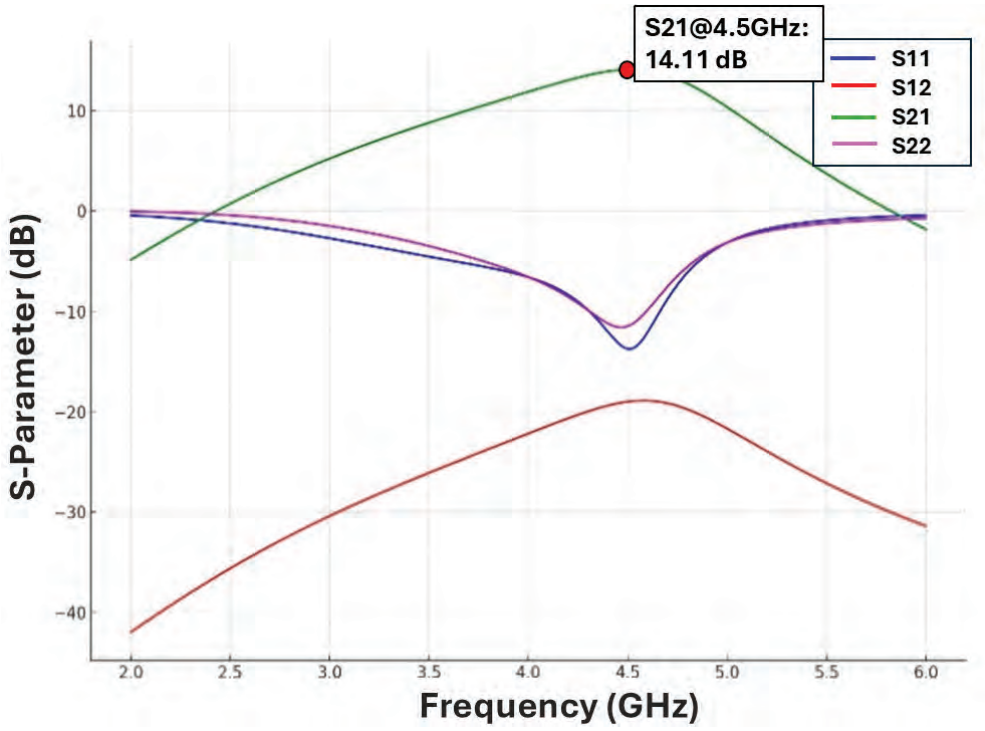


Figure 10. Final circuit of the LNA design.

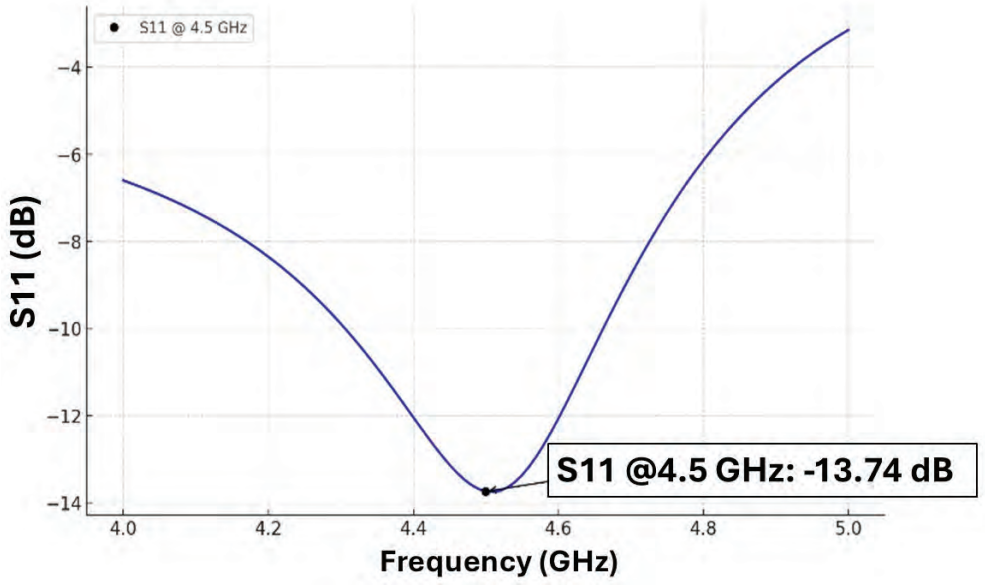
RESULTS AND DISCUSSION

Linear and Noise Performance

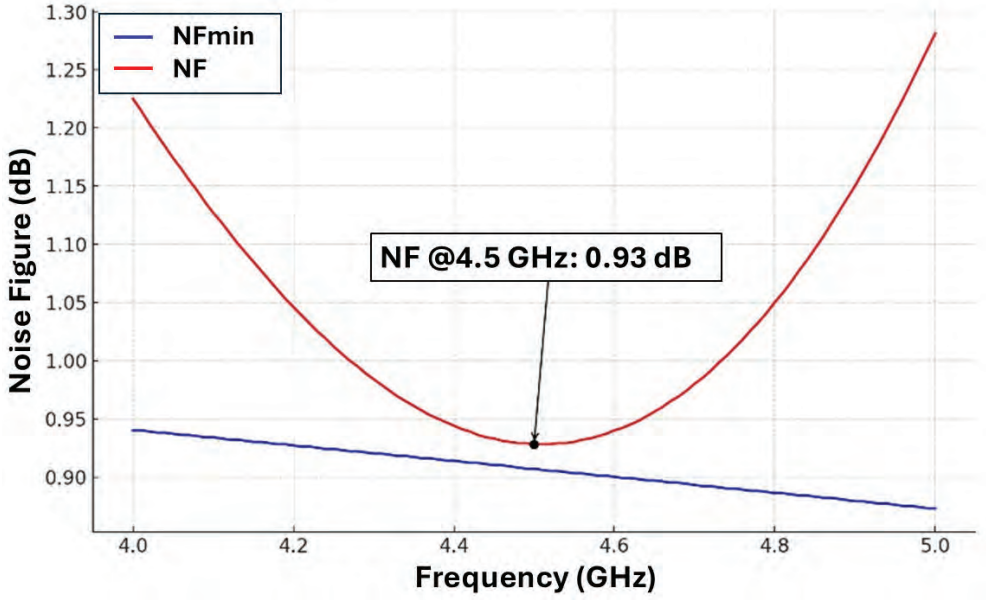
Simulations are created both for the linear and nonlinear characteristics of the LNA. In order to check the linear characteristics, S parameter analysis results can be observed in Figure 11(a). The gain is measured as 14.105 dB. To measure the performance of the LNA input matching, the S11 value was checked if it is below -10 dB. Detailed analysis of S11 can be seen in Figure 11(b). From the measurement, it can be said that the LNA input matching bandwidth is around 360 MHz. Noise figure simulation results can be seen in Figure 11(c). The noise figure acquired for 4.5 GHz is almost equal to the minimum noise figure, 0.928 dB, which is lower than the design objective.



(a)



(b)



(c)

Figure 11. Linear and noise performance of the LNA: (a) S parameters of the LNA (b) S11 analysis of the LNA. (c)Noise figure analysis of the LNA.

Matching Network Analysis

In order to observe the effectivity of the designed matching networks, the impedance points shown in Figure 8. are analyzed and can be seen in Figure 12. It is shown that a successful discrimination between the optimal impedances of frequencies is achieved.

Nonlinearity Analysis

In terms of nonlinearity, P1dB and IIP3 parameters are analyzed. The gain over input power graph is shown in Figure 13. From the compression point, P1dB is found close to 0 dBm.

Multi-tone simulations are also analyzed to measure the IIP3 of the LNA. The results can be seen in Figure 14. OIP3 of the LNA is measured as 19.047 dBm for lower tone. Regarding the gain (14.105 dB) IIP3 can be calculated as 4.942 dBm. Results gathered from simulations are compared to similar studies in Table 3.

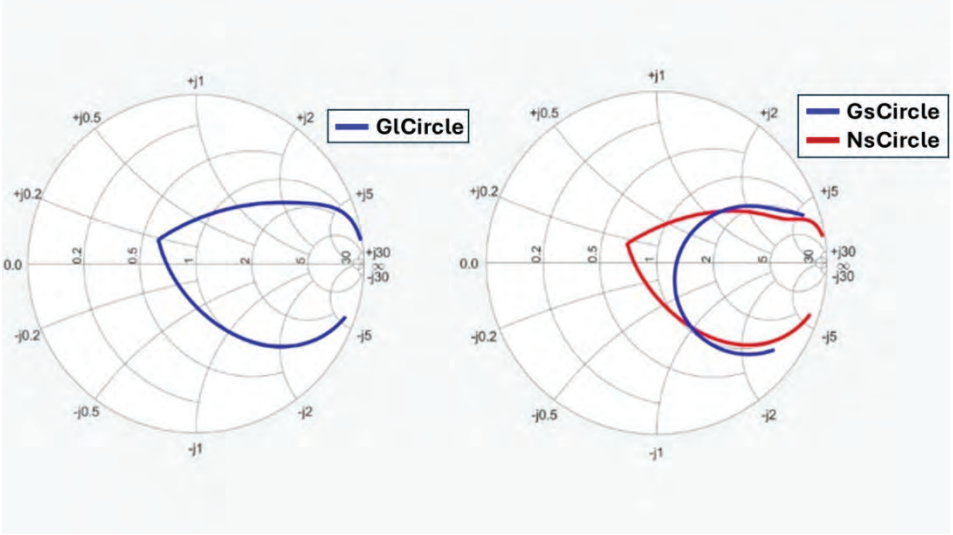


Figure 12. Smith Chart Analysis of the LNA.

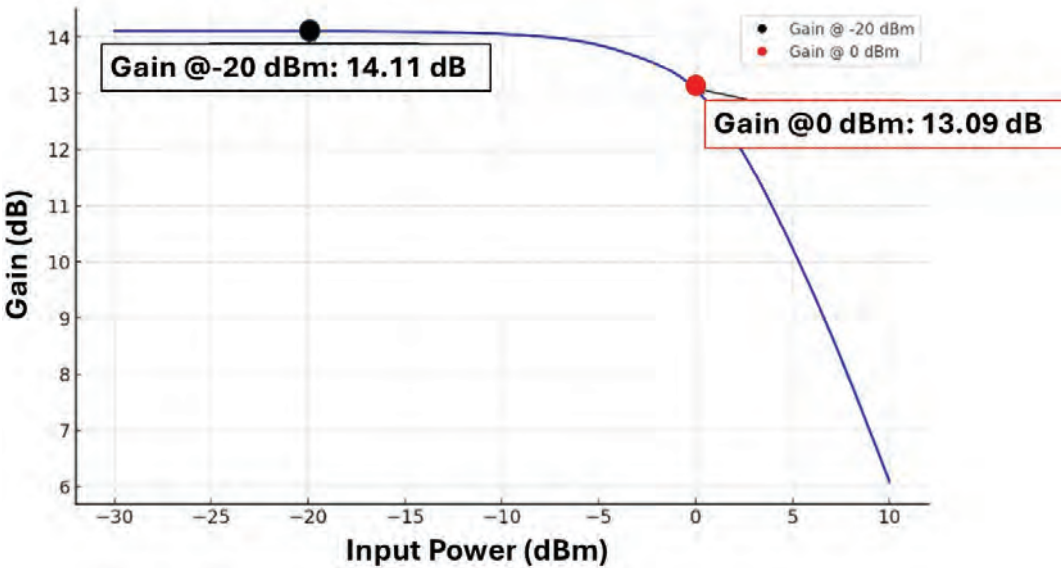


Figure 13. P1dB Analysis of the LNA.

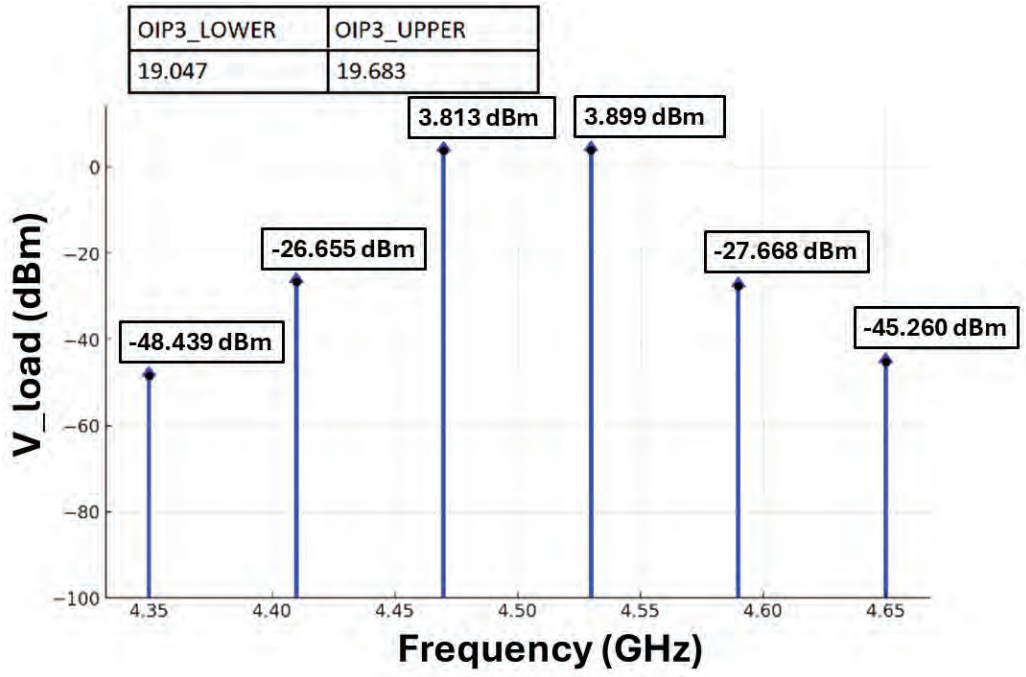


Figure 14. Multi tone analysis of the LNA.

The gain, noise figure, IIP3, and P1dB values from the LNA studies are visually compared in Figure 15(a), 15(b), 15(c), and 15(d), respectively.

Comparison with Other Studies

Results from this work and earlier studies [24-35] can be compared in various types of specifications.

By comparing the parameters provided, it can be said that the proposed design offers significantly higher linearity compared to most of the prior studies. Only two studies offer better linearity than the proposed design [33, 34].The study that stands to be better among all [33], has an IIP3 of 21 dBm and an IP1dB of 9 dBm. However, it exhibits a relatively high noise figure of 1.3 dB compared to this work, which offers 0.928 dB of noise figure. A similar situation applies to [34], which also demonstrates relatively good linearity. Although this work offers much higher gain

Table 3. Comparison of the Lna with other studies

Ref	Year	GPDK	Topology	Freq. (GHz)	Bandwidth (MHz)	Gain (dB)	NF (dB)	IIP3 (dBm)	IP1dB (dBm)
[24]	2021	180 nm	Cascaded CSA	2.4	~400 MHz	19	0.5	-15.87	-26
[25]	2019	180 nm	Cascaded CSA	5	Not Given	24.65	0.794	-26.81	-36.11
[26]	2022	180 nm	Cascode CSA	2.4	~600 MHz	22.75	1.75	-13	-19
[27]	2023	180 nm	CSA	5.5	~200 MHz	12.S1	2.8	-18.7	-14
[28]	2022	180 nm	CSA	1-7	~6000 MHz	17.95	1.11	Not Given	Not Given
[29]	2023	90 nm	CSA	4.4–5.0	~600 MHz	20.6–19.9	1.57–1.73	–15.4	Not Given
[30]	2019	180 nm	Cascode CSA	6	~200 MHz	15.9	0.92	-1.1	-8
[31]	2022	180 nm	Cascode CSA	2.4	~820.3 MHz	21.5	0.506	0.513	-8.17
[32]	2021	180 nm	Differential CSA	1.9-4.5	~2300 MHz	12.3	2.6	-2.5	Not Given
[33]	2019	250 nm	CSA	5-6	~2000 MHz	14	1.3	21	9
[34]	2021	22 nm	Cascode CSA	2.4-5.25	~2850 MHz	30	1.49	9	3
[35]	2024	150 nm	Cascaded CSA	2.8–3.8	~1000 MHz	35.5	1-1.8	Not Given	-25.5
This work	**2024**	**180 nm**	**CSA**	**4.5**	**~360 MHz**	**14.1**	**0.928**	**4.942**	**0**

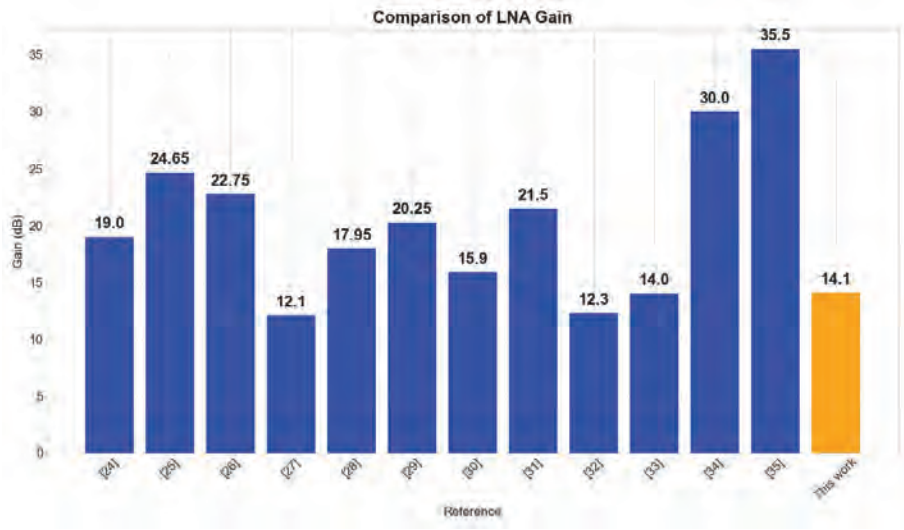
compared to the proposed design, it suffers from a wideband response, whereas the proposed design is more suited for narrowband applications.

In terms of noise figure (NF), there are some studies that offer better performance [24, 25, 30, 31] then this work does. Yet all of them fall short in terms of linearity. The best among these shows an IIP3 of -8 dBm and an IP1dB of -1.1 dBm, which are considerably worse than the proposed design.

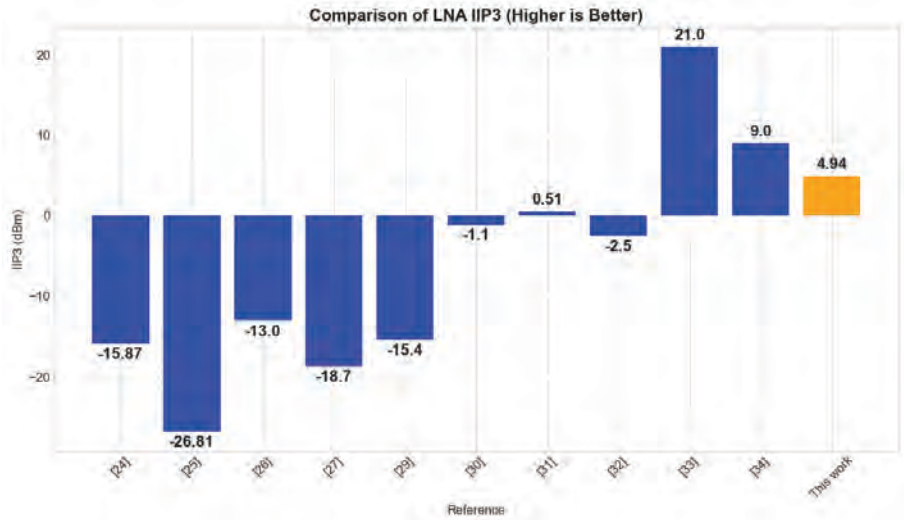
In terms of gain, which is not the primary specification for this work to dominate over, there are various studies that offers better performance than the proposed design. Yet, none of these studies can offer the linearity and NF performance of the proposed design offers. From that perspective, it is fair to say that the proposed design successfully balances all critical performance parameters linearity, noise figure, and gain while offering simplicity through its topology.

Simulation Limitations

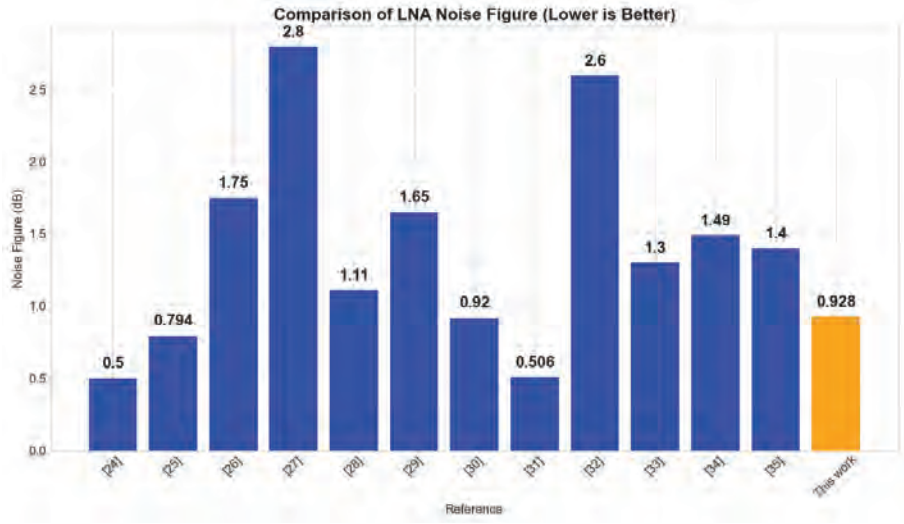
It must be noted that due to the nature of GPDK (Generic Process Design Kit), post-layout simulation results cannot be presented. This is due to GPDK not offering all layout components that are essential for layout design. This also caused the inability to make PEX (layout parasitic extraction) for this design. It is also not possible to make an accurate silicon area estimation due to this limitation. Yet, as mentioned earlier in the paper, parasitic values of



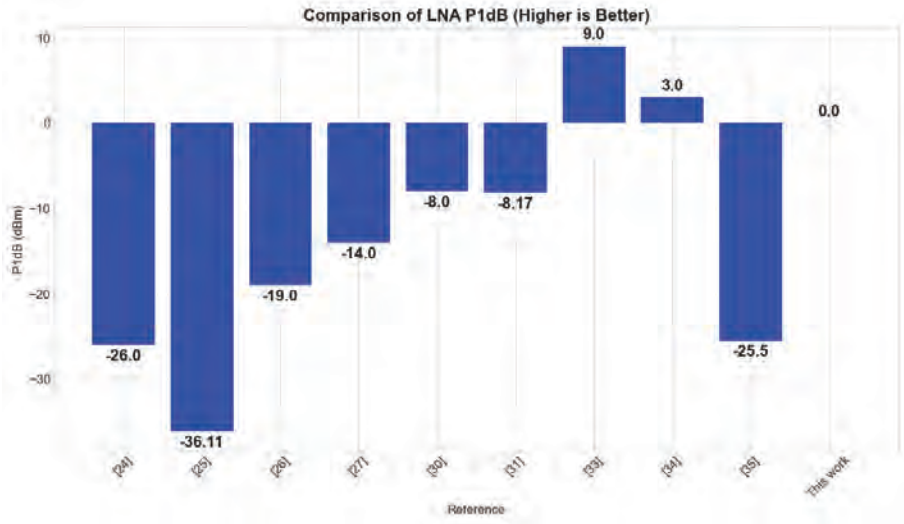
(a)



(b)



(c)



(d)

Figure 15. Performance comparison of this work with other LNA studies: (a) Comparison of Gain Values. (b) Comparison of IIP3 Values. (c) Comparison of Noise Figure Values. (d) Comparison of P1dB Values.

inductors, resistors and capacitors are checked in the early phases in the design and observed to be ineffective in terms of performance of the circuit.

CONCLUSION

In this paper, a narrowband and highly linear low-noise amplifier was designed within the common source topology by using the GPDK 180 nm process in ADS. Resistive feedback and source inductive degeneration methods were employed to improve stability and linearity. The topology was designed in such a way as to simplify the circuit by minimizing the number of components. Simulations indicated that the proposed design successfully covered 360MHz of bandwidth in 4.5 GHz while offering 0.912 dB noise figure, 14.1 dB gain, 0 dBm IP1dB and 4.942 dBm IIP3. These values show that the design offers better overall performance than the relative studies.

It must be emphasized that these results are based solely on pre-layout simulations. The limitations of the Generic Process Design Kit (GPDK) prevented post-layout extraction and analysis, meaning that parasitic effects from physical layout, process variations, and packaging are not accounted for in these results.

Despite this limitation, the results of this study successfully highlight the possibility of designing an LNA that combines both high linearity and simplicity. This combination is valuable for designers, as achieving good performance with fewer components contributes to creating cost-effective and robust systems.

NOMENCLATURE

f_T	*Cut-off Frequency*
V_a	*Early Voltage of The Transistor*
C_{total}	*Total Capacitance of The Transistor*
g_m	*Transconductance*
Ω	*Ohm*

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Duman M. Low noise amplifier design for hidden wireless LAN applications using band pass filter with geometric mean prototype element values. Sigma J Eng Nat Sci 2024;42:383–389. [CrossRef]

[2] Paek JS. Analysis and design of low-noise radio-frequency power amplifier supply modulator for frequency division duplex cellular systems. Electronics 2024;13:4635. [CrossRef]

[3] Bowick C, Blyler J, Ajluni C. RF circuit design. 2nd ed. Newnes; 2008.

[4] Pekau H, Haslett JW. Cascaded noise figure calculations for radio receiver circuits with noise-aliasing properties. IET Circuits Devices Syst 2007;1:517–524. [CrossRef]

[5] Pozar DM. Microwave engineering. 2nd ed. John Wiley & Sons; 1998.

[6] Lenka MK, Agrawal A, Khatri V, Banerjee G. A wideband receiver front-end with programmable frequency selective input matching. In: Proc 29th Int Conf VLSI Des 15th Int Conf Embed Syst; 2016; Kolkata, India. p. 168–173. [CrossRef]

[7] Garcia-Vazquez H, Khemchandani SL, Dualibe C, del Pino J. A selectable bandwidth LNA based on current conveyors. In: Proc Argent Sch Micro-Nanoelectron Technol Appl; 2015; Villa Maria, Argentina. p. 33–36. [CrossRef]

[8] Bagga S, Mansano AL, Serdjin WA, Long JR, Philips K, Pekarik JJ. A frequency-selective nested dual-loop broadband low noise amplifier in 90 nm CMOS. In: Proc Eur Solid-State Circuits Conf; 2012; Bordeaux, France. p. 121–124. [CrossRef]

[9] Fernandes M, Oliveira LB, Oliveira JP. A widely tunable narrowband balun-LNA with integrated filtering. In: Proc Int Conf Mix Des Integr Circuits Syst; 2014; Lublin, Poland. p. 160–165. [CrossRef]

[10] Lee J, Han S, Lee J, Kang B, Bae J, Jang J, et al. A sub-6GHz 5G new radio RF transceiver supporting EN-DC with 3.15Gb/s DL and 1.27Gb/s UL in 14nm FinFET CMOS. In: Proc IEEE Int Solid-State Circuits Conf; 2019; San Francisco, CA, USA. p. 354–356. [CrossRef]

[11] Yeh EHV, Lo AH, Chen WS, Yeh TJ, Chen M. A 16 nm FinFET 0.4 V inductor-less cellular receiver front-end

with ultra-low power and ultra-small area for 5G system in sub-6 GHz band. In: Proc Symp VLSI Circuits; 2017; Kyoto, Japan. p. C72–C73. [CrossRef]

[12] Verma AK, Kumar S. Resistive feedback low noise amplifier for software defined radio. In: Proc Int Conf Green Eng Technol; 2016; Coimbatore, India. p. 1–6. [CrossRef]

[13] Kanth MS, Shanthala S, Raj CP. Performance analysis of inductive source degeneration low noise amplifier using multi-finger technique. Int J Innov Technol Explor Eng 2019;8:1636–1642. [CrossRef]

[14] Sedra AS, Smith KC. Microelectronic circuits. 7th ed. Oxford University Press; 2014.

[15] Razavi B. Design of analog CMOS integrated circuits. 2nd ed. McGraw Hill; 2020.

[16] Han K, Han JH, Je M, Shin H. RF characteristics of 0.18-μm CMOS transistors. J Korean Phys Soc 2002;40:45–48.

[17] Sehmi S, Haddad F, Ben Hammadi A. Impact of multifinger MOSFET geometry on the electrical performance of RF circuits. Microelectron Reliab 2022;129:114445. [CrossRef]

[18] Razavi B. RF microelectronics. 2nd ed. Hoboken, New Jersey: Prentice Hall; 2011.

[19] Huang XH, Chen LJ, Zhou JF, Chen KS. Miller effect analysis and noise optimization of CMOS low noise amplifier. J Zhejiang Univ Sci C 2011;12:424–428.

[20] Tiwat P, Tingting Y, Guoguo L, Xiaojuan C, Xinyu L. MIM series capacitor model for MMIC design application. In: Proc Int Workshop Microw Millim Wave Circuits Syst Technol; 2013; Chengdu, China. p. 475–478. [CrossRef]

[21] Norhapizin K, Ismail MA, Ismat ARA, Marzuki A, Yahya MR, Awang Mat AF. Parasitic effects of spiral inductors on the performance of GaAs-based MMIC low noise amplifier. In: Proc IEEE Int Conf Semicond Electron; 2008; Johor Bahru, Malaysia. p. 134–137. [CrossRef]

[22] Ye F, Chiang JS, Chen CW, Sung YC. Dynamic bias circuits for efficiency improvement of RF power amplifier. In: Proc IEEE Asia-Pac Conf Circuits Syst; 2004. p. 821–824.

[23] Gonzalez G. Microwave transistor amplifiers: Analysis and design. 2nd ed. Hoboken, New Jersey: Prentice Hall; 1997.

[24] Babu DS, Reddy MM, Ramesh G, Babu DS, Daula SMS, Khan GA. A 2.4 GHz LNA design for 802.11 WLAN applications. In: Proc Conf Inf Commun Technol; 2021; Kurnool, India. p. 1–5. [CrossRef]

[25] Kushwaha DS, Rai AK, Agrawal S, Pal PK, Singhal HK. A narrow band cascode source degeneration low noise amplifier with pi-input matching at 5 GHz frequency. In: Proc Int Conf Innov Electron Signal Process Commun; 2019; Shillong, India. p. 301–304. [CrossRef]

[26] Jaradat RO, Shahroury FR, Ahmad HH, Abuishmais I. Design methodology for narrow-band low noise amplifier using CMOS 0.18 μm technology. Jordan J Comput Inf Technol 2022;8:98–111. [CrossRef]

[27] Chou JX, Lin YS. A compact 2.45/5.5-GHz dual-band LNA design using bridged-T coils. In: Proc IEEE Radio Wireless Symp; 2024; San Antonio, TX, USA. p. 115–117. [CrossRef]

[28] Goel R, Kumar A, Kumar M, Kumar S. Design of 1–7 GHz high gain wideband filter LNA for wireless applications. In: Proc Global Conf Adv Technol; 2022; Bangalore, India. p. 1–5. [CrossRef]

[29] Kim MS, Yoo SS. RF-SOI low-noise amplifier using RC feedback and series inductive-peaking techniques for 5G new radio application. Sensors 2023;23:5808. [CrossRef]

[30] Vanukuru V, Ghosal K, Godavarthi N, Dutta A, Narasimhamoorthy S, Swaminathan B, et al. A 6 GHz 0.92 dB NF cascode LNA in 180 nm SOI CMOS technology. In: Proc IEEE MTT-S Int Microw RF Conf; 2019; Mumbai, India. p. 1–4. [CrossRef]

[31] El Meligy AO, Albasha L. Design and analysis of various narrow-band LNA topologies. In: Proc Int Conf Commun Signal Process Their Appl; 2022; Cairo, Egypt. p. 1–7. [CrossRef]

[32] Joshi K, Mandal MK, Mandal P. Design of a wideband differential LNA based in CMOS 180 nm technology. In: Proc IEEE MTT-S Int Microw RF Conf; 2021; Kanpur, India. p. 1–4. [CrossRef]

[33] Abounemra AME, Helaoui M, Ghannouchi FM. A highly survivable C-band GaN HEMT LNA with resistive feedback technique. In: Proc Mediterr Microw Symp; 2019; Hammamet, Tunisia. p. 1–4. [CrossRef]

[34] Wang Z, Li Z, Li J, Wang X, Li Z. A 2.4–5.25 GHz balun-LNA in 22 nm CMOS technology. In: Proc Int Conf Integr Circuits Microsyst; 2021; Nanjing, China. p. 197–200. [CrossRef]

[35] Wang J, Wen J, Chi PL, Yang T. A 2.8–3.8 GHz reconfigurable GaAs low-noise amplifier with improved blocker tolerance. IEEE Microw Wireless Technol Lett 2024;34:419–422. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2017

Research Article

Missense single nucleotide polymorphisms analysis using bioinformatics tools in *SURF1* and *NDUFV1* genes associated with leigh syndrome

Aleyna ÖZTÜRK[1], Ömer Faruk KARASAKAL[2,*], Muhsin KONUK[3]

[1]*Department of Biotechnology, Graduate School of Science, Üsküdar University, İstanbul, Türkiye*
[2]*Medical Laboratory Techniques, Vocational School of Health Services, Üsküdar University, İstanbul, Türkiye*
[3]*Department of Molecular Biology and Genetics, Faculty of Engineering and Natural Sciences, Üsküdar University, İstanbul, Türkiye*

ARTICLE INFO

Article history
Received: 16 August 2024
Revised: 17 October 2024
Accepted: 13 February 2025

Keywords:
In silico; Leigh Syndrome, NDUFV1; Single Nucleotide Polymorphisms; SURF1

ABSTRACT

Leigh syndrome is a mitochondrial condition that causes severe neurological damage beginning early in life. By determining the possible harmful consequences of missense single nucleotide polymorphisms in the *SURF1* and *NDUFV1* genes linked to Leigh syndrome, this work seeks to clarify the molecular process of the illness. In the study, deleterious single nucleotide polymorphisms were identified using bioinformatics tools such as SIFT, PolyPhen-2, PredictProtein, PHD-SNP, SNP&GO, PANTHER and Meta-SNP, then their changes on protein integrity were analyzed using I-Mutant 2.0 and MUpro. The findings indicate that certain single nucleotide polymorphisms may significantly disrupt protein structure and stability, increasing the risk of Leigh syndrome. Using STRING and GeneMANIA databases, the interaction networks of these proteins with other proteins were examined, providing a more comprehensive understanding of the biological basis of Leigh syndrome. In our study, 6 single nucleotide polymorphisms in the *SURF1* gene and 11 single nucleotide polymorphisms in the *NDUFV1* gene were identified as potentially deleterious. Our study extends the previous literature on the genetic basis of Leigh syndrome and provides new information on the role of *SURF1* and *NDUFV1* gene mutations in disease pathogenesis. These findings may contribute to the creation of diagnostic and treatment strategies for Leigh syndrome in the future.

Cite this article as: Öztürk A, Karasakal ÖF, Konuk M. Missense single nucleotide polymorphisms analysis using bioinformatics tools in *SURF1* and *NDUFV1* genes associated with leigh syndrome. Sigma J Eng Nat Sci 2026;44(2):908–918.

INTRODUCTION

Leigh syndrome (LS) is a rapidly progressive neurological condition most often occurring in the first years of life. The condition presents with symmetric necrotic lesions in specific regions of the brain, including the brainstem, basal ganglia, and thalamus [1]. Mitochondrial dysfunction is the underlying cause of this disease. Symptoms include growth

***Corresponding author.**
*E-mail address: omerfaruk.karasakal@uskudar.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

retardation, muscle weakness (myopathy), respiratory distress (dyspnea), lactic acidosis and progressive brain damage (encephalopathy) [2]. Single nucleotide polymorphism (SNP) analyses can be used to investigate the genetic foundations of mitochondrial diseases such as LS. These analyses are crucial for determining the pathogenesis, clinical features and genetic characteristics of the disease. The SURF1 cytochrome c oxidase assembly factor *(SURF1)* gene is located on chromosome 9 and encodes a cytochrome c oxidase (COX) assembly factor. This gene is required for the assembly of the cytochrome c oxidase (COX; complex IV) enzyme, which is the terminal enzyme in mitochondrial respiratory chain [3]. Mutations in the SURF1 gene not only disrupt the assembly of COX and its enzymatic activity, resulting in a functional deficiency but also involving molecular mechanisms of Leigh syndrome (LS) [4]. The Human NADH: ubiquinone oxidoreductase core subunit V1 (NDUFV1) is encoded by the NDUFV1 gene on chromosome 11, and is a 51kDa component of nicotinamide adenine dinucleotide hydride-ubiquinone reductase that is essential for complex-I of mitochondrial respiratory chain, and plays an important role in NADH oxidation and ROS production [5]. Mutations in NDUFV1 result in phenotypes including myoclonic epilepsy, leukodystrophy and developmental arrest [6]. In addition to these, mutations in NDUFV1 were reported to be associated with a spectrum of diseases ranging from late-onset LS, ataxia and macrocephaly [7]. The *SURF1* and *NDUFV1* genes were associated with LS by genetic studies in the late 1990s and early 2000s. In particular, mutations in the *SURF1* gene have been directly linked to disease progression [8]. Mutations in the *NDUFV1* gene were first identified at least in 2000 association with LS and it has been observed that this gene is crucial for energy production and can increase the severity of symptoms [9].

This study used bioinformatics tools to examine the impact of SNPs leading to changes in amino acid residues in the *SURF1* and *NDUFV1* genes associated with LS on protein structure, function and stability. The SNPs are the most common genetic polymorphisms, which there is one nucleotide change in a specific position along genome. They tend to segregate into two alleles, which are the most numerous kind of genetic variants in the human genome [10].

Variants within genes can negatively or positively affect a gene's function, protein structure and predisposition to diseases [11]. Public databases record more than 10 million SNPs that are located in the genome about 100–300 base pairs apart. These polymorphisms add to genetic diversity, influencing characteristics, disease risk and response to drugs [12]. *In silico* simulation is the artificial replication of biological functions used to predict the organization and stability of living systems. These tests are often used for the stability of proteins, enzymes, genes, or drugs. *In silico* analysis is a realistic method to study protein, enzyme and gene stability, molecular basis of diseases and also it can be applied in drug development for understanding the molecular interaction. These analyses yield a wealth of useful information for researchers and concomitantly decrease the expense and time involved in laboratory studies [13]. Studies on variants of these genes in the literature are limited. This study is based solely on bioinformatic analyses and does not include experimental validation. Furthermore, our analyses do not include a population-based genetic diversity assessment. It may be useful to validate *in silico* analyses with laboratory experiments to better understand protein stability and biological effects. The distribution of these variants in different populations and their disease associations should be examined. Additionally, developing 3-D models of the SURF1 protein may contribute to comprehensive biological impact analyses.

This study provides an innovative contribution to the literature by comprehensively addressing the effects of SNPs on protein stability, functions and structural properties through multidimensional analyses. In addition, gene-gene and protein-protein interactions were analyzed to reveal new information about the molecular mechanisms of the disease. It was emphasized what role missense mutations in the *NDUFV1* and *SURF1* genes would play in the pathogenesis of Leigh syndrome and how the effects of these mutations could be characterized. These findings may enable the development of new approaches for early diagnosis and treatment strategies of LS in the future.

MATERIALS AND METHODS

Data Collection

Genes for analysis were selected based on a comprehensive literature review of LS. SNP identities, locations, nucleotide changes and their associated amino acid changes were obtained using resources from National Center for Biotechnology Information (NCBI). The UniProt platform was consulted to collect amino acid sequences in FASTA format and obtain UniProtKB protein identifiers.

In Silico Software Tools and Evaluation Criteria

Missense SNPs may contribute to changes in protein structure, stability and function, thus the impact of selected genes on proteins was analyzed *in silico* [14]. A number of software utilities were used to examine the potential for missense SNPs in SURF1 and NDUFV1 genes to disrupt protein structure. Only the SNPs which were potentially deleterious were chosen for further analysis. SIFT, PolyPhen-2 (HumVar and HumDiv), PredictProtein, PHD-SNP and SNPs&GO were the primary tools employed in the study followed by PANTHER and Meta-SNP. The software packages use a range of bioinformatics methods to find SNPs that are predicted to have an impact on the conformation or function of the encoded protein.

SIFT predicts whether an amino acid substitution in a protein will have a phenotypic effect and works by

calculating sequence homology, the properties of amino acids and evolutionary information to determine how changes from one state to another impact function [15]. PolyPhen-2, estimates the impact of amino acid substitutions on protein structure and activity using basic physical and evolutionary information. HumVar Model: Used for diagnosing Mendelian diseases and evaluating rare alleles, HumDiv Model: Effective for distinguishing significant mutations within human variations [16]. PredictProtein, enables the prediction of structural and functional characteristics of proteins. This is a system that brings together many programs for interactive protein structure and function studies [17]. Web-based PHD-SNP is designed to predict the possible impact of SNPs in protein coding regions. It is intended to distinguish the disease-related polymorphism from the neutral one [18]. SNPs&GO is a bioinformatics tool that traverses protein functional annotation (Gene Ontology terms) to infer the pathogenicity of single-point variations. It examines protein sequences to determine if a mutation is associated with disease [19]. PANTHER is a publicly available resource that classifies genes and gene products under ontologies by tracing their evolutionary relationship. This resource classifies proteins based on function [20]. Meta-SNP is a computational tool used to predict the pathogenicity of SNPs, by combining several models. It refines the prediction for disease-associated non-synonymous SNVs by integrating the predictions of SNAP, SIFT, PANTHER and PhD-SNP [21].

Bioinformatics Analysis Steps

The analyses conducted using these tools and algorithms were divided into the following steps:

SNP identification

SNPs in *SURF1* and *NDUFV1* genes were identified using NCBI and UniProtKB databases and their possible impacts on protein structure were examined.

Protein Function and Structural Analysis: SIFT and PolyPhen-2 were employed to predict the impact of amino acid changes on protein function. The predictions from these tools were analysed independently with HumVar and HumDiv to compare the outcome.

Pathogenicity assessment

Web-based software PHD-SNP and SNPs&GO were utilized to predict whether the single nucleotide polymorphisms (SNPs) in the present study had potential associations with disease. These analyses included estimating SNP effects with the contribution of Gene Ontology terms.

Integrated predictions

Meta-SNP integrated the results of SNAP, SIFT, PANTHER and PhD-SNP, allowing for a more accurate estimation of the disease-causing potential of SNPs. The data obtained as a result of the analyses were filtered and common deleterious SNPs were identified.

Gene-Gene and Protein-Protein Interactions

Gene-gene interactions were analyzed utilizing GeneMANIA, an online resource designed to evaluate gene functions and relationships. GeneMANIA is a software tool that provides information about the function of each gene, including pathways, co-localization, genetic interactions, physical relationships and common protein domains [22]. This tool facilitates the study of possible gene interactions.

Protein-protein interactions were evaluated utilizing another platform STRING. STRING improves the accuracy of predicted protein interactions by integrating various data sources, such as experimental datasets and co-expression analyses, thus enabling more reliable bioinformatic analyses [23]. Based on the provided data, possible functional partners of genes were identified.

Effects of Predicted Deleterıous Single Nucleotide Polymorphisms on Protein Stabilization

I-Mutant 2.0 and MUpro were employed to assess the effects of potentially damaging SNPs on protein stability. Support vector machine (SVM) technology is the basis of the I-Mutant 2.0, which predicts changes in protein stability induced by mutations. It calculates the change in free energy ($\Delta\Delta G$), which reflects the stabilizing or destabilizing effect of the mutation, based on the protein sequence and structure [24]. MUpro employs SVM to predict changes in protein stability and offers important information about protein behavior under mutational stress by predicting changes in protein stability induced by mutations [25]. These tools are specifically designed to predict the effects of variations at different positions on protein stability.

Construction of Three-Dimensional Models of Predicted Deleterious Variations

Three-dimensional models of the *NDUFV1* gene were created using the project HOPE platform, which provides tools to visualize and analyze protein structures [26]. The Project HOPE tool is a platform designed to analyze the structural and functional effects of SNPs or point mutations in proteins. Project HOPE combines data from various web services and databases, including UniProt and the Protein Data Bank (PDB), to generate detailed reports on how a mutation can affect the three-dimensional structure of a protein and its interactions. Project HOPE provided results on amino acid size, charge, hydrophobicity and conservation [27]. However, modeling information for the *SURF1* gene could not be obtained by Project HOPE because the three-dimensional structure of the *SURF1* gene is not available.

RESULTS AND DISCUSSION

Prediction Results of Deleterious Single Nucleotide Polymorphisms Using *In-Silico* Methods

5321 SNPs were detected in the *SURF1* gene, of which 734 were identified as missense variations. Among these

missense SNPs, 602 amino acid substitutions were evaluated. 3417 SNPs were detected in the *NDUFV1* gene, of which 431 were missense SNPs. Within this subset, 547 amino acid changes were analyzed.

Using all software tools, 6 deleterious SNPs (rs28933402, rs121918658, rs145615218, rs201492662, rs373551988, rs369707095) were identified in the *SURF1* gene (Table 1). These 6 missense SNPs include a total of 6 variants (G124E, Y274D, Y117S, R118W, R137W and G159W) predicted to be deleterious. In the *NDUFV1* gene (Table 1), 11 deleterious SNPs (rs121913661, rs138526825, rs138583785, rs150966634, rs201787156, rs11540007, rs139966879, rs143866203, rs201382784, rs375322599, rs375897089) were identified, with a total of 12 variants (E214K, R88G, R88C, R224C, R386C, E408K, G87S, G146V, R88H, G306C, G321S and E246K) predicted to be deleterious.

Results of Protein Stabilization

The effect of SNPs classified as deleterious by all the software tools on the stability of proteins was evaluated utilizing I-Mutant 2.0 and MuPro platforms. An overview of these results is provided in Table 1 and Table 2.

Results of Gene-Gene and Protein-Protein Interactions

Gene-gene and protein-protein interactions involving the *SURF1* and *NDUFV1* genes, as well as their connections with other genes, were examined using GeneMANIA. Figure 1 depicts these gene interactions, highlighting aspects such as co-expression, localization, genetic and physical interactions, predicted pathways and common protein domains. Additionally, the STRING database was used to map out the connections between *SURF1* and *NDUFV1* and ten related proteins, as illustrated in Figure 1.

Results of Construction of Three-Dimensional Models of Predicted Deleterious Variations

Each amino acid has its own characteristics, including size, charge, and hydrophobicity. Each of these characteristics can be affected by mutations. Hydrophobicity mutations can influence hydrogen bonding, while charge mutations can affect ionic bonds. Size variation between

Table 1. Possible prediction results of single nucleotide polymorphisms in *SURF1* gene using software tools

Analysis	rs28933402	rs121918658	rs145615218	rs201492662	rs373551988	rs369707095
Nucleotide Change	C>T	A>C	T>A / T>G	T>A / T>C	G>A	C>A
Amino Acid Change	G124E	Y274D	Y117S	R118W	R137W	G159W
PredictProtein Result	E	E	E	E	E	E
PredictProtein Score	90	95	81	68	91	83
SIFT Result	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt
SIFT Score	0.000	0.000	0.001	0.008	0.043	0.001
PolyPhen-2 HumDiv Result	PH	PH	PH	PH	PH	PH
PolyPhen-2 HumDiv Score	1000	1000	1000	1000	0.999	1000
PolyPhen-2 HumVar Result	PH	PH	PH	PH	PH	PH
PolyPhen-2 HumVar Score	1000	1000	0.995	0.998	0.948	1000
SNP&GO Result	Ds	Ds	Ds	Ds	Ds	Ds
SNP&GO RI	9	9	9	7	8	9
PHD-SNP Result	Ds	Ds	Ds	Ds	Ds	Ds
PHD-SNP RI	8	10	9	9	8	8
PANTHER Result	PH	PH	PH	PH	PH	PH
META-SNP Result	Ds	Ds	Ds	Ds	Ds	Ds
META-SNP Score	0.854	0.875	0.807	0.824	0.803	0.834
META-SNP RI	7	8	6	6	6	7
I-Mutant Result	Dec	Dec	Dec	Dec	Dec	Dec
I-Mutant Score	1	4	8	8	7	7
MUpro Result	Dec	Dec	Dec	Dec	Dec	Dec
MUpro DDG Value	-0,86483608	-1,3281959	-1,5154971	-0,98251135	-0,39536249	-0,065944524

E: Effective; Dlt: Deleterious; Ds: Disease; Dec: Decrease; PH: Potentially harmful.

Table 2. Possible prediction results of single nucleotide polymorphisms in *NDUFV1* gene using software tools

Analysis	s121913661	s138526825	s138526825	s138583785	s150966634	S201787156	s11540007	s139966879	s143866203	s201382784	s375322599	s375897089
Nucleotide change	G>A	C>G / C>T	C>G / C>T	C>A / C>T	C>T	G>A	G>A	G>A / G>T	G>A	G>C / G>T	G>A	G>A
Amino Acid change	214K	88G	88C	224C	386C	408K	87S	146V	88H	306C	321S	246K
Predict protein result	E	E	E	E	E	E	E	E	E	E	E	E
PredictProtein score	70	93	80	73	87	57	73	71	73	57	79	88
SIFT result	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt	Dlt
SIFT score	0.013	0.001	0.000	0.003	0.000	0.002	0.000	0.003	0.000	0.000	0.000	0.000
PolyPhen-2 HumDiv result	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH
PolyPhen-2 HumDiv score	1000	1000	1000	1000	1000	1000	1000	0.977	1000	1000	1000	0.999
PolyPhen-2 HumVar result	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH
PolyPhen-2 HumVar score	0.986	1000	1000	1000	0.999	0.999	1000	0.931	1000	0.997	1000	0.999
SNP&GO result	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds
SNP&GO RI	0.93	0.89	0.92	0.95	0.95	0.54	0.72	0.93	0.86	0.72	0.89	0.93
PHD-SNP result	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds
PHD-SNP RI	9	8	8	9	9	1	4	9	7	4	8	9
PANTHER result	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH	PH
META-SNP result	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds	Ds
META-SNP score	0.715	0.875	0.912	0.885	0.889	0.725	0.884	0.771	0.875	0.932	0.853	0.888
META-SNP RI	4	8	8	8	8	5	8	5	8	9	7	8
I-Mutant result	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec
I-Mutant score	5	6	5	6	4	8	1	4	5	6	8	7
MUpro result	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec	Dec
MUpro DDG value	-1.0879523	-1.5640386	-0.72560616	-0.2860499	-0.47840161	-1.8670333	-0.97483618	-0.60336513	-1.1489642	-0.83887817	-0.61125374	-0.99532374

E: Effective; Dlt: Deleterious; Ds: Disease; Dec: Decrease; PH: Potentially Harmful

normal and mutant-type amino acids can impair the formation of a hydrogen bond. These characteristics can impair protein function by causing the loss of ligand binding. Mutations can also influence local structure and binding sites. Moreover, mutations can impair interactions between crucial domains of proteins, thereby influencing protein function. The differences in hydrophobicity, electric charge, and molecular size of normal and mutated amino acids in the *SURF1* and *NDUFV1* genes were evaluated using Project HOPE. Details of these outcomes are provided in Table 3. Moreover, models and annotations for deleterious mutations caused by changes in amino acids in the *SURF1* and *NDUFV1* genes and three-dimensional structures for the *NDUFV1* gene were created using Project HOPE (Table 4).

Biological Implications of the Findings

The findings indicate that certain SNPs identified in the *SURF1* and *NDUFV1* genes may negatively affect protein stability, which may lead to impaired mitochondrial functions and contribute to the progression of LS. Specifically, variations in functional protein interactions can be triggered by some of the identified deleterious SNPs. These can lead to problems in energy production in mitochondria and the emergence of symptoms associated with cellular energy deficiency [28]. The results can be evaluated to develop more specific treatment and follow-up strategies for LS and to provide important clues regarding the potential impact of genetic differences on the underlying molecular mechanisms of the condition.

Studying SNPs is vital for elucidating the genetic foundations of diseases, assessing disease risk and predicting individual responses to treatment. Since there are so many known SNPs in the literature, conducting wet-lab experiments aimed at understanding the biological effects of each SNP is quite labor-intensive. Therefore, bioinformatics analyses are used as an effective pre-screening strategy to prioritize potentially pathogenic SNPs before proceeding to wet lab-based experimental validation studies [29].

It is predicted that the E214K and G146V variants identified in the NDUFV1 gene may negatively affect the NADH oxidation process by disrupting the structural stability and enzymatic activity of mitochondrial Complex I. Such structural and functional disruptions can lead to a decrease in oxidative phosphorylation efficiency and a reduction in cellular energy production. This disruption may lead to disruptions in electron transport system and accumulation of oxidative species [5,7]. Increased oxidative imbalance and metabolic deficiency may contribute to the pathogenesis of the disease by causing cellular damage and neurodegeneration.

It has been reported in the literature that P119L mutation in *SURF1* gene causes defects in COX mechanism and is associated with LS [4]. This mutation affects the structural stability of the COX complex, similar to the G124E and G159W mutations detected in our study. All three mutations lead to a decrease in COX activity and cause disruptions in energy metabolism. R88C and R386C mutations previously reported in *NDUFV1* gene are also related

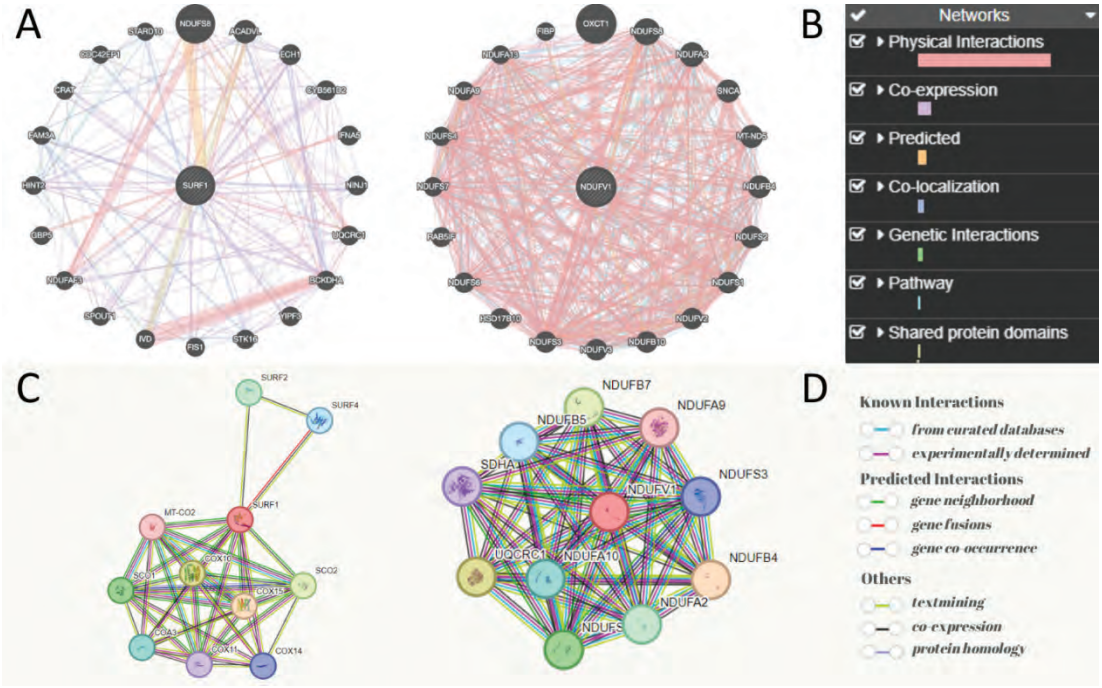


Figure 1A. *SURF1*, **B.** *NDUFV1* gene interaction map (GeneMANIA), **C.** *SURF1*, **D.** *NDUFV1* protein interaction map (STRING).

to complex I deficiency and LS [6,30]. These mutations may disrupt protein function by mechanisms similar to the E214K and G146V mutations we detected. In both cases, the decrease in complex I activity may affect cellular energy production and contribute to the clinical symptoms of the disease.

The study by Özkan et al., examining the effects of genetic variations on the disease, reveals that there is no significant relationship between prostate cancer and the rs7501939 polymorphism. This finding shows that SNP analyses are important for understanding both disease risk and disease characteristics [31].

The literature reported that Y204C and E214K variants in the *NDUFV1* gene have deleterious effects [32]. Another study mentioned the disease-associated effects of R88G, R386C and R199P variants in the *NDUFV1* gene; however, our study did not identify R199P as a common deleterious variant [30]. For the *SURF1* gene, the P119L variant has been documented in the literature, but was not found in our data in the NCBI database. In this context, the results of this study are consistent with the data in the literature and reveal the deleterious effects of SNPs in more detail.

When examining other mitochondrial diseases, MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes) disease is associated with the A3243G mutation in the MT-TL1 gene. This mutation disrupts the structure of mitochondrial transfer ribonucleic acid (tRNA) and negatively affects energy production. Mutations in LS are generally of nuclear deoxyribonucleic acid (DNA) origin and affect complexes I and IV of the mitochondrial respiratory chain, leading to disruptions in adenosine triphosphate (ATP) production. While *in silico* analyses predict the structural effects of this mutation by examining the tRNA structure in MELAS, mutations that disrupt protein function in LS are examined with protein modeling [33].

LHON (Leber's Hereditary Optic Neuropathy) is associated with mutations in the *MT-ND1*, *MT-ND4* and *MT-ND6* genes in mitochondrial DNA. One of the most common mutations in this disease, G11778A, disrupts the function of mitochondrial complex I and prevents energy production in optic nerve cells, leading to vision loss. *NDUFV1* mutations in LS similarly disrupt complex I function, but unlike LHON, neurological symptoms are more common and usually appear in infancy [34].

The difference in the size of mutant amino acids and wild-type amino acids can greatly influence protein structure and function. Variations in the size of amino acids can influence the overall structure and function of proteins [35].

Glycine, the smallest amino acid with a side chain chain made up of only one hydrogen atom, enhances the flexibility

Table 3. Amino acid changes of specific single nucleotide polymorphisms in *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool, with effects of these changes on attributes like such as size, charge and hydrophobicity for wild-type and mutant-type

Gene name	SNP ID	Amino acid substitution	Wild-type			Mutant-type		
			Size	Charge	Hydrophobicity	Size	Charge	Hydrophobicity
SURF1	**rs28933402**	**G124E**	<	Neutral	>	>	Negative	<
	rs121918658	**Y274D**	>	Neutral	>	<	Negative	<
	rs145615218	**Y117S**	>	-	-	<	-	-
	rs201492662	**R118W**	<	Positive	<	>	Neutral	>
	rs373551988	**R137W**	<	Positive	<	>	Neutral	>
	rs369707095	**G159W**	<	-	<	>	-	>
NDUFV1	**rs121913661**	**E214K**	<	Negative	-	>	Positive	-
	rs138526825	**R88G**	>	Positive	<	<	Neutral	>
	rs138526825	**R88C**	>	Positive	<	<	Neutral	>
	rs138583785	**R224C**	>	Positive	<	<	Neutral	>
	rs150966634	**R386C**	>	Positive	<	<	Neutral	>
	rs201787156	**E408K**	<	Negative	-	>	Positive	-
	rs139966879	**G146V**	<	-	<	>	-	>
	rs143866203	**R88H**	<	Positive	-	>	Neutral	-
	rs201382784	**G306C**	<	-	<	>	-	>
	rs375322599	**G321S**	<	-	-	>	-	-
	rs375897089	**E246K**	<	Negative	-	>	Positive	-

of protein structures. The flexibility of protein structures is crucial for the function of the proteins encoded by the SURF1 and *NDUFV1* genes, and the substitution of glycine with another amino acid can impair this flexibility [26]. For example, G124E, G159W in *SURF1* and G87S, G146V, G306C and G321S in *NDUFV1* represent substitutions of glycine with other amino acids that may affect protein function. Similarly, changes to or from glycine may affect the required rigidity at certain positions, thus affecting both function and structure. Table 4 shows the variations in electrical charge among wild and mutant-type residues. Proteins may form ionic interactions between oppositely charged residues, known as salt bridges, due to electrostatic forces. These salt bridges contribute to the protein's structure and interactions with other biomolecules. In an aqueous environment, certain ionic interactions occur on the protein surface, while others happen on the inside [26]. For example, in the E214K variant of the *NDUFV1* gene, the hydrogen bond between the lysine at position 219 and the arginine at position 224 may not reoccur due to the smaller size and different electrical charge of the new residue. Variations in hydrophobicity among the residues are also presented in Table 4.

As a result of HOPE analyses, Y274D variant in the *SURF1* gene, numbered rs121918658, identified in our study, was found to cause the substitution of aspartic acid instead of tyrosine and COX deficiency. These findings are consistent with those reported by [36] study, which is consistent with the destabilizing effects of the mutation on the COX complex.

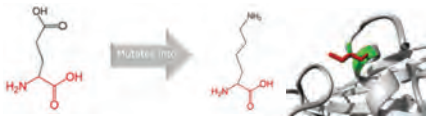
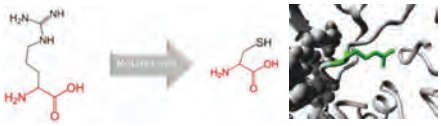
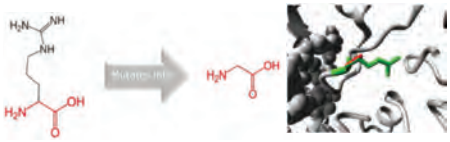
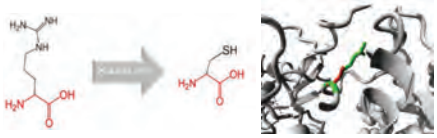
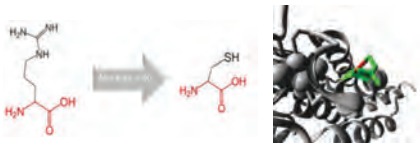
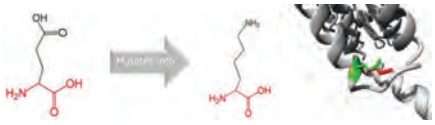
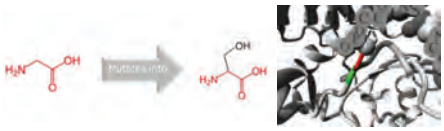
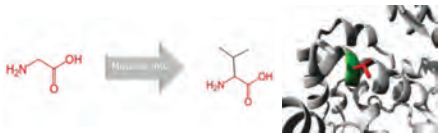
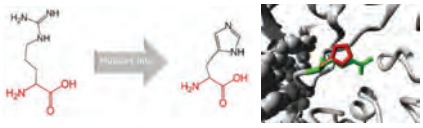
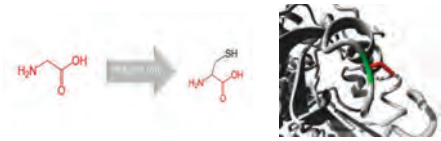
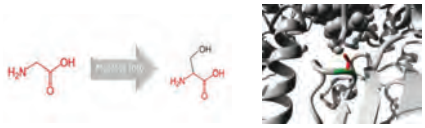
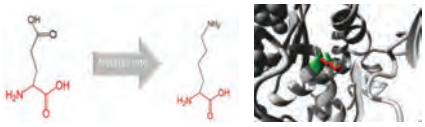
E214K variant numbered rs121913661 in the *NDUFV1* gene, identified in our study, was determined to be deleterious as a result of our analysis. Additionally, this variant was described by [37] study, it was detected in Leigh syndrome patients with complex 1 deficiency. The data in this experimental study are parallel to the data we obtained in our study.

The findings show that these SNPs may disrupt protein stability and lead to disruptions in mitochondrial functions, which may induce the neurological symptoms of the disease. The study can be an important data source for determining genetic risk factors and personalized treatment approaches. The information obtained can shed light on new research areas for the understanding of LS pathogenesis and mitochondrial diseases.

Table 4. Models and explanations of the *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool

Genes	SNP ID	Modelling	Explanation
SURF1	**rs28933402**		Glycine → Glutamic acid at position 124
	rs121918658		Tyrosine → Aspartic acid at position 274
	rs145615218		Tyrosine → Serine at position 117
	rs201492662		Arginine → Tryptophan at position118
	rs373551988		Arginine → Tryptophan at position 137
	rs369707095		Glycine → Tryptophan at position 159

Table 4. Models and explanations of the *SURF1* and *NDUFV1* genes obtained using the Project HOPE software tool

Genes	SNP ID	Modelling	Explanation
NDUFV1	**rs121913661**		Glutamic acid → Lysine at position 214
	rs138526825		Arginine → Glycine at position 88
	rs138526825		Arginine → Cysteine at position 88
	rs138583785		Arginine → Cysteine at position 224
	rs150966634		Arginine → Cysteine at position 386
	rs201787156		Glutamic acid → Lysine at position 408
	rs11540007		Glycine → Serine at position 87
	rs139966879		Glycine → Valine at position 146
	rs143866203		Arginine → Histidine at position 88
	rs201382784		Glycine → Cysteine at position 306
	rs375322599		Glycine → Serine at position 321
	rs375897089		Glutamic acid → Lysine at position 246

Green colour: Wild-type residue; Red colour: Mutant-type residue.

CONCLUSION

In this study, we evaluated the effects of SNPs causing amino acid substitutions in the SURF1 and NDUFV1 genes on the protein structure, function, and stability in LS, using bioinformatics tools for analysis. The findings obtained from our study are expected to serve as a valuable data source for both experimental and bioinformatics research. Future studies should include experimental studies in cellular and animal models to confirm the effects of these single nucleotide polymorphisms on protein function. In particular, their effects on protein-protein interactions, cellular energy metabolism and mitochondrial functions may be detailed with biochemical experiments. In addition, genotype-phenotype analyses may be performed in large patient cohorts to investigate the clinical biomarker potential of these deleterious single nucleotide polymorphisms. Another crucial step is the identification of targeted treatment strategies for these variants. For instance, the identification of small molecules or gene therapy strategies that could enhance the stability of mutant proteins could help reduce the symptoms of the patients. In this case, the clinical interpretation of the results and the identification of targeted treatment approaches could offer major breakthroughs in the treatment of mitochondrial disorders like LS.

ACKNOWLEDGEMENTS

This study was produced from the master's thesis of the first author, which was accepted at Üsküdar University, Department of Biotechnology.

AUTHORSHIP CONTRIBUTIONS

AÖ: Data collection, *in silico* analysis, writing—original draft preparation. KÖF: Organizing the research, designing the research and methodology, writing (review and editing). MK: Organizing the research, designing the research and methodology, writing (review and editing).

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Chae JH, Lee JS, Kim KJ, Hwang YS, Hirano M. Biochemical and genetic analysis of Leigh syndrome patients in Korea. Brain Dev 2008;30:387–390. [CrossRef]

[2] Johnson SC, Yanos ME, Kayser EB, Quintana A, Sangesland M, Castanza A, et al. MTOR inhibition alleviates mitochondrial disease in a mouse model of Leigh syndrome. Science 2013;342:1524–1528. [CrossRef]

[3] Péquignot MO, Desguerre I, Dey R, Tartari M, Zeviani M, Agostino A, et al. New splicing-site mutations in the SURF1 gene in Leigh syndrome patients. J Biol Chem 2001;276:15326–15329. [CrossRef]

[4] Pecina P, Čapková M, Chowdhury SKR, Drahota Z, Dubot A, Vojtíšková A, et al. Functional alteration of cytochrome c oxidase by SURF1 mutations in Leigh syndrome. Biochim Biophys Acta Mol Basis Dis 2003;1639:53–63. [CrossRef]

[5] Zanette V, Valle D, Telles BA, Robinson AJ, Monteiro V, Santos MLSF, et al. Ndufv1 mutations in complex I deficiency: Case reports and review of symptoms. Genet Mol Biol 2021. [CrossRef]

[6] Tsang WY, Sayles LC, Grad LI, Pilgrim DB, Lemire BD. Mitochondrial respiratory chain deficiency in Caenorhabditis elegans results in developmental arrest and increased life span. J Biol Chem 2001;276:32240–32246. [CrossRef]

[7] Incecik F, Herguner O, Besen S, Bozdoğan S, Mungan N. Late-onset Leigh syndrome due to NDUFV1 mutation in a 10-year-old boy initially presenting with ataxia. J Pediatr Neurosci 2018;13:205. [CrossRef]

[8] Péquignot MO, Dey R, Zeviani M, Tiranti V, Godinot C, Poyau A, et al. Mutations in the SURF1 gene associated with Leigh syndrome and cytochrome c oxidase deficiency. Hum Mutat 2001;17:374–381. [CrossRef]

[9] Lake NJ, Compton AG, Rahman S, Thorburn DR. Leigh syndrome: One disorder, more than 75 monogenic causes. Ann Neurol 2016;79:190–203. [CrossRef]

[10] Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res 2004;14:414–425. [CrossRef]

[11] Wang Z, Huang C, Lv H, Zhang M, Li X. In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. PLoS One 2020;15:e0227859. [CrossRef]

[12] Zhu J, Qiu C, Palla M, Nguyen T, Russo JJ, Ju J, et al. A microfluidic device for multiplex single-nucleotide polymorphism genotyping. RSC Adv 2014;4:4269–4277. [CrossRef]
[13] Merlino A, Graziano G, Mazzarella L. Structural and dynamic effects of α-helix deletion in Sso7d: Implications for protein thermal stability. Proteins 2004;57:692–701. [CrossRef]
[14] Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–3814. [CrossRef]
[15] Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–1082. [CrossRef]
[16] Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009;19:1553–1561. [CrossRef]
[17] Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein – An open resource for online prediction of protein structural and functional features. Nucleic Acids Res 2014;42:W337–W343. [CrossRef]
[18] Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum Mutat 2009;30:1237–1244. [CrossRef]
[19] Bhardwaj A, Dhar YV, Asif MH, Bag SK. In silico identification of SNP diversity in cultivated and wild tomato species: Insight from molecular simulations. Sci Rep 2016;6:38715. [CrossRef]
[20] Alizadehmohajer N, Zahedifar S, Sohrabi E, Shaddel Basir S, Nourigheimasi S, Falak R, et al. Using in silico bioinformatics algorithms for the accurate prediction of the impact of spike protein mutations on pathogenicity, stability, and functionality of SARS-CoV-2. Biochem Genet 2023;61:778–808. [CrossRef]
[21] Ali AAH, Ibrahim M. In silico analysis of single nucleotide polymorphisms in human CCR5 gene. bioRxiv 2020. Preprint. doi: https://doi.org/10.1101/2020.11.14.382739.
[22] Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, et al. GeneMANIA update 2018. Nucleic Acids Res 2018;46:W60–W64. [CrossRef]
[23] von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, et al. STRING: Known and predicted protein-protein associations integrated and transferred across organisms. Nucleic Acids Res 2005;33:D433–D437. [CrossRef]
[24] Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 2005;33. [CrossRef]
[25] Pan Q, Nguyen TB, Ascher DB, Pires DEV. Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures. Brief Bioinform 2022;23. [CrossRef]
[26] Venselaar H, te Beek TAH, Kuipers RKP, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases: An e-Science approach with life scientist friendly interfaces. BMC Bioinform 2010;11:bbac025. [CrossRef]
[27] Li H, Wetten S, Li L, St Jean PL, Upmanyu R, Surh L, et al. Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. Arch Neurol 2008;65:45–53. [CrossRef]
[28] Lee IC, Chiang KL. Clinical diagnosis and treatment of Leigh syndrome based on SURF1 genotype and phenotype. Antioxidants 2021;10:1950. [CrossRef]
[29] Bénit P, Chretien D, Kadhom N, De Lonlay-Debeney P, Cormier-Daire V, Cabral A, et al. Large-scale deletion and point mutations of the nuclear NDUFV1 and NDUFS1 genes in mitochondrial complex I deficiency. Am J Hum Genet 2001;68:1344–1352. [CrossRef]
[30] Marin SE, Mesterman R, Robinson B, Rodenburg RJ, Smeitink J, Tarnopolsky MA. Leigh syndrome associated with mitochondrial complex I deficiency due to novel mutations in NDUFV1 and NDUFS2. Gene 2013;516:162–167. [CrossRef]
[31] Özkan E, Erdemir A, Törer BD, Taşçı AI, Baskın Y, Ellidokuz H, et al. Genotyping and analysis of rs7501939 polymorphism for prostate cancer. Sigma J Eng Sci 2015;6:101–107.
[32] Kose M, Canda E, Kagnici M, Aykut A, Adebali O, Durmaz A, et al. SURF1 related Leigh syndrome: Clinical and molecular findings of 16 patients from Turkey. Mol Genet Metab Rep 2020;25. [CrossRef]
[33] El-Hattab AW, Adesina AM, Jones J, Scaglia F. MELAS syndrome: Clinical manifestations, pathogenesis, and treatment options. Mol Genet Metab 2015;116:4–12. [CrossRef]
[34] Engvall M, Kawasaki A, Carelli V, Wibom R, Bruhn H, Lesko N, et al. A novel mutation in the mitochondrial MT-ND5 gene associated with Leber hereditary optic neuropathy. Front Neurol 2021;12:652590. [CrossRef]
[35] Biro JC. Amino acid size, charge, hydropathy indices and matrices for protein structure analysis. Theor Biol Med Model 2006;3. [CrossRef]
[36] Munnich A, Rustin P. Cytochrome c oxidase deficiency. Am J Med Genet Semin Med Genet 2001;106:46–52. [CrossRef]
[37] Quaio CRDAC, Moreira CM, Novo-Filho GM, Sacramento-Bobotis PR, Groenner Penna M, Perazzio SF, et al. Diagnostic power and clinical impact of exome sequencing in a cohort of 500 patients with rare diseases. Am J Med Genet C Semin Med Genet 2020;184:955–964. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 919–953, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2018

Research Article

Driving style recognition with machine learning for intelligent control of vehicles

Ahmet BEŞKARDEŞ[1], Yakup HAMEŞ[1,*]

[1]*Department of Electrical and Electronics Engineering, İskenderun Tecnical University, İskenderun, Hatay, 31200, Türkiye*

ARTICLE INFO

Article history
Received: 09 September 2024
Revised: 14 October 2024
Accepted: 30 January 2025

Keywords:
Classification Algorithms; Data-Driven Approach; Driving Style Recognition; Machine Learning Models

ABSTRACT

In recent years, extensive research and development efforts have been increasingly carried out in both academia and industry to optimize the safety, efficiency, and environmental impact of vehicles traveling on roads. This study aims to contribute to these efforts by focusing on determining the driver's driving style. The approach of our study is crucial for enhancing driving safety and optimizing fuel efficiency. Data science and machine learning techniques were employed to identify the driver's driving style based on data collected during specific driving activities. Comprehensive driving data from various vehicle types were gathered as speed-time series using a global positioning system-based mobile application. The extracted features underwent necessary preprocessing and transformation to ensure their suitability for machine learning models. The prepared dataset was applied to k-nearest neighbors, support vector machine, decision tree, artificial neural network, logistic regression, naive Bayes, random forest, and light gradient boosting machine models. Three distinct driving styles were predicted: aggressive, normal, and calm. The accuracy rates achieved were 98.6% on residential district roads, 95.5% on urban roads, and 100% on motorways, with decision tree-based algorithms proving to be the most effective.

Cite this article as: Beşkardeş A, Hameş Y. Driving style recognition with machine learning for intelligent control of vehicles. Sigma J Eng Nat Sci 2026;44(2):919–953.

INTRODUCTION

With the significant advancements in the digitalization of road vehicles, accurately determining and adapting to the driving style has become crucial for optimizing safety, efficiency, and environmental impact. Correctly estimating the driving style and adapting the vehicle to this style is essential for enhancing vehicle efficiency, traffic safety, and driver comfort. The driving style of individuals significantly influences the efficient utilization of vehicles. Driving style is influenced by various factors, including individual traits like gender, age, attitude, and personality; environmental conditions such as the type of road, weather, journey distance, and traffic; and vehicle characteristics like type, model, and power [1]. To design vehicle control systems that accommodate individual driving styles, accurate prediction of these styles is necessary. Driving aggressively leads to higher energy consumption, whereas driving calmly helps to lower energy consumption [2]. If

***Corresponding author.**
*E-mail address: yakup.hames@iste.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

a vehicle's intelligent control system can adapt to the driver's driving style, it can optimize energy management by prioritizing performance when desired and fuel economy when calmness is preferred. Accurate assessment of driving style enhances driver safety [3,4], comfort, and enjoyment [5], fuel economy [6], and reduces emissions. Therefore, the ability to recognize driving styles has been integrated into new-generation advanced driver assistance systems (ADAS), significantly enhancing fuel consumption and driver safety [7].

The degree of acceleration and deceleration, turning speed, and frequency of lane changes when driving are typically considered when describing a driving style. However, it should not be forgotten that the same driver will exhibit different driving styles in different vehicles and environmental conditions. Three different driving styles can be defined as aggressive, calm, and moderate. Aggressive drivers try to minimize travel time, drive impatiently and excessively, do not maintain a safe following distance, make sudden and unexpected maneuvers, and change lanes frequently [8]. This situation causes dangerous situations in traffic as well as high fuel consumption [9]. Calm drivers avoid risky reactions and activities, do not accelerate or decelerate excessively, and drive more carefully [10]. Moderate drivers, on the other hand, are defined as an intermediate group that shows both aggressive and calm characteristics [7]. Table 1 shows the factors affecting driving styles according to vehicle, road, environment, and human factors.

Driving style prediction algorithms are developed based on the features extracted from the collected signals and the classification methods that will use these features. Martinez et al. reported that driving style recognition applications are developed using rules, models, and learning algorithms [7].

In the studies conducted through the rules, an output is obtained from the rules written according to the value of events such as acceleration, deceleration, left-right turn, or lane change, and the driving style is categorized accordingly [11]. Rule-based algorithms are simple to use and easy to understand, but they limit the number of parameters that can be studied. Therefore, the robustness and accuracy of the results are also very limited [7]. Working with more parameters is only possible with more advanced control systems such as fuzzy logic. The fuzzy system, which can be modeled with more parameters, can give results with higher accuracy while preserving the simplicity, robustness and easy understanding of the rule-based system. Dorr et al. conducted an online study by adding the road type and the distance between the vehicles into the input variables to determine the driving style [2]. Gilman et al., on the other hand, developed an application based on 17 factors with a fuzzy system in their study where they prioritized fuel consumption [12]. Fernandez and Ito developed a fuzzy system to predict traffic accidents and optimize route management. The inputs of the system were vehicle speed, acceleration, deceleration, and the driver's age, and the output were five different driver profiles very passive, passive, normal, aggressive, and dangerous [13]. Rule-based applications have the disadvantage of requiring expert knowledge as well as being straightforward and practical. In order to eliminate this negativity, the proposed model should be supported by looking at the previous data.

In model-based algorithms, driving style definitions are made based on certain formulas by making the necessary parameter settings using data-based methods [7]. With these models, it may be possible to achieve faster results in a simulation environment or in real-world applications.

In learning algorithms, a driving definition is made based on data. In this method, more robust and highly accurate results can be produced with the learning algorithms chosen in accordance with the problem after comprehensive data analysis. Machine learning algorithms are unsupervised, supervised, and hybrid. In unsupervised learning, there are no output tags that must be assigned in response to previous inputs. Therefore, there is no need to understand the process of the problem. This process is done through statistical analysis in these algorithms, which produces more clustering solutions. A dataset containing output tags is used to train the algorithm in supervised learning. When the model trained with historical data is given new data that it has not seen before, it is expected to predict the output. In supervised learning algorithms used for regression and classification, methods such as k-nearest neighbor (k-NN), decision trees, random forests, SVM,

Table 1. Influencing factors on driving style

Influencing factors on driving style		
Vehicle and road factors	**Environmental factors**	**Human factors**
Vehicle type and model	Traffic	Gender
Vehicle design	Season	Age
Vehicle power	Geographic region	Driving experience
Road type	Time of the day	Condition (alcohol, fatigue, stress)
Sight distance		Driving distance
Lighting conditions		Other drivers

neural networks, Naive Bayes, and logistic regression are most commonly used.

Studies with learning algorithms in defining driving style are much more than the other two algorithms. Van Ly et al. evaluated the characteristics of acceleration events, braking, and turning events from the data collected from inertial sensors and classified driver behavior with two different classification algorithms [14]. In this study, instead of being satisfied with two classifiers, increasing the variety of methods can lead researchers to better results. Zhang et al. extracted features describing driving habits from the data obtained from phone and car sensors and performed driver classification with support vector machines [15]. Quintero and colleagues modeled driver behavior with an artificial neural network by evaluating position, speed, acceleration, and steering angle data obtained from a global positioning system (GPS) data recording system [16]. Although a high success rate was achieved with neural networks in this study, if other algorithms were applied to the same data, comparing them in terms of processing time, interpretability, and practicality would be possible. Shahverdy and others used data on speed, acceleration, gravity, throttle, and engine speed (RPM). Using a deep learning model with these data, they predicted five different driver types with very high accuracy: normal, aggressive, distracted, sleepy, and drunk [17]. However, in this study, only three drivers and a single vehicle were tested, and the drivers drove according to the instructions they received before driving. Therefore, this study needs to be expanded with more realistic and larger dataset. Vaitkus et al. extracted statistical features such as mean, median, mode, variance, standard deviation, range, minimum, maximum, skewness, and kurtosis from triaxial accelerometer data for driving on a repetitive route. From the values of these features, they estimated two different driving styles with k-NN, either aggressive or normal [18]. However, as they stated, the high accuracy rate obtained in the tests performed in the same season, in the same traffic conditions, and on the same route can be misleading. Performing these tests with different and more driving data will increase the model's reliability. In addition, conducting tests with more diverse classification algorithms and increasing the number of categories of the output class (such as aggressive, normal, and calm) will further increase the value of the study. Kedar-Dongarkar and Das collected the acceleration, braking, and throttle values of the vehicle, and estimated the driver of the vehicle as one of the aggressive, moderate, and conservative profiles using a simple classifier method from these data [19]. After applying the sensor data to two different neural networks, Brombacher et al. detected defensive and sporty driving maneuvers and calculated a driving style classification score in five categories [20]. Saleh et al. extracted three different driving behaviors as normal, aggressive, and sleepy with the time series classification model from nine different sensor data obtained from the smartphone [21]. Wang et al. collected 400 driving data from 20 drivers in a driving simulator and applied the vehicle speed and throttle opening properties from this dataset to a semi-supervised support vector machine. With this method, he estimated three different driving styles, normal, vague, and aggressive [22]. Meteier et al. used physiological signals and machine learning techniques to predict sleep deprivation, driving environment, and sleepiness in experiments with 63 different drivers in a driving simulator. Although they could not predict sleepiness well, they achieved very good results in sleep deprivation [23]. Seraj used a rule-based classification method to analyze the driving data obtained from different roads such as highways, main arteries and ramps to analyze the longitudinal and lateral controls of drivers. His classification algorithm-based study revealed three different classes of drivers: calm, rational, and aggressive [24]. This study is distinguished by its use of real-world data across various road types, implementation of both short-term and long-term classification methods, and integration Advanced Driver Assistance Systems (ADAS). Khan et al. created a dataset containing 15 features from CAN-Bus data collected via OBD-II. They were able to distinguish two drivers almost perfectly with their classification algorithm, but their accuracy dropped significantly for ten drivers [25]. Benterki et al. identified two driving styles, calm and aggressive, using the Next Generation Simulation dataset with spectral clustering and K-means algorithms. This study is an example of unsupervised learning methods in driving style diagnosis [26].

Similar methods have been applied for the prediction of driving skills, aggression status, or road type. Chandrasiri et al obtained driving data from a driving simulator and tagged the profiles of these rides. From these data, they created a model within the scope of the driving environment, the behavior of the drivers, and the reactions of the vehicles to the environment. With this model they developed based on SVM and k-NN, they estimated a driver's driving skills [27]. Yu and others identified 16 key features from driving data collected over 6 months from real driving environments. By applying this data to neural networks and support vector machines, they predicted and detected abnormal conditions such as side slip, rapid U-turn, large radius turn, and sudden braking [28]. Bernardi and others extracted a comprehensive set of features from the data they took from the monitoring system they built into the vehicles. Applying these features to the time series classification model, they identified different driving styles, drivers' familiarity with their vehicles, and road types [29]. Moukafih et al used the UAH-DriveSet dataset, which consists of data from speed and acceleration sensors and a vehicle front view camera. They applied this data to a time series classification called the Long Short-Term Memory Fully Convolutional Network (LTSM-FCN). They estimated driver aggression status with this classification model [30]. Sun et al. reported that studies conducted under real traffic conditions would yield better results [31], while Eckert et al. reported that driving cycles derived from various road

and traffic conditions would further improve these results [32].

The studies described above fill important gaps in the prediction of driving style and its contribution to driving. However, for a reliable driving prediction, features such as the number of drives, driving time, driving purpose, distance traveled and the fact that these were done under real traffic conditions are extremely important. Therefore, in this study, we developed an approach based on real data that we produced with meticulous care and dedication, considering realistic uncertainties. Recognizing the crucial role of real-world factors like driving experience, purpose, and environmental conditions, we meticulously collected a comprehensive dataset encompassing diverse scenarios. We conducted both a large number and a wide variety of driving experiments to create this dataset. We used different vehicles such as cars, minibuses, and buses. We used vehicles on three different road types: residential, urban, and highway. We drove in all four seasons of the year and at different times of the day. These drives were in different geographical regions such as sea level, and mountainous areas, and were carried out for different purposes such as commuting to work or traveling for vacation. Our professional drivers who made the drives were of different ages, gender, and driving skill categories. While conducting our experiments, we paid particular attention to the following issues:

i. All drivers were in good mental and physical condition. Drivers did not drive while drowsy or intoxicated.
ii. We did not give drivers any instructions regarding vehicle operation. We left the driving choices up to the drivers' judgment: calm, normal, or aggressive. After the driving, we labeled the style of driving according to the personal evaluations of the drivers.
iii. We did not allow distracting activities such as talking to the person next to them or making a phone call while driving.
iv. Residential and urban driving activities were carried out in full compliance with traffic rules.

The technical contributions of this document are summarized below:

i. Driving data of 610 hours and 38 thousand km was collected with nine professional drivers.
ii. The speed-time information of these drives is recorded simply and practically. We extracted 17 meaningful features solely from speed-time data, demonstrating the potential for robust driving style characterization using readily available information.
iii. Driving data was evaluated using an extensive data processing and formatting methodology.
iv. With k-NN, support vector machines (SVM), decision trees, neural networks (NN), logistic regression, Naive-Bayes, random forest, and light gradient boosting machine (LightGBM) classifiers, high-accuracy models for discrete path types have been developed.
v. These models were compared with each other in many ways and the reasons for preference were explained.
vi. We have provided a clear and accessible framework to address classification problems in the context of driving style prediction. All the steps of how to solve any classification problem can be easily understood if the flow in this article is followed.

In light of the above contributions, we show that we can predict driving style with very high accuracy in a real traffic environment for various driving events such as drivers, driving purposes, driving times, and road types.

DRIVING CYCLES AND DATA MINING

When classifying drivers or driving, as Bouhsissin [33] stated, it is necessary to ask the following research questions and plan with an appropriate methodology:

-What is the purpose of the driving process, and what is its variety, duration, and range? What kind of driving cycle will be used?

-How will the driving style problem be addressed?

-Which features should be selected?

-Which techniques will be used for data preprocessing? Which necessary transformation operations will be performed on the data?

-What types of models will be used and how will these models be optimized?

We have organized our study under subheadings according to the above-mentioned items.

Driving Cycles

By performing a meticulous study to define the driving style, we stored the driving cycles of 9 professional drivers across various roads, using different vehicles, in diverse geographical regions and times, throughout the day, and under varying traffic conditions. To do this, we used a successful and widely used mobile application for speed tracking. This application was run before the vehicle started moving and remained running while the vehicle was moving. The smartphone application determined the location of the vehicle at certain time intervals with the help of the GPS receiver and calculated the distance between two consecutive locations, the time it took to cover this distance and the speed of the vehicle. Even though the vehicles were different during the tests, the mobile application was the same. In addition, by comparing the rides made with different vehicles on the same route, we observed that the same distance was covered in total, and the speed and location changed similarly every second. In this way, we ensured the data quality and security. We saved the speed and time information of the driving cycles that we exported from the application's database to a text file in a relational database. Table 2 shows a summary of these driving cycles.

Table 2. Statistical information on driving cycles

Driving cycles summary	
Region	**Many different regions of Turkey**
Total drivers	9
Total evening drives	789
Total noon drives	524
Total morning drives	619
Total drives in the residential district	977
Total drives in the urban roads	750
Total drives in the motorway	204
Total drives in calm traffic	1881
Total drives in heavy traffic	51
Total count of records	1931
Total duration of all driving	611.4 hour
Total driving distance	37808.3 km.
Collection time for all driving cycles	19 months

Professional drivers who performed these drives, summarized in Table 2, reported that this drive was calm, normal, or aggressive after each drive was finished. However, they received no guidance or instructions on driving before the drive.

Data Mining

In systems developed for predicting driving styles, there may be some difficulties in classifying data into certain categories or classes. There may be missing, noisy, or inconsistent data in the collected data. There may be smooth transitions between the defined classes. The number and distribution of samples may be unbalanced. If the features are not selected correctly, overlaps may occur between classes. We performed detailed data preprocessing to overcome the uncertainties that reduce the classification accuracy and model reliability and cause wrong decisions. We paid attention to the uncertainty and overlap issues when selecting or eliminating features. We tried different classification methods, including ensemble models that combine the results of multiple classifiers.

We completed this study in five stages as shown in Figure 1. First, we analyzed the issue we discussed and identified the necessary information for a solution. To achieve this, we collected driving data in the direction we planned. Then, we extracted meaningful attributes from this data that would provide insights about driving style. We performed basic data preprocessing on this data set. We transformed this data so that we could apply it to classification algorithms and moved on to the application of driving style recognition with machine learning methods.

Problem Understanding

In the first stage, we collected two different types of data during the test drives. The first of these was information including the time of the drive, traffic information, destination, and who the driver was. We collected them in a table called 'drive_info' in our relational database. Secondly, we obtained the driving cycle data, which consists of the speed-time values of this drive. We stored these in another table called 'velocity'.

Feature Extraction

As we mentioned in the introduction section, there are important features that distinguish, for example, aggressive driving style from other driving styles. Aggressive drivers usually drive at higher speeds, make frequent and sudden speed changes, have high acceleration and deceleration intensity, and experience many speed fluctuations during driving. In calm drivers, on the contrary, these values are much smoother and more regular. For this purpose, we determined 13 features, such as maximum, minimum, and average speed values, speed value change, positive and negative acceleration rates, and stop-start rate. Explanations of these features are given in Table 3. We calculated the values of these features using an application we created in the .NET environment using the C# programming language.

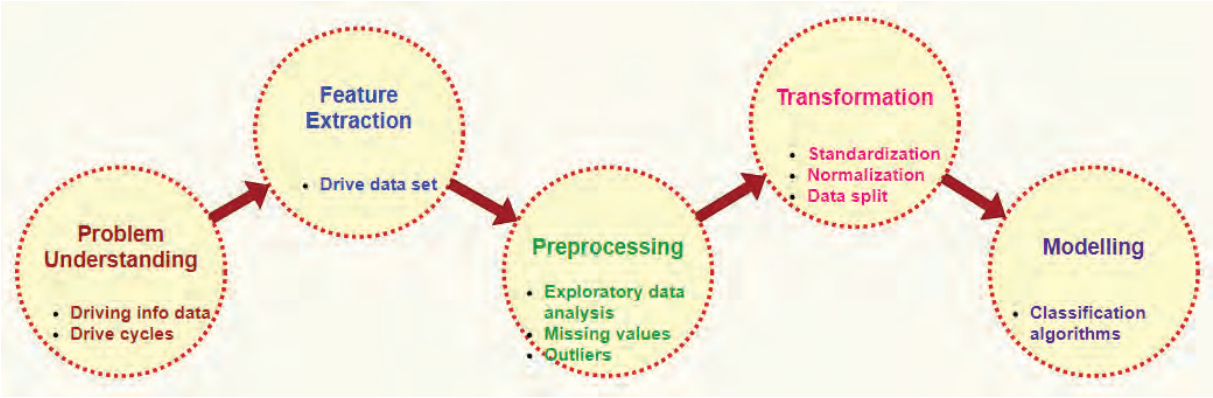


Figure 1. Stages to predict driving style.

Table 3. Attributes identified for classifying driving

Feature	Data type	Description
Max speed [km/h]	Numeric	The maximum speed of the driving
Mean speed [km/h]	Numeric	Average speed of the driving
Standard deviation [km/h]	Numeric	The standard deviation of the speed
Skewness	Numeric	Explains the distribution of velocity values. A right-skewed distribution suggests that outliers are due to high velocities, whereas a left-skewed distribution suggests that outliers are due to low velocities.
Kurtosis	Numeric	Kurtosis shows whether the tails of the normal distribution curve of velocity values are heavy or light.
Positive slop avg [m/s2]	Numeric	Mean value of positive accelerations in driving
Negative slop avg [m/s2]	Numeric	Mean value of negative accelerations in driving
Max point ratio [%]	Numeric	Ratio of peak speed points to overall time
Min point ratio [%]	Numeric	Ratio of lowest speed points to overall time
Hard acc ratio [%]	Numeric	Rate of rapid acceleration
Hard dec ratio [%]	Numeric	Rate of rapid deceleration
Overspeed [%]	Numeric	Rate of exceeding specified speeds for different road types (eg 120 km/h for motorway)
Stop and go ratio [%]	Numeric	Ratio of stop-start time to total time
Mean band ratio [%]	Numeric	Proportion of time spent traveling at average speed to total time
Driving style	Categorical	Driving style assigned at the end of the driving

There is no need to explain how some of the attributes (max, mean) in Table 3 are calculated, as the calculations of the others are given in Equations 1-12. Figure 2 illustrates how some features are calculated through visual figures.

$$Sample\ Standard\ Deviation = s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad (1)$$

$$Skewness = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1).s^3} \quad (2)$$

$$Kurtosis = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1).s^4} \quad (3)$$

In Equation 1-3, s is the sample standard deviation, $\bar{x}$ is the sample mean, n is the total number of sample elements.

$$Positive\ Slop\ Avg = \frac{\sum_{i=0}^{m} f(i+1) - f(i)}{m}, \quad f(i+1) > f(i) \quad (4)$$

$$Negative\ Slop\ Avg = \frac{\sum_{i=0}^{n} f(i+1) - f(i)}{n}, \quad f(i+1) < f(i) \quad (5)$$

In Equations 4 and 5, *f(i)* is the velocity of the driving at the time *i*, *m* is the positive acceleration number, and *n* is the negative acceleration number.

$$Max\ Point\ Ratio = 100 \times \frac{Number\ of\ \max point}{Total\ time} \quad (6)$$

$$Min\ Point\ Ratio = 100 \times \frac{Number\ of\ \min point}{Total\ time} \quad (7)$$

In Figure 2, the maximum points are shown in red circles and the minimum points are shown in green circles.

$$Hard\ Acc\ Ratio = 100 \times \frac{Number\ of positive\ slops}{Total\ time}, \quad positive\ slop > 0.5\ m/s^2 \quad (8)$$

$$Hard\ Dec\ Ratio = 100 \times \frac{Number\ of\ negative\ slops}{Total\ time}, \quad negative\ slop < -0.5\ m/s^2 \quad (9)$$

The ±0.5 m/s² value used as a parameter while calculating the hard acceleration and deceleration rates was determined as a result of the tests performed while driving.

$$Overspeed = 100 \times \frac{Number\ of\ speeds}{Total\ time}, \quad speed > 120\ km/h \quad (10)$$

In this feature, which is valid for highway driving, the ratio of the time traveled at speeds above 120 km/h to the total time is calculated.

$$Stop\ and\ Go\ Ratio = 100 \times \frac{Number\ of\ \text{zero}\ point}{Total\ time} \quad (11)$$

This parameter, which is defined mostly to determine the traffic situation, is calculated from the ratio of the zero points shown with orange dots in Figure 2 to the total time.

$$Mean\ Band\ Ratio = 100 \times \frac{Time\ spent\ on\ the\ mean\ band}{Total\ time} \quad (12)$$

While calculating the average speed band, speeds that are 20% above and 20% below the mean speed were considered. The value of 20 was determined in consultation with the experts on the subject. Figure 2 illustrates the mean speed band.

By calculating the attribute values shown in Table 3, we generated distinct data sets for the three road types. Tables 4-6 show the summary of the datasets we created for the residential district, urban road, and motorway, respectively.

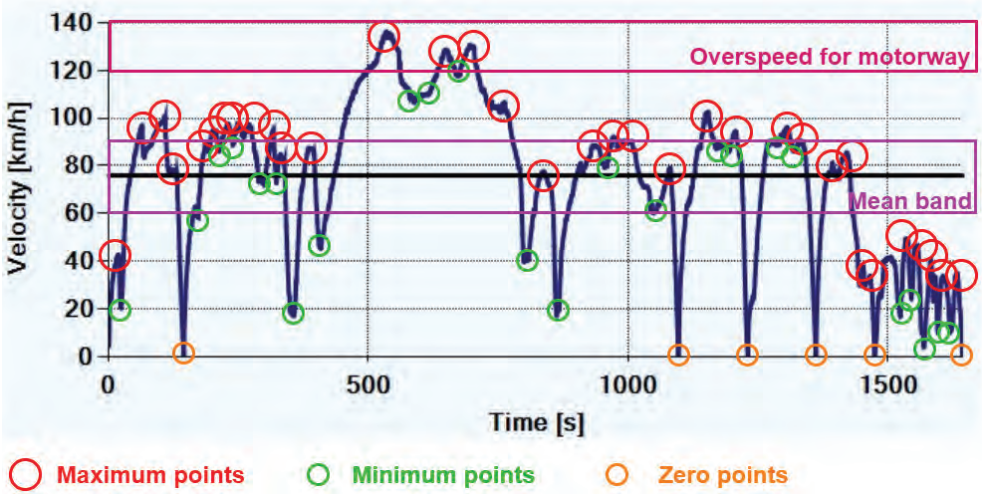


Figure 2. The maximum-minimum points and stop-start points used in the calculation of some features are shown in a sample velocity-time graph.

Table 4. Residential district dataset summary

	Max speed	Mean speed	Standard deviation	Skewness	Kurtosis	Positive slop avg	Negative slop avg	Max point ratio	Min point ratio	Hard acc ratio	Hard dec ratio	Overspeed	Stop and go ratio	Mean band ratio
Count	977	977	977	977	977	977	977	977	977	977	977	977	977	977
Mean	59.95	34.54	13.54	-0.48	2.62	0.41	-0.41	7.66	7.58	2.08	2.01	0	0.84	35.87
Std dev	8.18	4.28	2.01	0.30	0.44	0.08	0.09	1.57	1.59	0.79	0.78	0	1.24	8.90
Min	34.20	11.40	6.38	-1.43	1.45	0.22	-0.78	3.60	3.24	0.37	0	0	0	8.20
25%	53.60	31.70	12.21	-0.68	2.33	0.36	-0.45	6.48	6.37	1.53	1.50	0	0.45	29.80
50%	60.00	34.70	13.44	-0.51	2.56	0.40	-0.39	7.44	7.38	2.02	1.94	0	0.72	35.40
75%	65.20	37.40	14.70	-0.33	2.85	0.45	-0.35	8.69	8.64	2.50	2.46	0	1.05	41.40
Max	87.80	50.20	28.24	1.45	5.20	0.86	-0.21	12.81	13.30	5.91	5.01	0	35.06	69.40

Table 5. Urban road dataset summary

	Max speed	Mean speed	Standard deviation	Skewness	Kurtosis	Positive slop avg	Negative slop avg	Max point ratio	Min point ratio	Hard acc ratio	Hard dec ratio	Overspeed	Stop and go ratio	Mean band ratio
Count	750	750	750	750	750	750	750	750	750	750	750	750	750	750
Mean	86.68	43.69	22.72	-0.04	2.18	0.47	-0.48	6.21	6.14	2.17	2.26	0	1.09	22.76
Std dev	19.19	12.28	6.19	0.48	0.69	0.11	0.10	1.48	1.49	1.01	0.91	0	1.05	10.49
Min	38.20	14.70	7.94	-1.82	1.32	0.20	-1.18	2.86	2.48	0.09	0	0	0	3.30
25%	72.93	32.82	17.99	-0.34	1.73	0.39	-0.54	5.14	5.09	1.49	1.63	0	0.62	15.03
50%	90.85	44.70	24.04	-0.05	1.96	0.45	-0.48	6.02	5.96	1.98	2.17	0	0.86	20.6
75%	100.9	53.48	27.43	0.25	2.41	0.53	-0.42	7.11	6.97	2.65	2.75	0	1.31	28.1
Max	137.1	74.70	41.53	1.96	7.34	1.06	-0.23	12.77	12.77	6.97	6.36	9	21.17	80.6

Table 6. Motorway dataset summary

	Max speed	Mean speed	Standard deviation	Skewness	Kurtosis	Positive slop avg	Negative slop avg	Max point ratio	Min point ratio	Hard acc ratio	Hard dec ratio	Overspeed	Stop and go ratio	Mean band ratio
Count	204	204	204	204	204	204	204	204	204	204	204	204	204	204
Mean	132.45	76.85	30.84	-0.67	3.34	0.37	-0.41	4.72	4.71	1.23	1.34	10	0.48	37.73
Std Dev	24.73	16.53	6.80	0.75	2.43	0.08	0.09	1.80	1.80	0.69	0.65	13	0.32	20.86
Min	91.30	48.40	12.18	-3.27	1.56	0.11	-0.66	0.91	0.91	0	0	0	0	5.00
25%	112.70	63.75	26.43	-1.20	1.94	0.33	-0.46	3.55	3.53	0.76	0.85	0	0.25	22.50
50%	133.70	75.25	30.24	-0.39	2.30	0.38	-0.40	4.43	4.44	1.19	1.42	4	0.43	30.35
75%	151.28	89.78	36.06	-0.10	3.67	0.43	-0.35	5.96	5.96	1.68	1.77	15	0.66	46.47
Max	200.6	116.4	47.51	0.58	18.62	0.58	-0.18	9.42	9.38	3.31	3.12	57	2.39	94.50

Preprocessing

Before applying machine learning techniques to a data, it is necessary to ensure that the data is reliable. In this process, features that show high correlation or do not contribute to the performance of the model are removed from the data set. In addition, a data set may have missing, inconsistent, outlier or noisy data. Data preprocessing is performed to complete missing data, eliminate inconsistencies, discard outlier data, delete incorrect data if any, and clean noisy data [34].

As described in the previous section, after extracting the features, we performed exploratory data analysis (EDA) on our dataset. In this context, we first removed the overspeed feature, which is always 0 in the residential district data set. Then, by performing correlation analysis on all three data sets, we identified the features that were correlated with each other. Figures 3-5 show the correlation analysis results of the three data sets.

In correlation analysis, the pairwise relationships among all features are evaluated. The correlation coefficient between two variables ranges from -1 to 1. A value closer to -1 or 1 signifies a stronger relationship. Positive values indicate a direct relationship, whereas negative values indicate an inverse relationship. In the correlation analysis, we eliminated all but one of the features with a higher or lower correlation value than ±0.8, which we determined to be the

Residential District Dataset	Max Speed	Mean Speed	Standard Deviation	Skewness	Kurtosis	Positive Slop Avg	Negative Slop Avg	Max Point Ratio	Min Point Ratio	Hard Acc Ratio	Hard Dec Ratio	Stop and Go Ratio	Mean Band Ratio
Max Speed	1	0.70	0.73	0.32	0.05	0.20	-0.36	-0.45	-0.45	0.07	0.14	-0.20	0.19
Mean Speed	0.70	1	0.44	-0.32	0.25	-0.08	-0.07	-0.37	-0.38	-0.19	-0.13	-0.36	0.43
Standard Deviation	0.73	0.44	1	0.31	-0.44	0.32	-0.43	-0.47	-0.46	0.16	0.22	0.00	-0.38
Skewness	0.32	-0.32	0.31	1	-0.37	0.36	-0.39	-0.12	-0.12	0.32	0.36	0.17	-0.32
Kurtosis	0.05	0.25	-0.44	-0.37	1	-0.28	0.25	0.11	0.10	-0.25	-0.25	-0.14	0.84
Positive Slop Avg	0.20	-0.08	0.32	0.36	-0.28	1	-0.72	-0.01	-0.01	0.80	0.65	0.16	-0.31
Negative Slop Avg	-0.36	-0.07	-0.43	-0.39	0.25	-0.72	1	0.22	0.23	-0.58	-0.80	-0.07	0.24
Max Point Ratio	-0.45	-0.37	-0.47	-0.12	0.11	-0.01	0.22	1	1	0.34	0.19	0.08	0.02
Min Point Ratio	-0.45	-0.38	-0.46	-0.12	0.10	-0.01	0.23	1	1	0.34	0.19	0.09	0.02
Hard Acc Ratio	0.07	-0.19	0.16	0.32	-0.25	0.80	-0.58	0.34	0.34	1	0.71	0.14	-0.29
Hard Dec Ratio	0.14	-0.13	0.22	0.36	-0.25	0.65	-0.80	0.19	0.19	0.71	1	0.10	-0.28
Stop and Go Ratio	-0.20	-0.36	0.00	0.17	-0.14	0.16	-0.07	0.08	0.09	0.14	0.10	1	-0.29
Mean Band Ratio	0.19	0.43	-0.38	-0.32	0.84	-0.31	0.24	0.02	0.02	-0.29	-0.28	-0.29	1

Figure 3. Correlation analysis results in the residential district data set.

Urban Road Dataset	Max Speed	Mean Speed	Standard Deviation	Skewness	Kurtosis	Positive Slop Avg	Negative Slop Avg	Max Point Ratio	Min Point Ratio	Hard Acc Ratio	Hard Dec Ratio	Overspeed	Stop and Go Ratio	Mean Band Ratio
Max Speed	1	0.83	0.90	-0.14	-0.28	-0.02	-0.14	-0.42	-0.41	-0.15	-0.14	0.21	-0.15	-0.35
Mean Speed	0.83	1	0.75	-0.61	-0.28	-0.22	0.04	-0.57	-0.56	-0.40	-0.36	0.20	-0.33	-0.12
Standard Deviation	0.90	0.75	1	-0.11	-0.55	0.06	-0.20	-0.41	-0.40	-0.08	-0.09	0.18	-0.05	-0.61
Skewness	-0.14	-0.61	-0.11	1	0.11	0.30	-0.19	0.34	0.34	0.40	0.34	-0.03	0.28	-0.23
Kurtosis	-0.28	-0.28	-0.55	0.11	1	-0.09	0.16	0.22	0.22	0.00	0.01	-0.03	0.04	0.74
Positive Slop Avg	-0.02	-0.22	0.06	0.30	-0.09	1	-0.70	0.25	0.24	0.78	0.65	0.06	0.21	-0.22
Negative Slop Avg	-0.14	0.04	-0.20	-0.19	0.16	-0.70	1	-0.10	-0.08	-0.60	-0.74	-0.06	-0.15	0.22
Max Point Ratio	-0.42	-0.57	-0.41	0.34	0.22	0.25	-0.10	1	1	0.63	0.60	-0.06	0.13	0.08
Min Point Ratio	-0.41	-0.56	-0.40	0.34	0.22	0.24	-0.08	1	1	0.63	0.59	-0.06	0.13	0.07
Hard Acc Ratio	-0.15	-0.40	-0.08	0.40	0.00	0.78	-0.60	0.63	0.63	1	0.81	0.02	0.22	-0.18
Hard Dec Ratio	-0.14	-0.36	-0.09	0.34	0.01	0.65	-0.74	0.60	0.59	0.81	1	0	0.20	-0.12
Overspeed	0.21	0.20	0.18	-0.03	-0.03	0.06	-0.06	-0.06	-0.06	0.02	0	1	-0.04	0.01
Stop and Go Ratio	-0.15	-0.33	-0.05	0.28	0.04	0.21	-0.15	0.13	0.13	0.22	0.20	-0.04	1	-0.16
Mean Band Ratio	-0.35	-0.12	-0.61	-0.23	0.74	-0.22	0.22	0.08	0.07	-0.18	-0.12	0.01	-0.16	1

Figure 4. Correlation analysis results in the urban road data set.

Motorway Dataset	Max Speed	Mean Speed	Standard Deviation	Skewness	Kurtosis	Positive Slop Avg	Negative Slop Avg	Max Point Ratio	Min Point Ratio	Hard Acc Ratio	Hard Dec Ratio	Overspeed	Stop and Go Ratio	Mean Band Ratio
Max Speed	1	0.64	0.62	0.00	0.09	0.12	-0.34	-0.46	-0.46	-0.26	-0.16	0.72	-0.29	0.14
Mean Speed	0.64	1	0.04	-0.71	0.62	-0.44	0.25	-0.65	-0.65	-0.64	-0.66	0.69	-0.66	0.69
Standard Deviation	0.62	0.04	1	0.54	-0.58	0.64	-0.68	0.03	0.03	0.32	0.38	0.50	0.26	-0.59
Skewness	0.00	-0.71	0.54	1	-0.85	0.69	-0.63	0.45	0.45	0.62	0.72	-0.15	0.55	-0.85
Kurtosis	0.09	0.62	-0.58	-0.85	1	-0.70	0.57	-0.49	-0.49	-0.60	-0.66	0.10	-0.53	0.90
Positive Slop Avg	0.12	-0.44	0.64	0.69	-0.70	1	-0.80	0.49	0.48	0.81	0.81	0.01	0.63	-0.72
Negative Slop Avg	-0.34	0.25	-0.68	-0.63	0.57	-0.80	1	-0.14	-0.14	-0.51	-0.68	-0.12	-0.42	0.57
Max Point Ratio	-0.46	-0.65	0.03	0.45	-0.49	0.49	-0.14	1	1	0.83	0.76	-0.34	0.64	-0.54
Min Point Ratio	-0.46	-0.65	0.03	0.45	-0.49	0.48	-0.14	1	1	0.83	0.76	-0.34	0.64	-0.54
Hard Acc Ratio	-0.26	-0.64	0.32	0.62	-0.60	0.81	-0.51	0.83	0.83	1	0.91	-0.24	0.76	-0.67
Hard Dec Ratio	-0.16	-0.66	0.38	0.72	-0.66	0.81	-0.68	0.76	0.76	0.91	1	0	0.71	-0.71
Overspeed	0.72	0.69	0.50	-0.15	0.10	0.01	-0.12	-0.34	-0.34	-0.24	0	1	-0.28	0.14
Stop and Go Ratio	-0.29	-0.66	0.26	0.55	-0.53	0.63	-0.42	0.64	0.64	0.76	0.71	-0.28	1	-0.63
Mean Band Ratio	0.14	0.69	-0.59	-0.85	0.90	-0.72	0.57	-0.54	-0.54	-0.67	-0.71	0.14	-0.63	1

Figure 5. Correlation analysis results in the motorway data set.

Residential District Dataset	Urban Road Dataset	Motorway Dataset
Max Speed	Max Speed	Max Speed
Mean Speed	Skewness	Mean Speed
Standard Deviation	Kurtosis	Standard Deviation
Skewness	Positive Slop Avg	Positive Slop Avg
Min Point Ratio	Negative Slop Avg	Negative Slop Avg
Hard Acc Ratio	Min Point Ratio	Max Point Ratio
Hard Dec Ratio	Hard Dec Ratio	Overspeed
Stop and Go Ratio	Overspeed	Stop and Go Ratio
Mean Band Ratio	Stop and Go Ratio	Mean Band Ratio
	Mean Band Ratio	

Figure 6. List of features for three datasets.

threshold value. In this case, the features of our three datasets are shown in Figure 6.

One of the operations we performed during the data preprocessing phase was the control of missing data. However, no such imperfect data was found in our three data sets. Then, we checked for outliers. We used the local outlier factor (LOF) method because we aimed to reduce skewness. In the outlier analysis we performed by looking at the 20 neighbors around each point, we removed the data points that were above the threshold value we determined. By removing outliers, we ensured that our models did not suffer from over-learning. The results of the outlier analysis for the three data sets are shown in Figures 7-9.

In Figures 7-9, black dots show data points, red circles show the calculated outlier value for each point, and blue dots show samples with outlier scores exceeding the specified threshold value. At the end of this process, we removed 10 data points from the residential district data set and 7 data points from the urban road data set. No outliers were found in the motorway data set. After removing outliers, our sample counts were 967 for the residential district, 743 for the urban road, and 204 for the motorway.

Transformation

Transformation techniques are used to analyze raw data in a machine learning project and make it suitable for modeling. Methods such as scaling, standardization, normalization, and dimensionality reduction are among the common transformation techniques. After the transformation process, the performance of the model is expected to increase, especially linear models produce better results for scaled data. The computational cost is reduced with dimensionality reduction techniques. The effect of feature values with different units of measurement on the model is balanced. Transformation techniques also contribute visually and provide better interpretation of the obtained results.

In this section, we apply the standardization process after separating the data set as training and test data. Splitting the dataset will allow comparison of the accuracy values in the training and test data while finding the best parameters in the classification algorithms. In all three datasets, we split the dataset into 70% training

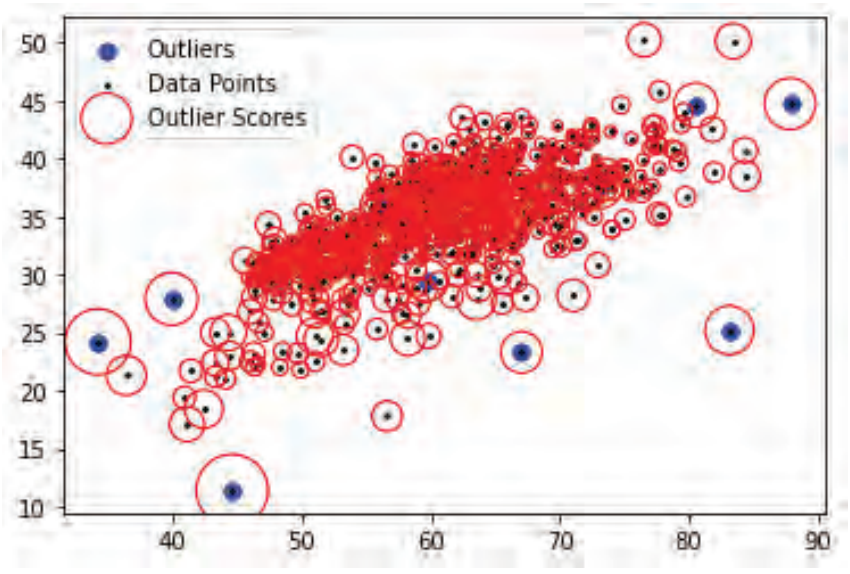


Figure 7. Outlier analysis for residential district.

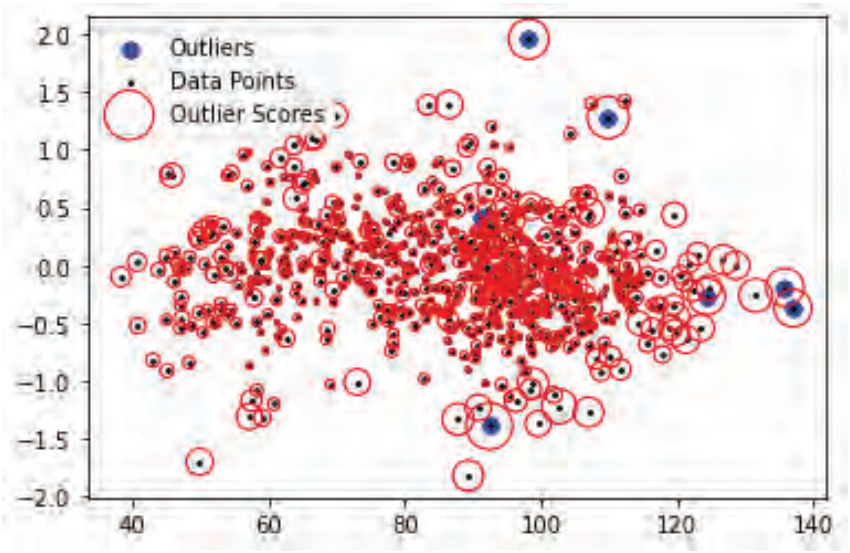


Figure 8. Outlier analysis for urban road.

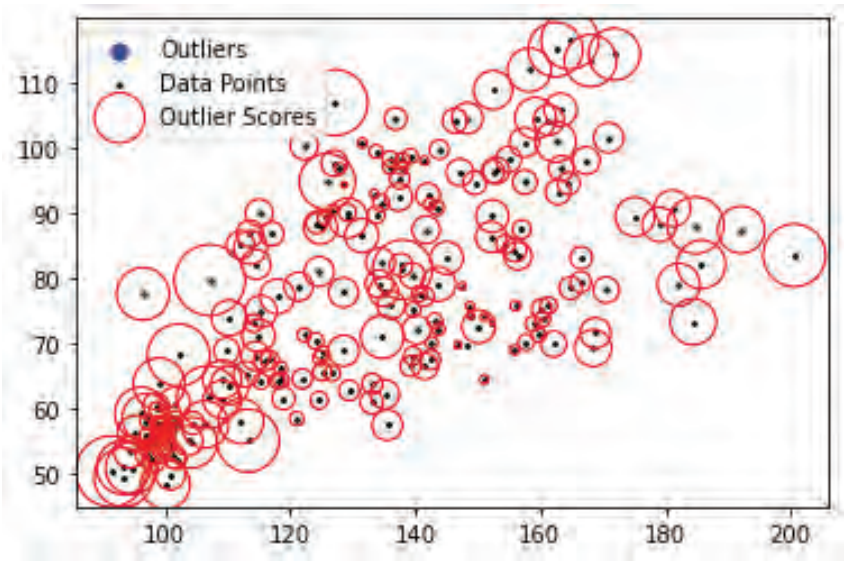


Figure 9. Outlier analysis for motorway.

data and 30% test data. Then we performed the standardization process to eliminate the scale difference between the features. With the standardization process shown in Equation 13, we ensure that the mean of the data of a feature is 0 and the standard deviation is 1. This process makes a significant contribution to the success of models that use distance-based calculations in particular.

$$z = \frac{x - \mu}{\sigma} \tag{13}$$

Here x is the variable value, μ is the mean value and σ is the standard deviation.

Modelling

After the pre-processing phase of the data is completed, the next step is to select an appropriate algorithm for the problem and run the machine learning process. Figure 10 shows the separation of the data into training and test data sets, training the model with training data, determining the hyperparameters that will give the best results in this model, and testing the model with these parameters with the test data.

We used the GridSearchCV [35] algorithm to find the best parameters of a model. While using this algorithm, as shown in Figure 11, we divided all the data into 10 pieces and used a different part as a test and the others as training

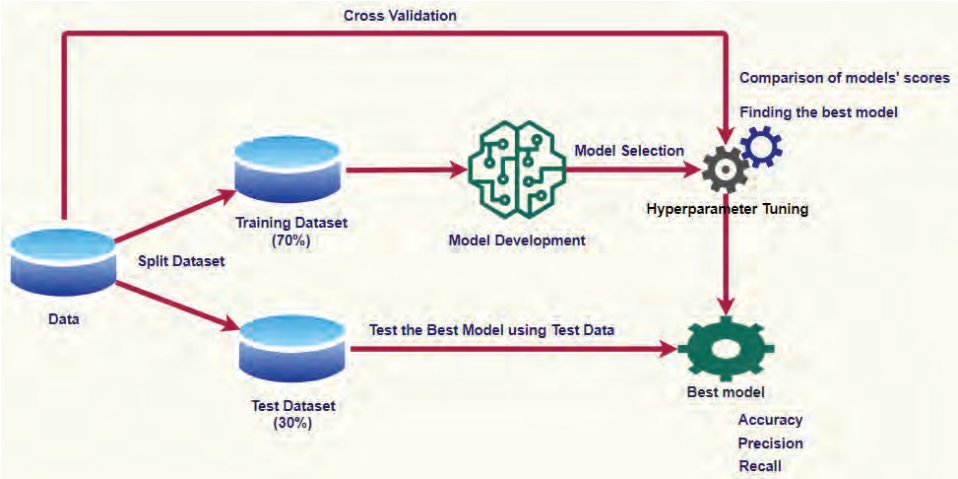


Figure 10. Finding the best model and testing it with test data.

data in each iteration. We determined the parameters that gave the highest accuracy value by trying the parameter to be adjusted for each classification model in turn.

On the one hand, the best parameters are found with GridSearchCV; on the other hand, it is important to ensure that the accuracy rates achieved in the training and test data sets are close to each other so that the model with these parameters does not overfit or underfit. As shown in Figure 12, if the model performs very well in the training dataset but not in the test dataset, there is an overfitting problem. That is, the variance of the model is high, it has memorized the training data and cannot generalize itself for new data. If the results are bad in both datasets, the model was over simplistic, produced high biases, and was not trained enough to generalize. When deciding on the best parameters of the model, it is necessary to find the best trade-off between variance and bias, as shown in Figure 13.

The model complexity shown in Figure 13 expresses the learning ability of a model and its fit to the data. Simple models have a high bias value because they do not learn well and cannot adapt to the data. Complex models, on the other hand, fit the data much better and react to even the

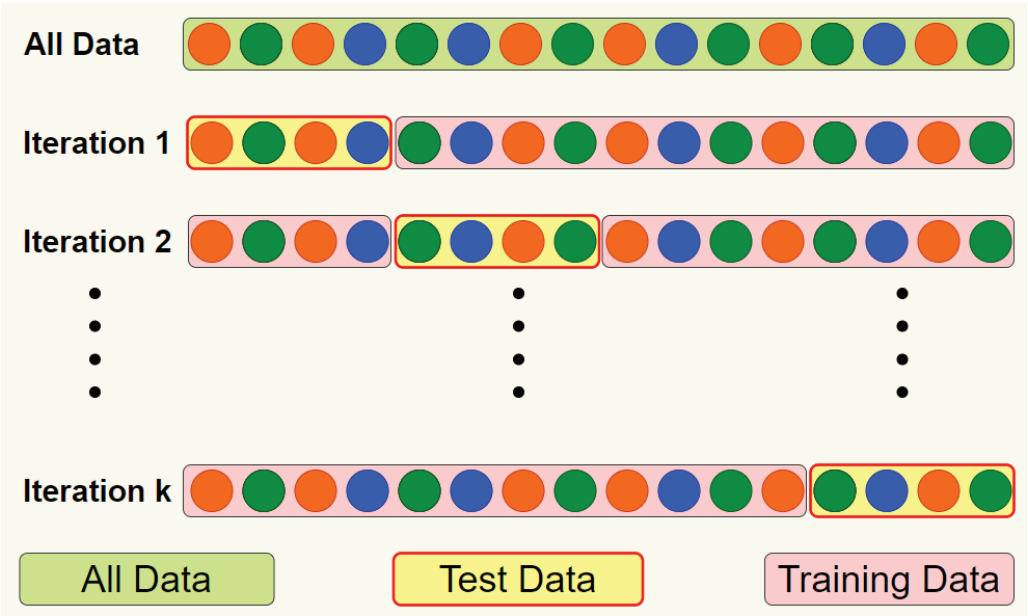


Figure 11. In the cross-validation method, all data is divided into k folds, each time a different fold is separated as test data, while the other folds are used as training data. The success of the model is calculated by taking the average of the result obtained in all iterations.

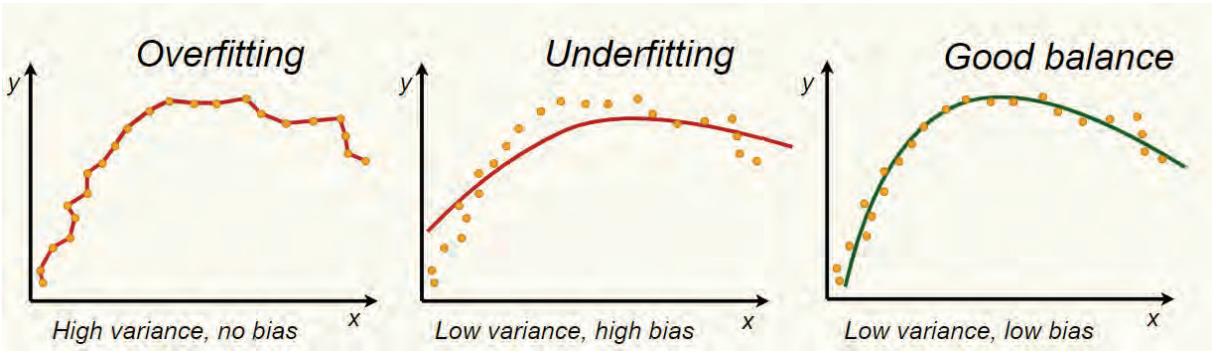


Figure 12. There is a problem of generalization in case of overfitting and inability to learn in case of underfitting. When the variance and biases are reduced, the ideal learning called good balance is reached.

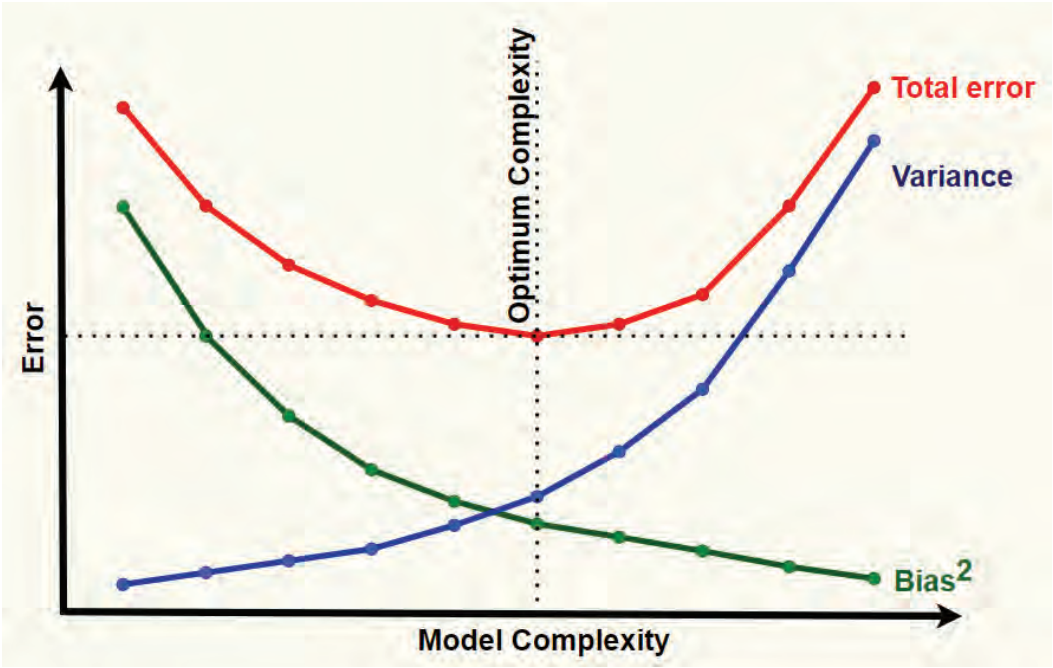


Figure 13. Optimal balance is achieved by making a trade-off between variance and bias when finding the optimal parameters of a model.

smallest changes in the data, which leads to poor performance on the test data. For optimum complexity, it is necessary to establish this balance and trade-off well.

Classification Algorithms

In situations where a clear link between outputs and inputs is not evident, machine learning algorithms employ a range of statistical, probabilistic, and optimization techniques to provide solutions by learning from past experiences, to draw meaningful conclusions from unstructured complex data stacks, and to detect patterns in large data sets [36]. These algorithms, categorized as supervised, unsupervised, and semi-supervised, have been successfully applied in areas such as intrusion detection [37], e-mail filtering [38], customer purchasing behavior detection [39], and manufacturing process [40]. In supervised learning, the model is trained to establish a relationship between the input set and the output in a labeled data set, and the most appropriate outputs are produced for the new data. In this process, input and output data are matched so that the model can make accurate predictions. Supervised learning is used especially in tasks such as classification and regression. Classification algorithms are used to separate data into specific categories, and these algorithms are used in various applications, such as determining whether an e-mail is spam or diagnosing a disease. In this study, which determines the driving style of a driver, supervised machine learning algorithms were used for the classification task.

When selecting machine learning models, it is necessary to pay attention to the type of data in the data set, the size of the data set, the complexity of the model to be designed, the generalization capabilities of the models, the resource consumption of the models, that is, the computational costs, and the interpretability of the outputs they produce. The fact that our data set consists of only numerical values, the size of our data set is not very large, and the computers on which the models work have large resources did not cause any restrictions in terms of computational costs. For this reason, it was possible to conduct experiments with many models of different complexities. The drive style was predicted with k-NN, support vector machines, decision trees, artificial neural networks, logistic regression, naive Bayes, random forest and LightGBM algorithms.

K-NN Classifier

The k Nearest Neighbor algorithm is a machine learning method that is straightforward, easy to grasp, and simple to implement [41]. Nearest neighbor methods find previously defined training examples at the closest distance to the new point and estimate the class of the new point accordingly. The k parameter specifies the number of nearest neighbors to be considered when classifying test data [42]. The distance calculation is determined by methods such as Manhattan, Euclidean, Minkowski as shown in Equation 14.

$$d(x,y) = \left(\sum_{i=1}^{n}(|x_i - y_i|)^q\right)^{\frac{1}{q}} \tag{14}$$

In this equation, referred to as the Minkowski method, setting (q = 1) results in the Manhattan distance, while setting (q = 2) uses the Euclidean distance for calculations.

Neighbor-based methods are considered non-generalizing machine learning methods because they can simply evaluate all the training data. Despite its simplicity, k-NN has been successful in many classification problems. Since it does not require parameters, it is often effective in classification scenarios where the decision boundary is highly irregular. It stores training data samples but does not create a model with that data. In classification using this method, the class of each point is determined by simple majority voting of its nearest neighbors, as shown in Figure 14.

Since k-NN, being one of the lazy algorithms, does not create any model, the training phase is costless and does not take time. However, the testing phase requires more time and memory. If the data set contains a large amount of data, scanning and storing all of them may strain the computer's memory usage. The success rate decreases in data sets with very high dimensionality and uneven distribution. To get better results with k-NN, outliers in the data set should be cleaned, and scaling should be done.

Support Vector Machines Classifier

Support vector machines (SVM) are a technique devised by Vapnik, grounded in statistical learning theory [43]. SVM is capable of classifying both linear and non-linear datasets. In this approach, the input space is transformed into a multidimensional inner product space known as the feature space, where the optimal planes are identified to enhance the classifier's generalization capability. Optimization theory and related statistical learning principles are utilized to identify the optimal planes. The marginal distance for a class refers to the gap between the decision hyperplane and the closest instance belonging to that class. To carry out the classification, we must identify the hyperplane that maximally separates the two classes, as illustrated in Figure 15. SVM offers several advantages: it can handle a large

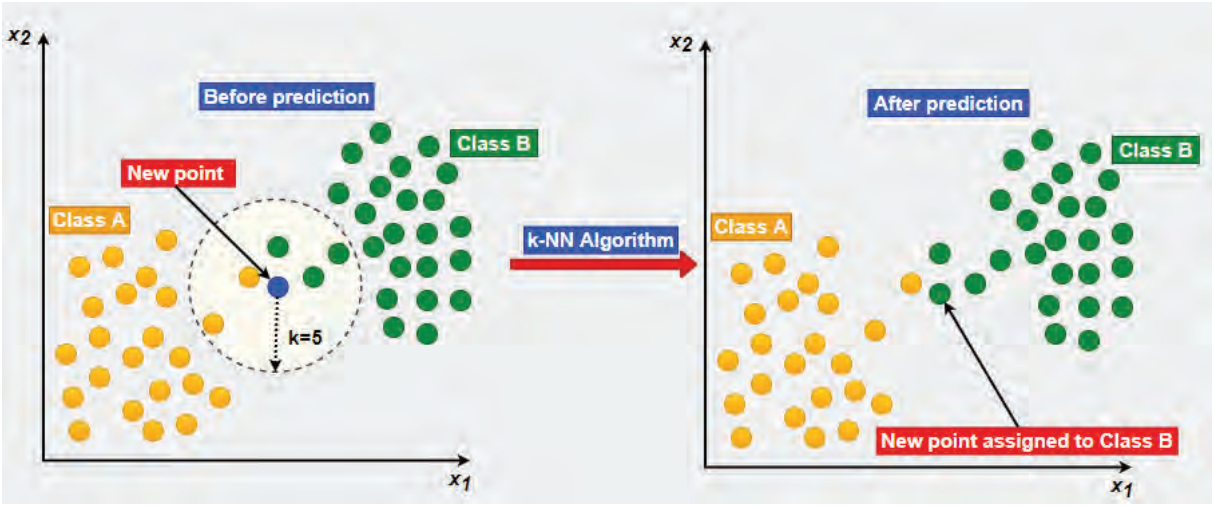


Figure 14. Assigning a new sample shown in blue to the green class with the k-NN algorithm with the nearest neighbor parameter k=5.

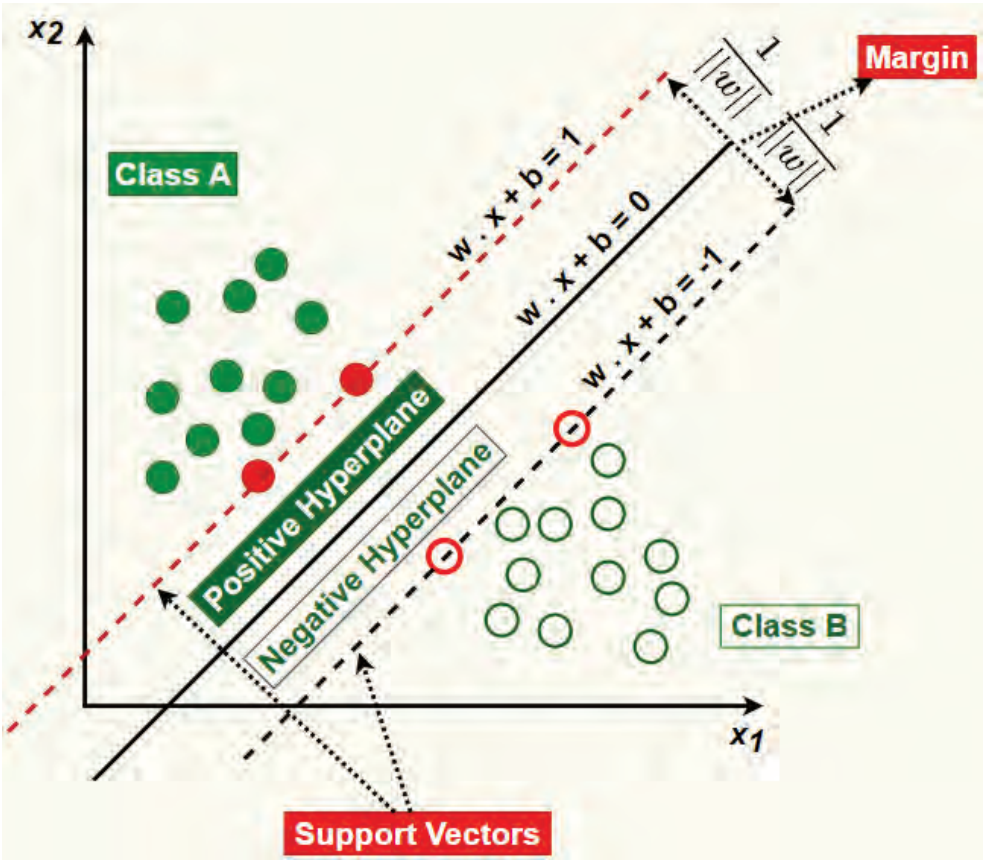


Figure 15. A hyperplane is determined by the SVM classifier that provides the furthest distance between the two classes.

number of independent variables, requires minimal input for learning, applies to both linearly separable and non-separable data, and delivers high accuracy results.

Let $y_i \in \{-1,+1\}$ be the class label corresponding to N data vectors, each of which consists of a data pair in n-dimensional real space. In this instance, x_i belongs to the first class if y_i = -1 and the second class if y_i = +1. Equations 15,16 provide a hyperplane of f(x) that separates linearly separable data, where f(x) > 0 if x_i belongs to the positive class and f(x)<0 if x_i belongs to the negative class. This hyperplane can be characterized as:

$$f(x) = w.x + b = \sum_{j=1}^{n} w_i.x_i + b \tag{15}$$

$$y_i f(x_i) = y_i(w.x + b) \geq 0 \quad i = 1,2,\dots N \tag{16}$$

Here, b is the threshold, and w is an n-dimensional vector. Furthermore, Equation 17 can be expressed as follows if w and x are the points that are closest to the extreme plane at a distance of 1/w:

$$y_i(w.x + b) \geq 1, \qquad i = 1,2,\dots N \tag{17}$$

The ideal parser extreme plane is the plane farthest from the extreme plane to the closest point. Maximizing this parser plane gives the SVM its capacity for generalization.

Decision Tree Classifier

Decision trees are one of the oldest and leading machine learning algorithms. The structure of this technique is based on the hierarchical decomposition of data [44]. The decision tree as a classifier was introduced by D. Morgan [45] and developed by JR Quinlan [46]. A decision tree models a decision logic over previously realized data. All internal nodes in the decision tree represent input variables or tests on the relevant attribute. Depending on the test result, the classification algorithm continues the test and branching process until it reaches a leaf node [46]. The extreme nodes, called leaves, indicate the decision of the model. The decision tree is a white-box model, offering a complete explanation of why a sample belongs to a particular class rather than another. Additionally, this method allows us to identify which feature is most significant for the output. This algorithm creates a decision-making process based on information entropy. Entropy shows the likelihood and uncertainty of an unexpected event. If all samples are identical, the entropy is 0. If the values are evenly distributed,

the entropy is 1. E(S) represents the information entropy of the dataset S. It is calculated as in Equation 18;

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \tag{18}$$

Here, p represents the probability of success. The entropy for all input attributes is computed as shown in Equation 19;

$$E(T,x) = \sum_{c \in X} P(c)E(c) \tag{19}$$

Where P(c) is the probability of the potential data point in x, E(c) is its value, T is the output attribute, and x is the input attribute. Figure 16 shows the classification of the driving style with the decision tree according to the attributes extracted from the drive cycle in a residential area.

Neural Network Classifier

To carry out a particular activity or function, machine learning algorithms known as artificial neural networks (ANNs) are inspired by the human brain. Warren McCulloch and Walter Pitts laid the foundations of this algorithm by describing a simple mathematical model for a neuron that receives inputs, processes these inputs, and

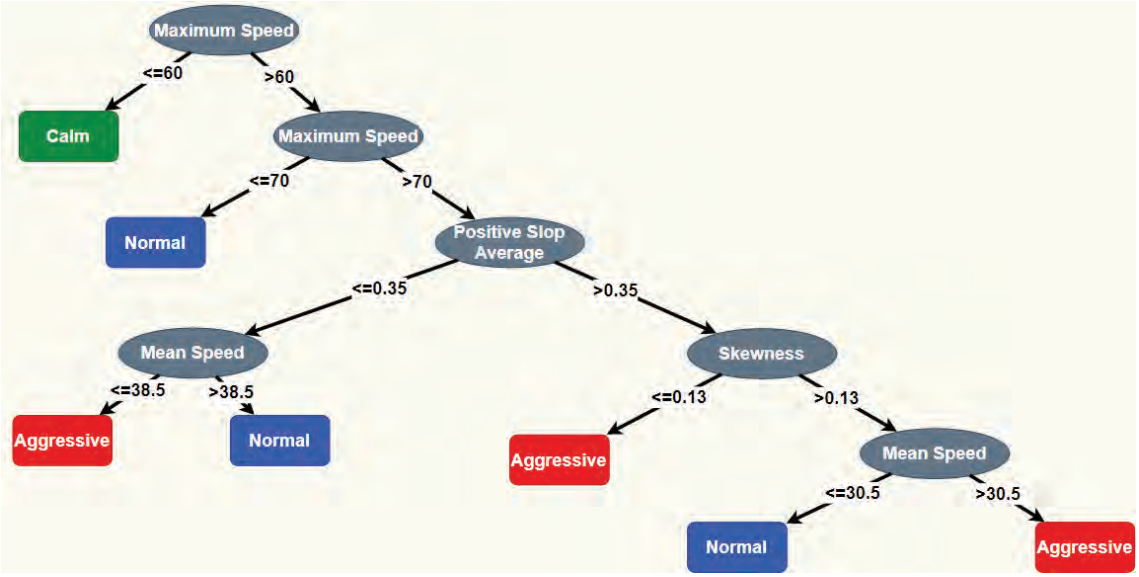


Figure 16. Classification of driving style with decision tree based on features extracted from drive cycle.

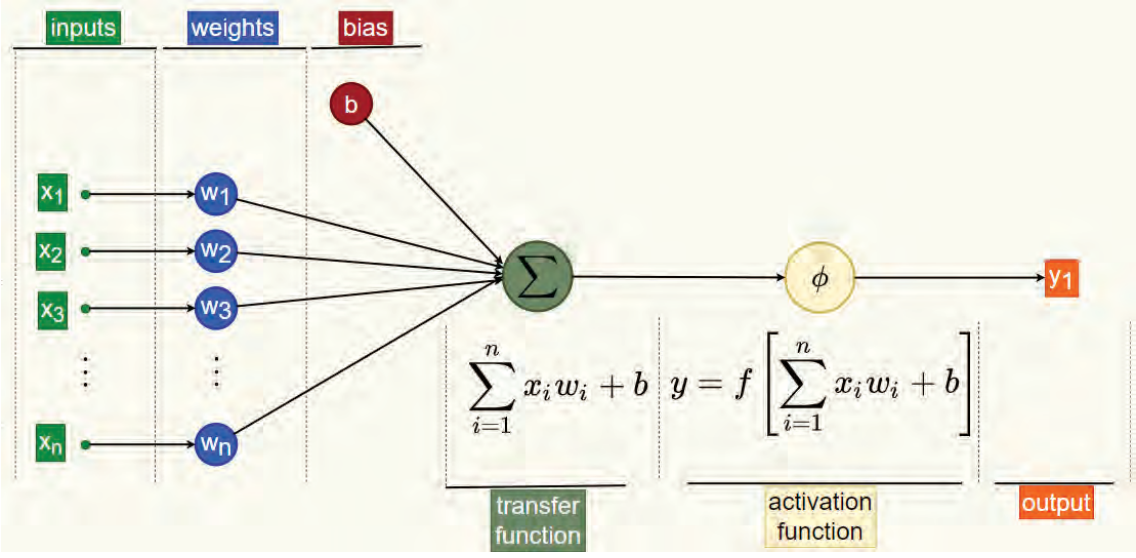


Figure 17. Calculations in a neuron.

returns an output [47]. A collection of input/output units with a weight assigned to each neuron is called a neural network. In order to forecast the right class label of the inputs, the ANN continuously modifies the weights during the learning phase.

In an artificial neural network, the inputs are multiplied by a weighting factor and summed and transferred to a transfer function. Figure 17 shows the input and output parameters, weight values, transfer, and activation functions of an artificial neuron.

The artificial neural model's mathematical expression shown in Figure 17 is shown in Equation 20.

$$y = f\left[\sum_{i=1}^{n} x_i w_i + b\right] \tag{20}$$

Here, x_i represents the inputs, w_i the weights, b the bias, and y the output. f indicates the activation function. The difference between the expected output and the obtained output is called the error. The weighting factors should be changed when the error is greater than the specified threshold. With this process, called backpropagation, the neurons are trained once again for new calculations and the errors are reduced [48]. The first phase of backpropagation is the feed-forward phase, in which the input information at the node is forwarded to calculate the output information. The second stage involves adjusting the connection strengths in response to variations between the output units' estimated and actual values. The main goal here is to reduce the discrepancy between the calculated and actual outputs through repeated processes [49]. This process is called artificial neural network learning.

There are two types of neural networks, feedforward and feedback. A feedforward neural network is not iterative. Neurons in one layer are only connected to neurons in the next layer and do not form a loop, signals only travel towards the output layer. In a feedforward neural network, the neurons are organized in layers and the outputs of the neurons in one layer are given as inputs to the next layer [50]. The mathematical expression of the structure shown in Figure 18 is given in Equation 21. Feedback neural networks, also known as recurrent neural networks, contain loops. Signals move in both directions, forming loops in the network. Feedback loops can cause the behavior of the network to change over time, depending on its input.

$$y_{i,k} = \varphi_{i,k}\left(\sum_{j=1}^{n_{k-1}} w_{j,k-1}^{i,k}\, y_{j,k-1} + b_{i,k}\right) \tag{21}$$

In Figure 18, n is the number of inputs, m is the number of outputs, k is the number of layers, and $h(k)$ is the number of neurons in the k-numbered layer. In Equation 21, $y_{i,k}$ is the output of neuron i in layer k. $n_{(k-1)}$ is the number of neurons in the (k-1) layer. $w_{j,k-1}^{i,k}$ is the connection weight between neuron i in layer k and neuron number j in layer (k-1). $b_{i,k}$ is the residual value for neuron i in layer k. $\varphi_{i,k}$ is the activation process applied to neuron i in layer k. Linear, threshold, sigmoid or tangent hyperbolic functions are used as activation functions.

Logistic Regression Classifier

Despite its name, logistic regression is used more for classification problems than regression in machine learning applications. In the literature, logistic regression is also

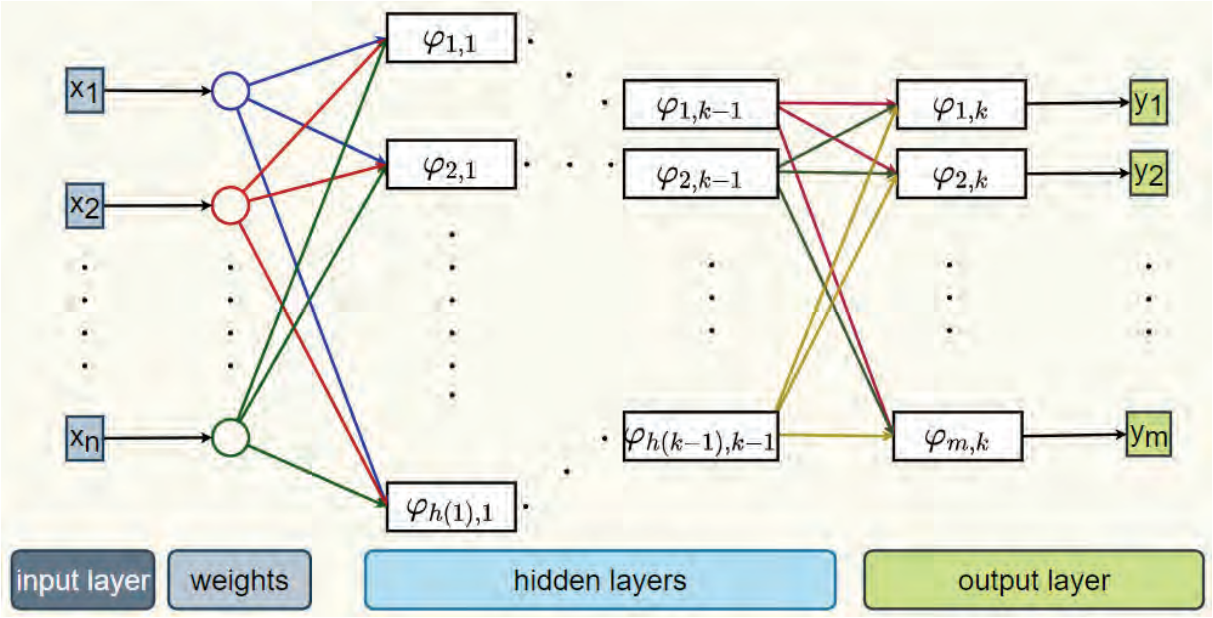


Figure 18. Structure of feedforward neural network.

known as maximum entropy classification or log-linear classifier. In logistic regression, the probability of an event can be statistically modeled using the values of categorical or numerical independent variables [51]. Logistic regression is a statistical model in mathematics that finds the relationship between x and y using a logistic or logit function. The logit function maps y as the sigmoid function of x as in Equation 22 and produces a value between 0 and 1 as shown in Figure 19. In linear regression, we use the equation shown in Equation 23, where each unit change in X influences $p(X)$ by β_1, resulting in a continuous variable. Since logistic regression aims to determine the probability of the dependent variable falling into a specific category, we will modify the left side of the equation in Equation 23 to yield a value between 0 and 1, as shown in Equations 24-26.

$$f(x) = \frac{1}{1+e^{-x}} = \frac{e^{x}}{1+e^{x}} \tag{22}$$

$$p(X) = \beta_0 + \beta_1 X \tag{23}$$

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \tag{24}$$

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 X} \tag{25}$$

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \tag{26}$$

As illustrated in Equation 26, the relationship between X and p(X) is no longer linear. The value of P(X) will be between 0 and 1.

Naive Bayes Classifier

Naive Bayes classifier is a simple, fast, accurate, and reliable supervised learning algorithm. The naive Bayes classifier is based on Bayes' theorem with the assumption of independence, which assumes that a particular feature in a class is independent of other features. Even though these features are dependent on each other, they are considered independent, thus simplifying the calculation. This is where the name Naive comes from. In this algorithm, statistical and probabilistic inferences are made between previous inputs and outputs [52]. The probability of a previous example is used to approximate each particular class.

According to Bayes' rule, the probability of both events A and B occurring is the product of the probability of A and the conditional probability of B given A, as demonstrated in Equation 27.

$$P(A \cap B) = P(A) \times P(B|A) \tag{27}$$

Here, P(A) represents the probability of event A occurring on its own, and P(B|A) denotes the probability of event

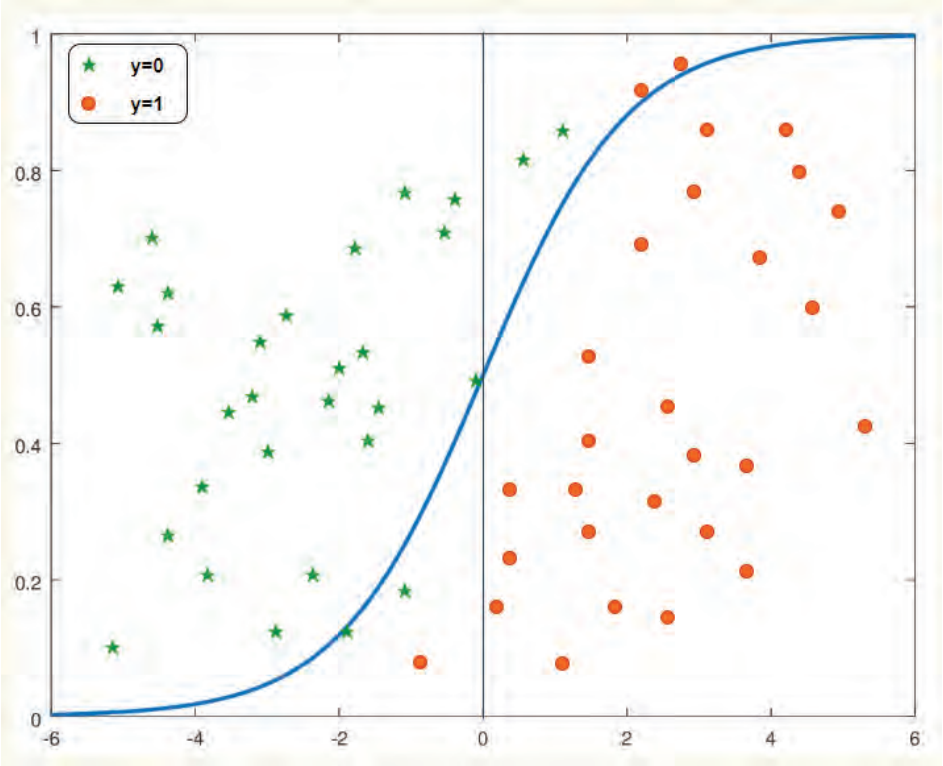


Figure 19. In logistic regression, the y values that will determine the output class take values between 0 and 1.

B occurring given that A has already occurred. This equation can also be written as shown in Equation 28:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{28}$$

Conditional probability distributions show the cause-and-effect linkages among a Bayesian Network model's inputs. Equation 29 represents the fully integrated probability distribution of a Bayesian network from the arguments $X_1, X_2, \ldots, X_n$.

$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_1|X_2, X_3, \ldots X_n)P(X_2|X_3, \ldots X_n) \\ &\quad \ldots P(X_{n-1}|X_n)P(X_n) \\ &= \prod_{i=1}^{n} P(X_i|X_{i+1}, \ldots, X_n) \end{aligned} \tag{29}$$

The Naive Bayes algorithm, the theory of which is explained in more detail in [52], produces both fast and successful results in matters such as document classification, medical diagnosis, monitoring of system performance, and spam filtering, despite its simplified assumptions [53].

Random Forest Classifier

Random forests are a supervised machine-learning method that uses many decision trees for classification and regression problems. In random forest applications, subsets of the data set are applied to various decision trees. The outputs from these trees are aggregated to generate the final output. As shown in Figure 20, random forests take the predictions of all trees, instead of relying on a single decision tree, and produce the final output using a technique called majority voting. Assessing the outcomes of decision trees executed with various and random feature selections collectively enhances accuracy and mitigates the overfitting problem [54].

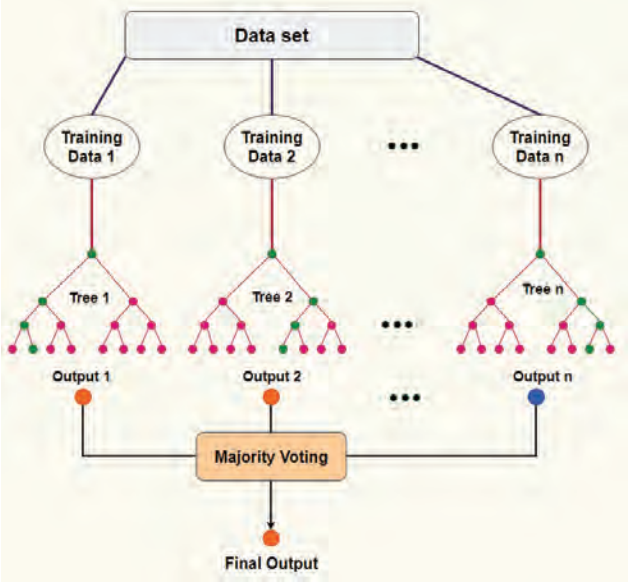


Figure 20. Random forest produces more accurate and effective outputs by re-evaluating the predictions from many decision trees.

The random forest method produces flexible and highly accurate predictions for both classification and regression problems. It works well with both categorical and numerical values. Since it uses a rule-based algorithm, it is not necessary to normalize or standardize the data set. However, it is important to pay attention to the number of decision trees used; as the number of trees increases, the training time will increase, and finding the final result will require more computational power. Random forest is not good at interpretability and determining the variables affecting the

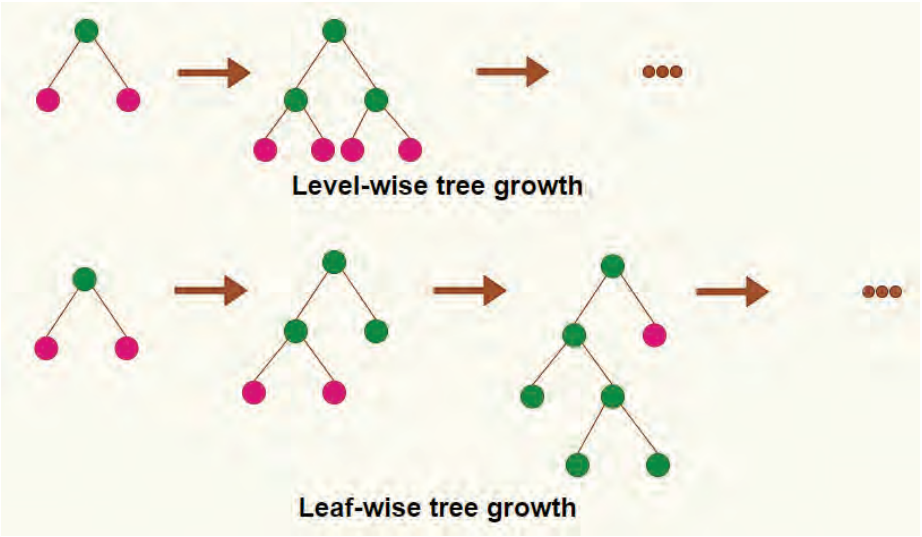


Figure 21. In the leaf-wise growth method, the leaf with the highest loss is split.

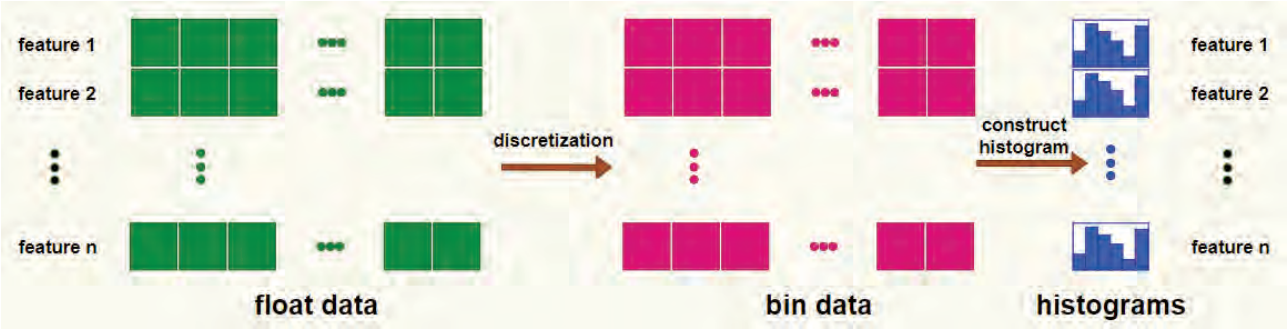


Figure 22. Histogram algorithm of LightGBM.

result; at this point, it differs from decision trees. Since a large number of decision tree outputs are converted to the final output by voting or averaging, this method loses its feature of being a white-box model.

Light GBM Classifier

LightGBM (Light Gradient Boosting Machine) is a machine learning method developed by Microsoft for classification, regression and time series problems. It has a similar structure to Gradient Boosting Decision Trees (GBDT) and XGBoost algorithms, which are efficient implementations of decision trees. The biggest difference between LightGBM and methods such as GBDT, XGBoost, and CART is that it uses a leaf-based splitting strategy instead of a level-based one [55]. As seen in Figure 21, the level-wise growth method keeps the trees balanced, while the leaf-wise growth method results in a more unbalanced growth pattern. Leaf-wise minimizes the total loss by splitting the leaf with the most loss. Since the tree grows depending on the number of leaves, it can cause over-learning when working with small data sets. To prevent this, it is necessary to adjust the tree depth parameter well.

Another feature of LightGBM is the histogram algorithm. As shown in Figure 22, each attribute dataset is divided into k integer intervals, and a histogram is created. The gradient value and sample count of each divided data sample are stored in the histogram. In this way, the data is made ready for use in a simpler way, less memory is consumed, and the training speed and efficiency of the model are higher [56,57].

Comparison of Classification Algorithms

When comparing machine learning models, in addition to metrics such as accuracy, precision, recall, f-score, ROC-AUC score, it is also necessary to consider features such as training and prediction time, resource usage, generalization capability, i.e. not overfitting, understandability, easy optimization, and the interpretability of the outputs. In this section, the advantages and limitations of the eight classification algorithms listed above are presented together in Table 7.

Accuracy Measurement

The performance of a classification algorithm is measured by finding some success metric from the results that the classification model calculates for the test data. In studies using classification algorithms, it is not sufficient to consider only the accuracy rate. Especially in unbalanced datasets, the accuracy rate alone cannot give accurate information about the classification performance. For this, a confusion matrix is created, as shown in Table 8, which compares the actual target values with those predicted by the model to evaluate the classification performance [58]. According to this table, a 'True Positive' (TP) situation occurs if a situation that actually exists positively in the forecasting process is predicted positively. If the existing condition is negative and the model's prediction is negative, a 'True Negative' (TN) condition occurs. If the existing condition is negative but the model predicts the result as positive, a 'False Positive' (FP) condition with a Type1 error occurs. A 'False Negative' (FN) condition with a Type 2 error occurs if the existing condition is positive and the model predicts the result as negative. Accuracy, precision, sensitivity (recall), true positive rate, false positive rate, and F1 score values are calculated from these values written in this matrix. In equations 30-34, the mathematical expressions of the success criteria mentioned above are given. The metric named support shows the number of samples found in each class of the test dataset.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{30}$$

$$Precision = \frac{TP}{TP + FP} \tag{31}$$

$$Recall = True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{32}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{33}$$

Table 7. Advantages and limitations of the classification algorithms described in this article

Classification algorithm	Advantages	Limitations
k-NN	-Understandable and easy to apply. -No training time, fast. -Does not contain many parameters. -Easy to optimize. -Can process samples with missing data.	-Very high dimensional data will reduce success. - Vulnerable to inconsistent data. -Scaling required. -Classes in the dataset must be balanced. -Classification performance may decrease as attributes are given equal importance.
Support vector machines	-Can work with a large number of variables. -High accuracy and performance. -Can learn with little input. -Applicable to non-linear data. -It is also effective in cases where the number of dimensions is more than the number of samples.	-The computational cost increases as the dataset gets larger or more complex. -Performance degrades in noisy or incorrect data. -It is difficult to understand the input-output relationship in the model. -In cases where the number of features is more than the number of samples, care should be taken to avoid overfitting and regularization terms when choosing kernel functions.
Decision tree	-The classification tree is easy to understand and interpret. Trees can be visualized. -Can handle both numeric and categorical data. -It is possible to validate a model using statistical tests. -Learning and prediction time is low.	-The model can generate overly complex trees that do not generalize data well with over-learning. This requires mechanisms such as pruning, setting the minimum number of specimens required per leaf, or setting the maximum depth of the tree. -The outputs of decision trees are piecewise fixed approaches. Therefore, they are not good at extrapolation. -The decision tree model creates biased trees if some classes are dominant. Therefore, it is recommended to balance the dataset before working with the model. - It is very sensitive to minor distortions in data.
Neural networks	-Can detect non-linear relationships between dependent and independent variables. -It can be applied to both classification and regression problems. -Gives successful results for datasets containing many entries.	-It is a black box model. Because the interaction between input and output is unpredictable, it requires more time and computational power. -Can only work with numeric data. -For complex problems, the computational cost of the training process increases.
Logistic regression	-Easy to understand and apply. -Performs very well on the linearly separable dataset. -Does not require any settings. -The computational cost is low.	-The accuracy of the model decreases if the input variables have complex relationships. -In cases where the number of features is more than the number of samples, an overfitting problem occurs.
Naive bayes	-It is fast and useful especially in large datasets. -Suitable for multiple classification problems. -Suitable for numerical and categorical data.	- If the classes are close to each other, success will decrease. -Normal distribution is taken as a basis when converting numeric data into categorical data.
Random forest	-Since it combines the predictions of multiple decision trees, it reduces the variation that may occur in a single tree and provides higher accuracy. -Because it integrates the predictions of several trees, it can withstand noisy, missing, or outlier data.	-Using many trees causes excessive memory usage and increased training time. -Since random forest combines the predictions of multiple trees, it becomes difficult to understand the underlying reason for the decision made by the model.
LightGBM	-Histogram-based algorithm that separates continuous feature values into separate bins increases speed and reduces memory usage. -Provides higher accuracy with leaf-based binning approach.	-The leaf-based splitting strategy may lead to overfitting, especially on small data sets, when very complex trees are produced.

Table 8. Confusion matrix

	Predicted	
Actual	True Positives (TP)	False Negatives (FN)
	False Positives (FP)	True Negatives (TN)

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (34)$$

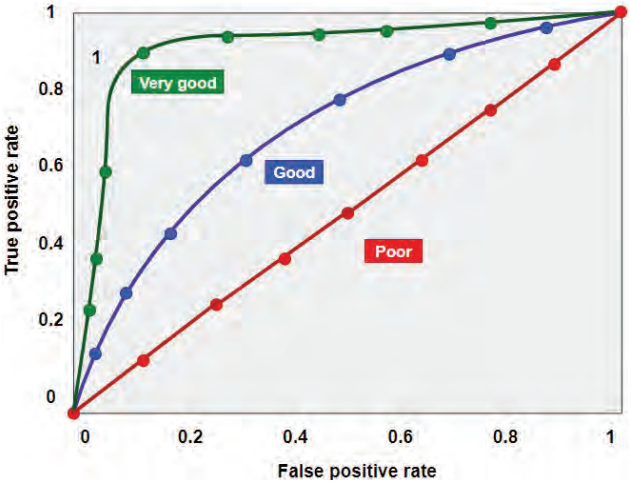


Figure 23. ROC curve showing classifier performance.

One of the criteria that indicates the performance of the classifier is the receiver operating characteristic (ROC). A ROC curve is formed by plotting the true positive rate versus the false positive rate at various threshold settings, as shown in Figure 23 [59]. The area under the ROC curve (AUC) is commonly used to evaluate the performance of a classifier. The larger the AUC value, the better the classifier predicts. Figure 23 represents three ROC curves within an abstract dataset. Since the area under the red ROC curve is half of the rectangle, its AUC value is 0.5. The blue and green curves cover a larger area, and their AUC values are higher. It can be said that the green curve, whose AUC value is very close to 1 in the figure, has a very high predictive accuracy.

As shown in Figure 23, a poor model making random predictions will have equal TPR and FPR values. The ideal scenario is to maximize TPR while minimizing FPR. In this case, true positives will be correctly classified and false positives will not be incorrectly classified.

RESULTS AND DISCUSSION

In this section, first, detailed test results of eight classification algorithms will be presented. Second, the success of these eight algorithms will be compared in terms of accuracy and processing time. Third, all the results we obtained in this study will be compared with other driving style prediction studies in the literature.

Classification Model Results

This section includes performance evaluation to compare classification algorithms to recognize driving style. This study was carried out using Scikit-learn library [35] in Anaconda Spyder environment on a 64-bit computer with an 11th Gen Intel® Core™ i7-1185G7 processor at 3.00 GHz. Three different driving style diagnoses were made with eight different classifiers using three different data sets consisting of features extracted from real driving data.

Before analyzing results, it is essential to choose the best-suited architecture for effective machine learning algorithm application. In most cases, this optimal architecture isn't immediately apparent and requires experimentation with various hyperparameter settings. Hyperparameters are adjustable parameters that control a machine learning model's behavior. Some key hyperparameters include:

-Nearest neighbor distance criterion
-Decision tree maximum depth
-Decision tree minimum leaf node samples
-Random forest number of trees
-Neural network layer and neuron counts

To identify the optimal hyperparameters, a common approach involves defining potential value ranges for each parameter, systematically trying different combinations, and evaluating their performance using appropriate metrics (e.g., accuracy, precision, recall). This process helps to balance training and testing accuracy, avoiding overfitting. In other words, when choosing hyperparameters, it is not enough to focus on achieving the highest accuracy on the training data; the hyperparameters that provide the highest accuracy should be selected such that the accuracies on both the training and test data are as close to each other as possible.

In this section, the results obtained with eight classification algorithms will be shown. The parameters of these algorithms were adjusted by choosing the most appropriate complexity value to avoid overfitting and underfitting situations. The selected parameters for each algorithm are explained in their respective sections.

K-NN Results

The results shown in Figure 24 were obtained with the k-NN model, which we preferred because it is straightforward to comprehend, optimize, and apply. Neighbors, weights, and metric parameters were adjusted to optimize the model. Selecting the 'distance' parameter as the weight parameter means that neighbors closer to the query point will have a greater influence than those farther away. Choosing Euclidean as the metric parameter means that the distances are determined by taking the square root of the sum of the squares of the distances in each dimension.

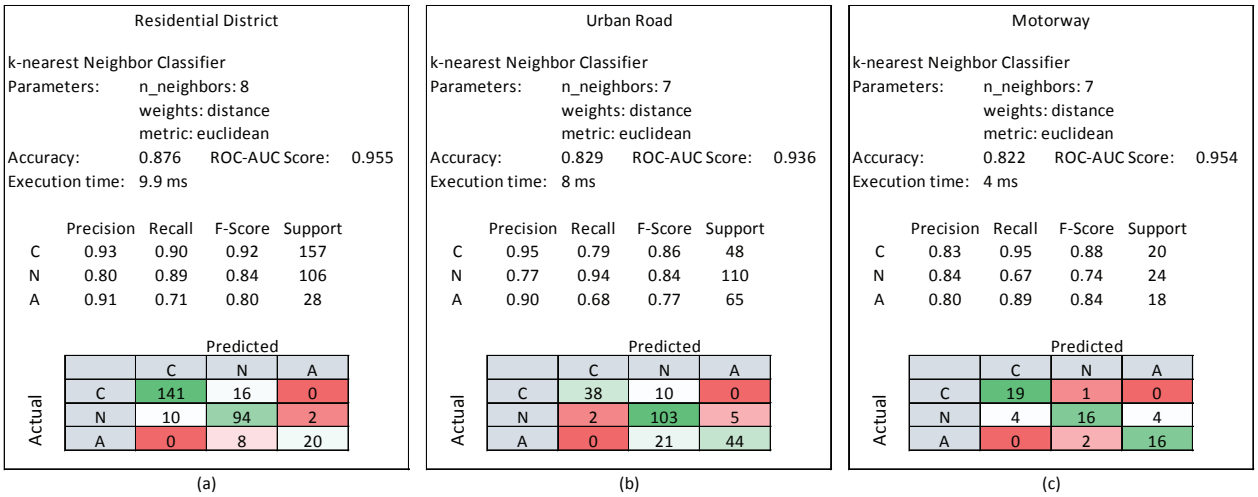


Figure 24. Results for a) residential district, b) urban road, c) motorway datasets with the k-NN algorithm.

Cleaning outliers, addressing the multidimensionality of the data set, and scaling affected the model results positively, and an average accuracy of 84.2% was reached for the three data sets. The confusion matrix shows that it cannot separate the aggressive class from the normal class. We evaluate that higher accuracy can be achieved if the classes in the data set are more balanced.

SVM Results

The results obtained with the SVM, which we prefer because it is not affected by the number of samples and dimensions, and has high accuracy and performance, are shown in Figure 25. We adjusted the regularization parameter, gamma, and kernel parameters to optimize the model. The regularization parameter, which takes only positive values, tells the model how much we want to avoid misclassifying each training sample. By determining the most appropriate value, we aim to maximize the margin between both classes and minimize the amount of misclassification. The gamma parameter is required when using the RBF kernel, which defines how far the effect of a single training sample reaches. The smaller the gamma value, the more linear the model behaves. When gamma is selected high, the model curvature will be high. As the gamma decreases, the regions separating the different classes become more generalized, causing overfitting as it gets larger. The kernel parameter specifies the kernel type to be used in the algorithm. The kernel function enables the linear separation of classes by transforming a low-dimensional input space into a higher-dimensional space. Since the data set is relatively small, the computational cost is low. An average of over 90% accuracy was achieved for the three data sets. The highest scores were obtained for the residential district dataset with 98.6%.

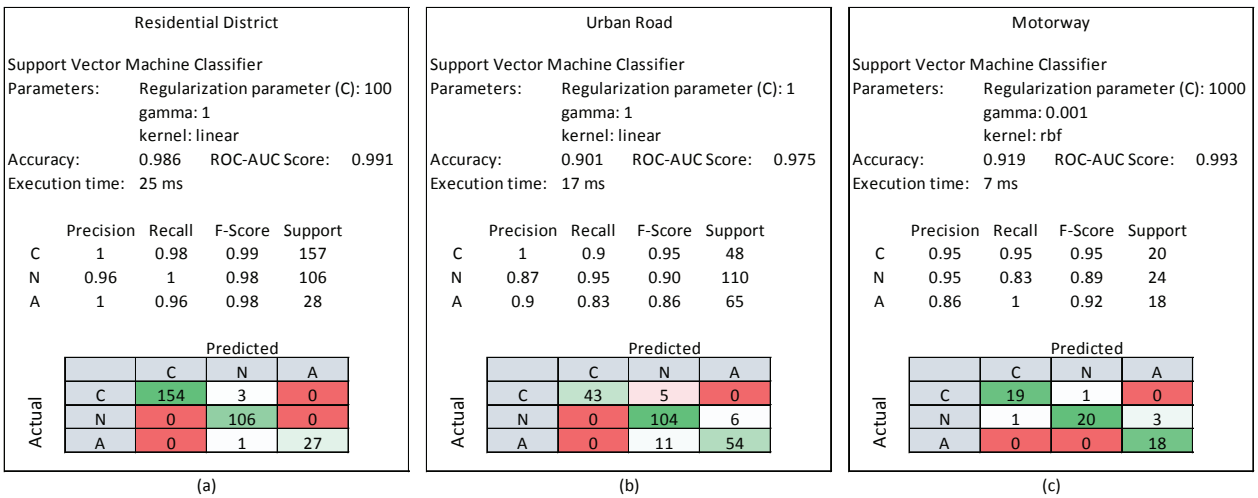


Figure 25. Results for a) residential district, b) urban road, c) motorway datasets with the SVM classifier.

Decision Tree Results

Figure 26 shows the results obtained with the decision tree, which is the most interpretable and visualizable among the classification algorithms. We used the Gini index for two datasets and information entropy for one dataset to decide which feature to split during model optimization. The "best" strategy for choosing the split at each node yielded the best result. For the most accurate classification, we tried different maximum depth and minimum sample number parameters. The response time of the decision tree is reasonable, and the average accuracy rate for all datasets is 96.86%. According to these results, the decision tree is one of the most suitable algorithms for classifying driving style.

MLP Results

The artificial neural network produces solutions to nonlinear problems with activation functions. The best learning is achieved with the tangent hyperbolic function, which serves as an activation function providing faster learning and a wider range for classification. Figure 27 shows the different regularization terms (alpha), the hidden layer and the number of neurons in these layers, the learning style, and the number of iterations. Although the success rate is high, this is the model with the highest processing time. In this study, the processing time is relatively low since the number of samples in the data set is not very large. However, when the data size increases, this should also be considered.

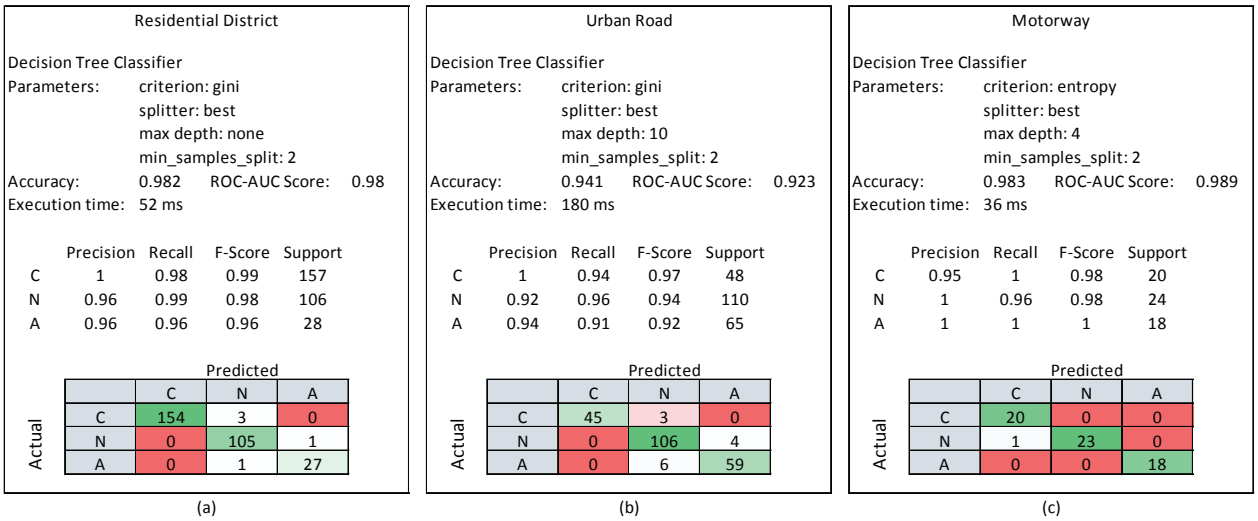


Figure 26. Results for a) residential district, b) urban road, c) motorway datasets with the decision tree algorithm.

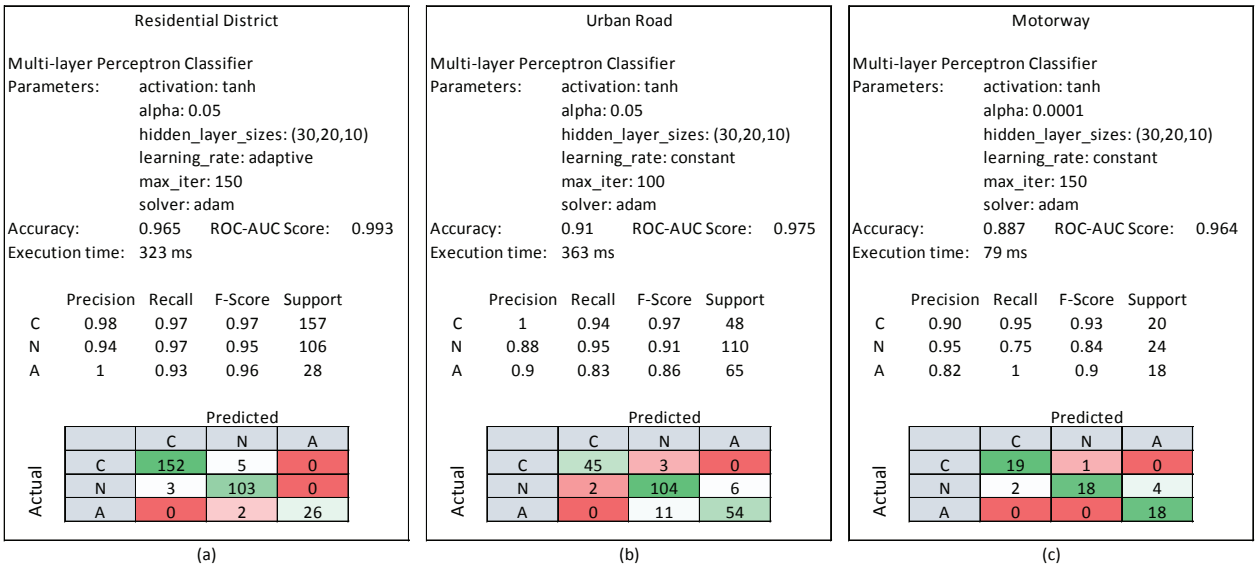


Figure 27. Results for a) residential district, b) urban road, c) motorway datasets with the MLP model.

Logistic Regression Results

We reached the optimum solution in logistic regression with newton-cg, which is preferred in multi-class problems. The highest success rate was obtained with the default penalty factor. The regularization factor (C) was set to 100 for two road types and 1 for one road type. When the results obtained in Figure 28 are examined, it is seen that logistic regression provides a favorable outcome in terms of transaction cost and accuracy.

Naive Bayes Results

The Naive Bayes classifier does not contain many tuning parameters. The only hyperparameter, the variance smoothing parameter, adds a user-defined value to the distribution's variance. This essentially broadens or smooths the curve so that the model can explain more samples that are further away from the distribution mean. Here, too, we calculated the best parameter (var_smoothing) values using the grid method and showed the results in Figure 29.

Random Forest Results

Random forest evaluates predictions made from many trees together instead of a single decision tree. It yields a final outcome from all the predictions produced according to the majority vote mechanism. The most appropriate selection of the number of trees and their features increases the accuracy and prevents the overfitting problem. Here, the hyperparameters of how many trees will be used (n_estimators), how many features will be considered in each tree (max_features), the depth of each tree, i.e., how many times it will be split (max_depth), and how many times the nodes in the tree should be split (max_leaf_nodes) are of great importance. The parameters in Figure 30 provide the

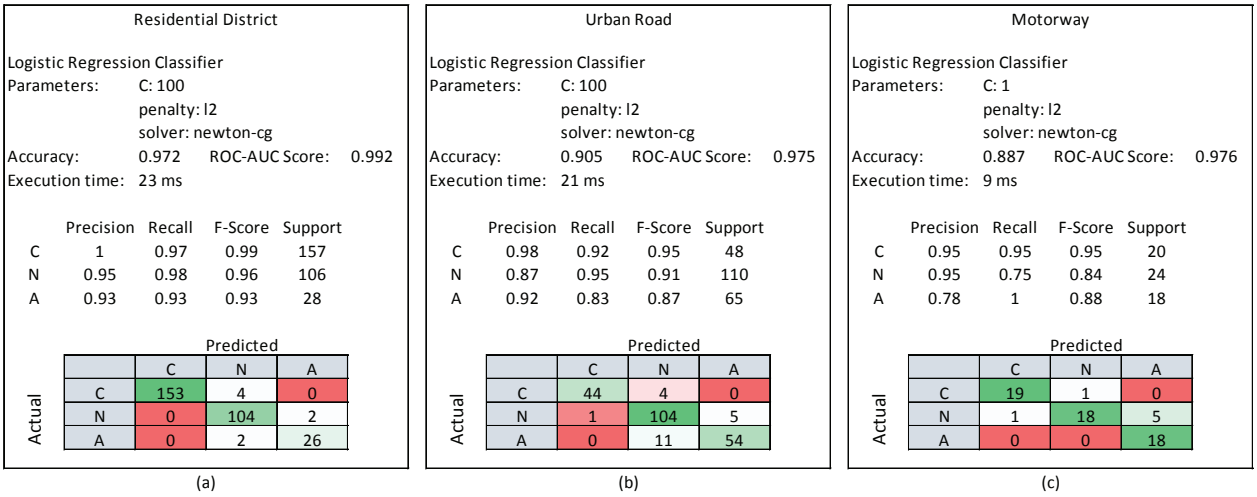


Figure 28. Results for a) residential district, b) urban road, c) motorway datasets with the logistic regression classifier.

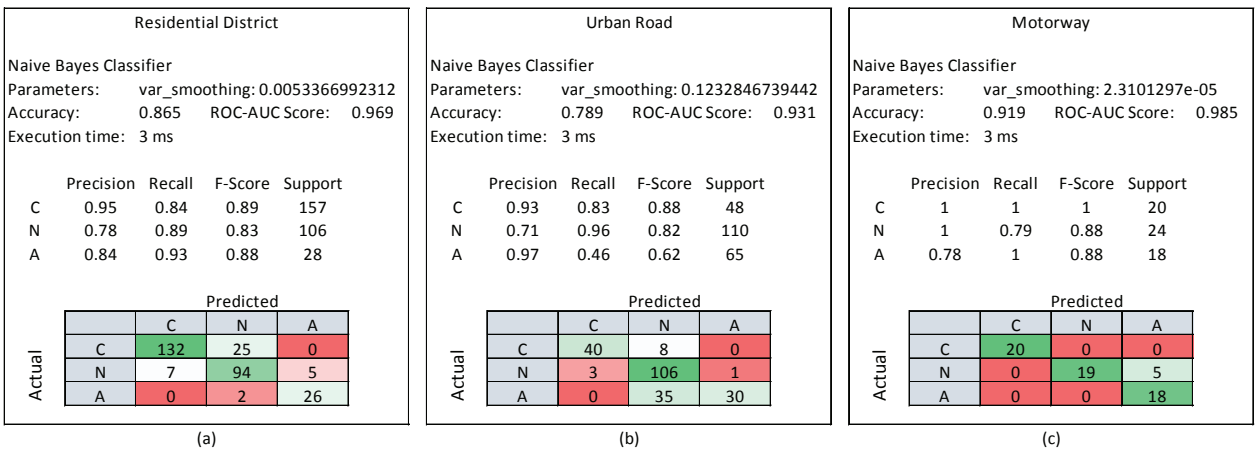


Figure 29. Results for a) residential district, b) urban road, c) motorway datasets with the Naive Bayes algorithm.

best strategies in the random forest algorithm, achieving an accuracy of over 95% in all types of paths.

Light GBM Results

LightGBM, which is designed to optimize model efficiency and memory usage, utilizes an optimal decision tree structure. Key components include leaf-based tree growth and histogram-based algorithms. Effective hyperparameter tuning is crucial for LightGBM's performance. We determined the hyperparameter settings to best classify the driving data, including the learning rate, the number of trees (n_estimators), the type and number of loss functions to be optimized (objective, num_class), and the maximum number of leaves in each tree (num_leaves), as shown in Figure 31. While determining these parameters, we ensured that the model did not fall into over-learning error. The results in Figure 31 show that we achieve very high accuracies, similar to other methods.

Comparison of Classification Algorithms

In this study, we comprehensively analyzed the performance of different machine learning algorithms in

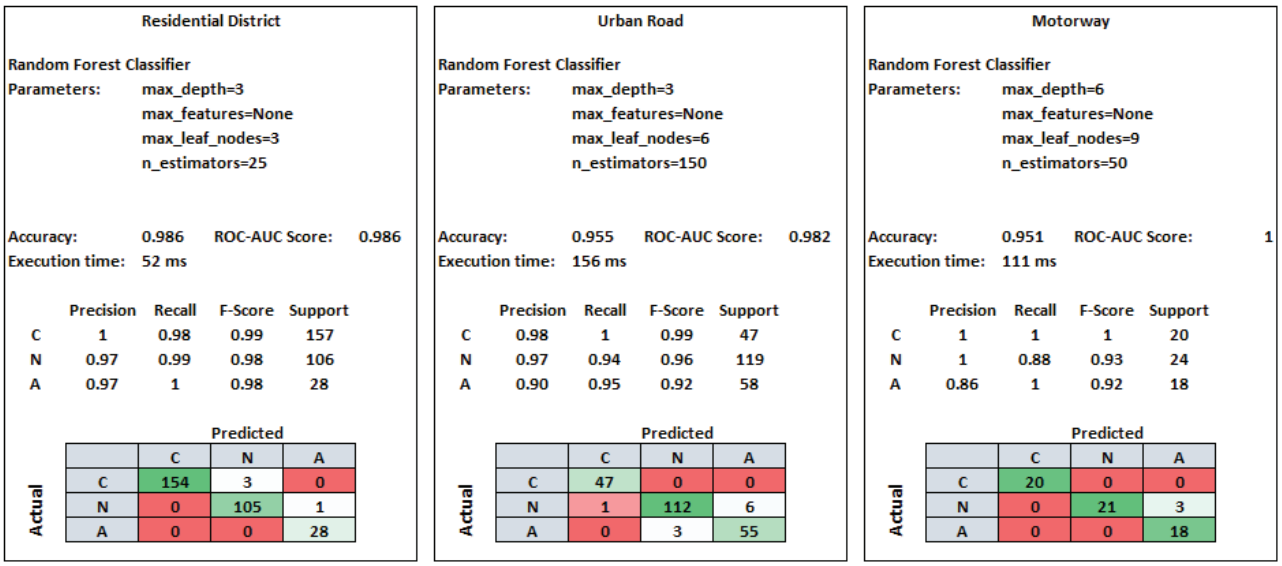


Figure 30. Results for a) residential district, b) urban road, c) motorway datasets with the random forest algorithm.

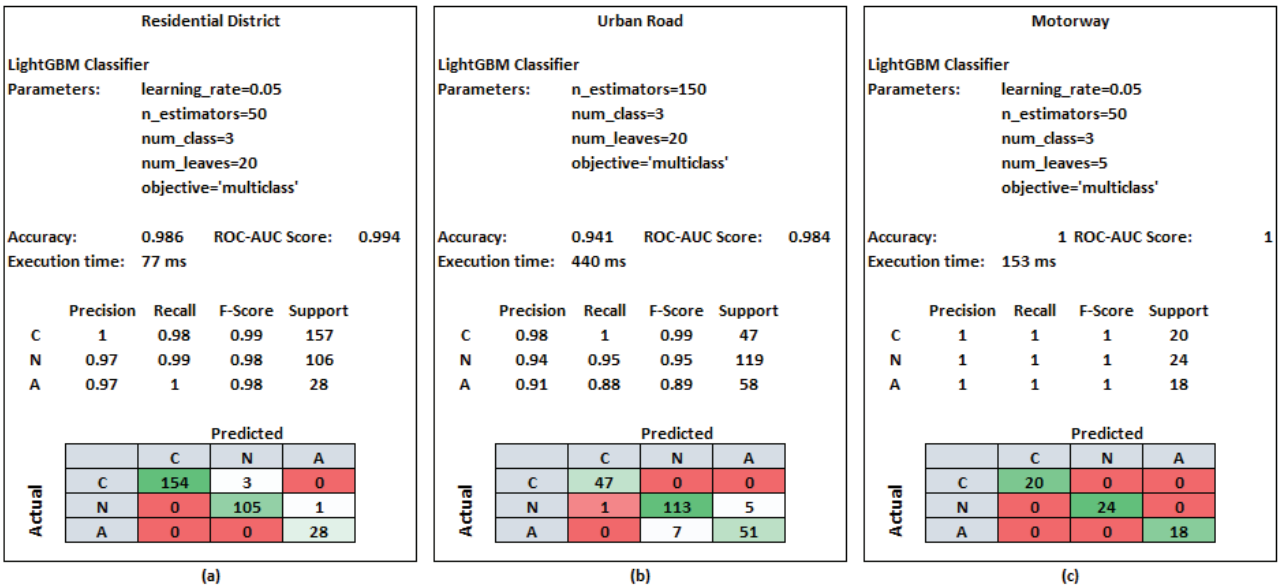


Figure 31. Results for a) residential district, b) urban road, c) motorway datasets with the LightGBM algorithm.

classifying driving data. Our findings reveal that each algorithm has its own advantages and disadvantages. For instance, k-NN, logistic regression, and naive Bayes exhibit very low processing times, although their accuracy rates may be slightly lower. On the other hand, algorithms such as SVM and artificial neural networks provide robust solutions for nonlinear problems. Decision trees offer significant benefits in terms of interpretability and visualization. Despite the disadvantages in processing time, LightGBM and random forest algorithms produce outputs based on the majority decision method and achieve higher accuracy values.

The above sections demonstrate that highly accurate solutions were obtained with all classification models. When comparing the performances of classification algorithms, we first looked at the accuracy rate. We supported this success criterion with the ROC_AUC score calculated from the confusion matrix metrics. We also added the response times to compare the resource usage of the models. Table 9 shows the results obtained from all classifiers. The performance comparison of the classification algorithms for each data set is graphically shown in Figure 32, Figure 33, and Figure 34, respectively.

When Figures 32-34 are examined, it will be seen that almost all models perform driving classification with high accuracy. Since the data of the overspeed feature can be separated more sharply in the motorway data set, greater accuracies were obtained in this data set. Similarly, in the residential district dataset, calm drivers were more easily separated from normal and aggressive drivers because they were driving at a slow speed, which positively affected the accuracies. In urban use, the traffic factor on the roads limited the speed of aggressive drivers, making it somewhat difficult to distinguish driving styles and having a negative effect on accuracy rates. Decision tree-based algorithms stand out in all data sets. We attribute this to decision

Table 9. Performances of classification algorithms for three different datasets

	Residential District			Urban Road			Motorway		
	Accuracy	ROC_AUC Score	Execution Time	Accuracy	ROC_AUC Score	Execution Time	Accuracy	ROC_AUC Score	Execution Time
k-NN	87.6 %	0.955	9.9 ms	82.9 %	0.936	207 ms	82.2 %	0.954	4 ms
Support vector machines	98.6 %	0.991	25 ms	90.1 %	0.975	17 ms	91.9 %	0.993	7 ms
Decision tree	98.2 %	0.980	52 ms	94.1 %	0.923	180 ms	98.3 %	0.989	36 ms
Multi layer perceptron	96.5 %	0.993	323 ms	91.0 %	0.975	363 ms	88.7 %	0.964	79 ms
Logistic regression	97.2 %	0.992	23 ms	90.5 %	0.975	21 ms	88.7 %	0.976	9 ms
Naive bayes	86.5 %	0.969	3 ms	78.9 %	0.931	3 ms	91.9 %	0.985	3 ms
Random forest	98.6 %	0.986	52 ms	95.5 %	0.982	156 ms	95.1 %	1	111 ms
LightGBM	98.6 %	0.994	77 ms	94.1 %	0.984	440 ms	100 %	1	153 ms

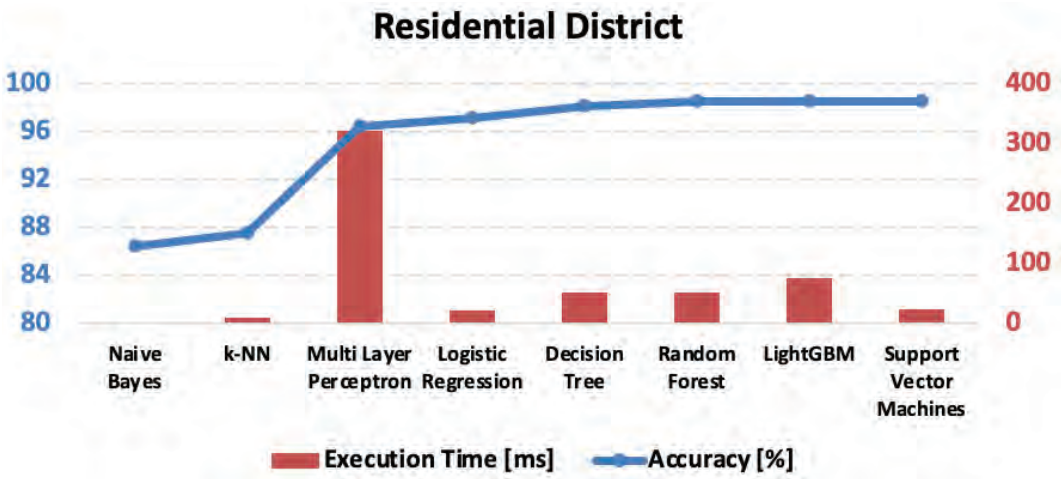


Figure 32. Comparison of classification algorithms for residential area dataset according to processing time and general accuracy criteria.

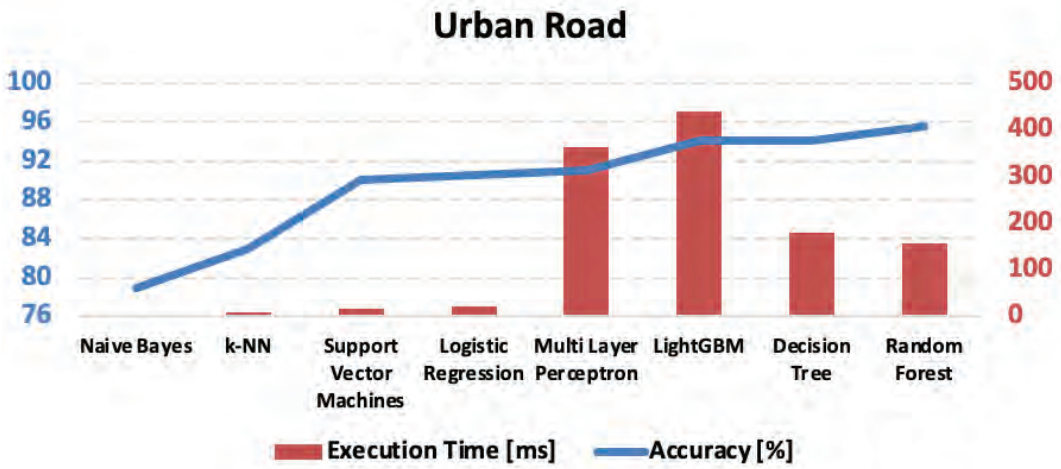


Figure 33. Comparison of classification algorithms for urban road dataset according to processing time and general accuracy criteria.

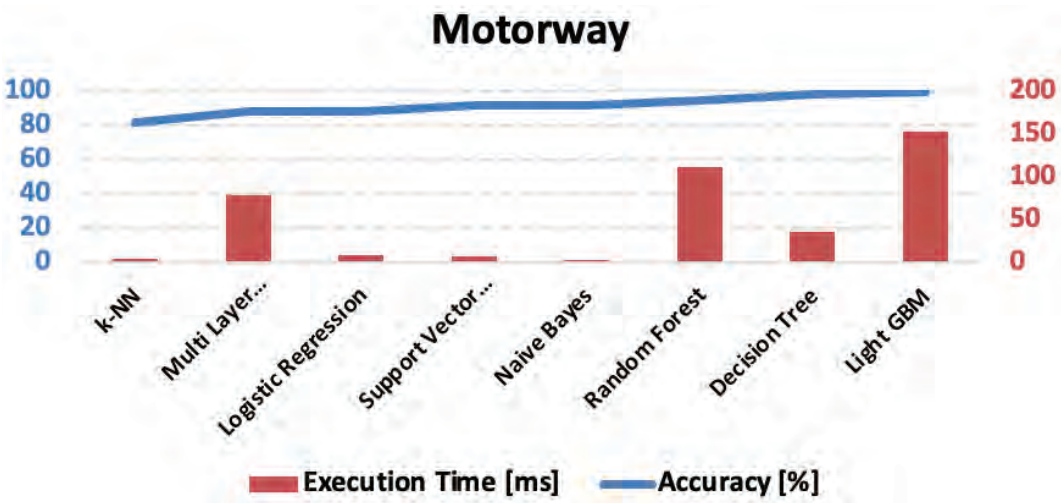


Figure 34. Comparison of classification algorithms for motorway dataset according to processing time and general accuracy criteria.

trees eliminating unimportant features and making faster and more accurate classifications according to the features they consider important. Due to the nature of our problem, decision tree-based models used prominent features such as maximum and average values of speed, sudden accelerations, and excessive speed ratio more effectively and made more accurate predictions. Our study also revealed the advantage of ensemble learning methods in coping with the overfitting problem. Although the processing times are reasonable for all models, it is necessary to consider this in cases where the data set is very large.

Comparison with Other Studies

We compare the methods proposed in this section with the studies mentioned in the literature summary of the introduction section, based on the criteria of data, algorithm, output information, and success rates. Information from 21 different studies is presented together in Table 10.

A careful examination of Table 10 reveals that driver or driving classification can be done successfully. New techniques have increased classification success for such problems. In particular, ensemble learning techniques are leading in this regard. Therefore, in our study, in addition to many traditionally used classification algorithms, we also used models based on ensemble learning such as random forest and LightGBM. As can be clearly seen in Table 10, our proposed methodology is highly effective in the following areas when compared to previous studies:

i. The models we developed were trained and tested with real road data, not simulation data.

Table 10. Comparison of studies on driving or driver behavior classification

Purpose of the study	Driving Data	Method	Features/Inputs	Outputs	Achievement	Authors
Classifying the driver's driving style	Speed profile of 11 standard drive cycles developed by Sierra Research	Rule-based classifier	Jerk profile	Calm, normal, aggressive	An accurate classification is shown depending on the fuel consumption.	Murphey et al.
Online driving style recognition	23 km track data simulated with IPG-Driver	Fuzzy logic	Longitudinal/ lateral acceleration, deceleration, speed, time gap	Sporty, normal, comfortable	They made driving classification with an average accuracy of 68%.	Dörr et al.
Classify drivers by their behavior	Data collected from a 3D virtual driving simulator	Fuzzy logic	Age, acceleration, braking, speed	Very passive, passive, normal, aggressive, dangerous	They tried to predict different undesirable situations such as traffic accidents, road congestion.	Fernandez and Ito
Diagnosing a driver profile to reduce the number of dangerous maneuvers	Driving data from two drivers on similar and different routes	k-means clustering, SVM	Acceleration, braking, and turning events	Differentiating between drivers	They were able to distinguish between 60-65% of similar drivers.	Van Ly et al.
Driver classification	1. Information on the driving of 3 different vehicles and 14 drivers on a predefined route 2. Driving data collected over 4 days from three couples sharing the same car	Support vector machine	Gyroscope, torque, GPS, accelerometer, pedal position, throttle positions, engine RPM	Information on which driver is driving which vehicle	Average classification accuracies were 75.83%, 85.83% and 86.67% using phone sensors only, car sensors only, and all sensors, respectively.	Zhang et al.
Driver classification	Data of 304 routes where 11 people drive with different vehicles on the same road	Neural networks	Velocity, acceleration, angle, GPS signals	1. Mishandling steer and pedals, speeding and getting out of the lane and road 2. Aggressive and moderate	The created model classified 97% of the drivers as good and 3% as bad drivers. The model was able to identify the driver with an accuracy of 97% to 98%.	Quintero et al.
Detection of driver behaviors to reduce the risk of traffic accidents	Data collected from the driving of three drivers in the same vehicle	Deep convolutional neural networks	Acceleration, gravity, throttle, speed, and engine RPM	Normal, aggressive, distracted, drowsy, and drunk driving	They reported that the optimal test results were 99.76% accurate.	Shahverdy et al.

Table 10. Comparison of studies on driving or driver behavior classification (continued)

Purpose of the study	Driving Data	Method	Features/Inputs	Outputs	Achievement	Authors
Driving classification	110 route driving information provided by the urban public transport system	k-NN	3-axis accelerometer signal statistical features	Normal and aggressive	In the experiments conducted in the same season, in traffic conditions and on the same route, 100% accuracy was achieved.	Vaitkus et al.
Driver classification for optimization of energy usage	Data of ten different drivers' rides on a predetermined route	Probability density function	Features extracted from the vehicle's powertrain signals	Aggressive, moderate and conservative	Based on the power demands placed on the vehicle powertrains, a 77% accuracy rate was recorded.	Kedar-Dongarkar and Das
Driving classification	58 test trips from different drivers	Neural networks	Attributes such as mean, max, min, variance extracted from longitudinal and lateral events	Very sporty, sporty, normal, defensive, very defensive	In the most relevant subset of 42 trips, a classification accuracy of 81% was achieved without misclassification.	Brombacher et al.
Driving behavior classification	UAH-DriveSet	Stacked long-short term memory recurrent neural networks	Accelerations along the x, y, z axes, Roll, pitch and yaw angles, Vehicle speed, Distance to vehicle ahead, Number of vehicles detected	Normal, aggressive and drowsiness	According to the confusion matrix data in classification skill, they reached an F1 score of 0.91.	Saleh et al.
Driving style classification	Data from a driving simulator with 20 licensed drivers	Semi supervised support vector machine	Vehicle speed, throttle opening	Aggressive, vague, normal	It has been reported that 86.6% was reached as the best classification rate with semi-supervised svm.	Wang et al.
Classification of driver fatigue	Data from a driving simulator with 63 adult participants	1. ANOVA 2. Random Forest 3. Neural Network 4. k-NN	Electrocardiogram, electrodermal activity signals, sleepiness and emotional state statement	Sleep deprivation, driving environment, sleepiness	Sleep deprivation, driving environment, and sleepiness predicted with an accuracy of 99%, 85%, and 73% respectively.	Meteier et al.
The classification of short and long-term driving behavior	13,792 real driving data in Ann Arbor, Michigan	1.Logistic regression 2. Support vector machine 3. Decision tree 4. Discriminant analysis	GPS data, brake status, headlight status, vehicle speed, acceleration, steering position, throttle position, yaw rate	Jerk, leading headway, yaw rate	It separated the classes with a confidence level of up to 99%.	Seraj

Table 10. Comparison of studies on driving or driver behavior classification (continued)

Purpose of the study	Driving Data	Method	Features/Inputs	Outputs	Achievement	Authors
Driver identification	KIA motors corporation dataset	k-NN	The 51 features extracted from vehicle, road and environmental parameters were reduced to 15.	Drivers	Accuracy for two drivers was 99.9%, and accuracy for 10 drivers was 76.36%.	Khan et al.
Classification of driving styles	Next Generation Simulation dataset	1. k-means 2. Spectral clustering	Features such as speed, acceleration, and yaw angle	Calm and aggressive drives	The K-means method gave slightly better results than spectral clustering.	Benterki et al.
Driving skill classification in curve driving scenes	Data from a driving simulator with 16 adult participants	1. k-NN 2. Support vector machine	Lateral and longitudinal controls	High-skilled driving, low/average-skilled driving	With SVM, classification accuracy was 95.7% in the full curve driving scene and 89% in the case of segmented curves.	Chandrasiri et al.
Abnormal driving behaviors detection	6-month driving data collected from real driving environments	1. Support vector machine 2. Neural network	16 key features were extracted from the driving behavior models.	Weaving, swerving, sideslipping, fast u-turn, turning with a wide radius, sudden braking	Average accuracy was 95.36 percent with the SVM classifier model and 96.88 percent with the NN classifier model.	Yu et al.
Driver and path detection	A real dataset of 292 observations	Neural networks	16 features related to road, speed and fuel consumption have been extracted.	City street, highway, dirt road	The best values obtained were 97% for determining the gender of the driver, 94% for determining the road type, and 97% for determining the driver familiarity.	Bernardi et al.
Aggressive driving detection	UAH-DriveSet	Long short term memory fully convolutional neural network	Speed, acceleration, roll, pitch, yaw, position, angle, road with, distance to ahead vehicle in current lane, time of impact to ahead vehicle	Normal driving, aggressive driving	They achieved an F score of 95.88% for a window length of 5 minutes.	Moukafih et al.
Object detection for road safety	4966 images	1. Faster R-CNN 2. RetinaNet, 3. YOLOv5	Images	Eye closure, yawning, smoking, mobile phone usage, and seatbelt compliance	As the best result, YOLOv5 recorded 125 FPS, 42 MB compact model size and 93.6% success.	Zia et al.

Table 10. Comparison of studies on driving or driver behavior classification (continued)

Purpose of the study	Driving Data	Method	Features/Inputs	Outputs	Achievement	Authors
Distracted driving detection system	Images and video data	CNN	Images and video frames of drivers	Using a mobile phone, eating, and looking away from the road	They detected driver distraction with 95% success.	Anitha et al.
Driving style classification	**38 thousand km driving data collected in 610 hours with nine professional drivers**	**1. k-NN** **2. SVM** **3. Decision tree** **4. NN** **5. Log. Regression** **6. Naive Bayes** **7. Random Forest** **8. LightGBM**	**Max/mean/st deviation of speed, skewness, kurtosis, max/min point ratio, acc/dec ratio, pos/neg slop avg, overspeed, stop an go ratio, mean band ratio**	**Aggressive, normal, calm driving.**	**The best results for three different roads: 1. 91.9% (motorway)** **2. 87.6% (residential district)** **3. 96.5% (residential district)** **4. 97.2% (residential district)** **5. 98.3% (motorway)** **6. 98.6% (residential district)** **7. 98.6% (residential district)** **8. 100 % (motorway)**	**Beskardes and Hames (this paper)**

ii. Our study has the longest range and the greatest total driving time with 9 professional drivers on different roads and with different vehicle types.
iii. It was studied with eight different classification algorithms. The success obtained with each of these algorithms is as high as the success obtained in other studies and is compatible with the literature.

CONCLUSION

Efforts to increase driving comfort and safety in road vehicles are continuously advancing. Driving style prediction systems, which are part of these efforts, can make important contributions to safety, such as detecting risky behaviors and preventing potential accidents caused by these behaviors. In vehicles integrated with driving style prediction systems, inefficient driving habits such as sudden acceleration or deceleration, and frequent stop-and-go can be detected and eliminated. More efficient driving can be achieved, resulting in lower fuel consumption and fewer harmful gas emissions. Driver fatigue can be detected in these vehicles and the driver can be warned. In this study, we estimated the driving style of drivers to contribute to the optimal utilization of these technologies. We used our meticulously compiled dataset, which reflects real driving characteristics as much as possible, rather than ready-made datasets. We worked on a driving style discrimination technique that can be used for purposes such as determining a driver's driving style from the data of their driving and offering more suitable vehicle options or adapting a vehicle to the driver's driving style. For this, we collected 38 thousand km of driving data with professional drivers. We extracted the necessary features from this data and calculated their value. After the necessary formatting and transformation processes with data mining techniques, we determined the driving style with eight classification algorithms. Using k-nearest neighbor, support vector machines, decision trees, neural networks, logistic regression, Naive-Bayes, random forest, and light gradient-boosting machine classifiers, we obtained consistent and highly accurate

results. We compared these methods extensively with each other and with similar studies in the literature. We have shown that the proposed methods are compatible with the literature, are suitable for determining driving style, and can be used for safe and enjoyable driving.

In this study, although the participants were of different ages, education levels, and genders, the number of participants was limited. Increasing the number of participants can advance the research further. Although the driving tests were conducted in many different regions, the majority were in the Mediterranean region. Evaluating more driving data from different regions could provide new insights into the differences in roughness, climate, and traffic conditions. The data evaluated to estimate the driving style was taken from a single source, namely an application on a smart device. Including data collected from devices such as in-car cameras in addition to this data source could show that information received from different sources supports each other and increases the prediction success rate.

In further studies, more detailed analyses could be performed by applying the data set not only to classification algorithms but also to time series-based algorithms. In the current situation, the estimation of the driving style is done after the drive is over. As a further stage, we can design a system that dynamically detects the changing driving style of the driver during a drive and adapts the vehicle to this current situation. For real-world applications, if the developed models are run on a tool that can be integrated into a car, a vehicle-human interaction can be established. This way, some warning, calming, or relaxing notifications can be sent to the driver from the vehicle's multimedia system. Moreover, adaptive systems can be designed to enable the vehicle to respond better to the driver's driving style to increase driving efficiency and comfort.

ACKNOWLEDGEMENTS

This work is part of Ahmet Beşkardeş's doctoral thesis.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Rolim C, Baptista P, Duarte G, Farias T, Pereira J. Real-Time Feedback Impacts on Eco-Driving Behavior and Influential Variables in Fuel Consumption in a Lisbon Urban Bus Operator. IEEE Trans Intell Transp Syst 2017;18:3061–3071. [CrossRef]

[2] Dörr D, Grabengiesser D, Gauterin F. Online driving style recognition using fuzzy logic. 2014 17th IEEE Int Conf Intell Transp Syst ITSC 2014 2014:1021–1026. [CrossRef]

[3] Zia H, Hassan I ul, Khurram M, Harris N, Shah F, Imran N. Advancing Road Safety: A Comprehensive Evaluation of Object Detection Models for Commercial Driver Monitoring Systems. Futur Transp 2025;5. [CrossRef]

[4] Anitha M, Namachivayam S, Christus ATA, Sb V. Real-Time Distracted Driver Detection and Alert System for Enhanced Road Safety 2024;2:196–213.

[5] Sim G, Ahn S, Park I, Youn J, Yoo S, Min K. Automatic longitudinal regenerative control of evs based on a driver characteristics-oriented deceleration model. World Electr Veh J 2019;10. [CrossRef]

[6] D'Agostino C, Saidi A, Scouarnec G, Chen L. Learning-Based Driving Events Recognition and Its Application to Digital Roads. IEEE Trans Intell Transp Syst 2015;16:2155–2166. [CrossRef]

[7] Marina Martinez C, Heucke M, Wang F-Y, Gao B, Cao D. Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. IEEE Trans Intell Transp Syst 2018;19:666–676. [CrossRef]

[8] Meiring GAM, Myburgh HC. A Review of intelligent driving style analysis systems and related artificial intelligence algorithms. Sensors 2015;15:30653–30682. [CrossRef]

[9] Xu L, Hu J, Jiang H, Meng W. Establishing style-oriented driver models by imitating human driving behaviors. IEEE Trans Intell Transp Syst 2015;16:2522–2530. [CrossRef]

[10] Imamura T, Yamashita H, Zhang Z, Othman MDR bin, Miyake T. A study of classification for driver conditions using driving behaviors. 2008 IEEE Int Conf Syst Man Cybern 2008. p. 1506–1511. [CrossRef]

[11] Murphey YL, Milton R, Kiliaris L. Driver's style classification using jerk analysis. 2009 IEEE Work Comput Intell Veh Veh Syst 2009, p. 23–28. [CrossRef]
[12] Gilman E, Keskinarkaus A, Tamminen S, Pirttikangas S, Röning J, Riekki J. Personalised assistance for fuel-efficient driving. Transp Res Part C Emerg Technol 2015;58:681–705. [CrossRef]
[13] Fernandez S, Ito T. Driver Classification for Intelligent Transportation Systems using Fuzzy Logic 2016:1212–1216. [CrossRef]
[14] Van Ly M, Martin S, Trivedi MM. Driver classification and driving style recognition using inertial sensors. 2013 IEEE Intell. Veh. Symp., 2013, p. 1040–1045. [CrossRef]
[15] Zhang C, Patel M, Buthpitiya S, Lyons K, Harrison B, Abowd GD. Driver Classification Based on Driving Behaviors. Proc. 21st Int Conf Intell User Interfaces. New York, NY, USA: Association for Computing Machinery; 2016, p. 80–84. [CrossRef]
[16] Quintero M CG, López JO, Cuervo Pinilla AC. Driver behavior classification model based on an intelligent driving diagnosis system. IEEE Conf Intell Transp Syst Proceedings, ITSC 2012:894–899. [CrossRef]
[17] Shahverdy M, Fathy M, Berangi R, Sabokrou M. Driver behavior detection and classification using deep convolutional neural networks. Expert Syst Appl 2020;149:113240. [CrossRef]
[18] Vaitkus V, Lengvenis P, Žylius G. Driving style classification using long-term accelerometer information. 2014 19th Int Conf Methods Model Autom Robot MMAR 2014 2014:641–644. [CrossRef]
[19] Kedar-Dongarkar G, Das M. Driver classification for optimization of energy usage in a vehicle. Proced Comput Sci 2012;8:388–393. [CrossRef]
[20] Brombacher P, Masino J, Frey M, Gauterin F. Driving event detection and driving style classification using artificial neural networks. 2017 IEEE Int Conf Ind Technol 2017. p. 997–1002. [CrossRef]
[21] Saleh K, Hossny M, Nahavandi S. Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks. 2017 IEEE 20th Int Conf Intell Transp Syst 2017. p. 1–6. [CrossRef]
[22] Wang W, Xi J, Chong A, Li L. Driving Style Classification Using a Semisupervised Support Vector Machine. IEEE Trans Human-Machine Syst 2017;47:650–660. [CrossRef]
[23] Meteier Q, Favre R, Viola S, Capallera M, Angelini L, Mugellini E, et al. Classification of driver fatigue in conditionally automated driving using physiological signals and machine learning. Transp Res Interdiscip Perspect 2024;26:101148. [CrossRef]
[24] Seraj M. The Classification of Short and Long-term Driving Behavior for an Advanced Driver Assistance System by Analyzing Bidirectional Driving Features. 2023.
[25] Khan M, Ali M, Haque F, Habib M. A machine learning approach for driver identification. Indones J Electr Eng Comput Sci 2023;30:276–288. [CrossRef]
[26] Benterki A, Maaoui C, Boukhnifer M, Judalet V. Driver Style Recognition Based on Vehicle Dynamic Data. 2024 10th Int Conf Control Decis Inf Technol 2024. p. 2566–25671. [CrossRef]
[27] Chandrasiri NP, Nawa K, Ishii A. Driving skill classification in curve driving scenes using machine learning. J Mod Transp 2016;24:196–206. [CrossRef]
[28] Yu J, Chen Z, Zhu Y, Chen Y, Kong L, Li M. Fine-Grained Abnormal Driving Behaviors Detection and Identification with Smartphones. IEEE Trans Mob Comput 2017;16:2198–2212. [CrossRef]
[29] Bernardi ML, Cimitile M, Martinelli F, Mercaldo F. Driver and Path Detection through Time-Series Classification. J Adv Transp 2018;2018:1758731. [CrossRef]
[30] Moukafih Y, Hafidi H, Ghogho M. Aggressive Driving Detection Using Deep Learning-based Time Series Classification. 2019 IEEE Int Symp Innov Intell Syst Appl 2019, p. 1–5. [CrossRef]
[31] Sun R, Chen Y, Dubey A, Pugliese P. Hybrid electric buses fuel consumption prediction based on real-world driving data. Transp Res Part D Transp Environ 2021;91:102637. [CrossRef]
[32] Eckert JJ, da Silva SF, Lourenço MA de M, Corrêa FC, Silva LCA, Dedini FG. Energy management and gear shifting control for a hybridized vehicle to minimize gas emissions, energy consumption and battery aging. Energy Convers Manag 2021;240:114222. [CrossRef]
[33] Bouhsissin S, Sael N, Benabbou F. Driver Behavior Classification: A Systematic Literature Review. IEEE Access 2023;11:14128–14153. [CrossRef]
[34] Alaa, T, Abdul M, Ali S, Alubady R, Adile M, Mohd K, et al. Data Management and Decision-Making Process Using Machine Learning Approach for Enterprises. Ournal Intell Syst Internet Things 2023:75–88. [CrossRef]
[35] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 2011;12:2825–2830.
[36] Mitchell TM, Mitchell TM. Machine learning. vol. 1. New York; McGraw-Hill; 1997.
[37] Sinclair C, Pierce L, Matzner S. An application of machine learning to network intrusion detection. Proc 15th Annu Comput Secur Appl Conf 1999. p. 371–377. [CrossRef]
[38] Sahami M, Dumais S, Heckerman D, Horvitz E. A Bayesian approach to filtering junk e-mail. Learn Text Categ Pap from 1998 Work, vol. 62, 1998. p. 98–105.
[39] Kim E, Kim W, Lee Y. Combination of multiple classifiers for the customer's purchase behavior prediction. Decis Support Syst 2003;34:167–175. [CrossRef]

[40] Kumar V, Sairam SDSS, Kumar S, Singh A, Nayak D, Sah R, et al. Prediction of Iron Ore Sinter Properties Using Statistical Technique. Trans Indian Inst Met 2017;70:1661–1670. [CrossRef]

[41] Erin K, Kutlu MÇ, Boru B. Comparison of gesture classification methods with contact and non-contact sensors for human-computer interaction. Sigma J Eng Nat Sci 2022;40:219–226. [CrossRef]

[42] Dogan D, Bogosyan S. Performance Analysis of SVM, ANN and KNN Methods for Acoustic Road-Type Classification. 2019 IEEE Int Symp Innov Intell Syst Appl, 2019, p. 1–6. [CrossRef]

[43] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer; 2000. [CrossRef]

[44] Morgan JN, Sonquist JA. Problems in the Analysis of Survey Data, and a Proposal. J Am Stat Assoc 1963;58:415–434. [CrossRef]

[45] Magerman DM. Statistical Decision-Tree Models for Parsing. 33rd Annual Meeting of the Association for Computational Linguistics. Cambridge, Massachusetts, USA: Association for Computational Linguistics; 1995. p. 276–283. [CrossRef]

[46] Quinlan JR. Induction of decision trees. Mach Learn 1986;1:81–106. [CrossRef]

[47] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115–133. [CrossRef]

[48] Tunçkaya Y, Köklükaya E. Comparative performance evaluation of blast furnace flame temperature prediction using artificial intelligence and statistical methods. Turk J Electr Eng Comput Sci 2016;24:1163–1175. [CrossRef]

[49] Gajic D, Savic-Gajic I, Savic I, Georgieva O, Di Gennaro S. Modelling of electrical energy consumption in an electric arc furnace using artificial neural networks. Energy 2016;108:132–139. [CrossRef]

[50] Ojha VK, Abraham A, Snášel V. Metaheuristic design of feedforward neural networks: A review of two decades of research. Eng Appl Artif Intell 2017;60:97–116. [CrossRef]

[51] Bhattacharjee P, Dey V, Mandal UK. Risk assessment by failure mode and effects analysis (FMEA) using an interval number based logistic regression model. Saf Sci 2020;132:104967. [CrossRef]

[52] Zhang H. The Optimality of Naive Bayes. Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004). Florida: AAAI Press; 2004.

[53] Badik ŞT, Akar M. Machine learning classification models for the patients who have heart failure. Sigma J Eng Nat Sci 2024;42:235–244. [CrossRef]

[54] Cihan P. The machine learning approach for predicting the number of intensive care, intubated patients and death: The COVID-19 pandemic in Turkey. Sigma J Eng Nat Sci 2022;40:85–94. [CrossRef]

[55] Tian L, Feng L, Yang L, Guo Y. Stock price prediction based on LSTM and LightGBM hybrid model. J Supercomput 2022;78:11768–11793. [CrossRef]

[56] Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU. A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting. IEEE Access 2019;7:28309–28318. [CrossRef]

[57] Zhang Y, Yu W, Li Z, Raza S, Cao H. Detecting Ethereum Ponzi Schemes Based on Improved LightGBM Algorithm. IEEE Trans Comput Soc Syst 2022;9:624–637. [CrossRef]

[58] Roshni Thanka M, Bijolin Edwin E, Ebenezer V, Martin Sagayam K, Jayakeshav Reddy B, Günerhan H, et al. A hybrid approach for melanoma classification using ensemble machine learning techniques with deep transfer learning. Comput Methods Programs Biomed Update 2023;3:100103. [CrossRef]

[59] Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019;19:281. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00053

Research Article

A study on numerical solutions of a fractional-order model for CAR T-cell therapy in leukemia using the laplace-adomian decomposition method

Rezaul KARİM[1,2,*], M. Ali AKBAR[2], M. A. BKAR PK[2], Pinakee DEY[1]

[1]*Department of Mathematics, Mawlana Bhashani Science and Technology University, 1902, Bangladesh*
[2]*Department of Applied Mathematics, University of Rajshahi, 6205, Bangladesh*

ARTICLE INFO

Article history
Received: 25 October 2024
Revised: 13 January 2025
Accepted: 13 February 2025

Keywords:
CAR T-cell; Fractional Order (FO); Mathematical Model (MM); Leukemia Disease; Stability Analysis

ABSTRACT

Leukemia is an aggressive form of blood cancer that develops in the bone marrow. In this study, we consider a fractional-order four compartmental mathematical model of leukemia which includes susceptible blood cells $S(t)$, infected blood cells $I(t)$, cancer cells $C(t)$, and immune blood cells $W(t)$, and we discuss the dynamics of transmission of the disease. We employ the Laplace-Adomian decomposition methods to obtain analytical solutions and the Runge-Kutta fourth-order approach to get numerical solutions for the mathematical model of leukemia. In order to illustrate the process, the convergent of the series solution is also given, and the corresponding plots against various orders of the differentiations are plotted using specific values for the model's parameters. We compared the solutions determined from the Laplace-Adomian decomposition methods and Runge-Kutta fourth-order. Graphical results demonstrate that the Laplace Adomian Decomposition method aligns closely with the Runge-Kutta fourth order method. For various fractional parameter values α the findings reveal that the fractional-order model offers more accuracy and stability compared to the conventional integer-order model. Some plots are presented to show the reliability and simplicity of the method. Leukemia virus is one of the many infectious diseases whose dynamics of spread can be better understood through mathematical modeling and underscores the urgent need to ensure global accessibility to this approach, with the potential to save countless lives worldwide.

Cite this article as: Karim R, Akbar MA, Bkar PK MA, Dey P. A study on numerical solutions of a fractional-order model for CAR T-cell therapy in leukemia using the laplace-adomian decomposition method. Sigma J Eng Nat Sci 2026;44(2):954–969.

INTRODUCTION

Leukemia originates from a cell in the bone marrow that undergoes a transformation into a leukemia cell. After this change, the leukemia cells may start to grow and thrive more effectively than healthy cells. Gradually, these leukemia cells can outnumber or inhibit the production of normal cells. Between 2009 and 2014, leukemia stayed the 5th leading cause of cancer passing away in males and the

*Corresponding author.
*E-mail address: rezaulmath11124@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

6th in females. By 2016, around 2.35 million people worldwide be situated living with leukemia, leading to about 353,500 passing away [1-2]. Scientific biological problems are modeled using the fascinating field of applied analysis known as fractional calculus, which is made up of free-order derivatives and integrals. A tumor is an uncontrolled growth of abnormal cells that can invade tissues. This treatment is altering the immune system and allows them to specifically target and destroy leukemia cells [3-4]. Numerous researchers have also reviewed studies using CAR-T in the leukemia model [5-6]. By delivering CAR T- cells, [7] examined the MM of the interaction between leukemia cells and immune cells. Our funding has been allocated to a broad range of cancer types, which are classified according to the predominant cell type impacted [8]. As numerous studies have shown [3,9-11], FO modeling is becoming more and more popular in the field of epidemiology because it can accurately represent intricate and nonlinear disease processes. The Atangana-Baleanu and Caputo-Fabrizio models are two well-known instances of fractional-order models that shed light on the dynamics of disease and the impact of healthcare capacity on disease transmission. As mentioned in references [12-15]. It presents the multistep LADM, which provides a more accurate approximation than conventional methods for modeling the dynamics of T-cells [16-19]. It discusses convergence and error while building method spaces and using Caputo's partial time derivative [20-22]. Maayah et al. explored approximate results and symmetrical attractors for an FO cancer-immune model using the method in [23,24]. Additionally, studies have investigated an in-host dengue contagion model with invulnerable comeback and proposed an innovative algorithm that integrates cubic uniform splines with limited difference techniques to solve FO diffusion singular wave models affected by damping-reaction forces [25-28]. Ponalagusamy et al. [29] explored the approximate solution of heat flow problems using a hybrid approach that combines the Rayleigh-Ritz method with STWS and RKHM techniques. In a separate study, Ponalagusamy [30] developed an inventive and effective computational procedure founded on STWS for comprehensive linear time-varying structures, including both singular and non-singular cases. Ponalagusamy and Senthilkumar [31] conducted a comparative analysis of various Runge-Kutta (RK4) orders and embedded methods in the simulation of multilayer raster CNNs. Additionally, Ponalagusamy et al. [32] introduced a new 5th-order, 5th -stage RK4 method based on the Heronian mean. Chandru et al. [33] proposed a 5th weighted Runge-Kutta algorithm, also derived using the Heronian mean, for solving initial value problems in ordinary differential equations. Furthermore, Ponalagusamy and Senthilkumar [34] developed a fourth-order embedded RK algorithm with Heronian mean incorporating error control for single-layer or raster cellular neural networks. Atalan et al. [35] developed optimization models to assess and improve the performance of healthcare systems. Dincer [36] analyzed mathematical modeling to determine the nickel inhibition constant in nitrification processes. Berrak and Ali [37] examined the stability of a neural field model incorporating small delays. Results are shown for convergence-error behavior and computational algorithms. A summary and suggestions for more research are included in the study's conclusion, which investigates the use of fractional differential equations in the modeling of diseases such as cancer and infectious diseases, with a particular emphasis on the relationship between cancer cells and the immune system. Important topics like the mathematical framework, error analysis, and solution representation are covered. Summaries of the findings and recommendations for further study are provided at the end. The work closes research gaps concerning the T-cells model by using a novel FO modeling system to realize how the virus spreads within a mass with adaptive invulnerability. The numerical solutions of the given FO model are analyzed using the ADM in combination with the Laplace transform. To validate the results, random values are assigned to the initial conditions and parameters. In this study, we develop and analyze a mathematical model (MM) for leukemia, represented by a system of FO differential equations. The model consists of four compartmental components: Fractional order $S(t)$, $I(t)$, $C(t)$, and $W(t)$ [38-39]. We aim to investigate the numerical simulations, graphical representations, parameter analyses, interpretations, and solutions of the fractional-order model using the RK4 method and the LADM. The results obtained from LADM are compared with those derived from numerical simulations. We present solution figures generated through LADM and simulate them alongside the numerical solutions for comparison. This analysis highlights the accuracy, simplicity, and effectiveness of the LADM approach. By including FO derivatives, the model can more exactly enlighten derivatives that are common in biological organizations but are often missed in conventional integer-order models, and many researchers have described [40-45].

This work is systematized as follows: Section 2 covers the preliminaries, providing foundational concepts. Section 3 offers a detailed description of the mathematical model (compartmental structure) for leukemia. Section 4 outlines the solution steps using the LADM method. Section 5 delves into the analysis of the fractional model, while Section 6 examines the solutions in detail. The Numerical simulations are presented in Section 7, which includes a subsection (7.1) focused on numerical simulations and result comparisons. Section 8 explores the stability criteria for the FO leukemia model, and Section 9 discusses the convergence analysis for the model. Lastly, Section 10 concludes the study and highlights possible directions for future research.

PRELIMINARIES

The Reimann-Liouville fractional (RLF) θ-order integral operator, where the function $q : R_+ \to R$, expressed J^{θ} q(t) is expressed as

$$J^{\theta}q(t) = \frac{1}{\Gamma(\theta)}\int_0^t \frac{q(\xi)}{(t-\xi)}d\xi.$$

Where $\theta \in R_+ : \theta \in (0,1)$ and $t > 0$. The gamma function $\Gamma(\theta)$is given by y

$$\Gamma(\theta) = \int_0^{\infty} x^{\theta-1} l^{\theta} dx$$

Mathematical Model (Compartment) of Leukemia

By integrating experimental data and theoretical analysis, the model can guide clinical practice and contribute to improving patient outcomes in cancer immune treatment. We will use a system of ODEs to describe the dynamics of (CAR) T Cell Treatment. Here $S(t)$ represents the number of susceptible blood cell class, $I(t)$ represent the number of infected blood cell class, $C(t)$ represent the number of leukemia cell class and $W(t)$ represent the number of immune blood cells. Here, Figure 1 represents the leukemia transmission with a compartment model. According to Khumaeroh et al. [4], Khatun and Biswas [8] model, we considered as:

The natural mortality rates of $S(t)$, $I(t)$, $C(t)$, and $W(t)$ are represented by the parameters α, μ, b, and τ, respectively. Due to the presence of $C(t)$ in the blood, the decay rate of $W(t)$ is indicated by the letter θ, v is the rate at which T cells and the proliferation rate of $W(t)$ is δ. Here, Figure 1 represents the Leukaemia transmission with the compartment model. The following set of ordinary differential equations governs our modified model Khumaeroh et al. [4], Khatun and Biswas [8] and karim et al. [36] is as:

$$\left.\begin{aligned} \frac{dS}{dt} &= A - \alpha S - \beta SC, \\ \frac{dI}{dt} &= \beta SC - \mu I - \gamma CI, \\ \frac{dC}{dt} &= r(1-bC)C - kCW, \\ \frac{dW}{dt} &= v + \delta C - \tau W - \theta WC. \end{aligned}\right\} \quad (1)$$

and $S(t) \geq 0$, $I(t) \geq 0$, $C(t) \geq 0$ and $W(t) \geq 0$. (2)

Steps of Solutions of LADM

The ADM, introduced by Adomian in 1980, is a robust approach for obtaining numerical and explicit solutions to systems of DEs arising in physical problems. The LT, widely recognized as a powerful tool in manufacturing and applied mathematics, complements ADM effectively. The integration of these two methods gives rise to the LADM, a highly efficient technique. In LADM, the LT is applied to convert DEs into algebraic equations, while nonlinear terms are expressed in terms of Adomian polynomials. This method is well-suited for solving deterministic and stochastic differential equations, including systems of linear and nonlinear ordinary and partial differential equations of both classical and fractional orders. Unlike other methods, LADM does not require perturbation, linearization, or a predefined step size, as in the RK4. Additionally, it does not depend on parameters like the Homotopy Perturbation Method (HPM). Although the solutions obtained through LADM align with those derived from the standard ADM, LADM is considered more powerful.

Let us consider the following FO differential equation given by

$$c_{J^{\propto}}V(t) = L_i(V_1, V_2, V_3 \dots\dots\dots V_i) + Z_i(V_1, V_2, V_3 \dots\dots\dots V_i).Y$$

$$Subject\ to$$

$$V^{\propto}(0) = C_k^i, for\ i = 1,2,3, \dots\dots m, and\ n_{i-1} \leq \propto \leq n_i.$$

Where $c_{J\propto}V_i(t)$ is Caputo–Fabrizio operator of i number of unspecified functions $P(t)$.

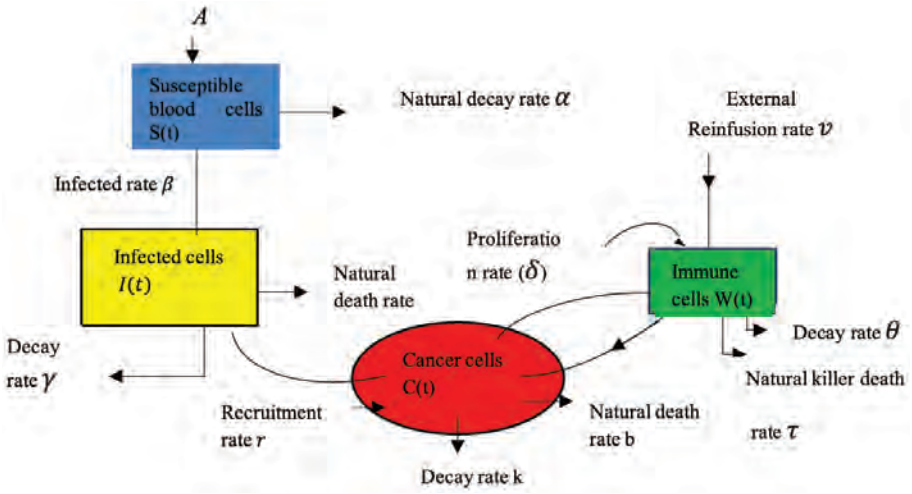


Figure 1. Leukaemia transmission with a compartment model.

Now from Equation (2), on (1) yields:

$$\frac{sL[V_i(t)] - V_i(0)}{s+\propto(1-s)} = L[L_i(V_1, V_2, V_3 \dots \dots V_i) + Z_i(V_1, V_2, V_3 \dots \dots V_i)].$$

Utilizing the ADM, the unknown function is decomposed as:

$$V_i(t) = \sum_{j=0}^{\infty} V_{ij}(t), \quad i = 1,2,3 \dots \dots \dots m.$$

Where, $Z_i(w_1 \dots \dots w_i) = \sum_{j=0}^{\propto} X_{ij}(t), i = 1,2,3, \dots \dots m.$ and A_{ij} is the using the definition of an adomian polynomial, one obtains:

$$[L\left[\sum_{j=0}^{\propto} V_{ij}(t)\right] = \frac{V_i^{k_i}(0)}{s} + \frac{s+\propto(1-s)}{s}\left(L[L_i(\sum_{j=0}^{\propto} V_{1j}(t), \dots \dots \dots \sum_{j=0}^{\propto} V_{mj}(t)\right)] + L[\sum_{j=0}^{\propto} X_{ij}(t)]).$$

Now from equation (4), w_{1j}, w_{2j},…… w_{nj}, $J \geq 0$:

$$V_{i(j+1)}(t) = L^{-1}\left(\frac{V_i^{ki}(0)}{s}\right) + L^{-1}\left(\frac{s+a(1-s)}{s} L \times \left[L_i\left(\sum_{j=0}^{\propto} V_{1j}(t), \dots \dots \sum_{j=0}^{\propto} V_{mj}(t)\right)\right] + L\left[\sum_{j=0}^{\propto} X_{ij}(t)\right]\right)$$

which simplifies to the equations be

$$V_{i(j+1)}(t) = C_k^i + L^{-1}\left(\frac{s+\propto(1-s)}{s} L\left[L_i\left(\sum_{j=0}^{\propto} V_{1j}(t), \dots \dots \sum_{j=0}^{\propto} V_{mj}(t)\right)\right] + L\left[\sum_{j=0}^{\propto} X_{ij}(t)\right]\right)$$

Applying FO and Caputo-derivative methods, we have

$$\left.\begin{aligned} \frac{{}^{c}d^{\alpha}S(t)}{dt} &= A - \alpha S - \beta SC, \\ \frac{{}^{c}d^{\alpha}I(t)}{dt} &= \beta SC - \mu I - \gamma CI, \\ \frac{{}^{c}d^{\alpha}C(t)}{dt} &= r(1-bC)C - kCW, \\ \frac{{}^{c}d^{\alpha}W(t)}{dt} &= v + \delta C - \tau M - \theta WC. \end{aligned}\right\} \quad (3)$$

where $\mu = \mu_1 + \frac{\eta\beta}{\delta}$ and $\delta = \delta_1 + \frac{\eta\tau}{\delta\theta}$

If the equations (3) be $\frac{{}^{c}d^{\alpha}S(t)}{dt} \geq 0, \frac{{}^{c}d^{\alpha}I(t)}{dt} \geq 0, \frac{{}^{c}d^{\alpha}C(t)}{dt} \geq 0$ and $\frac{{}^{c}d^{\alpha}W(t)}{dt} \geq 0$ respectively. Now from Equations (3) we get,

$$\left.\begin{aligned} C_{J^{\alpha 1}}S(t) &= A - \alpha S(t) - \beta S(t)C(t), \\ C_{J^{\alpha 1}}I(t) &= \beta S(t)C(t) - \mu I(t) - \gamma C(t)I(t), \\ C_{J^{\alpha 1}}C(t) &= r\big(1 - bC(t)\big)C(t) - kC(t)W(t), \\ C_{J^{\alpha 1}}W(t) &= v + \delta C(t) - \tau W(t) - \theta W(t)C(t). \end{aligned}\right\} \quad (4)$$

With initial conditions

$$S(0) = S_0, I(0) = I_0, C(0) = C_0, W(0) = W_0 \quad (5)$$

On the provided fractional-order model, numerical solutions are examined by utilizing the (ADM) in conjunction with the LT. Both the initial conditions and parameter values are randomly assigned in order to validate the generated results.

Variables and Parameter Analysıs of the Fractional Order Model

Table 1-2 in this section presents the definitions of each compartment along with detailed descriptions of the parameters used in the model (Altrock et al. [2], Khumaeroh et al. [4], and Khatun and Biswas [8]).

Table 1. Variables and parameter description of the fractıonal order leukemia model

Variables / parameters	Descriptions	Area	Unit
S	Susceptible blood cell class	$S \geq 0$	Concentration
I	Infected class	$I \geq 0$	Concentration
C	Leukemia blood cell class	$C \geq 0$	Concentration
W	Immune blood cell class	$W \geq 0$	Concentration
A	Recruitment rate of susceptible blood cells	$A \geq 0$	Concentration/day
μ	Natural death rate	$\mu \geq 0$	Concentration/day
α	Changing rate of susceptible class to infected class	$\alpha \geq 0$	1/day
γ	Infected cells natural death rate	$\gamma \geq 0$	Concentration/day
β	Decay rate of Infected cells	$\beta \geq 0$	Concentration/day
r	Progress rate of Leukemia blood cell	$r \geq 0$	1/day
θ	Carrying capacity of Leukemia blood cell	$\theta \geq 0$	Concentration/day
v	Exterior distillation rate of T-Cell Treatment	$v \geq 0$	Concentration/day
τ	Natural death rate of Recovered cell	$\tau \geq 0$	1/day
δ	Immune cell of proliferation rate	$\delta > 0$	1/day

Table 2. Purpose of the specific parameter of MM of leukemia (Khumaeroh et al. [4], Khatun and Biswas [8] and Karim et al. [36]).

Parameter	Cancer: endemic balance prior to intervention	After treatment, the endemic equilibrium is free of cancer	The cancer endemic equilibrium following treatment
A	1	1	1
α	0.001	0.01	0.001
β	0.00005	0.00001	0.00005
μ	0.0002	0.003	0.0002
γ	0.001	0.001	0.001
δ	0.01	0.03	0.01
r	0.18	0.019	0.18
b	0.001	0.001	0.001
k	0.005	0.04	0.005
v	0	0.5	1
τ	0.001	0.006	0.05
θ	0.002	0.01	0.002

Solutions

Taking the LT on both sides in Eq. (3), we attain the succeeding:

$$\left.\begin{aligned} L\{C_{J^\theta}S(t)\} &= L\{A-\alpha S(t)-\beta S(t)C(t)\},\\ L\{C_{J^\theta}I(t)\} &= L\{\beta S(t)C(t)-\mu I(t)-\gamma C(t)I(t)\},\\ L\{C_{J^\theta}C(t)\} &= L\{r(1-bC(t))C(t)-kC(t)W(t)\},\\ L\{C_{J^\theta}W(t)\} &= L\{v+\delta C(t)-\tau W(t)-\theta W(t)C(t)\}. \end{aligned}\right\} \quad (6)$$

$$\left.\begin{aligned} S^\theta S(t)-S^{\theta-1}S(0) &= L\{A-\alpha S(t)-\beta S(t)C(t)\},\\ S^\theta C(t)-S^{\theta-1}C(0) &= L\{r(1-bC(t))C(t)-kC(t)W(t)\},\\ S^\theta I(t)-S^{\theta-1}I(0) &= L\{\beta S(t)C(t)-\mu I(t)-\gamma C(t)I(t)\},\\ S^\theta W(t)-S^{\theta-1}W(0) &= L\{v+\delta C(t)-\tau W(t)-\theta W(t)C(t)\}. \end{aligned}\right\} \quad (7)$$

$$\left.\begin{aligned} S^\theta S(t) &= S^{\theta-1}S(0)+L\{A-\alpha S(t)-\beta S(t)C(t)\},\\ S^\theta C(t) &= S^{\theta-1}C(0)+L\{r(1-bC(t))C(t)-kC(t)M(t)\},\\ S^\theta I(t) &= S^{\theta-1}I(0)+L\{\beta S(t)C(t)-\mu I(t)-\gamma C(t)I(t)\},\\ S^\theta W(t) &= S^{\theta-1}W(0)+L\{v+\delta C(t)-\tau M(t)-\theta M(t)C(t)\}. \end{aligned}\right\} \quad (8)$$

$$\left.\begin{aligned} S(t) &= S^{-1}S(0)+\tfrac{1}{S^\theta}L\{A-\alpha S(t)-\beta S(t)C(t)\},\\ I(t) &= S^{-1}I(0)+\tfrac{1}{S^\theta}L\{\beta S(t)C(t)-\mu I(t)-\gamma C(t)I(t)\},\\ C(t) &= S^{-1}C(0)+\tfrac{1}{S^\theta}L\{r(1-bC(t))C(t)-kC(t)M(t)\},\\ W(t) &= S^{-1}M(0)+\tfrac{1}{S^\theta}L\{v+\delta C(t)-\tau M(t)-\theta M(t)C(t)\}. \end{aligned}\right\} \quad (9)$$

Assuming that the solution $S(t)$, $I(t)$, $C(t)$,$W(t)$ are in form infinite series by

$$\begin{aligned} S(t)=\textstyle\sum_{n=0}^{\infty}S_n\,, I(t)=\sum_{n=0}^{\infty}I_n,\\ C(t)=\textstyle\sum_{n=0}^{\infty}C_n,\ W(t)=\sum_{n=0}^{\infty}W_n, \end{aligned} \quad (10)$$

and non-linear term involved in the model are $S(t)$, $I(t)$, $C(t)$, $W(t)$ are decomposed by Adomaim.

$$S(t)C(t)=\sum_{n=0}^{\infty}A_n,\ C(t)I(t)=\sum_{n=0}^{\infty}B_n, W(t)C(t)=\sum_{n=0}^{\infty}D_n$$

where A_n, B_n and C_n are Adomain polynomial given by

$$\begin{aligned} A_n &= \frac{1}{\Gamma(n+1)}\frac{d^n}{dt}\left[\sum_{k=0}^{n}\lambda^k S_k\sum_{k=0}^{n}\lambda^n C_k\right]\lambda,\\ B_n &= \frac{1}{\Gamma(n+1)}\frac{d}{dt}^n\left[\sum_{k=0}^{n}\lambda^k C_k\sum_{k=0}^{n}\lambda^n I_k\right]\lambda,\\ D_n &= \tfrac{1}{\Gamma(n+1)}\tfrac{d^n}{dt}[\textstyle\sum_{k=0}^{n}\lambda^k W_k\sum_{k=0}^{n}\lambda^n C_k]\lambda. \end{aligned} \quad (12)$$

Substitute Equation (10)-(12) in the Eq. (9) we get,

$$\begin{aligned} S(t) &= S^{-1}S(0)+\frac{1}{S^\theta}L\{A-\alpha S_n(t)-\beta A_n\},\\ I(t) &= S^{-1}I(0)+\tfrac{1}{S^\theta}L\{\beta A_n-\mu I_n(t)-\gamma B_n\},\\ C(t) &= S^{-1}C(0)+\tfrac{1}{S^\theta}L\{r(1-bC_n(t))C_n(t)-kD_n\},\\ W(t) &= S^{-1}W(0)+\tfrac{1}{S^\theta}L\{v+\delta C_n(t)-\tau W_n(t)-\theta D_n\}. \end{aligned} \quad (13)$$

Using initial condition in Equation (13) we have,

$$\begin{aligned} S(t) &= \frac{n_1}{S}+\frac{1}{S^\theta}L\{A-\alpha S_n(t)-\beta A_n\}\\ I(t) &= \frac{n_2}{S}+\frac{1}{S^\theta}L\{\beta A_n-\mu I_n(t)-\gamma B_n\}\\ C(t) &= \frac{n_3}{S}+\frac{1}{S^\theta}L\{r(1-bC_n(t))C_n(t)-kD_n\}\\ W(t) &= \tfrac{n_4}{S}+\tfrac{1}{S^\theta}L\{v+\delta C_n(t)-\tau W_n(t)-\theta D_n\} \end{aligned} \quad (14)$$

We iterate term in Eq (14) we get,

$$\sum_{n=0}^{\infty} S_{(n+1)}(t) = L^{-1}\left[\frac{1}{S^{\theta}}L\{A-\alpha S_n(t)-\ \beta A_n\}\right],$$
$$\sum_{n=0}^{\infty} I_{(n+1)}(t) = L^{-1}\left[\frac{1}{S^{\theta}}L\{\beta A_n-\mu I_n(t)-\gamma B_n\}\right],$$
$$\sum_{n=0}^{\infty} C_{(n+1)}(t) = L^{-1}[\frac{1}{S^{\theta}}L[\ \{r(1-bC_n(t))C_n(t)-kD_n\}\ ]] \quad (15)$$
$$\sum_{n=0}^{\infty} W_{(n+1)}(t) = L^{-1}\left[\frac{1}{S^{\theta}}L\{v+\delta C_n(t)-\tau W_n(t)-\ \theta\ D_n\}\right]$$

The followings were obtained from equation (15) we get,

$$S_{(0)} = n_1, I_{(0)} = n_2, C_{(0)} = n_3, W_{(0)} = n_4 \quad (16)$$

When $n = 0$, from first Eq. of (15)

$$\begin{aligned} S_1 &= L^{-1}\left[\frac{1}{S^{\theta}}L\{A-\alpha S_0(t)-\ \beta A_0\}\right], \\ &= L^{-1}\left[\left[\frac{1}{S^{\theta}}L\{A-\alpha n_1,-\beta n_1 n_3,\}\right]\right], \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\frac{1}{S}\{A-\alpha n_1-\ \beta n_1 n_3,\}\right], \\ &= L^{-1}\left[\{A-\alpha n_1-\ \beta n_1 n_3,\}\frac{1}{S^{\theta+1}}\right], \\ S_1 &= \{A-\alpha n_1-\ \beta n_1 n_3\}\frac{t^{\theta}}{\Gamma(\theta+1)} \end{aligned} \quad (17)$$

When $n = 0$ in second Eq. of (15) we get,

$$\begin{aligned} I_1 &= L^{-1}\left[\left[\frac{1}{S^{\theta}}L\{\beta A_n-\mu I_n(t)-\gamma B_n\}\right]\right], \\ &= L^{-1}\left[\frac{1}{S^{\theta}}L\{\beta n_1 n_3-\mu n_3-\gamma n_2 n_3\}\right], \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\frac{1}{S}\{\beta n_1 n_3-\mu n_3-\gamma n_2 n_3\}\right], \\ &= L^{-1}\left[\{\beta n_1 n_3-\mu n_3-\gamma n_2 n_3\}\frac{1}{S^{\theta+1}}\right], \\ I_1 &= (\beta n_1 n_3-\mu n_3-\gamma n_2 n_3)\frac{t^{\theta}}{\Gamma(\theta+1)} \end{aligned} \quad (18)$$

Also, when $n = 0$ in third Eq. of (15), we get

$$\begin{aligned} C_1 &= L^{-1}[\frac{1}{S^{\theta}}L[\ \{r(1-bC_n(t))C_n(t)-kD_n\}\ ]] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}L\{r(1-bn_3)n_3-kn_3n_4\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\frac{1}{S}\{r(1-bn_3)n_3-kn_3n_4\}\right] \\ &= L^{-1}\left[\{r(1-bn_3)n_3-kn_3n_4\}\frac{1}{S^{\theta+1}}\right] \\ C_1 &= \{r(1-bn_3)n_3-kn_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)} \end{aligned} \quad (19)$$

Again when $n = 0$ in 4th Eq. of (15)

$$\begin{aligned} W_1 &= L^{-1}\left[\frac{1}{S^{\theta}}L\{v+\delta C_0(t)-\tau W_0(t)-\ \theta\ D_0\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}L\{v+\delta n_3-\tau n_4-\ \theta\ n_3 n_4\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\frac{1}{S}\{v+\delta n_3-\tau n_4-\ \theta\ n_3 n_4\}\right] \\ &= L^{-1}\left[\{v+\delta n_3-\tau n_4-\ \theta\ n_3 n_4\}\frac{1}{S^{\theta+1}}\right] \\ W_1 &= \{v+\delta n_3-\tau n_4-\ \theta\ n_3 n_4\}\ \frac{t^{\theta}}{\Gamma(\theta+1)} \end{aligned} \quad (20)$$

Similarly, when $n = 1$ then equation (10) in first Equation we get

$$\begin{aligned} S_2 &= L^{-1}\left[\frac{1}{S^{\theta}}L\{A-\alpha S_1(t)-\ \beta A_1\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}L\left\{A-\alpha(A-\alpha n_1-\ \beta n_1 n_3)\frac{t^{\theta}}{\Gamma(\theta+1)}\right.\right. \\ &\quad -\beta\left(n_3(A-\alpha n_1-\ \beta n_1 n_3)\frac{t^{\theta}}{\Gamma(\theta+1)}+n_1\{r(1-bn_3)n_3\right. \\ &\quad \left.\left.\left.-kn_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)}\right)\right\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\left\{A-\alpha(A-\alpha n_1-\ \beta n_1 n_3)\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}\right.\right. \\ &\quad -\beta\left(n_3(A-\alpha n_1-\ \beta n_1 n_3)\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}+n_1\{r(1-bn_3)n_3\right. \\ &\quad \left.\left.\left.-kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}\right)\right\}\right] \\ &= L^{-1}\left[\frac{1}{S^{\theta}}*\left\{A-\alpha(A-\alpha n_1-\ \beta n_1 n_3)\frac{1}{S^{\theta+1}}\right.\right. \\ &\quad -\beta\left(n_3(A-\alpha n_1-\ \beta n_1 n_3)\frac{1}{S^{\theta+1}}+n_1\{r(1-bn_3)n_3\right. \\ &\quad \left.\left.\left.-kn_3n_4\}\frac{1}{S^{\theta+1}}\right)\right\}\right] \\ &= L^{-1}\left[\left\{A-\alpha(A-\alpha n_1-\ \beta n_1 n_3)\frac{1}{S^{2\theta+1}}\right.\right. \\ &\quad -\beta\left(n_3(A-\alpha n_1-\ \beta n_1 n_3)\frac{1}{S^{2\theta+1}}+n_1\{r(1-bn_3)n_3\right. \\ &\quad \left.\left.\left.-kn_3n_4\}\frac{1}{S^{2\theta+1}}\right)\right\}\right] \\ S_2 &= L^{-1}\left[\left\{A-\alpha(A-\alpha n_1-\ \beta n_1 n_3)\frac{t^{2\theta}}{\Gamma(2\theta+1)}-\beta\left(n_3(A-\alpha n_1\right.\right.\right. \\ &\quad \left.\left.\left.-\ \beta n_1 n_3)\frac{t^{2\theta}}{\Gamma(2\theta+1)}+n_1\{r(1-bn_3)n_3-kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\right)\right\}\right] \end{aligned} \quad (21)$$

Similarly, when $n = 1$ then equation (15) in 2nd equation we get

$$I_2 = L^{-1}\left[\left[\frac{1}{S^\theta}L\{\beta A_1 - \mu I_1(t) - \gamma B_1\}\right]\right]$$

$$I_2 = L^{-1}\left[\left[\frac{1}{S^\theta} * L\left\{\beta\left(n_3(A-\alpha n_1-\beta n_1 n_3)\frac{t^\theta}{\Gamma(\theta+1)} + n_1\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)}\right) - \mu(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{t^\theta}{\Gamma(\theta+1)} - \gamma n_3(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{t^\theta}{\Gamma(\theta+1)} + n_2(\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)}\right\}\right]\right]$$

$$I_2 = L^{-1}\left[\left[\frac{1}{S^\theta} * \left\{\beta\left(n_3(A-\alpha n_1-\beta n_1 n_3)\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}} + n_1\{r(1-bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}\right) - \mu(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}} - \gamma n_3(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}} + n_2(\{r(1-bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}\right\}\right]\right]$$

$$I_2 = L^{-1}\left[\left[\frac{1}{S^\theta} * \left\{\beta\left(n_3(A-\alpha n_1-\beta n_1 n_3)\frac{1}{S^{\theta+1}} + n_1\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}} +\right) - \mu(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{1}{S^{\theta+1}} - \gamma n_3(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{1}{S^{\theta+1}} + n_2\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}}\right\}\right]\right]$$

$$I_2 = L^{-1}\left[\left[\left\{\beta\left(n_3(A-\alpha n_1-\beta n_1 n_3)\frac{1}{S^{2\theta+1}} + n_1\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}} +\right) - \mu(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{1}{S^{2\theta+1}} - \gamma n_3(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{1}{S^{2\theta+1}} + n_2\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}}\right\}\right]\right]$$

$$I_2 = L^{-1}\left[\left[\left\{\beta\left(n_3(A-\alpha n_1-\beta n_1 n_3)\frac{t^{2\theta}}{\Gamma(2\theta+1)} + n_1\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\right) - \mu(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{t^{2\theta}}{\Gamma(2\theta+1)} - \gamma n_3(\beta n_1 n_3 - \mu n_3 - \gamma n_2 n_3)\frac{t^{2\theta}}{\Gamma(2\theta+1)} + n_2\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\right\}\right]\right] \quad (22)$$

Also, when $n = 1$ in third Eq. of (15) we get

$$C_2 = L^{-1}\left[\frac{1}{S^\theta} * L\left[\ \{r(1-bC_1(t))C_1(t) - kD_1\}\ \right]\right.$$

$$C_2 = L^{-1}\left[\frac{1}{S^\theta} * \left[\ \left\{r\left(1-b\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)} * \{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)}\right) - kn_4\{r(1-bn_3)n_3 - kn_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)} + n_3(\{v+\delta n_3 - \tau n_4 - \theta\, n_3n_4\}\frac{t^\theta}{\Gamma(\theta+1)})\right\}\right]\right]$$

$$C_2 = L^{-1}\left[\frac{1}{S^\theta} * \left[\ \left\{r\left(1-b\{r(1-bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}} * \{r(1-bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}\right) - kn_4\{r(1-bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}} + n_3(\{v+\delta n_3 - \tau n_4 - \theta\, n_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}})\right\}\right]\right]$$

$$= L^{-1}\left[\frac{1}{S^\theta} * \left[\ \left\{r\left(1-b\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}} * \{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}}\right) - kn_4\{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}} + n_3(\{v+\delta n_3 - \tau n_4 - \theta\, n_3n_4\}\frac{1}{S^{\theta+1}}\right\}\right]\right]$$

$$= L^{-1}[\left[\left\{r\left(1 - b\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}}\right.\right.\right.$$
$$* \{r(1-bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}}\Big) - kn_4\{r(1-bn_3)n_3$$
$$- kn_3n_4\}\frac{1}{S^{2\theta+1}} + n_3(\{v + \delta n_3 - \tau n_4$$
$$\left.\left.\left. - \theta n_3n_4\}\frac{1}{S^{2\theta+1}}\right\}\right]$$

$$= L^{-1}[\left[\left\{r\left(1 - b\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\right.\right.\right.$$
$$* \{r(1 - bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\Big)$$
$$- kn_4\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}$$
$$\left.\left.\left. + n_3(\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}\right\}\right] \quad (23)$$

Similarly, when $n = 1$ then equation (15) in fourth Equation we get

$$W_2 = L^{-1}\left[\frac{1}{S^{\theta}} * L\{v + \delta C_1(t) - \tau W_1(t) - \theta D_1\}\right]$$
$$W_2 = L^{-1}\left[\frac{1}{S^{\theta}} * L\left\{v + \delta(\, b\{r(1 - bn_3)n_3\right.\right.$$
$$- kn_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)}) - \tau\{v + \delta n_3 - \tau n_4$$
$$- \theta n_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)} - \theta n_4\{r(1 - bn_3)n_3$$
$$- kn_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)} + n_3(\{v + \delta n_3 - \tau n_4$$
$$\left.\left. - \theta n_3n_4\}\frac{t^{\theta}}{\Gamma(\theta+1)})\right\}\right]$$

$$= L^{-1}\left[\frac{1}{S^{\theta}} * \left\{v + \delta(\, b\{r(1 - bn_3)n_3\right.\right.$$
$$- kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}})$$
$$- \tau\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}$$
$$- \theta n_4\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}}$$
$$\left.\left. + n_3(\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{\Gamma(\theta+1)}{\Gamma(\theta+1)S^{\theta+1}})\right\}\right]$$

$$= L^{-1}\left[\frac{1}{S^{\theta}} * \left\{v + \delta(\, b\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}})\right.\right.$$
$$- \tau\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{1}{S^{\theta+1}}$$
$$- \theta n_4\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{1}{S^{\theta+1}}$$
$$\left.\left. + n_3(\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{1}{S^{\theta+1}})\right\}\right]$$

$$= L^{-1}\left[\left\{v + \delta(\, b\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}})\right.\right.$$
$$- \tau\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{1}{S^{2\theta+1}}$$
$$- \theta n_4\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{1}{S^{2\theta+1}}$$
$$\left.\left. + n_3(\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{1}{S^{2\theta+1}})\right\}\right]$$
$$= L^{-1}\left[\left\{v + \delta(\, b\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)})\right.\right.$$
$$- \tau\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}$$
$$- \theta n_4\{r(1 - bn_3)n_3 - kn_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)}$$
$$\left.\left. + n_3(\{v + \delta n_3 - \tau n_4 - \theta n_3n_4\}\frac{t^{2\theta}}{\Gamma(2\theta+1)})\right\}\right] \quad (24)$$

Similarly, the remaining terms can be derived, ultimately yielding the solution in the form of an infinite series be

$$S(t) = S_0 + S_1 + S_{22} + \cdots, I(t) = I_0 + I_1 + I_2 + \cdots,$$
$$C(t) = C_0 + C_1 + C_2 + \cdots, W(t) = W_0 + W_1 + W_2 + \cdots, \quad (25)$$

RESULTS AND DISCUSSION

We conduct a numerical simulation to gain deeper insights into the dynamics and control of T cell treatment. With the initial values $S_0 = 149.079$, $I_0 = 1.082$, $C_0 = 1.5971$, $W_0 = 252.26$, $\theta = \alpha$, and parameters given in Table 1, 2, the simulation's results were evaluated to produce the series solution of arbitrary order that follows.

$$S = 152.079 + \frac{1.75985t^{\alpha}}{\Gamma(\alpha+1)} - \frac{5.0453249775t^{2t^{\alpha}}}{\Gamma(2\alpha+1}, \quad (26)$$

$$I = 1.082 - \frac{0.039485t^{\alpha}}{\Gamma(\alpha+1)} - \frac{0.0000294916t^{2\alpha}}{\Gamma(2\alpha+1}, \quad (27)$$

$$C = 1.5971 - \frac{0.47729t^{\alpha}}{\Gamma(\alpha+1)} - \frac{.0035676t^{2\alpha}}{\Gamma(2\alpha+1}, \quad (28)$$

$$W = 252.26 - \frac{15.3218t^{\alpha}}{\Gamma(\alpha+1)} - \frac{5.705114t^{2\alpha}}{\Gamma(2\alpha+1}, \quad (29)$$

Numerical Simulation and Comparison of Results

This section will compare the numerical solutions determined using the RK4 method with the analytical solutions of the FO at CAR T cell therapy model for leukemia obtained using the LADM. T-cell parameters $\alpha = 0.001$, $\beta = 0.00005$, $\mu = 0.0002$, and $\gamma = 0.001$, $A = 1.5$, are using serve as the foundation for immune cell characteristics in the ensuing numerical simulations. It is assumed when determining the estimated value of Table 2. Here, Figure 2 illustrates the stability region of the FO system for $0 < \alpha < 1$, as determined by the criterion outlined in Eq. (8.1). Also Figure 3 represents the analytical

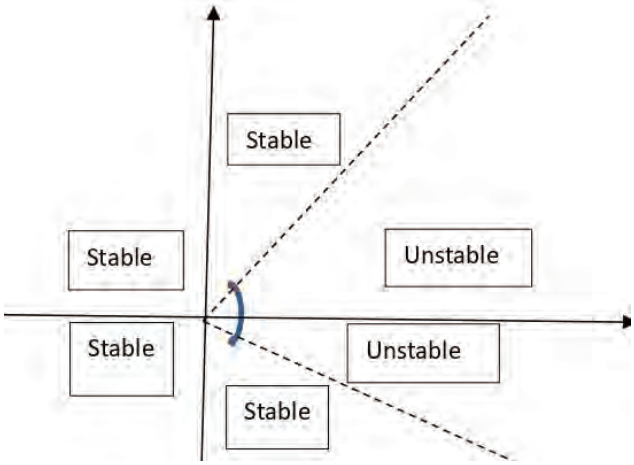


Figure 2. The stability region of the FO system for $0 < \alpha < 1$, as determined by the criterion outlined in Eq. (8.1)

solution for $S(t)$ and approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$. We execute simulation for the static ending time 50 days and plotted Figures 3-11 given below. Using the LADM, the study computed numerical results for S(t), I(t), C(t), and W(t). We have found that these results are studied in altered values of the FO parameter α. Numerical simulation of S(t), I(t), C(t), and W(t) plotted in Figures 3-11 given below. Over an interval of $0 < t < \alpha$ for various values of α=1, 0.75, 0.50, 0.25 respectively and all outcomes have been compared with the analytical solution of the considered problem. Figure 4, $I(t)$ highlight the significance of the effect of each parameter's sensitivity on the important reproduction quantity through two-dimensional plots that show the reactions to two different parameter values. Furthermore, Figure 5, $C(t)$ and, Figure 6, $W(t)$ are represents the analytical solution for approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$ respectively. Here, Figs. 3-6 illustrate that the fractional-order leukemia model offers greater flexibility, allowing for varied responses across different compartments of the proposed model (3). Notably, we assumed relatively small initial values, which justified the use of a short time interval. For extensive time intervals, larger initial values should be considered to ensure that the population values remain positive, and the reverse applies for shorter intervals. Figure 7 represents the plot comparing Susceptible, Infected, Cancer

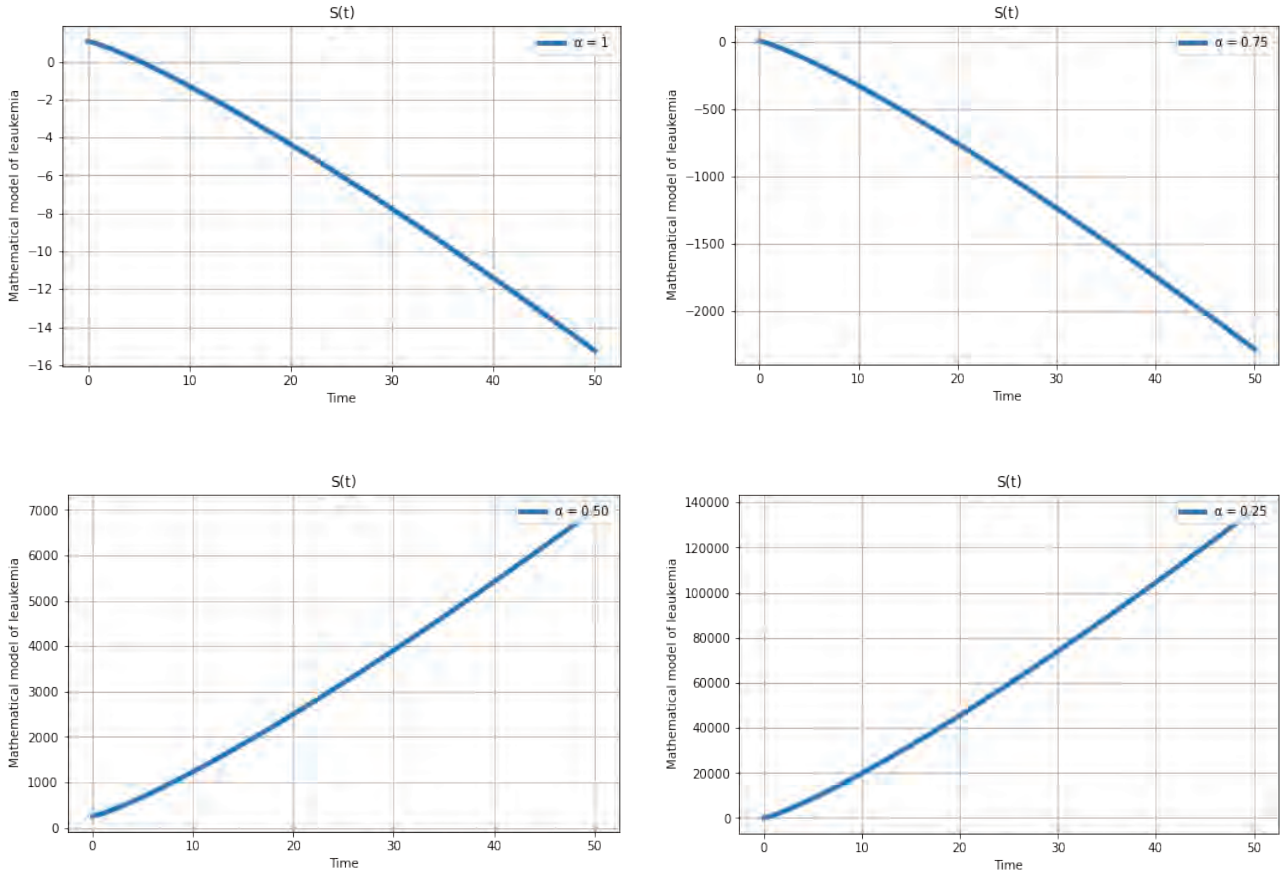


Figure 3. Analytical solution for $S(t)$ and approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$.

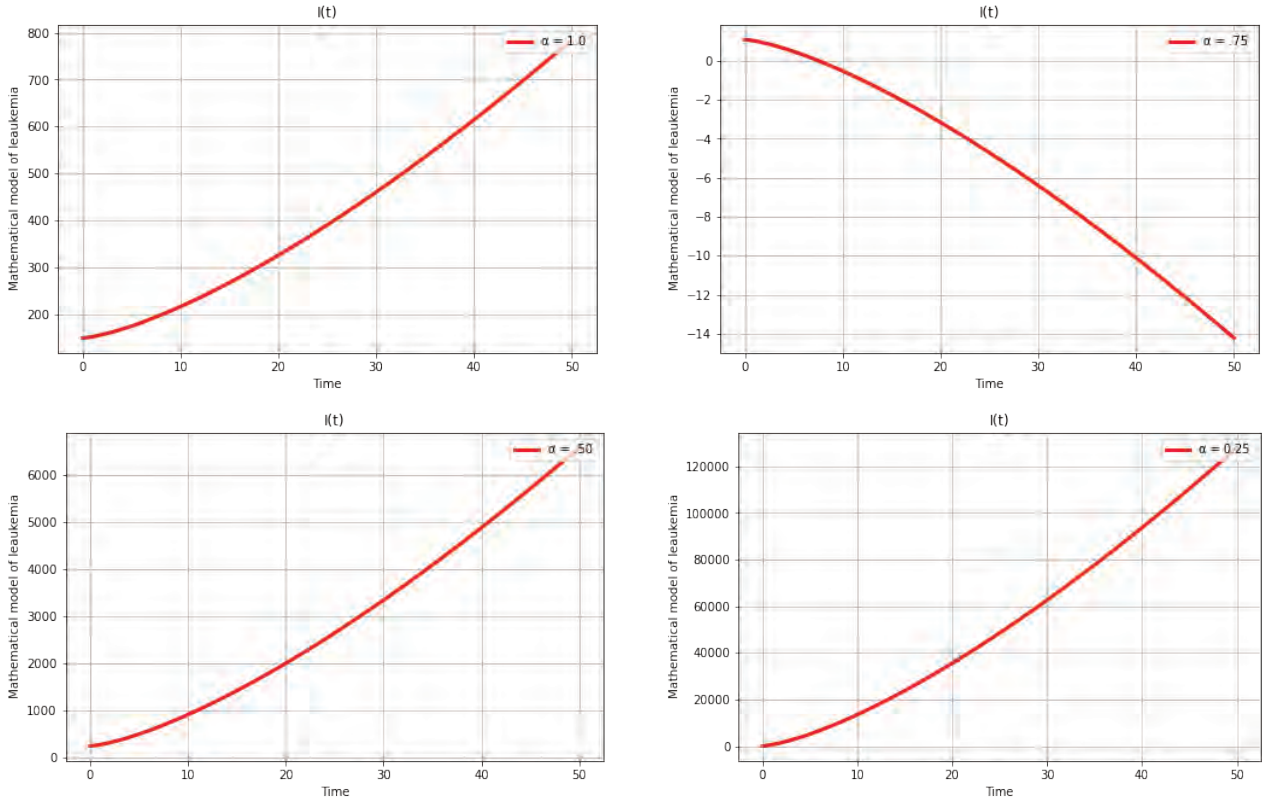


Figure 4. Analytical solution for $I(t)$ and approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$.

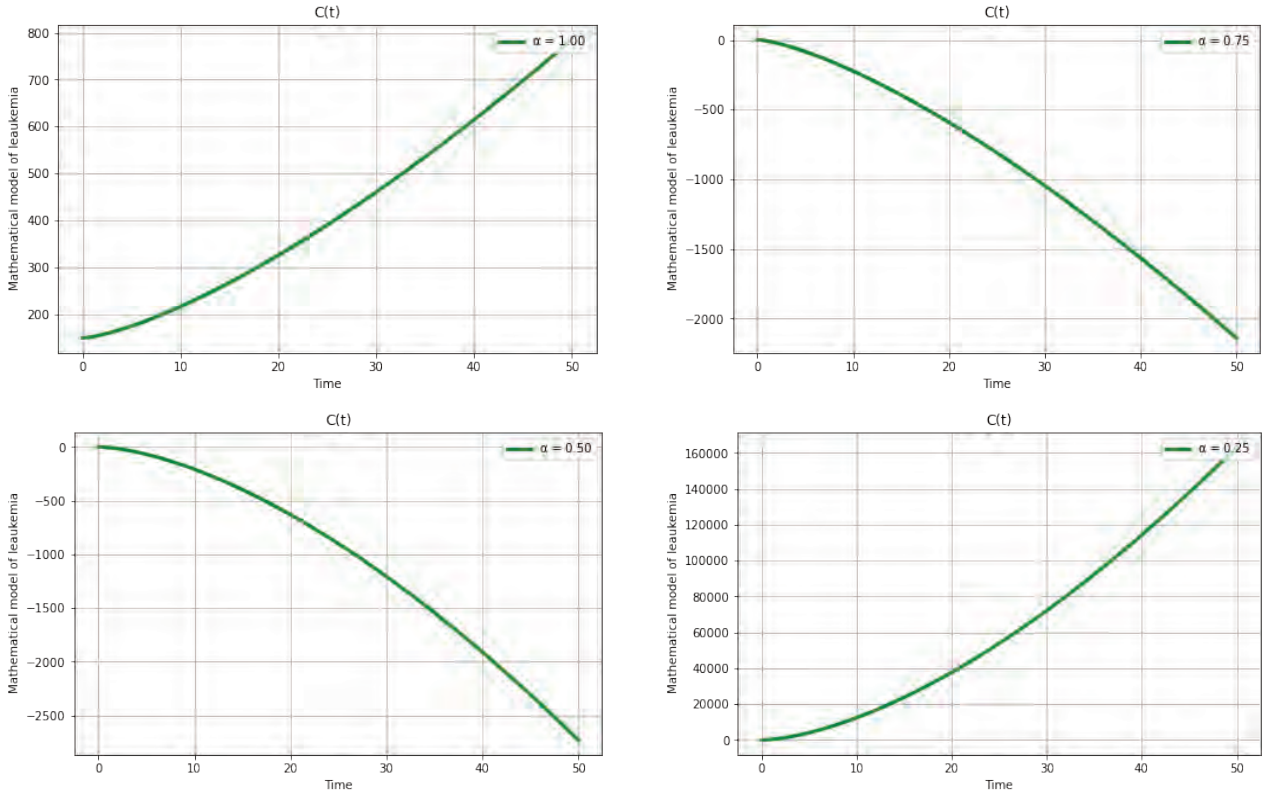


Figure 5. Analytical solution for $C(t)$ and approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$.

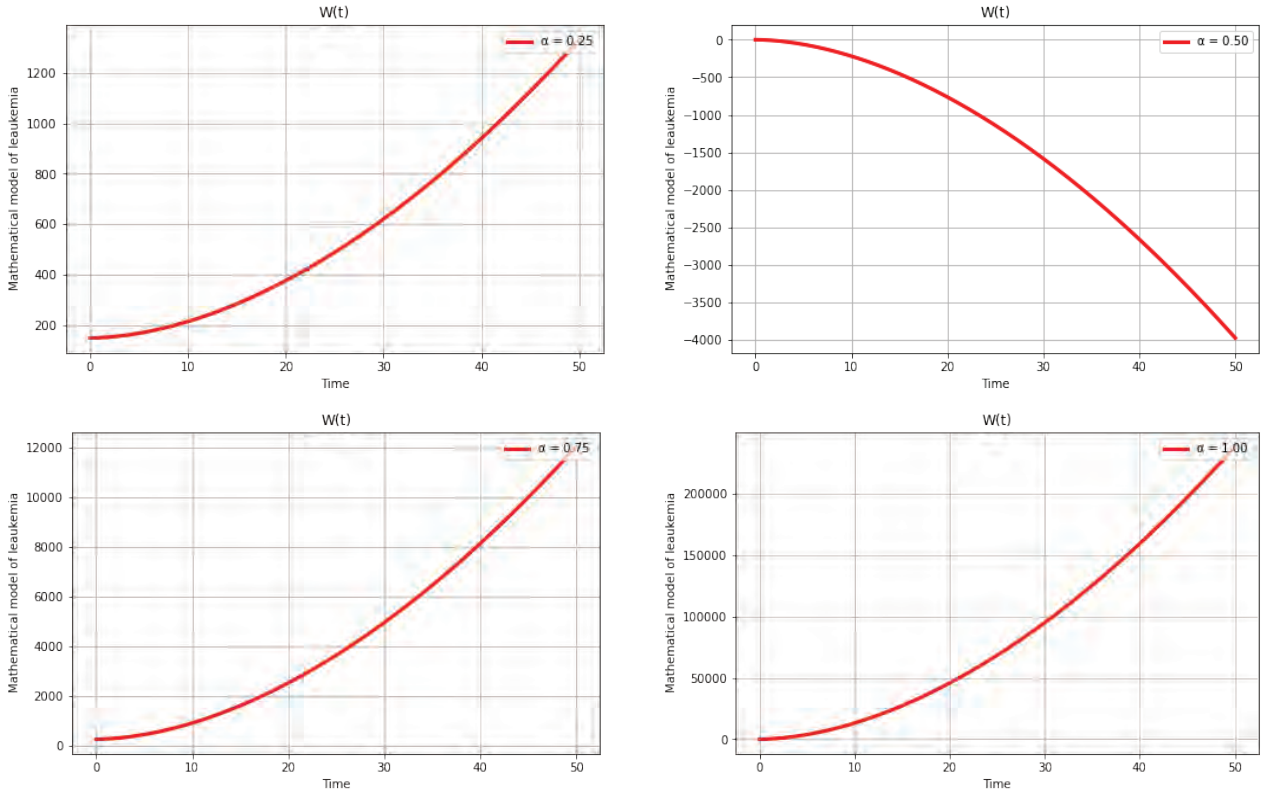


Figure 6. Analytical solution for $W(t)$ and approximate solutions obtained using the LADM for various fractional-order values of α within the range $0 < t < 1$.

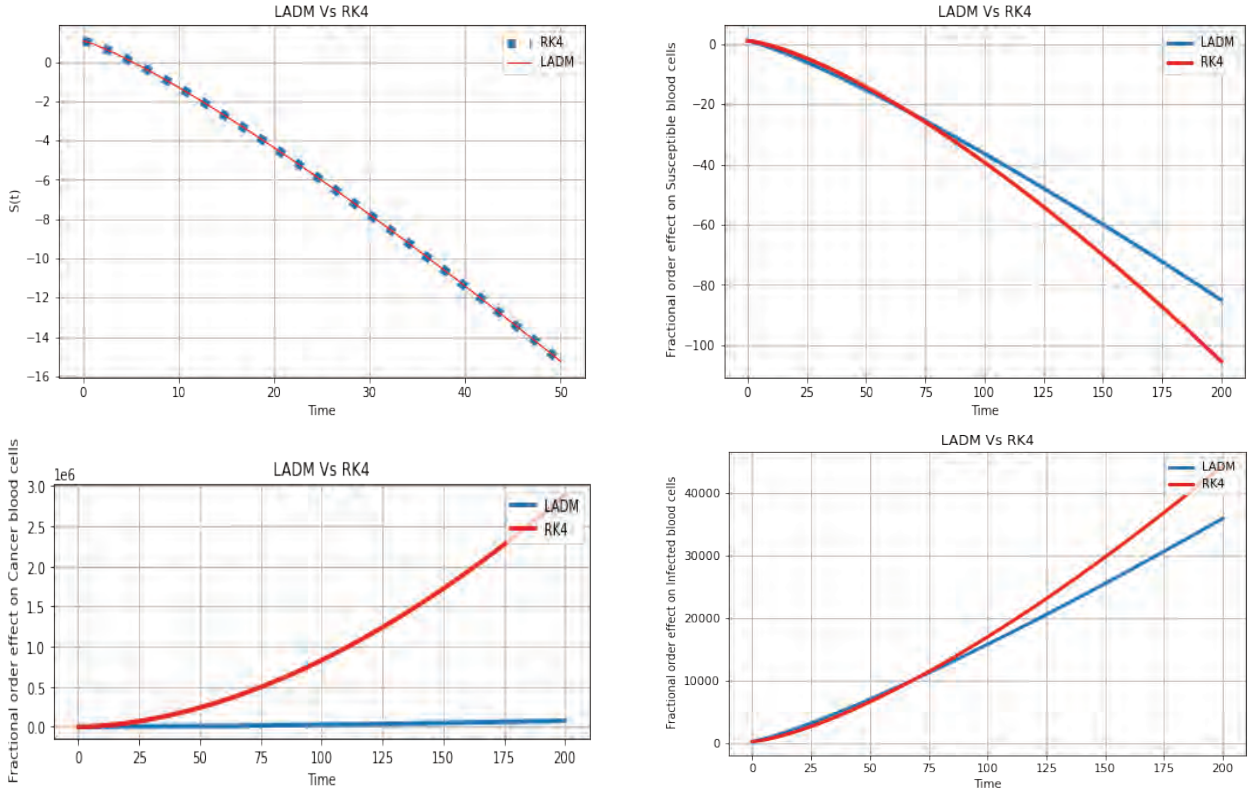


Figure 7. Plot comparing Susceptible, Infected, Cancer and immune blood cells LADM and RK4 method fractional order $\alpha_i = 1$ where $0 < t < 1$.

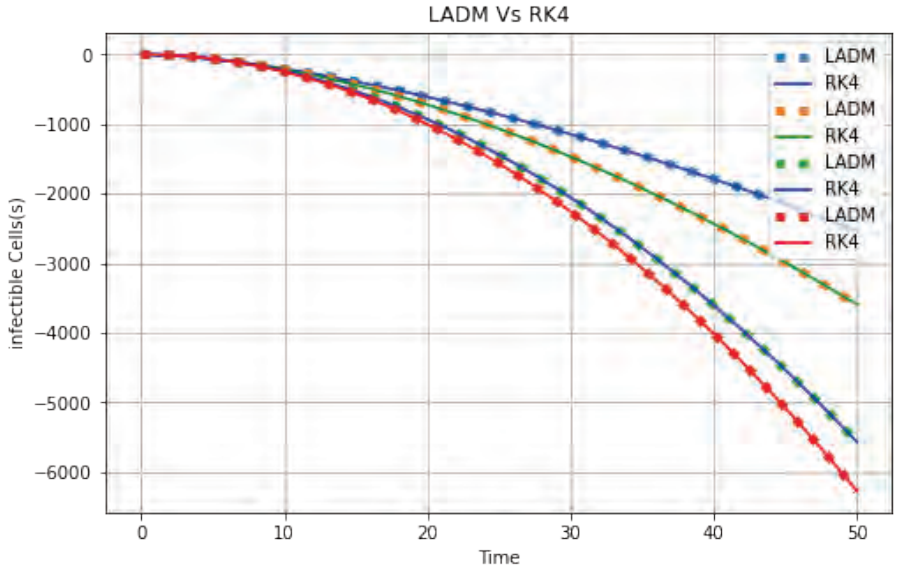


Figure 8. Plot comparing mathematical model of Leaukemia LADM and RK4 method FO $\alpha = 1$.

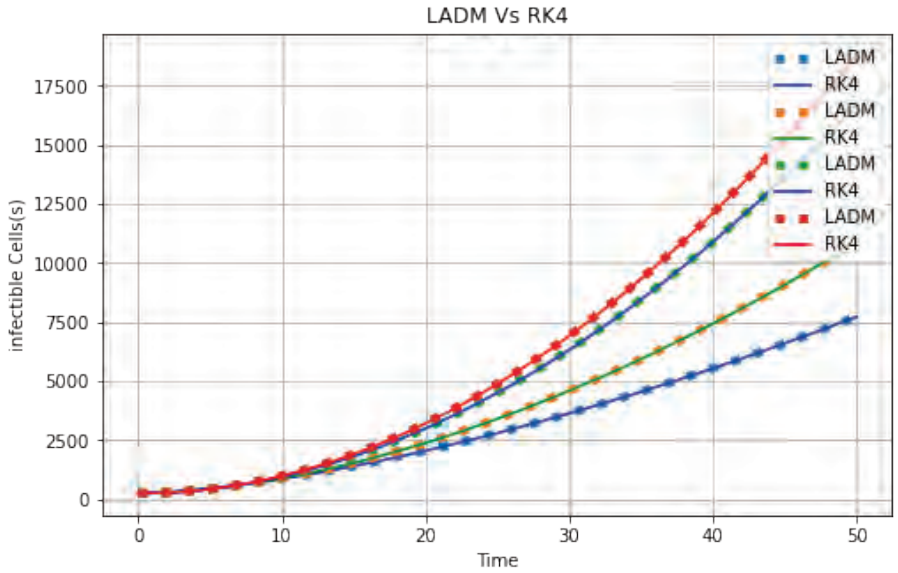


Figure 9. Plot comparing mathematical model of Leaukemia LADM and RK4 method FO $\alpha = 0.75$.

and immune blood cells LADM and RK4 method fractional order $\alpha_i = 1$ where $0 < t < 1$ Figure 8-11 present a comparison of the solutions obtained after three terms with those derived using the RK4 method for a classical order α. As shown in Figure 8 -11 represents the effect of the variation of parameter $\alpha = 1, 0.75, 0.50, 0.25$ respectively on the number S(t), I(t), C(t), and W(t). In this paper, red colors at dotted signed represents RK4 and blue colors represents LADM. We conclude that the analytical solution is much closer to numerical solutions. Hence, our proposed method is reliable and efficient. Numerical experiments are performed. The results indicate that, for the specified parameter values, the solutions generated by LADM using integer order align closely with those produced by RK4 at the given time. The model of leukemia dynamics within a host are investigated using a fractional-order model that emphasizes adaptive immune memory.

Stability Analysis

The disease-free equilibrium point of the leukemia model (1) is given as

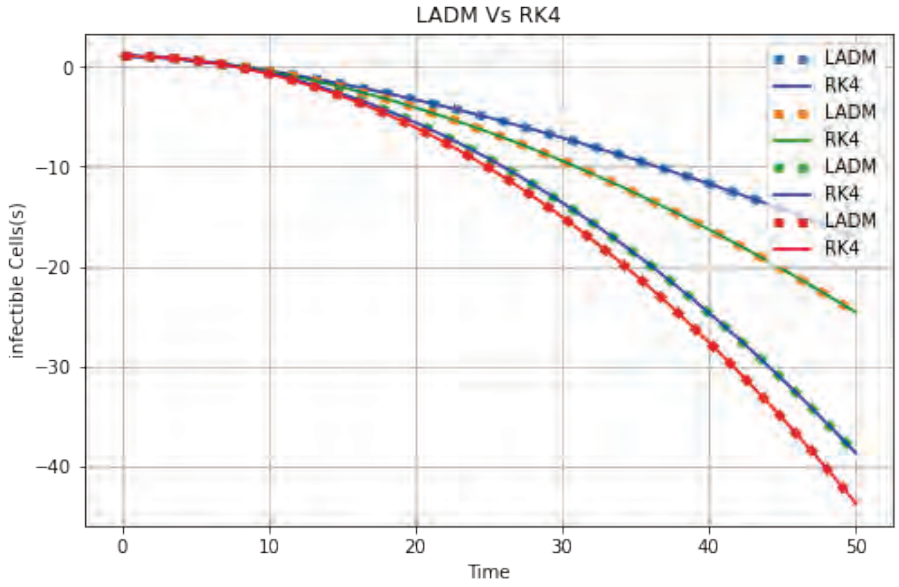


Figure 10. Plot comparing mathematical model of Leaukemia LADM and RK4 method FO $\alpha = 0.50$.

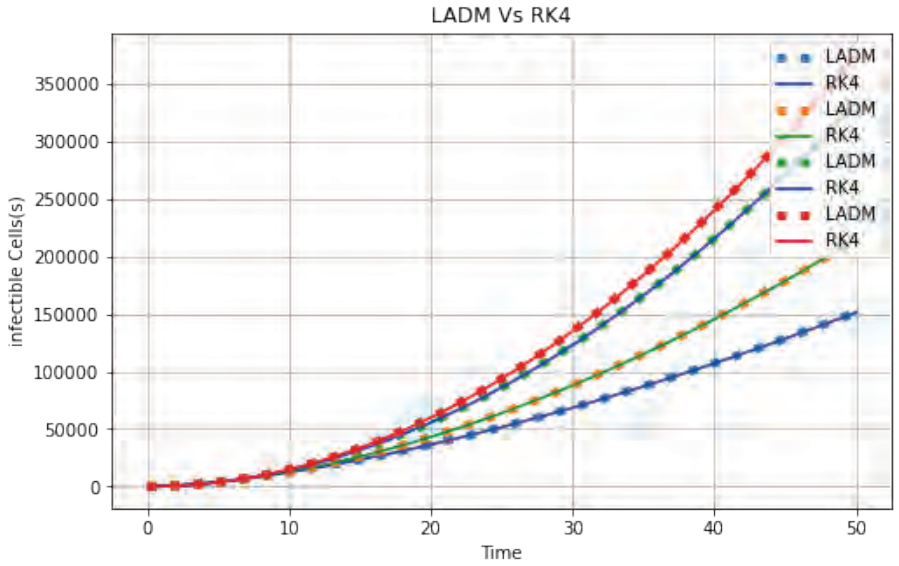


Figure 11. Plot comparing mathematical model of Leaukemia LADM and RK4 method FO $\alpha = 0.25$.

$P \equiv (\frac{A}{\alpha}, 0,0,0)$ and Endemic equilibrium with the innate immune response point

$$\left(\frac{Abr}{\alpha rb+\beta(r-kC^*)}, \frac{\beta Abr(r-kC^*)}{(\alpha rb+\beta(r-kC^*))((\mu rb+\gamma(r-kC^*))}, \frac{r-kC^*}{rb}, \frac{\delta k+r(\tau b+\theta) \pm \sqrt{\delta k+r(\tau b+\theta)^2-2\theta k\delta r}}{2\theta k}\right)$$

Theorem 1: Consider the following autonomous $\frac{d^\alpha v(t)}{dt^\alpha} = g(n(t); n(o); 0 < \alpha < 1$ nonlinear FO system:The following systems equilibrium points are solutions to the $g(n(t)=0$. If all eigen values (μ_j) of the jacobian matrix $J = \frac{\partial f}{\partial n}$ meet the condition n^* meet the condition

$$\left|arg(\boldsymbol{\mu_j})\right| > \frac{\alpha\pi}{2}, \tag{30}$$

Then the equilibrium point n^* is locally asymptotically stable.

Proof: Now, let us focus on the asymptotic stability of the endemic (positive) equilibrium for the model described

in (3). To analyze this, we evaluate the Jacobian matrix at the endemic equilibrium, which is expressed as follows:

The J is evaluated at the endemic equilibrium point P^*. [4] is given as:

$$J(P^*)=\begin{pmatrix} -\alpha-\beta C^* & 0 & -\beta C^* & 0 \\ \beta C^* & -\mu-\mu C & \beta S^*-\gamma I^* & 0 \\ 0 & 0 & r-2rbC^*-kW^* & -kC^* \\ 0 & 0 & \delta-\theta W^* & -\theta C^*-\tau \end{pmatrix} \quad (31)$$

Now from equation (8.2), we can write as:

$$\mu^4 + l_0\mu^3 + l_1\mu^2 + l_2\mu + l_3 = 0 \quad (32)$$

$$l_0 = V_1 + V_2 + V_3 + V_4 + V_5 + V_6,$$

$$l_1 = V_1V_2 + V_4V_6 + V_1V_4 + V_1V_6 + V_2V_4 + V_2V_6 + V_1pV_1^*,$$

$$l_2 = V_1V_4V_6 + V_2V_4V_6 + V_1V_2V_4 + V_1V_2V_6 + V_1V_1pC_1^* + V_1V_2pC_1^*,$$

$$l_3 = V_1V_2V_4V_6 + V_1V_2V_1pC_1^*,$$

$$V_1 = -a_1 - \beta C_1^*, V_2 = a_2 + \beta_1 C_1^*, V_3 = \beta S_1^* - \beta_1 I_1^*,$$

$$V_4 = (2peI_1^* + pW_1^* - p), V_5 = e - b_1W_1^*, V_6 = a_4 + \beta_1 C_1^*.$$

If $V(\varphi)$ denotes the discriminiant of polynomial $\varphi(\mu) = \mu^4 + l_0\,\mu^3 + l_1\,\mu^2 + l_2\,\mu + l_3$, where all the coefficients are real.

$$V(\varphi) = \begin{vmatrix} 1 & l_0 & l_1 & l_2 & l_3 & 0 & 0 \\ 0 & 1 & l_0 & l_1 & l_2 & l_3 & 0 \\ 0 & 0 & 1 & l_0 & l_1 & l_2 & l_3 \\ 4 & 3l_0 & 2\,l_1 & l_2 & 0 & 0 & 0 \\ 0 & 4 & 3l_0 & 2l_1 & l_2 & 0 & 0 \\ 0 & 0 & 4 & 3l_0 & 2l_1 & l_2 & 0 \\ 0 & 0 & 0 & 4 & 3l_0 & 2l_1 & l_2 \end{vmatrix} \quad (33)$$

We have the proposition [44].

Theorem 2. Assume that P^* exists in R^4

I. Proof: Put Ψ_1, Ψ_2, Ψ_3 are Routh-Hourwitz determinants:

$$\Psi_1 = l_0, \Psi_1 = \begin{vmatrix} l_0 & 1 \\ l_2 & l_1 \end{vmatrix}, \Psi_1 = \begin{vmatrix} l_0 & 1 & 0 \\ l_2 & l_1 & l_0 \\ 0 & l_3 & l_2 \end{vmatrix}.$$

When $\alpha = 1$, the equilibrium point P^* is locally asymptotically stable if

$$\Psi_1 > 0, \Psi_2 > 0, \Psi_3 = 0, l_3 > 0. \quad (34)$$

For all $\alpha \in [0,1)$, P^* is locally asymptotically stable, these (8.5) be sufficient but no essential. We will investigate the equation (8.1) from $n = 1$ to 4.

II. If $l_3 > 0$ be then point P^* is locally asymptotically stable.
III. If $V(\varphi) > 0$, $l_0 > 0$, $l_1 < 0$ and $\alpha > 2/3$ then P^* is equilibrium unstable.
IV. If $V(\varphi) < 0$, $l_0 < 0$, $l_1 > 0$, $l_2 < 0$, $l_3 > 0$, then P^* is equilibrium unstable.

Convergence Analysis

The obtained solution is a rapidly converging series that consistently approaches the exact solution. To verify the convergence of the series (15), we employ classical techniques, drawing on the approach outlined in [17].

Theorem 1: Let, R be a Banach space and N: $R \to R$ be a constructive nonlinear operator such that $\forall\ r, r^1 \in R$, $\| N(r) - N(r^1) \| \le p \| r - r^1 \|$, $0 < p < 1$. Then N has a unique point m such that $Nr = r$, where $r = (S, I, C, W)$. The series given in (15) can be written by LADM as: $R_y = NR_{y-1}, R_{y-1} = \sum_{i=1}^{y-1} R_i$, $y = 1,2,3 \ldots$., and assume that $r_0 \in B_j(r)$ where $B_j(r) = \{r^1 \in R: \|r - r^1\| < j\}$, t then we have

I. $r_y \in B_j(r)$;
II. $\lim_{y\to\infty} r_y = y$.

Proof: For (I), using mathematical induction for $y = 1$, we have

$$\| r - r^1 \| = \| N(r_0) - N(r) \| < p \| r_0 - r \|.$$

Put the result is true for k - 1 then

$\| r_0 - r \| \le p^{k-1} \| r_0 - r \|$. We have

$$\| r_0 - r \| = \| N(r_{k-1}) - N(r) \| \le p \| r_{k-1} - r \| \le p^k \| r_0 - r \|.$$

$$\| r_y - r \| \le p^y \| r_0 - r \| \le p^y j < j$$

which implies that $r_y \in B_j(r)$.

(II) since $\| r_y - r \| \le p^y \| r_0 - r \| \le p^y$ and $\lim_{y\to\infty} p^y = 0$, therefore we have $\lim_{y\to\infty} \| r_y - r \| = 0$ then $\lim_{y\to\infty} r_y = y$. Hence the theorem is verified.

CONCLUSION

We have successfully developed a scheme for numerical solutions mathematical model of CAR T-cell therapy of fractional order by using the Adomian decomposition method and the Runge-Kutta fourth order method. The results of the Laplace Adomian decomposition method and the Runge-Kutta fourth order method are compared with each other. Furthermore, the approximate solution by Laplace Adomian decomposition method is in complete agreement with the Runge-Kutta fourth order method for Figs.7-11 respectively. Further, this validates Laplace Adomian decomposition method for efficiency and accuracy in solving the proposed model for leukemia. In these simulations, the study demonstrates the impact on memory on invulnerable dynamics. The dynamic behavior of the structure was visualized as various parameters were varied using the LADM. Finally, this study made a significant insight into the T-cell therapy transmission dynamics. More compartments like as cytokine, vaccination may be added with the Mathematical model of leukemia fractional order differential equation to get more accurate solutions that can be treated as future research of this work.

AUTHOR CONTRIBUTIONS

The main study conception and design provided, Material preparation, Data collection provided by [RK].

Graphical Representation, Result discussion and analysis were performed by [RK], [MAA] and [ABK]. The first draft of the manuscript was written by [ABK], [RK], and [PD]. Conceptualization, research, methodology, supervision, visualization, and writing-review and editing-are all covered by [MAA],[ABK] and [RK].

CONFLICT OF INTEREST

The authors declare that they have no known conflicting financial interests or personal relationships that could have influenced the work disclosed in this publication.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Agarwal M, Bhadauria AS. Mathematical modeling and Analysis of Leukemia: Effect of External Engineered T Cells Infusion. Appl Applied Math Int J 2015;10:249–266.

[2] Altrock PM, Liu LL, Michor F. The mathematics of cancer: Integrating quantitative models. Nature 2015;15:730–745. [CrossRef]

[3] Mutlu B, Ozyoruk B. A research on mathematical model approaches in biomass supply chain. Sigma J Eng Nat Sci 2024;42:945–955. [CrossRef]

[4] Khumaeroh MS, Shalehah MA, Ilahi F. Mathematical model of Leukemia Treatment with Chimeric Antigen Receptor (CAR) T Cell Therapy. Mathline 2023;8:1077–1090. [CrossRef]

[5] Elsayad K, Oertel M, Haverkamp U, Eich HT. The effectiveness of radiotherapy for leukemia cutis. J Cancer Res Clin Oncol 2017;143:851–859. [CrossRef]

[6] Manoj P, Punith KB, Manjunath KC. Enhancing road safety through blackspots mitigative measures-A review. Sigma J Eng Nat Sci 2024;42:1670–1682. [CrossRef]

[7] June CH, O'Connor RS, Kawalekar OU, Ghassemi S, Milone MC. CAR T cell immunotherapy for human cancer. Science 2018; 359:1361–1365. [CrossRef]

[8] Khatun MS, Biswas MHA. Modeling the effect of adoptive T cell therapy for the treatment of leukemia. Comput Math Meth 2020;2:1–16. [CrossRef]

[9] Akdas E, Eren T. A sample application for scheduling search and rescue teams in an earthquake disaster. Sigma J Eng Nat Sci 2024;42:1555–1562. [CrossRef]

[10] Lee DW, Gardner R, Porter DL, Louis CU, Ahmed N, Jensen M, et al. Current concepts in the diagnosis and management of cytokine release syndrome. Blood 2014;124:188–195. [CrossRef]

[11] Gunerhan H, Kaabar MKA, Celik E. Novel analytical and approximate-analytical methods for solving the nonlinear fractional smoking mathematical model. Sigma J Eng Nat Sci 2023;41:331–343. [ABK], [RK],

[12] Gogoi UN, Saikia P, Mahanta DJ. Rapid parameter estimation of four non-linear growth models for analyzing the growth of Escherichia Coli. Sigma J Eng Nat Sci 2024;42:778–786. [CrossRef]

[13] Olayiwola MO, Alaje AI, Olarewaju AY, Adedokun KA. A caputo fractional order epidemic model for evaluating the efectiveness of high-risk quarantine and vaccination strategies on the spread of COVID-19. Healthc Anal 2023;3:100179. [CrossRef]

[14] Alaje AI, Olayiwola MO, Adedokun KA, Adedeji JA, Oladapo AO, Akeem YO. The modifed homotopy perturbation method and its application to the dynamics of price evolution in Caputo-FO Black Scholes model. Beni-Suef Univ J Basic Appl Sci 2023;12:93. [CrossRef]

[15] Olayiwola MO, Alaje AI, Yunus AO. A caputo fractional order financial mathematical model analyzing the impact of an adaptive minimum interest rate and maximum investment demand. Results Control Optim 2023;100349. [CrossRef]

[16] Alaje IA, Olayiwola MOO. A Fractional order MM for examining the spatiotemporal spread of COVID-19 in the presence of vaccine distribution. Healthc Anal 2023;4:100230. [CrossRef]

[17] Haq F, Shah K, Rahman G, Shahzad M. Numerical solution of fractional order smoking model via Laplace Adomian decomposition method. Alex Eng J 2018;57:1061–1069. [CrossRef]

[18] Yunus AO, Olayiwola MO, Adedokun KA, Adedeji JA, Alaje IA. Mathematical analysis of fractional-order Caputo's derivative of coronavirus disease model via Laplace Adomian decomposition method. Beni-Suef Univ J Basic Appl Sci 2022;11:144. [CrossRef]

[19] Dehingia K, Mohsen AA, Alharbi SA, Alsemiry RD, Rezapour S. Dynamical behavior of a FO model for withinhost SARS-CoV-2. Mathematics 2022;10:2344. [CrossRef]
[20] Cancer Today. Available at: https://gco.iarc.fr/today/home. Accessed on 17 Mar 2026.
[21] Arqub OA, Maayah B. Adaptive the Dirichlet model of mobile/immobile advection/dispersion in a time-fractional sense with the reproducing kernel computational approach: Formulations and approximations. Int J Mod Phys B 2023;37:2350179. [CrossRef]
[22] Maayah B, Arqub OA. Hilbert approximate solutions and fractional geometric behaviors of a dynamical fractional model of social media addiction affirmed by the fractional Caputo differential operator. Chaos, Solitons, Fract: X 2023;10:100092. [CrossRef]
[23] Maayah B, Arqub OA, Alnabulsi S, Alsulami H. Numerical solutions and geometric attractors of a fractional model of the cancer-immune based on the Atangana-Baleanu-Caputo derivative and the reproducing kernel scheme. Chin J Phys 2022;80:463–483. [CrossRef]
[24] Nuraini N, Tasman H, Soewono E, Sidarto KA. A with-in host dengue infection model with immune response. Math Comput Model 2009;49:1148–1155. [CrossRef]
[25] Yao SW, Arqub OA, Tayebi S, Osman MS, Mahmoud W, Inc M, et al. A novel collective algorithm using cubic uniform spline and finite difference approaches to solving fractional diffusion singular wave model through damping-reaction forces. Fractals 2023;31:2340069. [CrossRef]
[26] Shah K, Jarad F, Abdeljawad T, On a nonlinear FO model of dengue fever disease under Caputo–Fabrizio derivative. Alex Eng J 2020;59:2305–2313. [CrossRef]
[27] Borah M, Gayan A, Sharma JS, Chen Y, Wei Z, Pham VT. Is fractional-order chaos theory the new tool to model chaotic pandemics as COVID-19?. Nonlinear Dynam 2022;109:1187–1215. [CrossRef]
[28] Das K, Kumar GR, Ramesh K, Biswas HA. A Qualitative Analysis of Leukemia Fractional Order SICW Model. Jambura J Biomath 2024;5:46–53. [CrossRef]
[29] Ponalagusamy R, Murugesan K, Dhayabaran DP, Amirtharaj ECH. Numerical solution of heat flow problem by a combined method of rayleigh ritz with STWS and RKHM. Adv Model Anal A 2001;38:29–48.
[30] Ponalagusamy R. A Novel and Efficient Computational Algorithm of STWS for Generalized Linear Non-Singular/Singular Time Varying Systems. J Softw Eng 2008;2:1–9. [CrossRef]
[31] Ponalagusamy R, Senthilkumar S. A comparison of rk-fourth orders of variety of means and embedded means on multilayer raster CNN simulation. J Theor Appl Inf Technol 2007; 3.
[32] Ponalagusamy R, Alphonse PJA, Chandru M. Development of new fifth-order fifth-stage Runge Kutta method based on heronian mean. Int J Eng Sci 2011;2:162–197.
[33] Chandru M, Ponalagusamy R, Alphonse PJA. A New Fifth-Order Weighted Runge-Kutta Algorithm Based on Heronian Mean for Initial Value Problems in Ordinary Differential Equations. J Appl Math Inf 2017;35:191–204. [CrossRef]
[34] Ponalagusamy R, Senthilkumar S. A new fourth order embedded RKAHeM (4, 4) method with error control on single layer/raster cellular neural network. Springer 2009;3:303–305. [CrossRef]
[35] Atalan A, Donmez NFK, Donmez CC. Developing optimization models to evaluate healthcare systems. Sigma J Eng Nat Sci 2020;38:853–873.
[36] Karim, R., Akbar, M. A., Pk, M. B., & Dey, P. A study on fractional-order mathematical and parameter analysis for CAR T-cell therapy for leukemia using homotopy perturbation method. Partial Differential Equations in Applied Mathematics 2025; 14: 101152.
[37] Ozgur B, Demir A. The stability analysis of a neural field model with small delay. Sigma J Eng Nat Sci 2024;42:900–904. [CrossRef]
[38] Karim R, Bkar pk MA, Dey P, Akbar MA, Osman MS. A study about the prediction of population growth and demographic transition in Bangladesh. J Umm Al-Qura Univ Appl Sci 2024;11:91–103. [CrossRef]
[39] Maude SL, Laetsch TW, Buechner J, Rives S, Boyer M, Bader P, et al. Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. N Engl J Med 2018;3785:439–448. [CrossRef]
[40] Liao SJ. An approximate solution technique not depending on small parameters: a special example. Int J Non-Linear Mech 1995;30:371–380. [CrossRef]
[41] He JH. Homotopy perturbation technique. Comput Methods Appl Mech Eng 1999;178:257–262. [CrossRef]
[42] Ahmed E, El-Sayed AMA, El-Saka HA. On some Routh–Hurwitz conditions for fractional order differential equations and their applications in Lorenz, Rössler, Chua and Chen systems. Phys Lett A 2006;358:1–4. [CrossRef]
[43] Raghu A, Gajjela N, Garvandha M. MHD Maxwell dusty fluid in thermally stratified radiative flow with temperature-dependent thermal conductivity and Cattaneo-Christov model. Heliyon 2024;10:e30355. [CrossRef]
[44] Raghu A, Gajjela N, Aruna J, Niranjan H. Significance of modified Fourier heat flux on Maxwell hybrid (Cu-Al_2O_3/H_2O) nanofluid transport past an inclined stretching cylinder. J Therm Anal Calorim 2024;1–19. [CrossRef]
[45] Garvandha M, Gajjela N, Narla V, Kumar D. Thermodynamic entropy of a magnetized nanofluid flow over an inclined stretching cylindrical surface. J Therm Eng 2021;10:1253–1265. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 970–982, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2019

Research Article

A hybrid deep learning framework with attention mechanism for anti money laundering in cryptocurrency transactions

Emine CENGİZ[1,*], Murat GÖK[1]

[1]*Department of Computer Engineering, Yalova University, Yalova, 77200, Türkiye*

ARTICLE INFO

Article history
Received: 04 October 2024
Revised: 12 November 2024
Accepted 28 January 2025

Keywords:
Anti Money Laundering, Attention Mechanism; Bidirectional Long Short Term Memory; Convolution Neural Network; Deep Learning

ABSTRACT

Bitcoin and other cryptocurrencies are technological innovations that have transformed the world of finance. However, these developments introduce new risks. Because of the relative anonymity these systems offer, money laundering is one of the most significant risks connected to cryptocurrencies. This anonymity makes it more difficult to identify money obtained through illegal means, which in turn makes it possible for criminal operations to continue and grow. As a result, creating efficient techniques to identify and stop money laundering in cryptocurrency transactions has grown in importance as a research challenge. In this study, we suggest a hybrid model based on deep learning to detect illicit money transfers in cryptocurrency transactions. Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory Networks (BiLSTM), and an Attention Mechanism are all integrated in the suggested method. CNNs are used in the model's initial stage to extract significant features from the unprocessed data. The BiLSTM layer then received these features in order to identify the dependencies in the sequential data structures. The Attention Mechanism enhances the overall classification performance in the last step by giving the BiLSTM outputs importance weights. The Elliptic dataset was used to assess the suggested model's performance. The experimental results indicate that the model achieves superior performance compared to existing methods, with an accuracy of 97.9%, precision of 99.0%, recall of 98.0%, and F1-score of 98.0%. The findings of this study highlight the effectiveness of deep learning models enriched with Attention Mechanism for detecting illicit activities in cryptocurrency transactions. Beyond presenting a high performing model, its contribution to the literature lies in offering an novel approach to prevent crimes associated with the growing use of cryptocurrencies.

Cite this article as: Cengiz E, Gök M. A hybrid deep learning framework with attention mechanism for anti money laundering in cryptocurrency transactions. Sigma J Eng Nat Sci 2026;44(2):970–982.

***Corresponding author.**
*E-mail address: emine.cengiz@yalova.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Black money is defined as earnings obtained through illegal means. Money laundering represents the income obtained as a result of crimes such as drug trafficking, arm smuggling, and human trafficking, as if obtained from legal sources [1, 2]. Money laundering is the root cause of many crimes that use illicit money. It is difficult to estimate the total amount of money laundered each year because of the illegal nature of the transactions. However, [3] states that the revenue obtained through money laundering is between 800 billion and 2 trillion dollars worldwide. Money laundering poses a significant threat to global financial system [4]. Anti Money Laundering (AML) guidelines have been established to prevent harm caused by this activity. To combat money laundering, the European Union adopted the "Fifth Money Laundering Directive (5AMLD)." This law obliges maintaining customer history, monitoring transactions, and reporting any suspicious transactions to money laundering [5].

With the emergence of cryptocurrencies, money laundering evolved technologically. Cryptocurrency is a digital payment system designed in a virtual environment (without a bank) to verify and secure monetary transactions using cryptographic methods. Unlike paper money and financial systems, cryptocurrencies have a decentralized structure. The decentralization of each cryptocurrency originates from the blockchain structure, which acts as a database. It has been reported that money laundering, a criminal crime, is carried out through the blockchain [6, 7].

A rule based system is a basic method employed for detecting money laundering. It consists of a series of conditions that analyze whether specific events have occurred or if certain thresholds have been surpassed, indicating a suspicious state. However, this technique also has several disadvantages. One of its major disadvantages is that it leads to a high false positive rate and requires experts to create rules [8]. Machine learning (ML) methods, on the other hand, overcome the challenge of rule based systems by extracting complex models from historical data and can reduce high false positive and false negative rates [9]. With the emergence of big data, deep learning techniques can be considered as a solution to prevent money laundering [10].

Whereas feature extraction is performed manually in ML methods, the features of the input data are learned automatically in Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) [11]. Feature extraction plays a critical role in enhancing the performance of the proposed model and effectively analyzing the dynamic and complex nature of cryptocurrency transactions. Specifically, cryptocurrency transactions are often irregular, high dimensional, and contain time series data. When analyzed directly, such data can negatively impact both the learning process and the classification performance. Feature extraction allows the model to extract meaningful and summarized information from such complex, high dimensional data. The CNN model was applied to f synthetic financial transaction data by Kute et al. [12]. In terms of accuracy and false negative rates, the proposed approach performed better than other machine learning techniques like Random Forest (RF), Support Vector Machine (SVM), and XGBoost. The Shapley Additive exPlanations (SHAP) method was used to make the model outputs easier to understand. The visual analysis of each feature's contribution to the classification process was evaluated by anti-money laundering (AML) specialists. Rani et al. [13] suggested an LSTM-based model to detect and prevent money laundering. This model is designed to support the reporting of suspicious activity and the construction of client risk profiles. By modeling financial data as time series, it becomes possible to detect transaction patterns associated with money laundering. Bidirectional Long Short-Term Memory (BiLSTM) is a deep learning architecture that allows temporal dependencies to be learned from both forward and backward directions. The Attention Mechanism (AM) operates by modulating the contribution of hidden states within sequential representations, assigning greater emphasis to those deemed more informative [14]. In doing so, it reduces the impact of less relevant signals and directs the model's focus toward features that carry stronger discriminative value. Consequently, the learning process becomes more efficient and the overall classification performance improves. Tang et al. [15] proposed the BiLSTM4DPS model to detect phishing scam accounts on the Ethereum network. This model is based on a BiLSTM architecture enhanced with an Attention Mechanism. In their study, features such as transaction amounts and transaction counts were used to analyze the account behavior. This model converts account transaction records into sequences to extract the temporal and latent patterns of transactions. By integrating AM with masking techniques, a classification model is constructed to identify fraudulent accounts. Jainish and Alwin [16] proposed a deep learning based approach for financial fraud detection. The accuracy of fraud detection is improved by using Fisher score based feature selection and BiLSTM with AM. The model processes the data in matrix form to learn temporal dependencies and then focuses on important data through AM. A model for identifying unusual patterns in blockchain transactions was proposed by Wang et al. [17]. In this model, an inception structure was added to the CNN architecture to facilitate the learning of various patterns. To better account for the interactions among learned representations, a self-attention component was incorporated into the architecture, with the aim of refining overall model performance.

Although deep learning approaches have shown promise in the detection of money laundering, many existing implementations remain limited in their ability to represent the structural complexity and evolving dynamics of cryptocurrency transactions. Addressing this gap, the present work develops a hybrid CNN–BiLSTM–AM architecture that integrates convolutional feature extraction,

bidirectional sequence modeling, and an attention mechanism within a unified framework.

In this model, the extraction of local features using CNN and the extraction of sequence features using BiLSTM were combined. AM is then used to dynamically evaluate the importance of each feature in the sample input. The novelty of this study lies in the combination of deep learning techniques and AM to improve the detection of illegal cryptocurrency transactions. Graph based algorithms for AML detection were typically employed in earlier research. This study fills this gap in the literature by showing how deep learning can be effectively used in challenging areas, such as the detection of illegal activities.

The main contributions of this study are summarized as follows:

- In this study, after testing various models, including CNN, LSTM, BiLSTM, and their combinations, the CNN-BiLSTM-AM hybrid model is proposed. This model is designed to detect illegal money transfers and is enhanced by the AM.
- The hybrid model was tested on the Elliptic dataset and achieved superior results compared to existing methods in the literature.
- ·This study indicates that incorporating AM can meaningfully support AML efforts. These results offer insightful direction for further studies meant to reduce the dangers connected to the increasing use of cryptocurrencies.

The structure of this paper is as follows. Section 2 reviews previous studies and methods using an Elliptic dataset. Section 3 discusses the dataset, working principles, and the architecture of the proposed model. The functionality and interactions of each component of the model are described. Section 4 explains the model setup designed in this study and provides the experimental results and performance metrics. A comparison was conducted between the proposed model and alternative methods using an Elliptic dataset. Section 5 provides a discussion of the findings, implications, and limitations of this study. Finally, Section 6 summarizes the key findings of this study.

LITERATURE REVIEW

Recently, research in the field of AML has incorporated graph based algorithms and ML. In this section, a summary of the work conducted for AML using an Elliptic dataset is provided.

One kind of neural network intended for handling and evaluating graph-structured data is the Graph Neural Network (GNN). GNN are capable of handling dynamic structures made up of nodes and edges, whereas traditional neural networks usually work with fixed and regular data structures. GNN are able to model the structural relationships and connectivity patterns of a graph by learning the features of its nodes and edges through the use of neighborhood information. A vast variety of models based on different learning strategies are included in GNN architectures. These consist of the Graph Isomorphism Network (GIN), Graph Attention Network (GAT), Graph Convolutional Network (GCN), and GraphSAGE. GCN is a deep learning model used to recognize and predict patterns on graph based data structures. GCN is a structure adapted to graph data structures based on a CNN. First, Weber et al. [18] proposed an Elliptic dataset in their study and divided the dataset into 70% training and 30% test sets to detect illegal Bitcoin transactions. They used a GCN, Logistic Regression (LR), RF, and Multilayer Perceptron (MLP) to evaluate the dataset and perform binary classification. The results were compared for precision, sensitivity, and F1 score. The RF technique achieved precision, sensitivity, and F1 score of 0.95, 0.67, and 0.78, respectively. Alarab et al. [19] presented a new approach based on MLP and GCN to predict illegal transactions on the Elliptic dataset. In the proposed method, they obtained the result that the hidden representation of GCN derived features and a linear layer increased the performance of the model according to the study [18]. GAT is a deep learning model used to model the relationships between nodes in graph structured data. Pocher et al. [20] used the GCN and GAT methods. This is the first time that GAT has been applied to detect anomalies in Bitcoin. Their results showed that the GCN yielded better results than the GAT. GraphSAGE is a method used for node embedding learning in large scale graphs. Chen et al. [21] proposed a GraphSAGE model that can detect target node features from neighboring features and their time series features. They aimed to detect whether the target node was anomalous to its neighboring nodes. For this purpose, they computed the similarities between the features of neighboring nodes. They also used GraphSAGE to combine the features of neighboring nodes. Deep Graph Infomax (DGI) and GIN are two different GNN methods used for modeling graph data. Lo et al. [22] proposed Inspection-L, a self supervised GNN framework based on DGI and GIN. DGI is a general unsupervised learning approach for acquiring node representations in data with graph structure. The proposed Inspection-L method is based on the combined use of RF and DGI to detect illegal transactions. Additionally, the proposed model allows each node to access all graph structural models by capturing neighborhood information. They used the DGI along with the recommended GIN to learn node placements. Node placements were then evaluated using the RF algorithm for the training and test sets.

Deep learning is an effective tool for solving complex and multidimensional problems such as detecting and preventing money laundering. To address the spatiotemporal complexity of money laundering transactions, Wan and Li [23] proposed a deep learning based prediction model called MDGC-LSTM. The model aims to capture multilayer relationships within transaction data by combining a Dynamic Graph Convolutional Network and LSTM network. It achieved a 25% higher Macro-F1 score compared to the best existing models on the Elliptic dataset. Yang et

al. [24] proposed two methods to solve AML problems. The first method combines the LSTM-GCN algorithms, and the second method uses ensemble learning. A multi-label classification model is trained using the LSTM-GCN algorithm, which improves the detection accuracy of illegal transactions. In the second method, multiple models are trained to identify and separate the illegal transaction data. Xia et al. [25] proposed a hybrid spatiotemporal prediction model based on GCN and LSTM, called MGC-LSTM. Elliptic dataset is a structure consisting of 49 different time steps. The 49 time steps were learned independently and the output of the previous step served as the input for the next step. LSTM was used to obtain the temporal dependence of the dataset, and GCN was used to learn the spatial dependence of its processes. The results trained by the LSTM were provided as inputs to the GCN. Xiao et al. [26] presented a novel method for tuning GCN network parameters, called CTDM, which integrates EvolveGCN with Minimal Gated Unit (MGU). MGU was developed as an alternative to LSTM cells. MGU uses fewer input gates than LSTM cells. In addition, MGU offers a simpler structure by directly applying an additional logistic value to the memory cell and the forget gate. CTDM uses the MGU to improve the parameters of the GCN and requires fewer learning parameters through the MGU. Guo et al. [27] proposed LB-GLAT to capture the topological structure and attribute features of money laundering. LB-GLAT was designed to successfully identify the target of blockchain transactions using a transaction graph and an inverse transaction graph. The problem of over smoothing was reduced using a long term layer attention mechanism. They obtained an accuracy of 97.0%, precision of 93.0%, recall of 84.0%, F1-score of 88.87% and AUC performance of 0.98. Wang et al. [28] proposed the RMGANets method to overcome the limitations of GCN. This method is a reinforcement learning based approach that incorporates multi relational attention graph awareness. In their study, they constructed a large multi relational graph that includes transactions, addresses, users, and cash flows as nodes. Reinforcement learning was utilized to complete the missing node information caused by masking operations and to emphasize the differences between the various types of nodes.

Label scarcity refers to the fact that not all or some examples in the training dataset have class labels. Lorenz et al. [9] studied the lack of labels in an Elliptic dataset. In the dataset, 21% of Bitcoin transactions are labeled as legal and 2% as illegal. The remaining transactions do not contain tag data. The number of labels was increased using Active Learning to improve the performance of unsupervised learning. The classification was performed using XGBoost (76%), LR (45%), and RF (83%). Despite employing a distinct approach to address this issue, the authors did not intend to achieve superior outcomes. Monte Carlo dropout (MC-dropout) is a method used in uncertainty estimation for ML and especially deep learning models. This method is related to a technique known as dropout, which is used to organize a network during learning. Alarab et al. [29] and Alarab and Prakoonwit [30] applied the MC-Dropout method in their studies. In [30], an active learning framework for analyzing the Bitcoin dataset was introduced. They used active learning called Monte-Carlo Dropout (MC-Dropout) and Monte-Carlo Based Adversarial Attack (MC-AA) to calculate uncertainties. They later proposed a temporal-GCN by combining LSTM and GCN models. The results showed that they achieved an accuracy of 97.7%. Jatoth et al. [31] analyzed feature selection to improve classification. The purpose of feature selection is to select fewer or more appropriate features in order to obtain a better model. They applied ensemble and classical supervised learning methods to classify the legal and illegal transactions. Consequently, they achieved better results in community learning models.

A review of AML studies highlights the increasing complexity of money laundering methods and the significant challenges posed by the anonymity of virtual currency transactions to existing methods. The lack of labeled data makes this problem worse and restricts the effectiveness of supervised learning-based models [23]. A range of methods such as GCN, LSTM, and various hybrid or graph-based architectures has been examined in prior research. However, these models often have limited scalability, interpretability, and adaptability to the dynamic nature of cryptocurrency transactions.

This study proposes a novel hybrid model that combines CNN, BiLSTM, and an AM in order to get around these limitations. The attention component directs the model toward the most salient features during classification, combining convolutional feature extraction and bidirectional temporal modeling instead of depending on a single modeling paradigm. By effectively learning both temporal dependencies and structural relationships, the proposed model aims to increase the reliability and accuracy of illicit transaction detection. This method shows a strong performance on complex and large datasets by addressing the shortcomings of previous approaches. This study contributes a significant innovation to the AML field by filling critical gaps in the literature.

MATERIAL AND METHODS

In this section, the dataset used in this study is described. An overview of the existing CNN, BiLSTM, and AM approaches utilized to perform the work is then provided. Finally, the proposed method is described.

Dataset

Elliptic dataset used in this study is a Bitcoin graph, where nodes represent transactions and edges represent flows between transactions [18]. This dataset is recognized as one of the most extensively labeled datasets available for cryptocurrencies [32]. It comprises 49 graph structures sampled from the Bitcoin blockchain at various time

intervals. The dataset consists of 203,769 node processes and 234,355 edges. Each transaction is divided into three classes: legal (e.g., wallet providers, exchanges, legitimate services, and miners), illegal (e.g., terrorist organizations, ransomware, and Ponzi schemes), and unknown. Of the total number of transactions, 2% (4,545) were labeled as illegal and 21% (42,019) as legal. The remaining 77% (157,205) as unknown. The dataset comprises 166 features, with the first 94 features relating to transaction specific information, such as input number, output number, transaction fee, and time step. The remaining 72 features relate to the aggregate information about the direct neighbors of the process, providing the correlation coefficient, standard deviation, and maximum and minimum values of each process.

Convolutional Neural Networks

CNN is a widely utilized artificial neural network architecture within the domain of deep learning [33]. It uses special layers such as convolution and pooling to understand the structure and properties of data. CNN is composed of convolution, activation, pooling, and fully connected layers [34]. The convolution layers enable the generation of feature maps by applying sliding filters over the input data [35]. These feature maps represent different attributes of the input data. Activation layers apply a specific activation function to each generated feature map.. This activation function emphasizes or suppresses the significance of certain features in each feature map based on weights. Pooling layers were used to decrease the dimensions of the feature maps [36]. In practice, pooling operations rely on maximum or average sampling strategies to condense the feature maps. This reduction in dimensionality lowers the number of parameters that must be learned, easing computational requirements while also mitigating the tendency toward overfitting. Once the feature representations have been smoothed in this way, they are passed to fully connected layers, where they are processed within a conventional neural network structure for final classification or prediction. These layers ultimately generate the final outputs by leveraging the features learned by the CNN in earlier stages.

Bidirectional Long Short Term Memory

Recurrent Neural Networks (RNN) are deep learning models that operate based on prior knowledge of data [37]. The basic component of RNN is their cell structure. The cells process the inputs and previous outputs while maintaining their internal states. The RNN uses these cells at each time step, and the outputs of these cells are provided as inputs for the next step. Feedback algorithms are another important component of RNN. This algorithm identifies the source of the error and optimizes network parameters. Feedback is important to reduce errors during network training. Another RNN variation is the LSTM network [38]. LSTM uses sequential data as the input and makes predictions by remembering the information held at the previous time steps. Thus, LSTM is highly successful at processing complex sequential data. The LSTM cell consists of a mechanism comprising a forget, input gates, output gates, and cell state. Figure 1 shows the LSTM structure.

Forget gate in a RNN is responsible for determining which information from the previous hidden layer and current input will be retained or forgotten. This is achieved by passing the combined information from the previous hidden layer and current input through a sigmoid function.

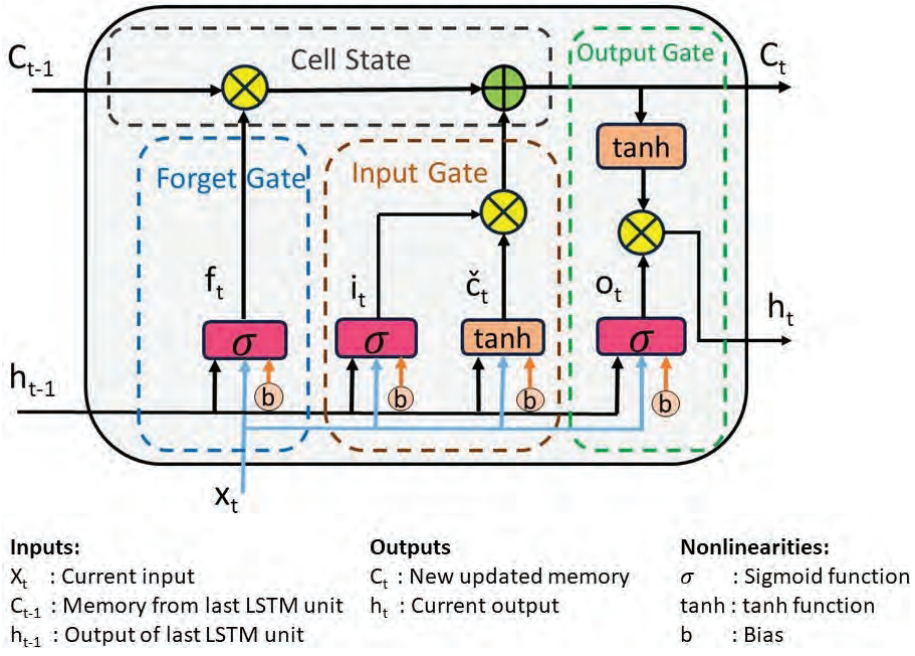


Figure 1. LSTM structure.

Information from the input. X_t and the previous output h_{t-1} are passed through the sigmoid function. The sigmoid function outputs a value between 0 and 1, where a value close to 0 indicates that the information will be largely forgotten and a value close to 1 indicates that the information will be retained. The forget gate [39, 40] f_t is given by Equation 1.

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \quad (1)$$

Input gate is used to update the cell state. First, a sigmoid function is applied, and the information that will be retained is determined. The output of the sigmoid function is reduced between -1 and 1 using the tanh function. This process normalizes the output of the cell and helps the network operate more stably. The cell state was updated by multiplying the two values obtained. This new memory was added to the C_{t-1} memory to obtain the C_t. In Equation 2, the sigmoid function, and in Equation 3, the tanh function steps are given.

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\check{C}_t = tanh\,(W_C . [h_{t-1}, x_t] + b_C \quad (3)$$

The most important task for the Cell State in a cell is to carry information. It receives data that must be passed on to the end of the cell, and then to other cells. First, the Forget Gate output was multiplied by the output of the previous layer. Next, it is added to the value from the input gate. These operations update the state of the cell and allow relevant information to be transmitted to the next cell. The new cell state [39, 40] C_t is given by Equation 4.

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t \quad (4)$$

Output Gate determines the value to be sent to the next layer. This value was used for the prediction. At this stage, the previous output (value) and the current input are fed into a third sigmoid function. The Cell State value passes through the tanh function. The output of the tanh function and the result of the sigmoid function were multiplied. The resulting value goes to the next layer as the "previous value." Equation 5 shows the output cell state o_t and Equation 6 shows the output value h_t [39, 40].

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o \quad (5)$$

$$h_t = o_t * tan\,h(C_t) \quad (6)$$

BiLSTM is an improved version of the LSTM algorithm [41]. Although the LSTM architecture is unidirectional, the BiLSTM architecture is bidirectional. BiLSTM consists of two LSTM layers: one processing the input in the forward direction, and the other in the backward direction. LSTM can make predictions using historical data, whereas BiLSTM can make predictions using historical and future data. BiLSTM network structure consists of four layers: Input Layer, Forward Layer, Backward Layer and Output Layer. Figure 2 shows the structure of the BiLSTM.

In BiLSTM, the input data sequence (x_{t-1}, x_t, x_{t+1}) is used as the input for both forward and backward LSTM. At each step t, the forward LSTM hidden state ($\vec{h}_t$) and backward LSTM hidden state ($\overleftarrow{h}_t$) were calculated simultaneously. The output h_t of the BiLSTM is calculated using both $\vec{h}_t$ and $\overleftarrow{h}_t$ information. Equations [7-9] calculate the values of $\vec{h}_t$, $\overleftarrow{h}_t$ and h_t [42].

$$\vec{h}_t = f((w_1 x_t + w_2 \vec{h}_{t-1})) \quad (7)$$

$$\overleftarrow{h}_t = f((w_3 x_t + w_5 \overleftarrow{h}_{t+1})) \quad (8)$$

$$h_t = g((w_4 \vec{h}_t + w_6 \overleftarrow{h}_t)) \quad (9)$$

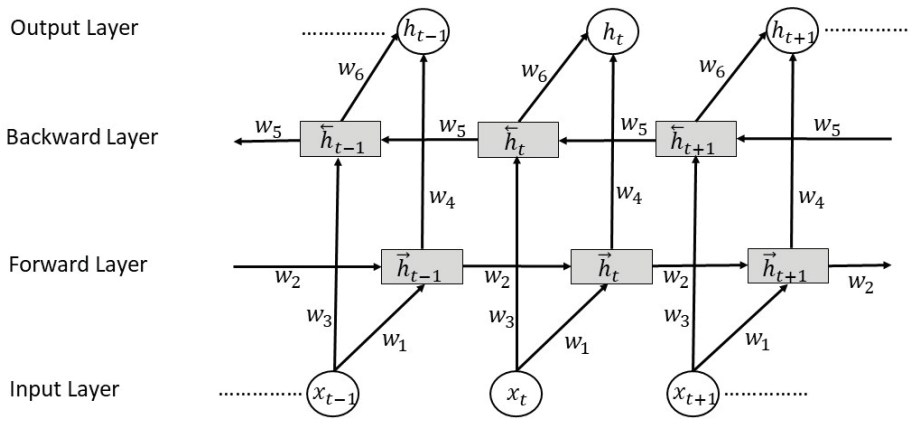


Figure 2. BiLSTM structure.

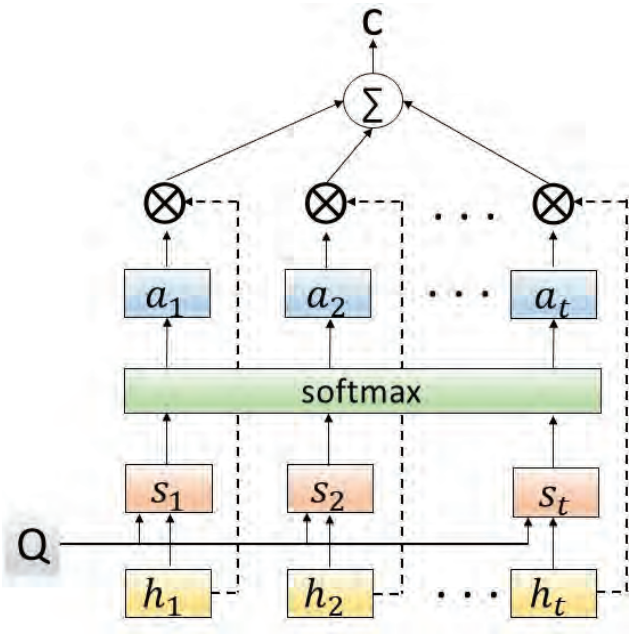


Figure 3. AM structure.

Attention Mechanism

AM refers to a mechanism in deep learning models that allows them to focus on a particular input [43]. This makes the model better able to perform a particular task, and helps it use fewer computational resources. As shown in Figure 3, the AM computational process consists of three stages.

1. Attention score (s_t): The similarity or relationship between input and output features is measured by the attention score. The model uses this score to decide which input features to prioritize. Thus, the attention score emphasizes which input features are more important to the model when generating the output. The attention score [44] is calculated using Equation 10.

$$s_t = tanh\,(W_h h_t + b_h) \tag{10}$$

Where W_h, h_t and b_h are the weight, input vector and deviation of the AM, respectively.

Attention scores determine which input items the model prioritizes, allowing the model to focus on its resources. Inputs associated with lower attention scores are treated as less influential, whereas those with higher scores receive greater emphasis during processing. This enables the model to concentrate on salient information while minimizing the impact of irrelevant, noisy, or redundant inputs.

2. Attention weight (at): Indicates the importance of a particular input item or component, and the attention weight value is calculated using Equation 11 [45].

$$a_t = \text{softmax}\big(s(h_t, Q)\big) = \frac{\exp(s(h_t,Q))}{\sum_t \exp(s(h_t,Q))'} \tag{11}$$

Softmax is commonly used to calculate attention weights. The sum of these weights, which have values ranging from 0 to 1, is 1. This property allows the mechanism to evaluate different input features systematically and assign stronger importance to those most relevant to the decision-making process.

Context vector (c): Context vector enables the model to focus on the input elements that are considered more important. Critical information in the data can be more effectively represented in this way. The context vector is obtained by combining the input elements weighted by their corresponding attention weights [46].

$$c = \sum_t a_t h_t \tag{12}$$

Proposed Model

In this study, the CNN-BiLSTM-AM model is proposed, which combines the features of CNN, BiLSTM, and AM to detect licit and illicit transactions in the Elliptic dataset. The structure of the CNN-BiLSTM-AM model is illustrated in Figure 4. This model consists of the input layer, CNN layer, BiLSTM layer, AM layer and output layer.

In the following, each block of the proposed method is described in detail:

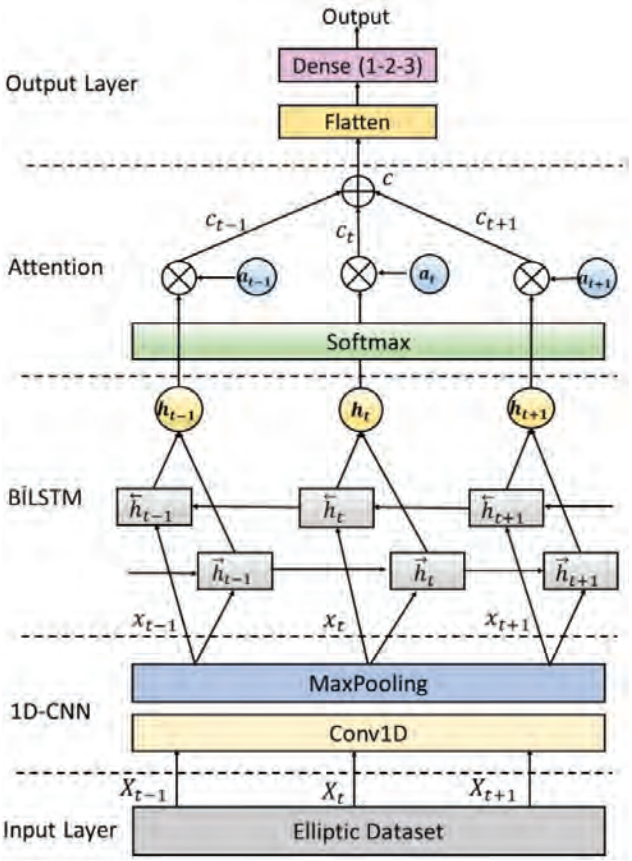


Figure 4. Proposed model.

Input Layer represents the data of cryptocurrency transactions. The data is expressed as time series and is represented as (X_{t-1}, X_t and X_{t+1}) over a specific time period.

CNN is utilized to extract features from the time series data. To extract important features from the input data, sliding filters are applied. The features taken from the CNN layer are further summarized using MaxPooling.

BiLSTM is employed to capture the sequential dependencies in the time series data. The outputs at each time step ($t-1$, $t+1$) are processed based on both past and future information. $\overrightarrow{h}$ and $\overleftarrow{h}_t$ represent the hidden states processed in the forward and backward directions, respectively. The goal of this bidirectional processing is to record the data's context in both directions.

Attention Layer added to the end of the BiLSTM block, this layer assigns different weights to hidden states that contribute differently to the cryptocurrency transactions. This makes it easier to create a high level vector that compiles all pertinent data.

Output Layer flattens the context vectors that the AM produced into a single vector. The final classification is then carried out by one or more fully connected layers using this flattened vector as input.

The parameters of the methods used in the model are provided in Table 1.

Table 1. Parameters of CNN-BiLSTM-AM method

Parameters	Value
Conv1D filters	64
Conv1D kernel size	1
Conv1D activation function	Relu
Conv1D padding	Same
Maxpooling pool_size	2
BiLSTM units	64
BiLSTM activation function	Relu
Attention	-
Dense1 neurons number	64
Dense2 neurons number	32
Dense3 neurons number	1
Training Parameters	**Value**
Optimizer	RMSprop
Epochs	100
Batch_size	64
Loss function	binary-crossentropy

EXPERIMENTAL RESULTS

In this section, the performance results of the CNN-BiLSTM-AM hybrid method applied to the Elliptic datasets used in this study are presented and compared with other studies. The impact of random data splits on performance outcomes was utilized by using the k-fold cross-validation technique. This method involves dividing the dataset into k parts, with each part serving as the test set in a rotation. The remaining data were used to train the model. This process was repeated by creating different combinations of the training and test sets in each cycle. Consequently, the model's performance consistency across various data splits was assessed more reliably. In this study, k is set to 10. The dataset was divided into ten equal parts, and one part was reserved as the test data. The nine parts that remained were trained. The test data was used to assess the models' performance.

Of the total number of transactions, 21% (42,019) are labeled legal, 2% (4,545) are labeled illegal, and the remaining 77% (157,205) are unknown. The dataset's imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE) [47]. By reproducing the samples from the minority class, SMOTE balances the dataset. The SMOTE process includes the following steps:

1. A sample belonging to the minority class is selected.
2. A random sample is selected and a vector is created between these two samples.
3. Using this vector to generate new instances, synthetic instances with similarities to the original instance are generated.
4. New synthetic instances are added as part of the original minority class.

SMOTE method was used for k-fold cross validation, where the dataset was balanced at each of the k-folds, and k-fold cross validation was performed on the dataset at each fold separately. This approach ensured that the model at each stage was trained on balanced data. Performance of the hybrid method was then evaluated using accuracy, sensitivity, precision, and F1-score metrics. These values were calculated using Equation 13–16. Performance metrics were obtained from the confusion matrix presented in Table 2.

In the confusion matrix, True Positive (TP) represents an illicit instance accurately identified as illicit. False Positive (FP) refers to a licit instance mistakenly classified as illicit. False Negative (FN) denotes an illicit instance incorrectly identified as licit, while True Negative (TN) represents a licit instance correctly identified as a licit. Accuracy, precision, recall, and F1-score were obtained using the confusion matrix.

Table 2. Confusion matrix

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	**TP**	**FN**
	Negative	**FP**	**TN**

Accuracy: The percentage of correctly classified samples.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (13)$$

Precision: The proportion of true positives among the samples predicted as positive. This value measures the success of the correct predictions.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

Recall: The metric shows the number of values that need to be estimated as positive and are correctly estimated as positive. If this value is extremely low, illicit transactions cannot be detected effectively.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (15)$$

F1 Score: The harmonic mean of the precision and recall values.

$$F_1 - \text{score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (16)$$

Table 3 shows the comparison of the proposed hybrid model with other standalone deep learning methods. CNN is effective in capturing local patterns within the data. BiLSTM is unique in its ability to comprehend relationships that change over time. Together, these two models allow for the simultaneous learning of the temporal dependencies and structural features in the data. By enabling the model to concentrate on the most crucial aspects of the data, the AM further improves the classification accuracy. This feature offers a significant benefit, particularly given the dynamic nature of cryptocurrency transactions. AM enables the model to discard irrelevant data and focus only on meaningful patterns. The proposed model in this study has a structure that enhances the performance of each of the CNN, BiLSTM, and hybrid methods.

The CNN-BiLSTM-AM model delivered the top performance in precision, recall, and F1-score, with values of 99%, 98%, and 98%, respectively, as shown in Table 4 Additionally, an accuracy of 97.9% was obtained, representing the second-highest result in the comparison. Although the Ensemble (Boosting-CFS) method [31] reported the highest overall accuracy, the CNN-BiLSTM-AM model achieved stronger precision, recall, and F1-score values, indicating a more balanced classification of both positive and negative classes.. Although the LB-GLAT method [27] showed 97.76% accuracy, the proposed model exceeded this success with 99.0% accuracy. GCN + MLP [19] and Temporal-GCN [30], which are among the graph based methods, achieved results close to those of the proposed model with 97.4% and 97.7% accuracy values, respectively. While RMGANets [28] has high performance in accuracy 97.96%, the CNN-BiLSTM-AM model still achieves a higher overall balance between all metrics, confirming its better performance in classifying both positive and negative instances. These results demonstrate the superiority of the AM and hybrid architecture of the proposed model for understanding complex data structures. In general, it is observed that the proposed model achieves more effective results than the methods in the literature.

DISCUSSION

In this study, the CNN-BiLSTM-AM hybrid model developed to detect illicit money transfers was tested on the Elliptic dataset. With a precision of 97.62%, the CNN model exhibited a high level of ability to extract features from raw data; however, its recall remained at 91.09%. With a recall of 90.18% and a relatively low precision of 86.49%, the BiLSTM model successfully captured the temporal dependencies in the sequential data.

The CNN-BiLSTM model achieved a high precision of 98.7% by combining the advantages of both methods; however, its recall was only 64.4%. In contrast, the proposed CNN-BiLSTM-AM model, which was improved by incorporating the AM, performed better and more

Tablo 3. Performance comparison of proposed hybrid model and standalone methods

Method	(%) Precision	(%) Recall	(%) F1-score	(%) Accuracy
CNN	97.62	91.09	83.79	96.53
LSTM	65.04	84.75	68.0	91.82
BİLSTM	86.49	90.18	78.10	94.32
CNN+LSTM	73.33	84.78	78.58	95.46
CNN+BİLSTM	98.7	64.40	77.9	96.45
CNN+AM	71.27	82.45	75.87	94.74
BİLSTM+AM	50.78	66.67	56.58	92.43
CNN-BiLSTM-AM	**98.0**	**98.0**	**97.9**	**99.0**

Table 4. The comparison results of CNN+BiLSTM-AM and other methods on elliptic dataset

Reference	Method	(%) Precision	(%) Recall	(%) F1-score	(%) Accuracy
Weber et al. [18]	GCN	81.2	51.2	62.8	
	Skip-GCN	81.2	62.3	70.5	
	EvolveGCN	85.0	62.4	72.0	
Alarab et al. [19]	GCN + MLP	89.9	67.8	77.3	97.4
Pocher et al. [20]	GCN (tx)	90.6	79.0	84.4	-
	GAT (tx)	89.7	60.5	72.3	
Chen et al. [21]	GraphSAGE	61.8	80.2	66.1	91.3
Lo et al. [22]	Inspection-L	97.2	72.1	82.8	-
Yang et al. [24]	LSTM+GCN	85.4	71.2	77.6	-
Xia et al. [25]	MGC-LSTM	96.9	84.4	90.2	
Xiao et al. [26]	CTDM	85.1	87.4	83.5	-
Guo et al. [27]	LB-GLAT	93.17	84.94	88.87	97.76
Wand et al. [28]	RMGANets	96.82	91.32	93.85	**97.96**
Alarab et al. [29]	MC-dropout	-	-	72.9	96.6
Alarab and Prakoonwit [30]	Temporal-GCN	92.7	71.3	80.6	97.7
Jatoth et al. [31]	Ensemble (Boosting-CFS)	98.0	93.0	95.0	**98.0**
	Ensemble (Stacking-CFS)	98.0	92.0	92.0	**98.0**
Proposed Model	**CNN+BiLSTM+AM**	**99.0**	**98.0**	**98.0**	**97.9**

evenly, achieving 97.9% F1-score, 98.0% precision, and 98.0% recall. When considered in relation to earlier studies, these results suggest performance at a state-of-the-art level.

Class imbalance was handled through the application of SMOTE within each fold of the cross-validation process, so that resampling was confined to the training data in every iteration. Embedding this step directly into the validation procedure helped prevent inflated performance estimates and supported a more stable assessment across folds. In this sense, the resulting performance patterns suggest that the model behaves consistently under both conceptual expectations and practical data conditions.

The empirical results show that the CNN–BiLSTM–AM framework maintains stable performance in identifying illicit cryptocurrency transactions, without introducing excessive computational burden. Rather than functioning solely as a high-performing predictive model, the framework also provides insight into how attention mechanisms interact with temporal and structural features in complex classification settings. Future research aimed at better understanding and reducing the risks associated with the growing adoption of cryptocurrencies is anticipated to be guided by this study's findings.

CONCLUSION

Cryptocurrencies represent technological advancements that have significantly altered the financial systems of nations. However, this transformation has also contributed to an increase in illicit activities in the financial domain. In particular, money laundering, terrorist financing, and similar crimes have become more serious threats as cryptocurrencies have been adopted more widely. Consequently, AML in cryptocurrency systems has become a priority issue in the financial sector. In this context, increasing transaction transparency is important. In addition, regulatory policies and compliance mechanisms aimed at strengthening identity verification processes must be developed. Artificial intelligence-based methods are considered effective tools for detecting suspicious transactions and money laundering activities.

This study introduces a method for identifying illegal activities within cryptocurrency transactions. At the same time, the study is intended to offer practical insight into addressing financial crime. The model developed in this study may serve as a reference structure for financial institutions seeking to strengthen their AML strategies. Given its compatibility with blockchain based environments, it also points to possible improvements in transparency and operational security within financial systems. These results are expected to shed light on future research that aims to meet the growing needs of the financial ecosystem and create a safer digital environment from ethical, societal, and economic perspectives.

The findings suggest that the proposed model is not only technically robust but also practically viable for detecting illicit activity. Nevertheless, evaluating its performance across diverse datasets remains important for assessing generalizability, particularly in view of constraints related

to dataset size and structural characteristics. Further investigation is needed to analyze optimization strategies and computational requirements more systematically, particularly in assessing the practicality of real-time deployment.

Against the backdrop of the growing transformation of Bitcoin within the financial ecosystem, this study explores how deep learning techniques can be applied to the detection of illicit activities. The CNN-BiLSTM-AM hybrid model developed in this study offers a novel approach for achieving this goal. As a result of the experiments on the elliptic dataset, it was observed that the model performed better than the other methods in the literature. The results indicate that deep learning techniques can be effective tools in the fight against AML and can make a significant contribution to combating financial crimes. It is also believed that this study will inspire further research and development efforts to understand and address the risks associated with the increasing use of cryptocurrencies.

ACKNOWLEDGEMENTS

This study was supported by Scientific and Technological Research Council of Turkey (TUBITAK, Grant No: 123E312) within the scope of the TUBITAK 1002-A support program.

AUTHORSHIP CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: study conception and design: Cengiz and Gök; data collection: Gök; analysis and interpretation of results: Cengiz and Gök; draft manuscript preparation: Cengiz and Gök. All authors reviewed the results and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The public dataset https://www.kaggle.com/datasets/ellipticco/elliptic-data-set

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Firmansyah W, Atmadja HT. Juridical analysis awareness of profession advocacy to financial transaction reports and analysis centre (PPATK) during prevent and eradicate money laundering crime. J Multidiscip Acad 2021;5:308–314.

[2] Soltani R, Nguyen UT, Yang Y, Faghani M, Yagoub A, An A. A new algorithm for money laundering detection based on structural similarity. In: Proceedings of the 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); 2016. p. 1–7. [CrossRef]

[3] Bieler SA. Peeking into the house of cards: Money laundering, luxury real estate, and the necessity of data verification for the Corporate Transparency Act's beneficial ownership registry. Fordham J Corp Finan Law 2022;27:193.

[4] Schneider F, Windischbauer U. Money laundering: Some facts. Eur J Law Econ 2008;26:387-404. [CrossRef]

[5] Ahmed AAA. Anti-money laundering recognition through the gradient boosting classifier. Acad Account Financ Stud J 2021;25:1–11.

[6] Bryans D. Bitcoin and money laundering: Mining for an effective solution. Indiana Law J 2014;89:441.

[7] Van Wegberg R, Oerlemans JJ, van Deventer O. Bitcoin money laundering: Mixed results? An explorative study on money laundering of cybercrime proceeds using bitcoin. J Financ Crime 2018;25:419–435. [CrossRef]

[8] Jullum M, Løland A, Huseby RB, Anonsen G, Lorentzen J. Detecting money laundering transactions with machine learning. J Money Laund Control 2020;23:173–186. [CrossRef]

[9] Lorenz J, Silva MI, Aparício D, Ascensão JT, Bizarro P. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. In: Proceedings of the 1st ACM International Conference on AI in Finance; 2020. p. 1–8. [CrossRef]

[10] Jensen RIT, Iosifidis A. Qualifying and raising anti-money laundering alarms with deep learning. Expert Syst Appl 2023;214:119037. [CrossRef]

[11] Aggarwal K, Mijwil MM, Al-Mistarehi AH, Alomari S, Gök M, Alaabdin AMZ, et al. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. Iraqi J Comput Sci Math 2022;3:115–123. [CrossRef]

[12] Kute DV, Pradhan B, Shukla N, Alamri A. Explainable deep learning model for predicting money laundering transactions. Int J Smart Sens Intell Syst 2024;17. [CrossRef]

[13] Rani K, Deepak B, Hemanthh P, Mohanasudhan B, Sathishkumar G. Spatio-temporal network based bank transactional behaviour analysis to detect

suspicious activities. In: Proceedings of the 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA); 2024. p. 1-6. [CrossRef]

[14] Wang Y, Huang M, Zhu X, Zhao L. Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing; 2016. p. 606–615. [CrossRef]

[15] Tang M, Ye M, Chen W, Zhou D. BiLSTM4DPS: An attention-based BiLSTM approach for detecting phishing scams in ethereum. Expert Syst Appl 2024;256:124941. [CrossRef]

[16] Jainish GR, Alwin Infant P. Attention layer integrated BiLSTM for financial fraud prediction. Multimed Tools Appl 2024;83:80613–80629.

[17] Wang Z, Ni A, Tian Z, Wang Z, Gong Y. Research on blockchain abnormal transaction detection technology combining CNN and transformer structure. Comput Electr Eng 2024;116:109194. [CrossRef]

[18] Weber M, Domeniconi G, Chen J, Weidele DKI, Bellei C, Robinson T, et al. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. arXiv [Preprint] 2019:arXiv:1908.02591.

[19] Alarab I, Prakoonwit S, Nacer MI. Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In: Proceedings of the 2020 5th International Conference on Machine Learning Technologies; 2020. p. 23-27. [CrossRef]

[20] Pocher N, Zichichi M, Merizzi F, Shafiq MZ, Ferretti S. Detecting anomalous cryptocurrency transactions: An AML/CFT application of machine learning-based forensics. Electron Mark 2023;33:37. [CrossRef]

[21] Chen C, Li Q, Chen L, Liang Y, Huang H. An improved GraphSAGE to detect power system anomaly based on time-neighbor feature. Energy Rep 2023;9:930–937. [CrossRef]

[22] Lo WW, Kulatilleke GK, Sarhan M, Layeghy S, Portmann M. Inspection-L: Self-supervised GNN node embeddings for money laundering detection in bitcoin. Appl Intell 2023;53:19406–19417. [CrossRef]

[23] Wan F, Li P. A novel money laundering prediction model based on a dynamic graph convolutional neural network and long short-term memory. Symmetry 2024;16:378. [CrossRef]

[24] Yang G, Liu X, Li B. Anti-money laundering supervision by intelligent algorithm. Comput Secur 2023;132:103344. [CrossRef]

[25] Xia P, Ni Z, Xiao H, Zhu X, Peng P. A novel spatiotemporal prediction approach based on graph convolution neural networks and long short-term memory for money laundering fraud. Arab J Sci Eng 2021;47:1921–1937. [CrossRef]

[26] Xiao L, Han D, Li D, Liang W, Yang C, Li KC, et al. CTDM: Cryptocurrency abnormal transaction detection method with spatio-temporal and global representation. Soft Comput 2023;27:11647–11660. [CrossRef]

[27] Guo C, Zhang S, Zhang P, Alkubati M, Song J. LB-GLAT: Long-term bi-graph layer attention convolutional network for anti-money laundering in transactional blockchain. Mathematics 2023;11:3927. [CrossRef]

[28] Wang Q, Tsai WT, Du B. RMGANets: Reinforcement learning-enhanced multi-relational attention graph-aware network for anti-money laundering detection. Complex Intell Syst 2025;11:5. [CrossRef]

[29] Alarab I, Prakoonwit S, Nacer MI. Illustrative discussion of mc-dropout in general dataset: Uncertainty estimation in bitcoin. Neural Process Lett 2021;53:1001–1011. [CrossRef]

[30] Alarab I, Prakoonwit S. Graph-based LSTM for anti-money laundering: Experimenting temporal graph convolutional network with bitcoin data. Neural Process Lett 2023;55:689–707. [CrossRef]

[31] Jatoth C, Jain R, Fiore U, Chatharasupalli S. Improved classification of blockchain transactions using feature engineering and ensemble learning. Future Internet 2021;14:16. [CrossRef]

[32] Alarab I, Prakoonwit S, Nacer MI. Competence of graph convolutional networks for anti-money laundering in bitcoin blockchain. In: Proceedings of the 2020 5th International Conference on Machine Learning Technologies; 2020. p. 23–27. [CrossRef]

[33] Mijwil MM, Aggarwal K, Doshi R, Hiran KK, Gök M. The distinction between R-CNN and Fast R-CNN in image analysis: A performance comparison. Asian J Appl Sci 2022;10. [CrossRef]

[34] Baysal K, Taskin D. Blocking harmful images with a deep learning based next generation firewall. Sigma J Eng Nat Sci 2024;42:1133–1147. [CrossRef]

[35] Cengiz E, Yaylak F, Gülbandılar E. Investigation of polyps in endoscopy images by using deep learning algorithm. Eskisehir Osmangazi Univ J Eng Archit 2022;30:441–453. [CrossRef]

[36] Cihan M, Uzbaş B, Ceylan M. Fusion and CNN based classification of liver focal lesions using magnetic resonance imaging phases. Sigma J Eng Nat Sci 2023;41:119–129. [CrossRef]

[37] Pascanu R, Gulcehre C, Cho K, Bengio Y. How to construct deep recurrent neural networks. arXiv [Preprint] 2013:arXiv:1312.6026.

[38] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D 2020;404:132306. [CrossRef]

[39] Chen L, Zhang W, Ye H. Accurate workload prediction for edge data centers: Savitzky-Golay filter, CNN and BiLSTM with attention mechanism. Appl Intell 2022;52:13027–13042. [CrossRef]

[40] Chen Y, Fang R, Liang T, Sha Z, Li S, Yi Y, et al. Stock price forecast based on CNN-BiLSTM-ECA model. Sci Program 2021;2021:2446543. [CrossRef]

[41] Schuster M. Acoustic model building based on non-uniform segments and bidirectional recurrent neural networks. In: Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing; 1997. p. 3249–3252. [CrossRef]

[42] Ren F, Jiang Z, Wang X, Liu J. A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network. Cybersecurity 2020;3:4. [CrossRef]

[43] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48–62. [CrossRef]

[44] Kavianpour P, Kavianpour M, Jahani E, Ramezani A. A CNN-BiLSTM model with attention mechanism for earthquake prediction. J Supercomput 2023;79:19194–19226. [CrossRef]

[45] Li R, Ye X, Yang F, Du KL. ConvLSTM-Att: An attention-based composite deep neural network for tool wear prediction. Machines 2023;11(2):297. [CrossRef]

[46] Tian Y, Wen J, Yang Y, Shi Y, Zeng J. State-of-health prediction of lithium-ion batteries based on CNN-BiLSTM-AM. Batteries 2022;8:155. [CrossRef]

[47] Gökçen T, Odabaş A. Cryptocurrency price prediction using GPR and SMOTE. Sigma J Eng Nat Sci 2024;42:1448–1458. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 983–993, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2020

Research Article

Investigating the effectiveness of IS-DOS prevention mechanisms for mitigating DOS attacks in VANET

Harsh Pratap SINGH[1], Nagendra SINGH[2,*], Sai Kiran Naik BANOTH[3], Vedika PATIL[4], Neha ARYA[5], Neeraj SHRIVASTAVA[6]

[1]*Department of Computer Science Engineering, Medicaps University, Indore, 453331, India*
[2]*Department of Electrical Engineering, Trinity College of Engineering & Technology, Karimnagar, 505001, India*
[3]*Technical Lead at Fannie Mae, 20005, USA*
[4]*Department of Computer Engineering, KC College of Engineering and Management Studies and Research, Thane, 400603, India*
[5]*Department of Electronic and Instrumentation, Shri Govindram Seksaria, Institute of Technology and Science, Indore, 452003, India*
[6]*Department of Electrical Engineering, Medicaps University, Indore, 45333, India*

ARTICLE INFO

Article history
Received: 10 February 2025
Revised: 13 April 2025
Accepted: 15 June 2025

Keywords:
Vehicular Networks; Mobile Ad-hoc Network; Routing Load; Black Hole; Security Threat; Ad Hoc On-Demand Distance Vector; IS-DoS (Improved Security in Denial of Service).

ABSTRACT

Ensuring secure and reliable communication in vehicular networks, a type of mobile ad hoc network is a major challenge due to the frequently changing topology. The dynamic nature of VANETs, where nodes can frequently join or leave the network, can lead to insecure communication and link disconnections caused by obstructions such as buildings, tunnels, bridges, and fossils. Disconnections can cause packet loss and negatively impact network performance. It is often difficult to determine the reason behind packet loss, whether it is due to node detection or security threats. VANETs are vulnerable to various attacks, including denial-of-service, black hole, and greyhole attacks. This study focuses on mitigating DoS flooding attacks to enhance network safety and efficiency of VANET communication. In this work, we propose a novel intrusion detection algorithm IS-DOS, which may help effectively identify and prevent DOS flooding attacks by leveraging the traffic pattern analysis with the AODV routing protocol framework. To evaluate the effectiveness of the proposed approach, use network simulator NS 2.34 for the conventional DOS, QRT-DOS and our proposed model IS-DOS. The quantitative performance metrics such as packet delivery ratio, throughput, end-to end delay, and routing overhead were analyzed under various network conditions. The simulation results demonstrated that IS-DOS improves the packet delivery ratio by 15% and throughput by 12% compared to existing DOS and QRT-DOS mitigation approach. Moreover, IS-DoS diminishes average packet delay by 20% and drops routing overhead by 18%, signifying enhanced network efficiency and resilience. These findings recommend that IS-DoS not solitary mitigates flooding attacks more effectually than prior approaches but also sustains network stability in the existence of high mobility and recurrent topology changes. The novelty of this work deceits in its adaptive detection mechanism tailored for VANET surroundings, which develop beyond existing solutions by suggesting real-time, low-overhead attack anticipation without

***Corresponding author.**
*E-mail address: ernsingh007@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

compromising normal network operations. This research contributes to improving the security and reliability of vehicular communications, which is indispensable for the prevalent adoption of intelligent transportation systems.

Cite this article as: Singh HP, Singh N, Banoth SKN, Patil V, Arya N, Shrivastava N. Investigating the effectiveness of IS-DOS prevention mechanisms for mitigating DOS attacks in vanet. Sigma J Eng Nat Sci 2026;44(2):983–993.

INTRODUCTION

The growing global demand for transportation, fueled by rapid urbanization and population growth, has spurred the advancement of intelligent and secure vehicular communication systems [1]. These systems rely on communication tools and internet services and can be based on dynamic wireless networks, static wired networks, or hybrid networks. While static infrastructure networks present many challenges, such as cell sites, access points, and physical and digital cables, wireless networks are easier to establish and maintain [2,3]. Vehicular Ad Hoc Networks (VANETs) ... have emerged as a pivotal component in modern Intelligent Transportation Systems (ITS). Contributing significantly to road safety, traffic management, and infotainment services. Additionally, they allow for seamless integration with the Intelligent Transport System and testing of the latest developments in smart car networks [4].

These advancements are expected to revolutionize the driving experience and characteristics and improve the safety of vehicles on highways and in cities. However, the open literature extensively explores. A critical challenge remains unresolved: maintaining secure and reliable communication under conditions of high mobility and frequent topological changes [5,6].

In order to ensure the safety of VANET communications, previous research has proposed various techniques based on the measures of cryptography and trust such as digital certificates, Public Key Infrastructure (PKI), and trust models [7-9]. Whereas these techniques are providing some degree of security, they are often ineffective in high speed scenarios, or when the adversaries are involved, and immediate trust checking along with low latency reaction is necessary. One of the major unexplored issues on this area is how the network can be exploited by Distributed Denial of Service (DoS) attacks, especially the flooding-based DoS attacks, which interfere with the network performance due to overloading of routing protocols. Such attacks can severely disrupt service, result in delays and loss of packets particularly in the busy urban environment. Moreover, as VANET connections are temporary, even now it is not easy to differentiate malicious behavior and usual disconnection [10,11].

This work suggests IS-DoS (Intelligent Suppression of DoS Flooding Attacks), a novel security approach created especially for dynamic vehicle situations, to fill these shortcomings. By employing behavior-based traffic pattern analysis to identify and prevent DoS flooding attacks instead of just cryptographic techniques, the suggested methodology improves VANET's security. Its capacity to function in real-time, adjust to changing network conditions, and reduce detection overhead is what makes it innovative. Additionally, this work presents an RSU-based addition that makes it possible for roadside units to efficiently monitor and react to unusual traffic patterns. We use the suggested IS-DoS method over the AODV routing protocol and test it with the NS-2.34 simulator. The algorithm is evaluated in comparison to the baseline. Using performance indicators including throughput, end-to-end latency, routing overhead, and packet delivery ratio (PDR), DoS and QRT-DoS techniques.

In summary, the novelty of this study lies in its development of a lightweight, real-time DoS mitigation strategy tailored for VANET environments. Unlike conventional cryptographic defenses, IS-DoS provides a proactive and scalable approach that improves both security and network efficiency under high-mobility conditions. Section II of the article presents a comprehensive review of related work on improving the safety and efficiency of VANETs. Section III describes our proposed algorithm for detecting DoS flood attacks in VANETs. Section IV presents the experiment setup and analysis of the results, while the conclusions of the proposed work are mentioned in Section V.

RELATED WORK

A method for detecting DDoS attacks that makes use of big data technology is presented in Article [1]. It uses Apache Spark to process massive amounts of traffic and HDFS to store important suspicious info [2]. We employ a micro-batch data processing methodology to enable real-time traffic gathering [3], and we analyse the acquired traffic using a Random Forest (RF) classifier. The NSL-KDD and UNSW-NB15 datasets were used to validate the suggested method, which produced remarkable detection accuracies of 99.95% and 98.75%, respectively [4].

To enable robust broadcasting on VANETs, a unique method for choosing multipoint relay (MPR) schemes is proposed in [5, 6]. It is an optimisation of the relay selection technique that utilises a weak 2-hop set of MPR nodes by considering each 2-hop node multiple times. The Fuzzy Cluster Management System forms vehicle clusters on VANETs based on two fuzzy models of FCMS1 and FCMS2. Article [7] introduces FCMS1 that takes into consideration the input characteristics in the form of vehicle centrality,

security, cluster affiliation, and relative speed in predicting the likelihood of vehicle remaining within a cluster.

The authors of [8] present a multivariate stream analysis method of detecting and evaluating the DDoS attacks on VANETs. To achieve more accuracy in making a detection, the technique considers the traffic in different time periods and conditions, and applies specific rules to the different traffic types. Article [9] presents ID-MAP, a special identity authentication method that exploits proxy vehicles and encrypting messages using elliptic curve that operates under the random oracle model to achieve message authenticity and to avoid forgery[10].

Article [11] introduces a non-parametric statistical method of detecting low-rate DDoS attacks in intelligent transportation systems. This method beat two parametric models in terms of average detection delay and false alarm rate when tested with the Simulation of Urban Mobility tool [12]. In [13], a Quick Response Table (QRT)-based secure routing strategy that logs route data and highlights unknown anomalies is described. According to the simulation data QRT can improve throughput, packet delivery

Table 1. Comparison table of existing literature

Article/Reference number	Technique/approach	Key results/findings	Comparison with proposed work
[1]	DoS attack prevention using general VANET routing security	Proposed heuristic method for general DoS prevention	Lacks big data scalability or ML accuracy used in proposed method
[2]	Apache Spark-based DDoS detection	Real-time mitigation architecture using Spark	Similar big data tech but lacks comparative ML evaluation
[4]	Distributed NIDS using RF classifier	Accuracy: 99.95% (NSL-KDD), 98.75% (UNSW)	Comparable to proposed method; confirms efficacy of RF on same datasets
[5,6]	IEEE 802.11p DoS detection, Secure auth	Real-time relay strategy for VANETs	Different layer focus (MAC/ authentication); does not include big data or ML models
[8]	Trust framework (TEAM)	Uses multivariate stream analysis	Focused on trust, not suitable for scalable DDoS
[9,10]	Fuzzy cluster-based MPR with trust/security	Improves clustering efficiency and relay trust	Complementary to our traffic-level detection but not core to DDoS prevention
[11]	Statistical low-rate DDoS detection	Reduced false positives, better avg. detection time	Doesn't scale to high-speed traffic as in Spark+RF architecture
[13]	Anonymous auth + privacy	Improved security + low overhead	Auth-layer focus; doesn't handle DDoS on transport/network layers
[17,18]	P-Secure proactive DoS detection	Improved delay, throughput	Suitable for DoS but lacks real-time big data handling or accuracy benchmarks
[22]	IDoS-CC (intelligent DDoS + congestion control)	Real-time congestion-aware DDoS detection	Comparable but lacks ML accuracy and dataset validation
[23]	Puzzle-based co-auth	Prevents DoS via pseudonym attack resistance	Authentication-level focus only
[24]	Fuzzy logic + fog-based 5G VANET detection	Effective at edge-layer DDoS detection	Similar but doesn't benchmark with standard datasets
[26]	ML-MDS using SVM, DT, etc.	Accuracy improved across classifiers	Proposed RF model performs better and is validated with two datasets
[27]	ML in SDN-WISE IoT	Reliable for SDN-based DDoS detection	Limited to SDN-IoT; doesn't address VANET big-data challenges
[29]	ML-based intrusion detection with RF	Multi-level IDS improved detection	Similar ML basis but proposed method uses big data + real datasets
[31]	Hybrid LogitBoost prediction model	Improved DDoS resilience	Uses hybrid model but lacks real-time big data validation
[35]	Real-time multi-stage IDS	Low false alarm rate, multi-stage detection	Aligns with proposed goal but lacks performance metrics on benchmark datasets

rate (PDR) and reduce routing load and delay in mitigating DoS attacks[14,15].

Article [16] recommends the use of an anonymous authentication procedure that allows tracking the entry of entities and ensures privacy to prevent the introduction of hostile vehicles. Article [17,18] presents the P-Secure method of proactive DoS attack detection that reduces the processing overhead in the system and enhances its network performance. Simulation results show that P-Secure is more effective in packet delivery ratio, delay, and data transfer rate compared to OBU-based VANET models.

Article [19] emphasises the necessity of strong management techniques to guarantee safe VANET connectivity. In [20], we introduce a node reliability strategy and local route redesign procedures to handle route creation collisions. To prevent serious accidents, a crash prevention protocol for vehicular networks is put forth in [21]. It consists of intelligent, roadside, and sensor devices.

The goal of the IDoS-CC strategy put forth in [22] is to identify DoS attacks and manage congestion. To prevent DoS attacks that target pseudonymous authentication, Article [23] proposes a co-authentication method that uses a hash puzzle system. For 5G-enabled smart cities, [24] proposes a fog-based DDoS detection method, based on fuzzy logic, to distinguish between malicious and genuine communication.

To solve scalability issues in SDNs, a lightweight framework for DoS detection and mitigation were presented in [25]. Using the UNSW_NB-15 dataset, Article [26] looks at the ML-MDS model by sorting vehicular traffic with methods like logistic regression, decision trees, SVMs, and neural networks. In [27], a machine learning-based technique in the SDN-WISE IoT controller uses a specialized testing environment to identify DDoS attacks. Article [28] adds to the research by proposing a way to find low-rate and hidden DDoS attacks through communication between the SDN controller and the system that helps to reduce the attacks.

Article [29] presents a machine learning-based multilevel intrusion detection system that uses random forest for feature selection and quick correlation-based filtering. Article [30] uses the K-Means and C5.0 algorithms to investigate anomaly identification in VANETs. To improve vehicular infrastructure resistance against DDoS assaults, Article [31] suggests predictive risk evaluation using a hybrid Logit-Boost model. Article [32] presents a trust-based approach for dynamic pseudonym modifications.

To prevent impersonation, Article [33] offers a safe, effective conditional privacy-preserving authentication scheme that uses hash functions and bilinear pair cryptography. Article [34] employs random forests, and posterior detection approaches to enhance the precision of intrusion detection [35]. Lastly, a real-time multi-stage intrusion detection system (IDS) with low false-alarm rates and automatic intrusion detection capabilities for intelligent transportation systems is described in [36]. Table 1 shows a deep comarisions of the existing works.

Proposed Algorithm

This work used an IS-DoS Prevention Mechanisms approach to tackling the issue of DoS flooding attacks in vehicular networks to explore desirable flooding schemes. The objective of the proposed mechanism is to ensure secure information sharing among vehicles while addressing challenges such as delivering data correctly, selecting appropriate nodes or vehicles to transmit data to the destination, and maintaining communication among vehicles [21]. In order to achieve this, Road Side Units (RSUs) are deployed at every terminal to monitor traffic conditions and identify any misbehavior or malicious activity in the network. The approach operates in two modes: Vehicle-RSU (V to RSU) and Vehicle-Vehicle (V to V) communication [22]. The RSU acts as a communication hub between vehicles and can detect malicious nodes or vehicles, decreasing their trust value and preventing them from secure communication within the network [23].

Working steps of the proposed algorithm

The suggested approach's algorithmic steps are explained in this section. It provides a series of easy-to-follow guidelines for putting the strategy into practice and reducing DoS flooding attacks in vehicle networks. Algorithm 1 describes a protocol for secure communication in a vehicular network using road site monitoring units. It includes steps for registration, authorization, and route packet transmission while considering the range of the monitoring units.

Algorithm 1: Algorithm for authorisation and communication in vehicular networks

Input: Cr: Number of cars in the network
RSUi: Road Site Monitoring Units
DOS: Denial of Service Attack in network
CMUi: Attack control and monitoring unit ∈ RSUi
Ψ: Radio range of RSUi is 15 Km
Rp: AODV (Ad hoc On-Demand Distance Vector)
lr: Vehicle location
sr: Vehicle speed

Output: Received packet, PDR, throughput, routing overhead, delay, DOS packet detection

```
Step 1:
Ci wants to communicate with Cj
If RSUi.authorize(Ci):
Step 2:
Ci.broadcast(Rp)
    Step 3:
      If not RSUi.is_in_range(Cj):
        RSUi.broadcast_to_neighbors(RSUi.RSUj, Rp)
    Step 4:
      If RSUi.RSUj.find(Cj):
        If RSUi.RSUj.check_db(Cj):
          Step 5:
            If RSUi.RSUj.authorize(Cj):
              RSUi.RSUj.send_packet(Cj, Rp)
            Else:
              RSUi.send_feedback(Ci, "Cj is not authorized.")
```

```
            Else:
              RSUi.send_feedback(Ci, "Cj is not registered.")
            Else:
              RSUi.send_feedback(Ci, "Cj is not in range.")
          Else If RSUi.check_db(Cj):
            Step 6:
              If RSUi.authorize(Cj):
                RSUi.send_packet(Cj, Rp)
              Else:
                RSUi.send_feedback(Ci, "Cj is not authorized.")
        Else:
          RSUi.send_feedback(Ci, "Cj is not registered.")
      Else:
        RSUi.send_registration_message(Ci)
        Ci.register_to_itself()
        Ci.authorize_for_future_communication()
```

Algorithm 2 is designed to monitor real-time communications systems and detect abnormal behaviour. It analyses the communication data, decreases trust value and blocks communication if abnormal behaviour is detected.

Algorithm 2: Detecting Abnormal Behavior in Real-Time Communications SystemInputs

Cr: number of cars in the network
Ci: a specific car in the network
Cj: Another specific car in the network
l1: initial location of Ci
l2: destination location of Ci
Ψ: radio range of RSUi
RSUi: Road Site Monitoring Unit
CMUi: Attack control and monitoring unit

Outputs

location: real-time location of Ci
speed: speed of Ci
communication_data: communication data of Ci
behavior: behavior of Ci and Cj (normal or abnormal)
temporary_block_authorization: authorization temporary block if Ci or Cj is untrusted.

```
Step 1:
    RSUi_sense() {
    track_locations(Cr)
    }
Step 2:
    while (Ci travels from location l1 to l2 and Ψ is under RSUi) {
    capture_data(Ci) {
    location = get_location(Ci)
    speed = get_speed(Ci)
    communication_data = get _communication_data(Ci)
    }
    }
Step 3:
    if (Ci communicates with Cj through RSUi) {
    CMUi_on()
    monitor_behavior(Ci, Cj)
    }
Step 4:
    if (data_not_normal(Ci) or data_not_normal(Cj)) {
    behavior = analyze_behavior(Ci, Cj)
    if (behavior == "abnormal") {
    decrease_trust_value(Ci)
    decrease_trust_value(Cj)
    block_communication(Ci, Cj)
    untrusted_detected()
    temporary_block_authorization()
    }
    } else {
    normal_behavior_monitored()
    assist_ci_cj()
    handover_communication(Ci, Cj)
    }
    Else {
    Ci_only_travels_using_RSUi_guidelines()
    provide_service_to_Ci()
}
```

Attacker Detection

Figure 1 shows the flow of security approach for detecting DoS attacks in VANETs. Detecting unwanted activity in the network is crucial for identifying malicious nodes that flood the network with false information packets and continuously drop legitimate data packets.

DOS Attack Prevention Scheme Flowchart

Figure 2 illustrates the proposed approach for identifying malicious vehicles or nodes and protecting the network against DoS attacks. When CMU messages are received, the approach blocks the malicious vehicles from selecting secure nodes for transmitting data packets and enhances the network's reliability [24-27].

EXPERIMENTAL SETUP AND RESULTS ANALYSIS

This section involves the analysis of designing and implementing experiments to test hypotheses and draw meaningful conclusions about the performance of the proposed algorithm. This process is crucial in validating the effectiveness and efficiency of the new proposed technologies.

Simulation Setup

We used the network simulator NS-2.34 to compare the suggested method with earlier studies in order to assess its quality. Data packets were routed via the AODV protocol in the simulation. Since DoS attacks have the potential to seriously hinder data packet transmission, the simulation was centered on identifying them in VANETs. The simulation was set up with 30 wireless sensor nodes/devices randomly deployed in an 1100 x 1100-meter simulation area. The physical medium used was wireless, and the nodes were allowed to move randomly during the simulation.

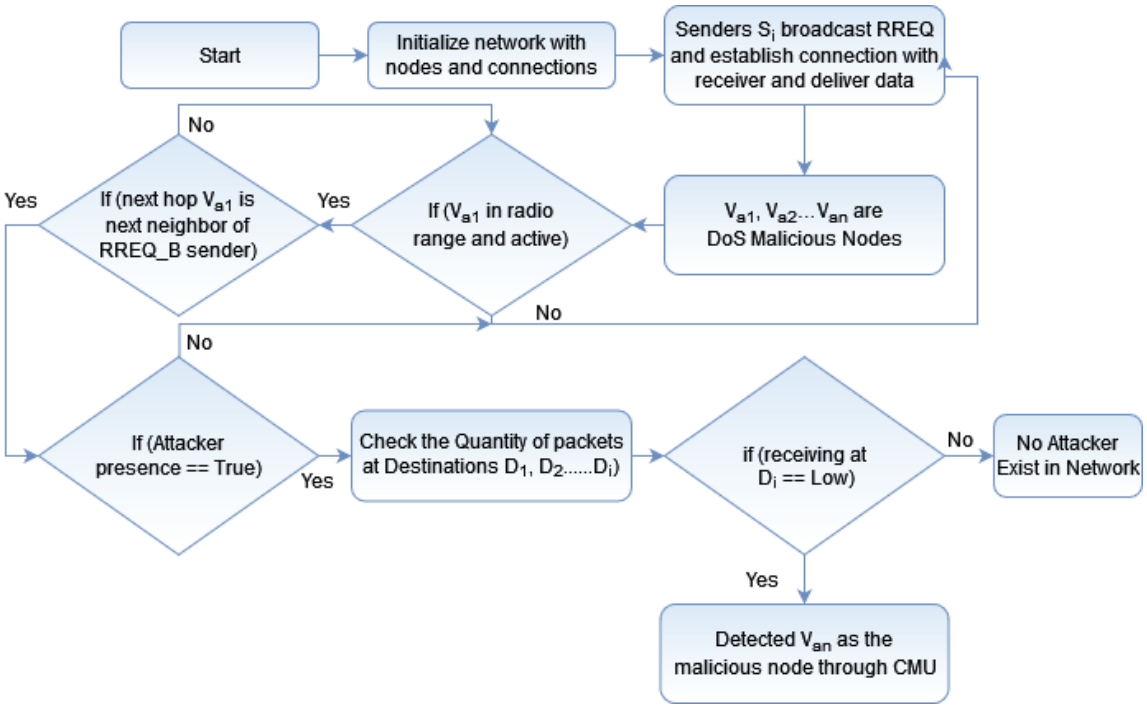


Figure 1. Attacker Detection.

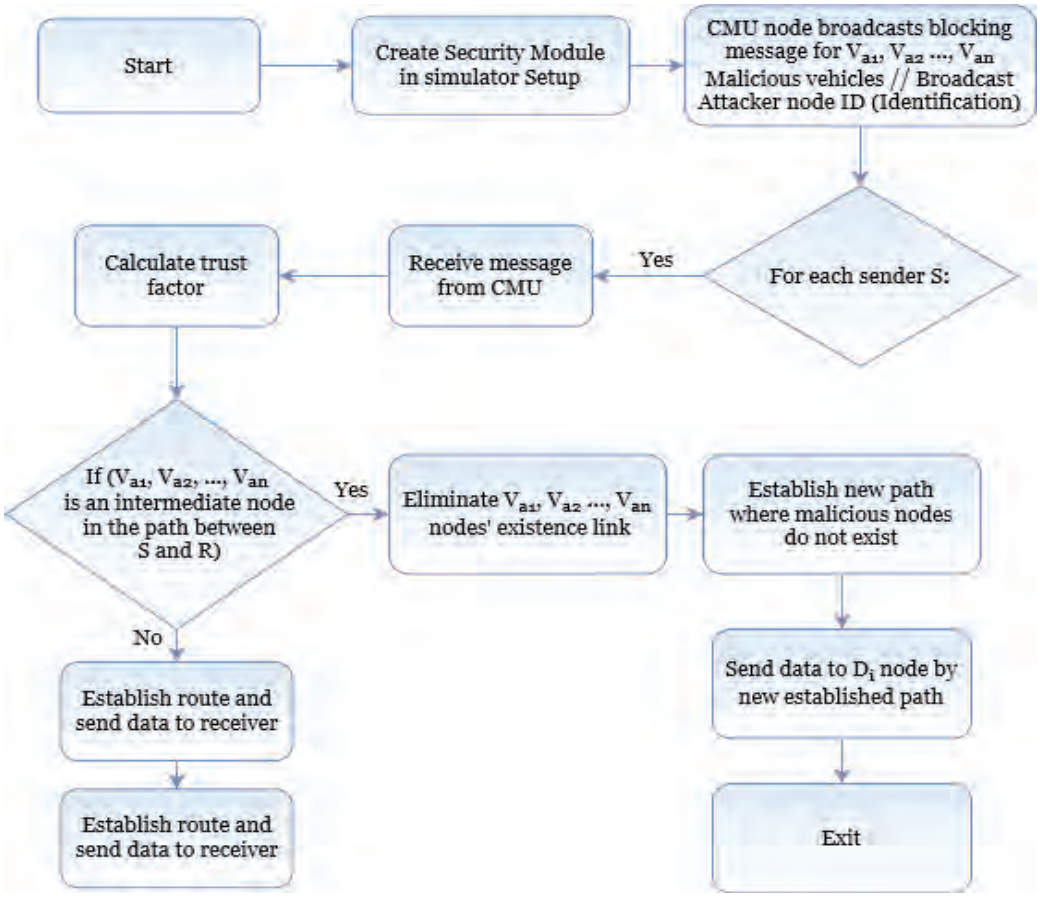


Figure 2. Prevention scheme.

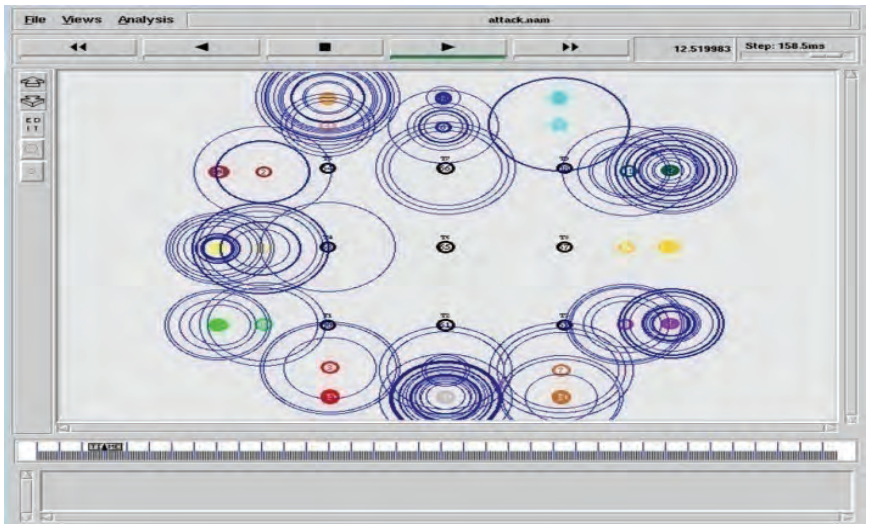


Figure 3. Dos attack prevention simulation result.

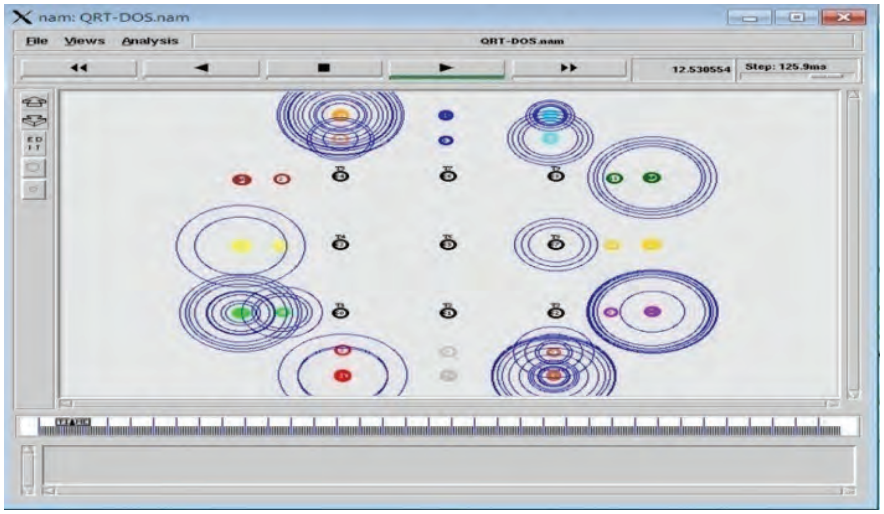


Figure 4. QRT-Dos attack prevention simulation result.

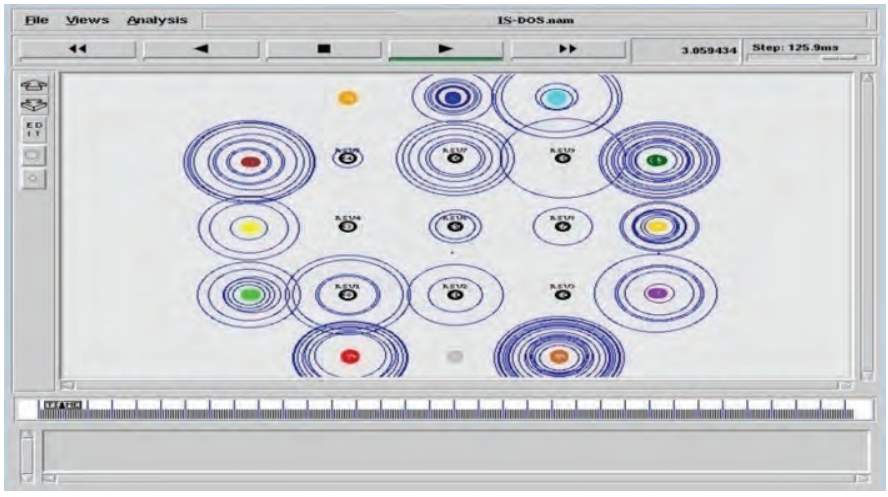


Figure 5. IS-Dos attack prevention simulation result**.**

The simulation results for the Dos attack prevention system are displayed in Figure 3. We simulated 300 iterations, and the queue length was set to 50. We used MAC 802.11p as the MAC layer and simulated Denial-of-Service (DoS) flooding as the attack type. We used two prevention mechanisms, QRT-DoS and IS-DoS, to prevent DoS attacks. Two different forms of traffic were used in the simulation: file transfer protocol and constant bit rate. The propagation radio model used was Two Ray Ground, and the data rate was set randomly. The QRT-Dos attack prevention system is displayed in Figure 4, and the IS-Dos attack prevention system is displayed in Figure 5.

Here it is clearly seen that the IS-Doc prevention system showing better performance compared to Dos and QRT-Dos. In the result analysis section 4.2 presented a complete comparative analysis is given between Dos, QRT-Dos and IS-Dos.

RESULTS AND DISCUSSION

The analysis involved performance evaluation of different prevention techniques.

Table 2. Simulation parameters for WSN DoS attack prevention

Parameters	Configuration Value
Simulation Area	1100 x 1100 meters
Network Type	Wireless Sensor Network (WSN)
Nodes/Devices	30
Material	Wireless
Movement of Node	Random
Number of Iteration	300
Length of Queue	50
MAC Layer	MAC 802.11p
Type of Attack	DoS flooding
Type of Prevention	QRT-DoS, IS-DoS
Type of Traffic	Constant Bit Rate, File Transfer Protocol
Propagation Radio Model	Two Ray Ground
Rate	Random

Packet Sends Analysis

Figure 6 displays the packet send analysis results for the suggested method. The analysis revealed that the proposed approach with the QRT-DoS prevention mechanism resulted in fewer packets being sent than the IS-DoS prevention mechanism when a DoS flooding attack was present in the VANET. Figure 6 illustrates how the suggested IS-DOS technique works to improve performance by preventing attacker flooding in the VANET. The different simulation parameters for preventing WSN DoS attacks are displayed in Table 2.

Analysis of Packet Receive

The impact of DoS flooding attacks on packet receive analysis is shown in Figure 7. The DoS attacker is very harmful forthe network because those vehicles that are in the range of attackersandthusattackers are directly or indirectly flooded flooding with fake message to all nearby nodes in the network. The fake packets are flooded by attacker node/s and the attacker is actually flooded with unwanted packets and drops in whole data packets. In the presence of DoS flooding attacks in VANET, the packet receive analysis with QRT-Dos was lower than that of the Improved Security (IS-dos) Approach. The attacker dropped data in the network but the proposed IS-DOS is stopping its malicious activities.

Throughput Analysis

The effects of the DoS flooding attack on the network's throughput were analyzed, and the results are presented in Figure 8. It is observed that the output of the QRT prevention mechanism is lower compared to the Improved Security (IS) approach when a DoS flooding attack is present in the vehicular network

Packet Drop Analysis

Figure 9 illustrates how DoS flooding attacks affect the AODV routing protocol's packet drop analysis. In the VANET, packet drop happens as a result of increased network traffic. The findings indicate that when DoS attacks

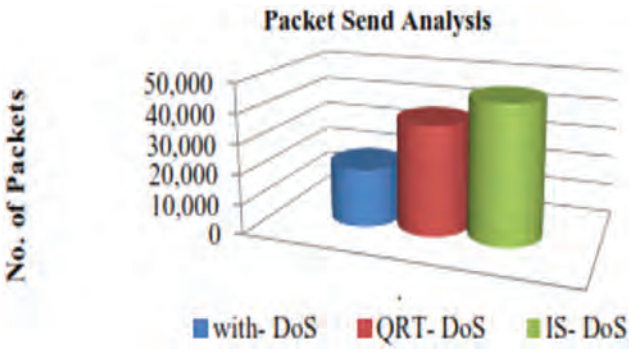


Figure 6. Packet send analysis using AODV protocol in VANET.

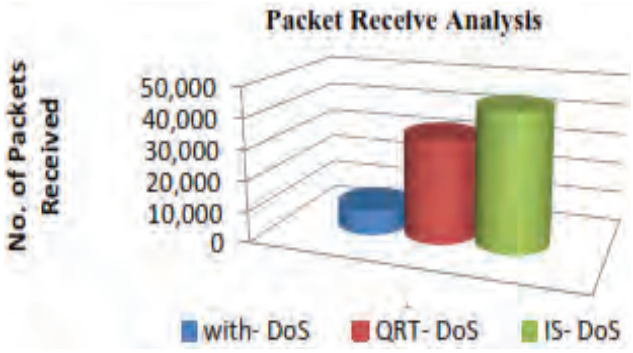


Figure 7. Packet Receive Analysis using AODV protocol in VANET.

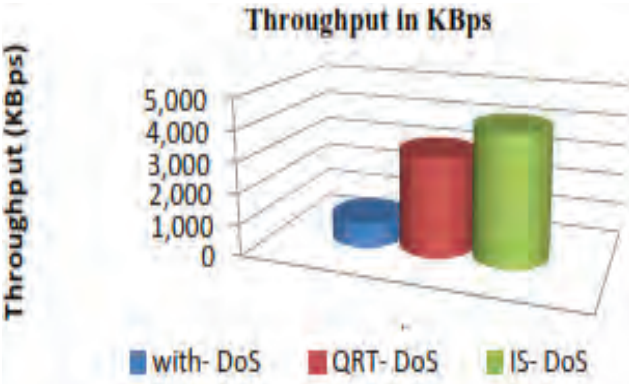


Figure 8. Throughput Analysis using AODV protocol in VANET.

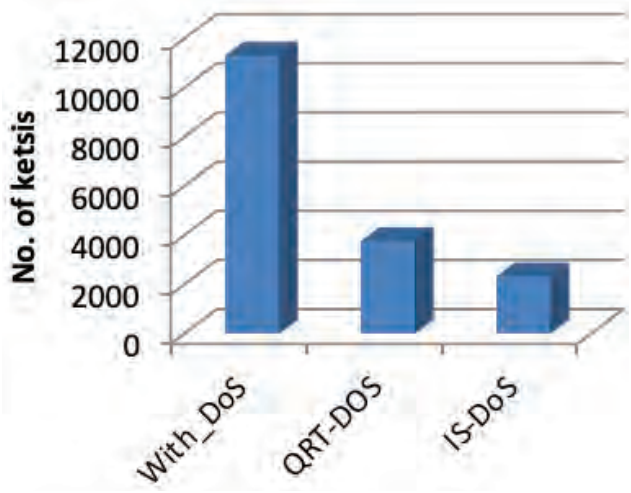


Figure 9. Packet drop analysis between proposed and existing techniques.

are present in the vehicular network, the packet drop analysis using QRT is higher than the Improved Security Approach.

Packet Delivery Ratio Analysis

The packet delivery ratio is an important metric that indicates the number of packets successfully delivered to their destination nodes compared to the total number of packets transmitted in the network. A higher packet delivery ratio shows better network performance. The impact of the DoS flooding attack on the packet delivery ratio in VANET is shown in Figure 10. It has been observed from figure, Figure 10 that the packet delivery ratio analysis with QRT is lower than the Improved Security (IS) Approach in the presence of a DoS flooding attack in the vehicular network. This means that the IS approach is more effective in preventing the impact of DoS attacks on the packet delivery ratio in VANET.

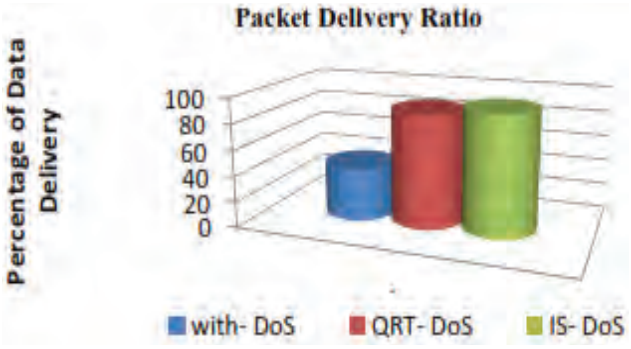


Figure 10. Packet Delivery Ratio analysis using AODV protocol in VANET.

Routing Load Analysis

Figure 11 illustrates the impact of DoS flooding attacks on routing load analysis. The routing load should be minimized to improve network performance, as it leads to delays. It has been observed that QRT results in a higher routing load than the Improved Security (IS) approach in the presence of DoS flooding attacks in vehicular networks.

Denial-of-Service Packets

This performance metric measures the packet loss due to the Denial-of-Service (DoS) attack in the VANET. Figure 11 displays the impact of DoS packets on VANET. The

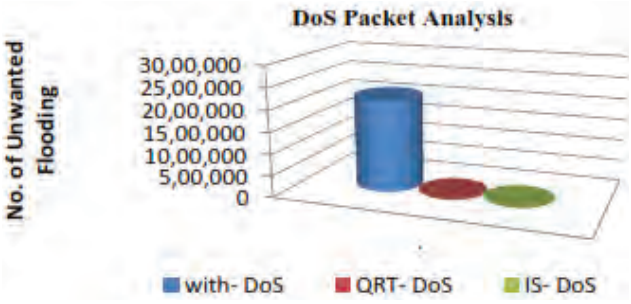


Figure 11. Denial of Service Packets using AODV protocol in VANET.

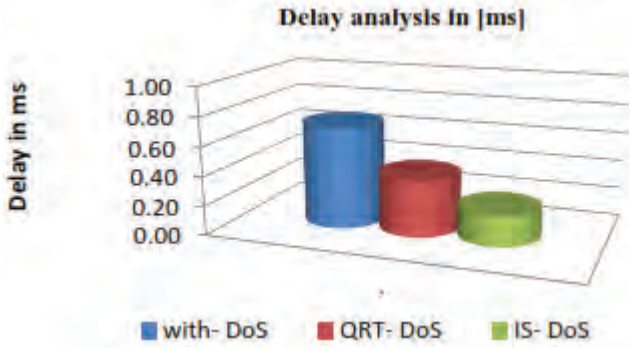


Figure 12. Delay analysis using AODV protocol in VANET.

results indicate that unwanted packet flooding is lower in the Improved Security (IS) approach than in the QRT-DoS approach.

Delay Analysis [ms]

This parameter shows the delay analysis due to a denial-of-service attack in VANET. Figure 12 shows the delay analysis, and it has been observed that the delay is very low in the Improved Security (IS) Approach compared to the QRT-DoS approach.

CONCLUSION

The Improved Security (IS) Approach is used for increasing security and improving the Vehicular network performance. Suppose abnormal behaviour of any vehicle is detected in the network. In that case, that vehicle's trust value is decreased and marked as untrusted by the control and monitoring unit (CMU) in its table. The following results were observed after implementing the IS Approach: (i) the Packet Delivery Ratio of the IS approach is nearly 5% more than the existing QRT approach; (ii) the throughput of the IS approach is about 1000 KBps more than the existing QRT approach; (iii) packet Drop Analysis of the IS approach is about 1400 packets less than the existing QRT approach. Other measuring parameters like packet send and receive analysis, routing load analysis, denial of service packets, and delay analysis also show better results with the IS approach.

Following future enhancements required for the improvement of the robustness, efficiency, and effectiveness of the IS-DOS prevention mechanisms:-

- Exploring advanced machine learning techniques that help improve efficiency, like improved deep learning techniques.
- In future studies, include 5G/ 6G enhanced high bandwidth communication technologies that help to reduce the latency and improve the efficiency of the VANET system.
- In the future, hybrid security systems may be implemented, like cryptographic-based methods.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Verma K, Hasbullah H, Kumar A. Prevention of DoS attacks in VANET. Wirel Pers Commun 2013;73:95-126. [CrossRef]

[2] Haydari A, Yilmaz Y. Real-time detection and mitigation of DDoS attacks in intelligent transportation systems. In: Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC); 2018. p. 157-163. [CrossRef]

[3] Singh D, Ranvijay, Yadav RS. A state-of-art approach to misbehaviour detection and revocation in VANET: Survey. Int J Ad Hoc Ubiquitous Comput 2018;28:77-93. [CrossRef]

[4] Gao Y, Wu H, Song B, Jin Y, Luo X, Zeng X. A distributed network intrusion detection system for distributed denial of service attacks in vehicular ad hoc network. IEEE Access 2019;7:154560-154571. [CrossRef]

[5] Lyamin N, Vinel A, Jonsson M, Loo J. Real-time detection of denial-of-service attacks in IEEE 802.11 p vehicular networks. IEEE Commun Lett 2013;18:110-113. [CrossRef]

[6] Asaar MR, Salmasizadeh M, Susilo W, Majidi A. A secure and efficient authentication technique for vehicular ad-hoc networks. IEEE Trans Veh Technol 2018;67:5409-5423. [CrossRef]

[7] Ahmad F, Franqueira VN, Adnane A. TEAM: A trust evaluation and management framework in context-enabled vehicular ad-hoc networks. IEEE Access 2018;6:28643-28660. [CrossRef]

[8] Usha M, Ramakrishnan B. Robust MPR: A novel algorithm for secure and efficient data transmission in VANET. Wirel Pers Commun 2020;110:355-380. [CrossRef]

[9] Bylykbashi K, Elmazi D, Matsuo K, Ikeda M, Barolli L. Effect of security and trustworthiness for a fuzzy cluster management system in VANETs. Cogn Syst Res 2019;55:153-163. [CrossRef]

[10] Kolandaisamy R, Md Noor R, Ahmedy I, Ahmad I, Reza Z'aba M, Imran M, et al. A multivariant stream analysis approach to detect and mitigate DDoS attacks in vehicular ad hoc networks. Wirel Commun Mob Comput 2018;2018:2874509. [CrossRef]

[11] Waraich PS, Batra N. Prevention of denial of service attack over vehicle ad hoc networks using quick response table. In: Proceedings of the 4th International Conference on Signal Processing, Computing and Control (ISPCC); 2017. p. 586-591. [CrossRef]
[12] Azees M, Vijayakumar P, Deboarh LJ. EAAP: Efficient anonymous authentication with conditional privacy-preserving scheme for vehicular ad hoc networks. IEEE Trans Intell Transp Syst 2017;18:2467-2476. [CrossRef]
[13] G R, C GK, Singh N, Patil VR. Autonomous forecasting of traffic in cellular networks based on long-short term memory recurrent neural network. Cybern Syst 2023;56:33-45. [CrossRef]
[14] Singh HP, Sheetlani J, Salimath N, Gopal KM. Design and implementation of an algorithm for mitigating the congestion in mobile ad hoc network. Int J Emerg Technol 2019;10:472-479.
[15] Singh HP, Singh R. Prevention mechanism of black hole and jamming attack in mobile ad-hoc network. J Harmon Res Eng 2020;8:1-7. [CrossRef]
[16] Krishna K, Vamshi, Reddy KG. VANET vulnerabilities classification and countermeasures: A review. Majlesi J Electr Eng 2022;16:63-83.
[17] Shahid KA. Secure trust-based blockchain architecture to prevent attacks in VANET. Sensors 2019;19:49-54. [CrossRef]
[18] Poongodi M. Intrusion prevention system for DDoS attack on VANET with reCAPTCHA controller using information based metrics. IEEE Access 2019;7:158481-158491. [CrossRef]
[19] Bashir RR, Saeed Y, Ali A, Algarni AD, Muthanna A, Hijjawi M, et al. 2cap: A novel curve crash avoidance protocol to handle curve crashes in vehicular ad-hoc network. IEEE Access 2024;12:60601-60619. [CrossRef]
[20] Gopi R. Intelligent DoS attack detection with congestion control technique for VANETs. Comput Mater Contin 2022;72(1). [CrossRef]
[21] Shiek Aalam S, Sivakumar M, Anjali Devi S, Teja PVS, Singh N. Emerging trends in vehicular communication and routing for intelligent transportation systems. In: Bhatia J, Tanwar S, Rodrigues JJPC, Kumhar M, editors. Deep Learning Based Solutions for Vehicular Adhoc Networks. Studies in Computational Intelligence, Vol. 1207. Singapore: Springer; 2025. [CrossRef]
[22] Gaurav A. DDoS attack detection in vehicular ad-hoc network (VANET) for 5G networks. In: Security and Privacy Preserving for IoT and 5G Networks: Techniques, Challenges, and New Directions; 2022. p. 263-278. [CrossRef]
[23] Singh N, Singh HP, Mishra A, Khare A, Swarnkar M, Almas SK. Blockchain cloud computing: Comparative study on DDoS, MITM and SQL injection attack. In: Proceedings of the IEEE International Conference on Big Data & Machine Learning (ICBDML); 2024; Bhopal, India. p. 73-78. [CrossRef]
[24] Nayak RP. MLMDS: Machine learning based misbehavior detection system for cognitive software-defined multimedia VANETs in smart cities. Multimed Tools Appl 2023;82(3):3931-3951. [CrossRef]
[25] Bhayo J, Shah SA, Hameed S, Ahmed A, Nasir J, Draheim D. Towards a machine learning-based framework for DDOS attack detection in software-defined IoT (SD-IoT) networks. Eng Appl Artif Intell 2023;123:1-17. [CrossRef]
[26] Singh HP, Patel H, Singh N, Rane P, Soni S. DLAR: Deep learning-based adaptive routing strategies for vehicular ad-hoc networks (VANETs). In: Bhatia J, Tanwar S, Rodrigues JJPC, Kumhar M, editors. Deep Learning Based Solutions for Vehicular Adhoc Networks. Studies in Computational Intelligence, Vol. 1207. Singapore: Springer; 2025. [CrossRef]
[27] Ali WA. Intrusion detection system for vehicular ad hoc network attacks based on machine learning techniques. Inf Secur J A Glob Perspect 2024;33(6):128-139. [CrossRef]
[28] Singh N, Suhane P, Singh HP, Rane P, Shukla D. Traffic prediction and modeling in vehicular ad hoc networks. In: Bhatia J, Tanwar S, Rodrigues JJPC, Kumhar M, editors. Deep Learning Based Solutions for Vehicular Adhoc Networks. Studies in Computational Intelligence, Vol. 1207. Singapore: Springer; 2025. [CrossRef]
[29] Verma A, Saha R. IPREVIR: Fortifying vehicular networks against denial of service attacks. IEEE Access 2024;12:48301-48320. [CrossRef]
[30] Memon I, Shaikh RA, Shaikh H. Dynamic pseudonyms trust-based model to protect attack scenario for internet of vehicle ad-hoc networks. Multimed Tools Appl 2024;83:13395-13426. [CrossRef]
[31] Al-Shareeda MA. SE-CPPA: A secure and efficient conditional privacy-preserving authentication scheme in vehicular ad-hoc networks. Sensors 2021;21:1-22. [CrossRef]
[32] Bangui H, Ge M, Buhnova B. A hybrid machine learning model for intrusion detection in VANET. Computing 2022;104(3):503-531. [CrossRef]
[33] Khan IA. An enhanced multi-stage deep learning framework for detecting malicious activities from autonomous vehicles. IEEE Trans Intell Transp Syst 2021;13(12):25469-25478. [CrossRef]
[34] Shah AS, Karabulut MA. Optimization of drones communication by using meta-heuristic optimization algorithms. Sigma J Eng Nat Sci 2022;40(1):108-117. [CrossRef]
[35] Kemal E. Cross-layer design in wireless AD HOC and sensor networks. Sigma 2010;28:66-73.
[36] Tonbul E, Takan MA, Büyükköse GT, Erginel N. Modeling open vehicle routing problem with real life costs and solving via hybrid civilized genetic algorithm. Sigma J Eng Nat Sci 2024;42(3):714-730. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2021

Research Article

Solving Hamiltonian systems based on a data-driven deep learning algorithm

Tayfun ÜNAL[1,2,*], Ayten İrem IŞIK[2], Ünver ÇİFTÇİ[2]

[1]*Department of Information Technologies, Kırklareli University, Kırklareli, 39100, Türkiye*
[2]*Department of Mathematics, Tekirdağ Namık Kemal University, Tekirdağ, 59030, Türkiye*

ARTICLE INFO

Article history
Received: 12 June 2024
Revised: 18 September 2024
Accepted: 02 February 2025

Keywords:
Hamiltonian Systems, Data-driven, Deep Learning, Energy Conserving

ABSTRACT

Hamiltonian systems possess important properties, such as the preservation of the symplectic structure and the conservation of energy. Traditional numerical iteration methods generally violate these properties. This paper proposes a data-driven deep learning algorithm to solve separable Hamiltonian systems. It is computationally efficient and scalable. The proposed algorithm tries to hold the balance between symplectic and energy-conserving. Additionally, it can work efficiently with few data points, and the obtained solution is continuous. The algorithm consists of three main steps: (i) obtaining data points from a solver, (ii) data augmentation, and (iii) ensuring energy conservation. The algorithm was evaluated on the simple harmonic oscillator, nonlinear oscillator, and Lotka-Volterra systems. The proposed algorithm has demonstrated successful performance across all these systems despite the few data. In conclusion, it has been experimentally demonstrated that the proposed method is effective even with few data points. Furthermore, the proposed algorithm can effectively work regardless of the solver chosen in all examples.

Cite this article as: Ünal T, Işık Aİ, Çiftçi Ü. Solving Hamiltonian systems based on a data-driven deep learning algorithm. Sigma J Eng Nat Sci 2026;44(2):994–1003.

INTRODUCTION

Solving differential equations with deep learning algorithms dates back approximately three decades [1, 2]. To date, two main approaches have been widely used in research for this purpose. One approach involves solving differential equations using equation-driven deep learning algorithms [1, 3, 4]. Another one is using data-driven deep learning algorithms [5, 6]. Recently, the utilization of both equation-driven and data-driven deep learning algorithms has gained popularity in the numerical study of dynamical systems [7]. One of the most significant motivations for this approach is that it can incorporate the laws of physics [8], as the time-dependent dynamics of a system are governed by physical laws such as energy and momentum. An example of this is Hamiltonian systems, which are dynamical systems that conserve energy [9]. To solve Hamiltonian systems, a significant amount of research has been conducted utilizing both approaches. However, they have employed automatic differentiation, which is more computationally expensive,

*Corresponding author.
*E-mail address: tayfununal39@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

or utilized complex deep learning models. In this context, this paper proposes a data-driven deep learning algorithm to solve separable Hamiltonian systems without automatic differentiation and complex deep learning architectures.

In recent years, an equation-driven approach for Hamiltonian systems has been introduced by Mattheakis et al. [4]. They proposed a Hamiltonian neural network architecture designed to approximate Hamiltonian trajectories. Motivated by [4], this paper utilizes their model but in a different manner. The first difference is that a data-driven model is employed here. The second point is that they employed automatic differentiation to obtain the derivatives of the model with respect to the time variable. However, the derivatives of the neural network were not computed here, as the data was obtained from an ordinary differential equation solver.

For Hamiltonian systems, there has been increased research that employs the data-driven approach. In particular, Greydanus et al. [10] presented Hamiltonian Neural Networks (HNN). They designed a neural network that represents the Hamiltonian function. Automatic differentiation was used in the training of HNN. The notable aspect of the study, and another difference from this paper, is the learning of the Hamiltonian function using pixel data. Subsequently, HNNs were developed to solve chaotic systems by Han et al. [11] and named adaptable HNN. Meanwhile, HNNs were utilized to learn the trajectories of Hamiltonian systems in [12]. Furthermore, there are various works concerning the representation of the Hamiltonian function using neural networks [13–15]. These works employed automatic differentiation, which, as previously mentioned, is computationally expensive. Moreover, they utilized advanced architectures, such as autoencoders [13] and symmetry-preserving extensions of HNNs [14], among others.

Physics-inspired neural networks have been introduced for solving differential equations [8, 16–22], where they incorporate the laws of physics. Karniadakis et al. [8] proposed the physics-informed neural network (PINN). They constructed a loss function to satisfy differential equations and added a new term to the loss function to ensure the conservation of physical laws. The derivatives of the neural network were obtained using automatic differentiation once again. Recent advancements in PINNs encompass improvements in training algorithms, architectural designs, and computational efficiency. A neural network-based architecture called SciNet was presented by Iten et al. [16] with the goal of simulating human physical thinking in order to extract basic physical principles from empirical data without the need for prior knowledge. SciNet demonstrates how these neural networks can independently discover physical notions. Yu et al. [17] developed Gradient Enhanced Physics Knowledge Neural Networks (gPINNs) to improve training performance and accuracy, using residual gradients of PDEs. To further improve performance, they also combined gPINNs with a residual-based adaptive refinement method. First-Order PINNs (FO-PINNs) were developed by Glasdstone et al. [18]. It reformulates higher-order PDEs into a series of first-order equations. Thus, FO-PINNs reduce training time per iteration by avoiding the computation of higher-order derivatives using automatic differentiation. Moreover, it enhanced accuracy compared to standard PINNs. Wang et al. [19] developed the NAS-PINN, which employs Neural Architecture Search (NAS) to automatically identify optimal neural network architectures for solving PDEs. Using PINNs, Norambuena et al. [20] suggested a computational approach to optimal quantum control problems. They show that PINNs are more flexible and energy efficient than traditional control methods when it comes to solving state-to-state transmission in open quantum systems. For conventional numerical integrators in Hamiltonian systems, maintaining phase space structures and energy conservation over extended time scales is a major difficulty. In this regard, SympFlow, a neural network-based symplectic integrator that incorporates time-dependent Hamiltonian functions, was developed by Canizares et al. [21]. By leveraging physics-informed machine learning techniques, SympFlow effectively addresses these challenges. Another work, called the physics-enhanced neural network, was presented by Choudhary et al. [22]. They applied HNN to chaotic systems.

Moreover, there is a research perspective focused on defining Hamiltonian systems using neural networks that preserve the symplectic structure [23–25]. Tong et al. [23] presented a fourth-order integrator using a neural network architecture called TaylorNet. Additionally, symplectic networks (SympNets) were introduced by [24]. Hamiltonian systems are learned using SympNet in the paper. Subsequently, SympNets were improved by [25], and the architecture was designed to locally preserve the symplectic structure.

Another aspect of the study involves learning dynamical systems using Poisson neural networks [26], as well as employing deep learning to represent the Koopman operator [27] and using thermodynamics-aware time integrators for learning dynamical systems [28].

In this paper, a data-driven deep learning algorithm is proposed to solve separable Hamiltonian systems. The algorithm is applied to the simple harmonic oscillator, nonlinear oscillator, and Lotka-Volterra as examples. The results were then evaluated and compared with the scipy.integrate.odeint method in Python. The algorithm outperforms its counterparts, even with few data. The requirements for the proposed algorithm are a separable Hamiltonian system and a traditional solver, such as Euler, symplectic Euler, Verlet, or Runge-Kutta. Numerical symplectic integrators cannot conserve energy. Conversely, numerical methods that conserve energy cannot preserve the symplectic structure [29]. Thus, the algorithm is designed to ensure conservation in both cases. To achieve this, a symplectic integrator, such as symplectic Euler or Verlet, is employed to obtain data. The contributions of this work are outlined as follows:

- The proposed algorithm solves Hamiltonian equations with few data.
- It provides a continuous solution and scales to high dimensions.
- It balances the preservation of symplectic structure and energy conservation.

The structure of this paper is organized as follows: In Section 2, Hamiltonian systems will be introduced. Subsequently, the algorithm will be described in Section 3. Section 4 includes the examples. In Section 5, the results will be presented. Finally, the results are evaluated, and the discussion will focus on future work.

Hamiltonian Systems

A Hamiltonian system is a dynamical system characterised by a function $\mathcal{H}$: $\mathbb{R}^{2d} \to \mathbb{R}$ that satisfies the following equations:

$$\frac{d\boldsymbol{q}}{dt} = \frac{\partial \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{p}}, \quad \frac{d\boldsymbol{p}}{dt} = -\frac{\partial \mathcal{H}(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{q}}, \tag{1}$$

where $q = (q_1, q_2, \ldots, q_d) \in \mathbb{R}^d$ and $p = (p_1, p_2, \ldots, p_d) \in \mathbb{R}^d$ are usually called as position and momentum vectors in the phase space, respectively. Additionally, t represents the time variable here.

In this paper, separable Hamiltonian systems of the form

$$\mathcal{H}(q,p) \quad = \mathcal{T}(p) + \mathcal{V}(q) \tag{2}$$

are considered, where $\mathcal{T}(p)$ represents the kinetic energy function and $\mathcal{V}(q)$ denotes the potential energy function. One of the most significant properties of Hamiltonian systems is the conservation of energy. This property arises from the fact that the derivative of $\mathcal{H}$ with respect to t is zero, as demonstrated by Equation .

$$\frac{\partial \mathcal{H}}{\partial t} = \frac{\partial \mathcal{H}}{\partial q}\frac{dq}{dt} + \frac{\partial \mathcal{H}}{\partial p}\frac{dp}{dt} = \frac{\partial \mathcal{H}}{\partial q}\left(\frac{\partial \mathcal{H}}{\partial p}\right) + \frac{\partial \mathcal{H}}{\partial p}\left(-\frac{\partial \mathcal{H}}{\partial q}\right) = 0 \tag{3}$$

Therefore, it can be concluded that $\mathcal{H}$ remains constant for all t. From this perspective, the deep learning algorithm proposed in this study ensures that the solution of the system adheres to the constraint of energy conservation.

MATERIALS AND METHODS

The proposed method consists of the following three steps:

1. Solve the system with few data points t values (denoted by t_s) within the specified time interval using a symplectic solver, such as symplectic Euler or Verlet.
2. Develop a deep learning model and train it on the data points obtained from the solver, utilizing the Mean Squared Error (MSE) loss function. After the training phase, augment the dataset by generating additional t points (denoted by t_m) using the trained model. Specifically, the model predicts the values of q and p corresponding to the additional t_m data points within the specified time interval.
3. Finally, the augmented dataset is trained using an energy-conserving loss function for the same deep learning model. Upon completion of this training, the model is capable of providing a continuous approximate solution for the specified time interval t.

Figure 1 provides an overview of the solution steps.

In this study, both the symplectic Euler and Verlet schemes are employed as solvers for all examples. For the Hamiltonian system, the symplectic Euler and Verlet schemes solve system using the algorithms provided in equations and , respectively:

$$\begin{aligned} q_{i+1} &= q_i + h\nabla_p \mathcal{H}(q_{i+1}, p_i) \\ p_{i+1} &= p_i - h\nabla_q \mathcal{H}(q_{i+1}, p_i) \end{aligned} \tag{4}$$

$$\begin{aligned} p_{i+1/2} &= p_i - \tfrac{h}{2}\nabla_q \mathcal{H}(q_i, p_{i+1/2}), \\ q_{i+1} &= q_i + \tfrac{h}{2}(\nabla_p \mathcal{H}(q_i, p_{i+1/2}) + \nabla_p \mathcal{H}(q_{i+1}, p_{i+1/2})), \\ p_{i+1} &= p_{i+1/2} - \tfrac{h}{2}\nabla_q \mathcal{H}(q_{i+1}, p_{i+1/2}). \end{aligned} \tag{5}$$

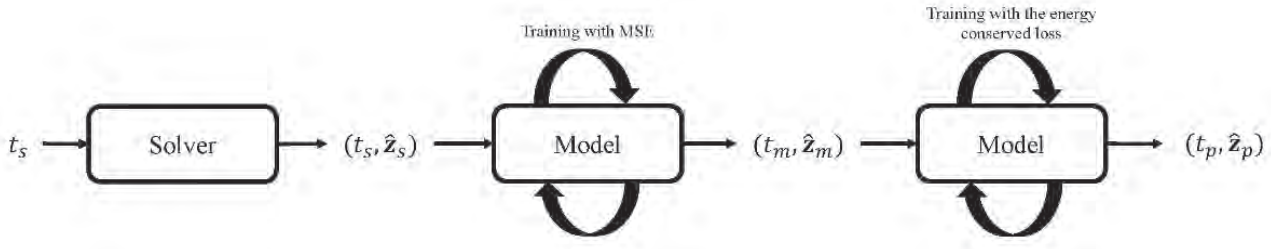


Figure 1. The proposed method consists of three steps. In the first step, the system is solved with a solver for a small number of points, t_s. In the second step, training is performed using the data points obtained, with a model that employs an MSE loss function, and additional data points are predicted for t_m using the model. In the final step, training is conducted with a model that employs an energy-conserving loss function, and predictions are made for t_p. As a result, the final model gains the capability to make predictions for any t.

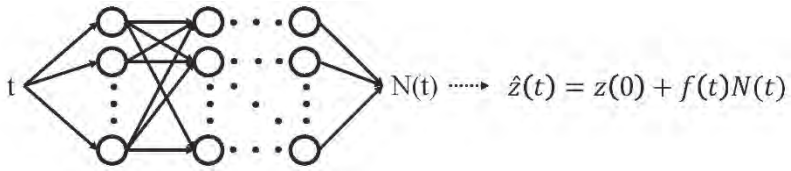


Figure 2. In both models, the structure is identical to that of a fully connected deep learning model, except for the final layer. In each model, the final layer is designed as $\hat{z}(t) = z(0) + f(t)N(t)$, where $N(t)$ represents the output of the last layer of the fully connected model, ensuring the initial conditions of the system are satisfied. The input is the time t, and the output consists of $\hat{z}(t) = (q, p)$ values.

In these equations (4) and (5), h represents the step size and q_i, p_i are the solutions at the i-th time point. Moreover, it is explicit that $\nabla_q \mathcal{H}(q, p) = \nabla_q \mathcal{V}(q)$ and $\nabla_p \mathcal{H}(q, p) = \nabla_p \mathcal{T}(p)$ for seperable Hamiltonian systems.

As mentioned above, given an initial point defined as $z_0 = z(t_0 = 0) = (q_0, p_0)$ and a time interval t, the system described in Equation is initially solved using either the symplectic Euler or Verlet scheme. As a result, a dataset is obtained that comprises $\{(t_0, \hat{z}_0), (t_1, \hat{z}_1), \ldots, (t_s, \hat{z}_s)\}$. Secondly, this dataset is used in the deep learning algorithm that has the architecture in Figure 1.

In this model, the architecture is similar to multilayer perceptron (MLP) except the last layer (6) where the output is:

$$\hat{\boldsymbol{z}}(t) = \boldsymbol{z}(0) + f(t)\boldsymbol{N}(t). \tag{6}$$

Here, $N(t)$ represents the output of model before the last layer. The last layer have been just designed for satisfying the given initial point [4]. Accordingly, f is defined as:

$$f(t) = 1 - e^{-t^2} \tag{7}$$

Equation so that $\hat{z}(0) = z(0)$ when $t = 0$. After training the model with the MSE loss function defined in Equation, predictions of $\hat{z}$ can be made for various t points. The model has just fitted to the dataset. In contrast to the Hamiltonian, the model has been employed to approximate a different trajectory so far. However, the model provides flexibility for augmenting data points. For $s < m$, m sample is taken by uniform distribution from given interval t. Then for each t_i, $i = 1, 2, \ldots, m$, model makes a prediction. Thus, an augmented dataset is obtained, denoted as $\{(t_1, \hat{z}_1), (t_2, \hat{z}_2), \ldots, (t_m, \hat{z}_m)\}$, which will be utilized by the model in the second training phase with the energy-conserved loss function defined in Equation. In the final step, the energy-conserved loss function was designed to enhance the model's performance.

As a result, this method guarantees excellent performance even when using sparse data points obtained from a symplectic solver. The proposed method outperforms competing strategies by avoiding neural network derivative calculations. It reduces computational load and increases numerical stability. The approach is also scalable. By simply changing the parameters, it can be applied to a wide variety of complex physical problems.

MSE Loss Functions

In the first training, the widely recognized loss function known as the Mean Squared Error (MSE) was utilized, defined as follows:

$$\mathcal{L}_{MSE}(\boldsymbol{z}, \hat{\boldsymbol{z}}(t)) = \frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{z}_i - \hat{\boldsymbol{z}}(t_i))^2. \tag{8}$$

The significance of this training lies in its capacity to learn from the dataset, thereby facilitating the augmentation of data points after the training process.

Energy-Conserving Loss Function

The second loss function, referred to as the energy-conserving loss function, is defined as follows:

$$\begin{aligned}\mathcal{L}_{Energy}(\boldsymbol{z}, \hat{\boldsymbol{z}}(t)) = \gamma \frac{1}{k}\sum_{i=1}^{k}(\boldsymbol{z}_i - \hat{\boldsymbol{z}}(t_i))^2 \\ + (1-\gamma)\frac{1}{m}\sum_{i=1}^{m}(\mathcal{H}(\boldsymbol{z}_0) \\ - \mathcal{H}(\hat{\boldsymbol{z}}(t_m)))^2\end{aligned} \tag{9}$$

This loss function consists of two components: the first part ensures that the model preserves the basic structure of the data, while the second part ensures the conservation of the Hamiltonian. Here, γ is a crucial hyperparameter because it controls the balance between fitting the data and ensuring the conservation of the Hamiltonian. Its value directly affects the degree to which the Hamiltonian is preserved throughout the optimization process. This implies that, if γ is chosen sufficiently small, the model places more emphasis on energy conservation, thus prioritizing the conservation of the Hamiltonian over the precise fitting of the data.

EXAMPLES

Deep learning algorithms are highly effective tools for learning invariant patterns, capturing complex patterns and relationships within a dataset. Even when trained on relatively small datasets, they have the ability to effectively generalize by identifying underlying patterns and extending beyond the given data. In this study, the proposed model is applied to Hamiltonian dynamical systems, especially the

simple harmonic oscillator, the nonlinear oscillator, and the Lotka-Volterra system. The performance of the proposed model is compared to that of the 'scipy.integrate.odeint' solver, which employed 1500 time points to solve the non-linear oscillator and Lotka-Volterra systems. However, for the simple harmonic oscillator, the analytical solution is utilized owing to its accessibility. The hyperparameters were selected based on insights gained from experimental observations. The common parameters and hyperparameters of the neural network are presented in Table 1.

The sine function is utilized as the activation function due to the enhancement of the model's predictive capability by trigonometric functions [4]. The experiments were conducted on virtual machines provided by Google Colab, equipped with 2 vCPUs and 13GB of RAM. Additionally, the Python programming language and the TensorFlow library were used in all examples[1].

Table 1. The common parameters and hyper-parameters

	Values		Values
Input dim	1	Activation function	sinus
Output dim	2	Augmented t (tm)	10000
Hidden layer	20	Learning rate	0.0001
Neurons	64	γ (Gamma)	0.0001
Epoch	5000		

Simple Harmonic Oscillator

The kinetic and potential energies are defined in Equation , respectively:

$$\mathcal{T}(p) = \frac{p^2}{2m}, \quad \mathcal{V}(q) = \frac{kq^2}{2}. \tag{10}$$

For simplicity, let $k = m = 1$; then the Hamiltonian is given by:

$$\mathcal{H}(q,p) = \frac{q^2}{2} + \frac{p^2}{2}. \tag{11}$$

The initial point, solution interval, sample size, and batch size for the simple harmonic oscillator are provided in Table 2.

Table 2. The parameters for simple harmonic oscillator

	Symplectic-Euler and Verlet
Initial point	(0, 1)
Interval of t	[0, 2π]
Samples of t	5
Batch Size	1000

Nonlinear Oscillator

The kinetic and potential energies are defined, respectively, as follows:

$$\mathcal{T}(p) = \frac{p^2}{2}, \quad \mathcal{V}(q) = \frac{q^2}{2} + \frac{q^4}{4}. \tag{12}$$

Thus, the Hamiltonian takes the form:

$$\mathcal{H}(q,p) = \frac{p^2}{2} + \frac{q^2}{2} + \frac{q^4}{4}. \tag{13}$$

For the nonlinear oscillator, the initial point, solution interval, sample size, and batch size are provided in Table 3.

Table 3. The parameters for nonlinear oscillator

	Symplectic-Euler and Verlet
Initial point	(1.3, 1)
Interval of t	[0, 4π]
Samples of t	25
Batch size	1000

Lotka-Volterra

The Hamiltonian of the Lotka-Volterra dynamical system is defined as:

$$\mathcal{H}(x,y) = \delta x - \gamma ln(x) + \beta y - \alpha ln(y), \tag{14}$$

where x and y do not represent canonical coordinates. Assuming $q = ln(y)$, $p = ln(x)$, and setting $\alpha = 2/3$, $\beta = 4/3$, $\gamma = \delta = 1$, the Hamiltonian takes the form:

$$\mathcal{H}(q,p) = e^p - p + \frac{4}{3}e^q - \frac{2}{3}q, \tag{15}$$

where

$$\begin{aligned} \mathcal{T}(p) &= \delta e^p - \gamma p = e^p - p, \\ \mathcal{V}(q) &= \beta e^q - \alpha q = \frac{4}{3}e^q - \frac{2}{3}q. \end{aligned} \tag{16}$$

For the Lotka-Volterra problem, the initial point, solution interval, sample size, and batch sizes are specified in Table 4.

Table 4. The parameters for Lotka-Volterra

	Symplectic-Euler and Verlet
Initial Point	(-0.105, -0.105)
Interval of t	[0, 50]
Samples of t	45
Batch Size	5000

1 Project page is available at https://tayfununal.github.io/Article-2/

RESULTS AND DISCUSSION

Initially, the symplectic Euler method is employed as the solver in the proposed model. Only 5 data points are utilized for the harmonic oscillator, 25 data points for the nonlinear oscillator, and 45 data points for the Lotka-Volterra system during the training phase. Subsequent to the training, the results are presented in Figures Figure 3, Figure 4, and Figure 5. Despite training the model with a minimal amount of data, high performance is achieved across all examples. In the cases of the simple harmonic oscillator, nonlinear oscillator, and Lotka-Volterra system, the model demonstrated significantly better energy conservation compared to the symplectic-Euler method and provided a closer approximation to the ground truth.

MSE values for energy conservation were computed by comparing both the symplectic-Euler method and the proposed model to the ground truth in Table 5. This

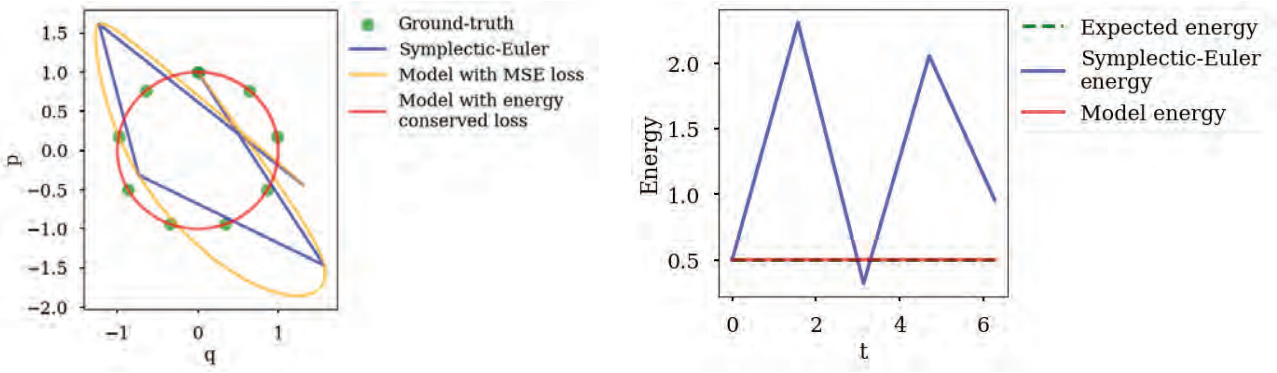


Figure 3. Phase diagram and energy graph of the harmonic oscillator using the symplectic-Euler solver.

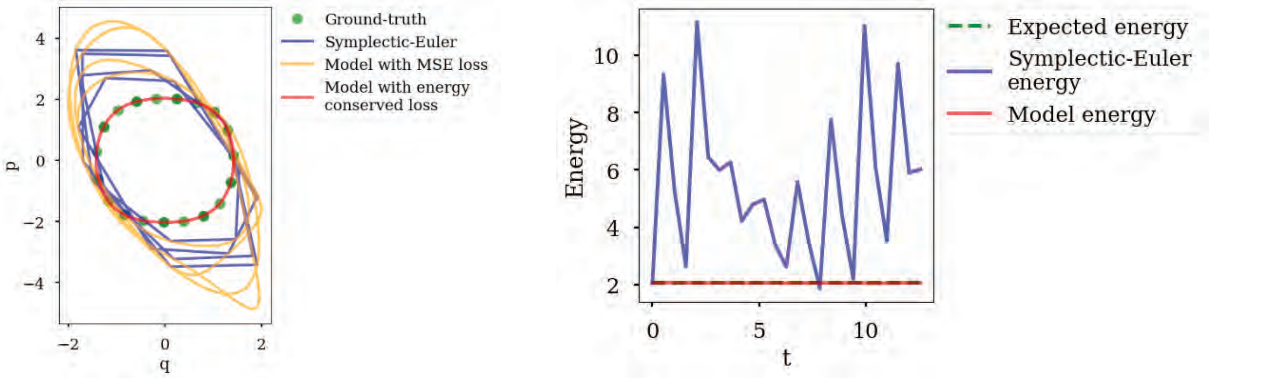


Figure 4. Phase diagram and energy graph of the nonlinear oscillator using the symplectic-Euler solver.

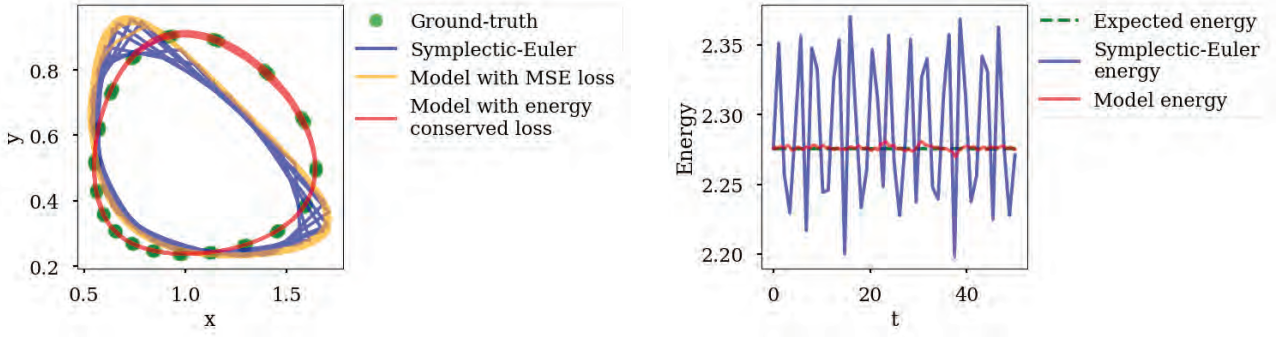


Figure 5. Phase diagram and energy graph of the Lotka-Volterra system using the symplectic-Euler solver.

Table 5. MSE values for both symplectic-Euler and the proposed model according to ground-truth

	MSE values	
Examples	**Symplectic -Euler**	**The proposed model**
Simple harmonic oscillator	1.1874350	0.0000010
Non-linear oscillator	18.471113	0.0000940
Lotka-Volterra	0.0029499	0.0000028

comparison highlights the effectiveness of the proposed model in maintaining energy conservation more accurately than the symplectic-Euler method.

The second solver employed in this model is the Verlet method. As previously mentioned, the examples utilize 5, 25, and 45 data points, respectively. The results are presented in Figures Figure 6, Figure 7, and Figure 8.

MSE values for energy conservation were computed by comparing both the Verlet method and the proposed model to the ground truth in Table 6. This comparison highlights

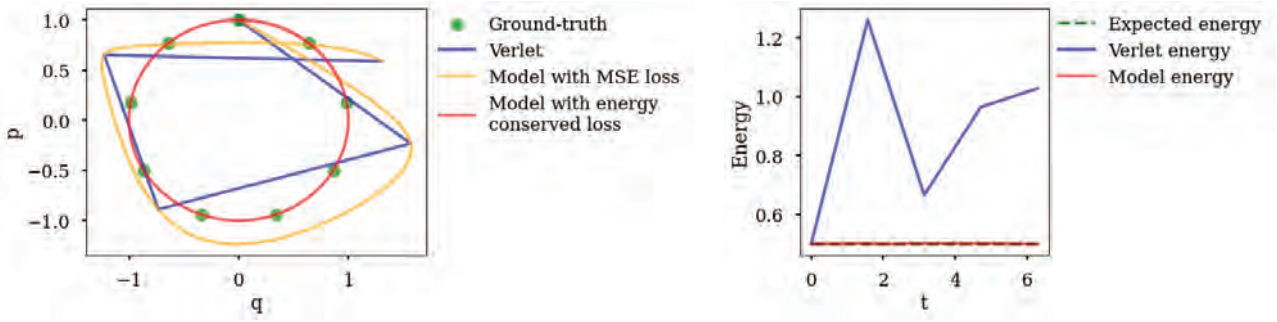


Figure 6. Phase diagram and energy graph of the harmonic oscillator using the Verlet method as the solver.

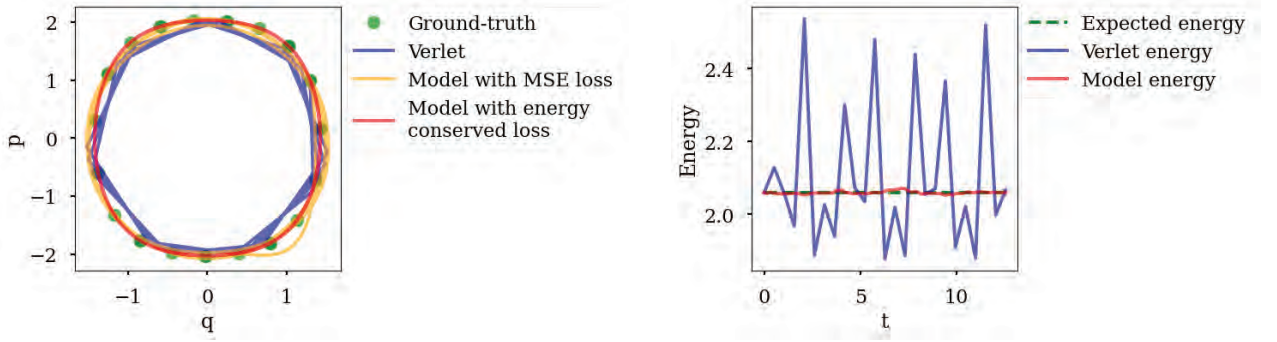


Figure 7. Phase diagram and energy graph of the nonlinear oscillator using the Verlet method as the solver.

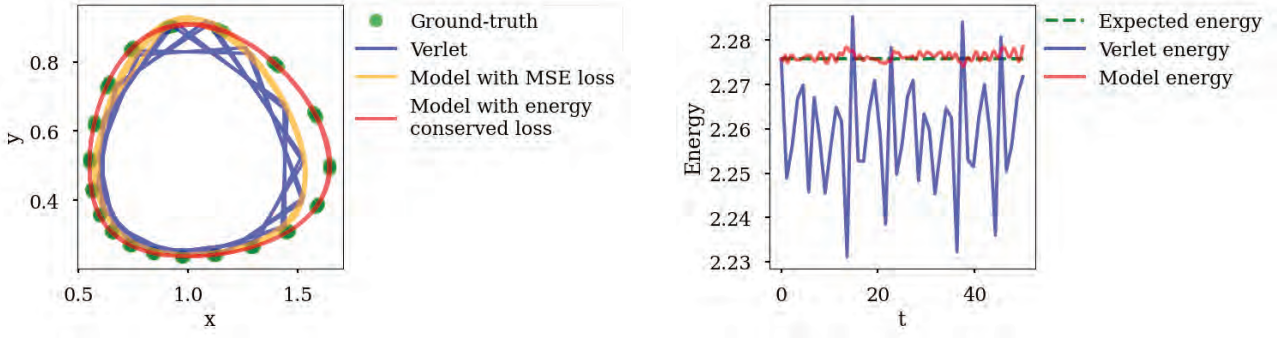


Figure 8. Phase diagram and energy graph of the Lotka-Volterra system using the Verlet method as the solver.

Table 6. MSE values for both Verlet and the proposed model according to ground-truth

	MSE values	
Examples	**Verlet**	**The proposed model**
Simple harmonic oscillator	0.220253	0.0000030
Non-linear oscillator	0.044066	0.0000250
Lotka-Volterra	0.000431	0.0000003

the effectiveness of the proposed model in maintaining energy conservation more accurately than the Verlet method.

Lastly, a comparison between the results obtained using the symplectic-Euler method and the Verlet method is presented for all examples, as illustrated in Figures Figure 9, Figure 10, and Figure 11. Although the Verlet method conserves the symplectic structure more effectively than the symplectic-Euler method, the results demonstrate that the model performs exceptionally well with both solvers.

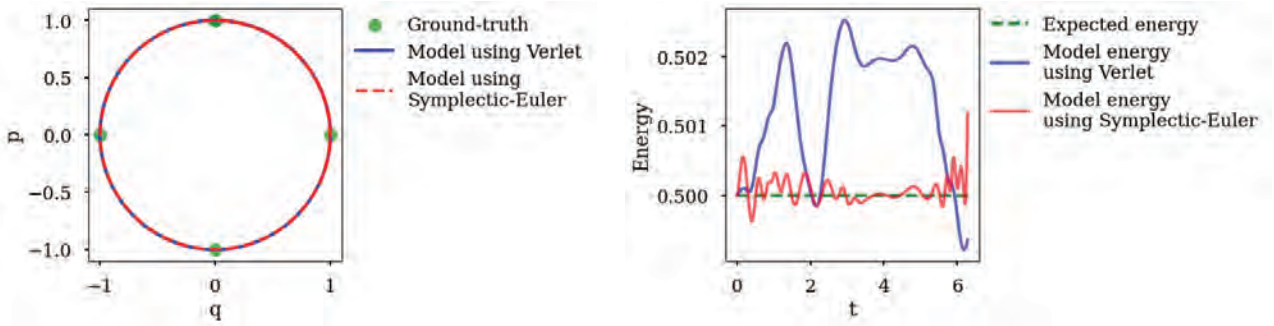


Figure 9. Symplectic-Euler versus Verlet methods in the phase diagram and energy graph of the harmonic oscillator.

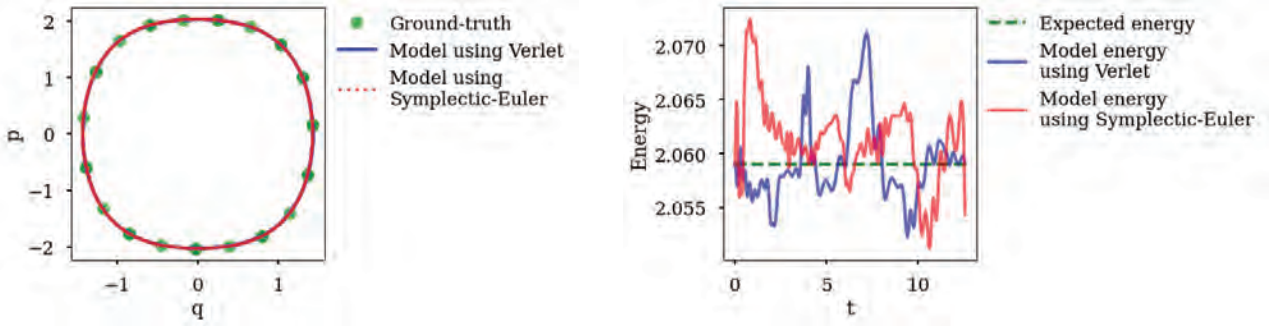


Figure 10. Symplectic-Euler versus Verlet methods in the phase diagram and energy graph of the nonlinear oscillator.

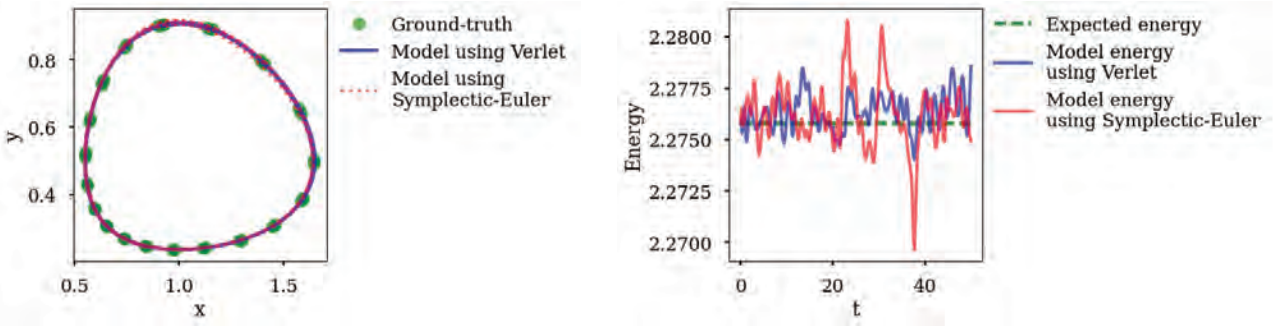


Figure 11. Symplectic-Euler versus Verlet methods in the phase diagram and energy graph of the Lotka-Volterra system.

CONCLUSION

A data-driven deep learning algorithm for solving separable Hamiltonian systems has been presented. With the proposed algorithm, Hamiltonian systems have been solved numerically. Since Hamiltonian systems are widely used in physics and engineering, the proposed algorithm will be beneficial in these fields. For instance, many models in mechanics can be expressed in Hamiltonian form. Therefore, the suggested algorithm can be applied in the study and control of mechanical systems. The proposed algorithm offers advantages over existing schemes by providing a continuous solution that balances the preservation of symplectic structure and energy conservation even with few data. There are no limitations on the choice of an ordinary differential equation solver in the initial step. In this study, Symplectic-Euler and Verlet methods were employed; however, alternative solvers such as Euler, Runge-Kutta, or neural ordinary differential equation solvers may also be utilized. Experimental results demonstrate that the algorithm performs effectively, even when utilizing few data points. An important distinction of the proposed method compared to its counterparts is that it does not utilize the differential of the neural network. This approach renders the method easy to implement, allowing it to be applied similarly to ordinary regression problems, which is advantageous in terms of numerical stability. In future work, the research will be extended to address partial differential equations utilizing the proposed method with certain modifications. Furthermore, other deep learning models, such as autoencoders and flow-based models, will be employed to augment the data.

REFERENCES

[1] Lagaris IE, Likas A, Fotiadis DI. Artificial neural netwoxrks for solving ordinary and partial differential equations. IEEE Trans Neural Netw 1998;9:987-1000. [CrossRef]

[2] Lee H, Kang IS. Neural algorithm for solving differential equations. J Comput Phys 1990;91:110-131. [CrossRef]

[3] Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural ordinary differential equations. Adv Neural Inf Process Syst 2018;31.

[4] Mattheakis M, Sondak D, Dogra AS, Protopapas P. Hamiltonian neural networks for solving equations of motion. Phys Rev E 2022;105:65305. [CrossRef]

[5] Yildiz S, Goyal P, Bendokat T, Benner P. Data-driven identification of quadratic symplectic representations of nonlinear Hamiltonian systems. arXiv [Preprint] 2023:arXiv:2308.01084. [CrossRef]

[6] Chen S, Kevrekidis PG, Zhang H, Zhu W. Data-driven discovery of conservation laws from trajectories via neural deflation. arXiv [Preprint] 2024:arXiv:2410.05445. [CrossRef]

[7] Aydın M, Yakut O, Alli H. Implementation of sliding mode control with artificial neural network to the rotary inverted pendulum system. Sigma J Eng Nat Sci 2013;5:1-7.

[8] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 2018;378:686-707. [CrossRef]

[9] Khoo Z, Wu D, Low JSC, Bressan S. Separable Hamiltonian neural networks. Phys Rev E 2024;110:44205. [CrossRef]

[10] Greydanus S, Dzamba M, Yosinski J. Hamiltonian neural networks. arXiv [Preprint] 2019:arXiv:1906.01563.

[11] Han CD, Glaz B, Haile M, Lai YC. Adaptable Hamiltonian neural networks. Phys Rev Res 2021;3:23156. [CrossRef]

[12] Haitsiukevich K, Ilin A. Learning trajectories of Hamiltonian systems with neural networks. arXiv [Preprint] 2022:arXiv:2204.05077. [CrossRef]

[13] Bertalan T, Dietrich F, Mezić I, Kevredikis IG. On learning Hamiltonian systems from data. Chaos 2019;29:121107. [CrossRef]

[14] Dierkes E, Flaßkamp K. Learning Hamiltonian systems considering system symmetries in neural networks. IFAC-PapersOnLine 2021;54:210-216. [CrossRef]

[15] Celledoni E, Leone A, Murari D, Owren B. Learning Hamiltonians of constrained mechanical systems. J Comput Appl Math 2023;417:114608. [CrossRef]

[16] Iten R, Metger T, Wilming H, Rio LD, Renner R. Discovering physical concepts with neural networks. Phys Rev Lett 2020;124:10508. [CrossRef]

[17] Yu J, Lu L, Meng X, Karniadakis GE. Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems. Comput Methods Appl Mech Eng 2022;393:114823. [CrossRef]

[18] Gladstone RJ, Nabian MA, Meidani H. FO-PINNs: A first-order formulation for physics informed neural networks. arXiv [Preprint] 2022:arXiv:2210.14320.

[19] Wang Y, Zhong L. NAS-PINN: Neural architecture search-guided physics-informed neural network for solving PDEs. J Comput Phys 2024;496:112603. [CrossRef]

[20] Norambuena A, Mattheakis M, González FJ, Coto R. Physics-informed neural networks for quantum control. Phys Rev Lett 2024;132:10801. [CrossRef]

[21] Canizares P, Murari D, Schönlieb CB, Sherry F, Shumaylov Z. Hamiltonian matching for symplectic neural integrators. arXiv [Preprint] 2024:arXiv:2410.18262.

[22] Choudhary A, Lindner JF, Holliday EG, Miller ST, Sinha S, Ditto WL. Physics-enhanced neural networks learn order and chaos. Phys Rev E 2020;101:62207. [CrossRef]

[23] Tong Y, Xiong S, He X, Pan G, Zhu B. Symplectic neural networks in Taylor series form for Hamiltonian systems. J Comput Phys 2021;437:110325. [CrossRef]

[24] Jin P, Zhang Z, Zhu A, Tang Y, Karniadakis GE. SympNets: Intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems. Neural Netw 2020;132:166-179. [CrossRef]

[25] Bajars J. Locally-symplectic neural networks for learning volume-preserving dynamics. J Comput Phys 2023;476:111911. [CrossRef]

[26] Jin P, Zhang Z, Kevrekidis IG, Karniadakis GE. Learning Poisson systems and trajectories of autonomous systems via Poisson neural networks. IEEE Trans Neural Netw Learn Syst 2022;34:8271-8283. [CrossRef]

[27] Lusch B, Kutz NJ, Brunton SL. Deep learning for universal linear embeddings of nonlinear dynamics. Nat Commun 2018;9:4950. [CrossRef]

[28] Hernandez Q, Badías A, González D, Chinesta F, Cueto E. Deep learning of thermodynamics-aware reduced-order models from data. Comput Methods Appl Mech Eng 2021;379:113763. [CrossRef]

[29] Zhong G, Marsden JE. Lie-Poisson Hamilton-Jacobi theory and Lie-Poisson integrators. Phys Lett A 1988;133:134-139. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00099

Research Article

Selection of optimal coating materials for bamboo reinforcement in concrete using entropy-fuzzy topsis model

Vivek PAHUJA[1,*], Pradeep Kumar GHOSH[1], Anil KUMAR[1]

[1]*Department of University Teaching, Chhattisgarh Swami Vivekanand Technical University, Chhattisgarh, 491107, India*

ARTICLE INFO

Article history
Received: 11 August 2024
Revised: 03 October 2024
Accepted: 06 March 2025

Keywords:
Bamboo; Concrete; Entropy; Fuzzy TOPSIS; Material Selection

ABSTRACT

Bamboo is a viable reinforcement material due to its high strength-to-weight ratio, cost-effectiveness, and eco-friendliness. However, raw bamboo poses challenges such as high water absorption and a weak bond with concrete, which can be mitigated by appropriate coating material. An entropy-driven fuzzy TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) is intended to analyze and determine the optimal coating material. Entropy method is used to compute criteria weights whereas fuzzy-TOPSIS is used to rank alternatives by solving uncertainties and subjective preferences. The data set used was literature-based to make a decision on thirteen coating materials, the properties of which were studied based on four parameters, namely concrete compressive strength, bamboo tensile strength, embedded length and the pullout load. The results show that the epoxy-based resin material Sikadur-752 had the highest relative clossness of 0.6348. To ensure that the model was reliable under different weight scenarios, sensitivity analysis was also carried out. The proposed model advances coating material selection for bamboo reinforcement, offering a framework for future research. This study contributes to the development of bamboo-reinforced concrete construction for improving structural performance and service life.

Cite this article as: Pahuja V, Ghosh PK, Kumar A. Selection of optimal coating materials for bamboo reinforcement in concrete using entropy-fuzzy topsıs model. Sigma J Eng Nat Sci 2026;44(2):1004–1015.

INTRODUCTION

The demand for housing continues to rise, driven by population growth and urban migration. In order to construct these houses, huge amounts of steel and concrete are needed, materials that are limited in supply and whose production is extremely polluting, releasing 1.91 tonnes of CO_2 per tonne of steel [1]. This has prompted the exploration of alternative materials that are affordable, sustainable, and possess adequate structural properties. A few natural and synthetic replacements for steel have been explored in the last few decades, and bamboo has emerged as one of the promising natural alternatives to steel for affordable housing [2–5].

However, raw bamboo presents significant challenges, including high water absorption and poor bonding with concrete. Bamboo absorbs and releases moisture during the

*Corresponding author.
*E-mail address: vivekpahuja97@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

casting and curing process, as shown in Figure 1; During the casting and curing process, bamboo absorbs water from concrete and expands; then releases moisture and contracts.. This process of swelling and shrinkage may affect the integrity of structure. Chemical treatments are provided to address this issue by creating an impermeable coating on the bamboo surface. These treatments also enhance the bond strength between bamboo and concrete by increasing adhesion [6–8]. While many treatments have been explored, the question of which is most effective remains unresolved.

The study of the bond behavior at the bamboo-concrete interface using pull-out tests has been performed on the influence of different chemical treatments, including bitumen asphalt emulsion [9], sulfur [10], varnish [11], sikadur-32 gel [7,12], negrolin [13], SJK-61 epoxy mortar [14], bondtite [15], tack coat [16], water-based epoxy coating [3], sikadur 752 [17], algicoat RC-104 [18], araldite [19], linseed oil [20], spar varnish [21], oil paint [22], Bituminous paint [8], synthetic resin [23], and sikadur-32 LP [24]. The findings highlight the extensive range of existing chemical treatments that could make the bond stronger, which shows the gap between the most efficient ones. Many studies, for instance, test different coating materials under varying conditions, making it challenging to establish a reliable comparison standard. This limitation emphasizes the need for a comprehensive, data-driven approach to material selection.

Multi-criteria decision-making (MCDM) methods have been established and extensively used in different areas to address the problem of choosing the appropriate material among numerous viable choices. For instance, Chanpuypetch et al. [25] utilized AHP and TOPSIS to select bamboo products whereas Özmen-Akyol [26] employed fuzzy TOPSIS in selecting the material used in marine settings. Likewise, YILDIZ [27] used fuzzy TOPSIS in selecting mobile RFID readers. Rahim et al. [28] utilized fuzzy TOPSIS for material selection in automotive body panels. Prasad et al. [29] proposed a model for selecting appropriate coating materials for magnesium alloys by integrating Fuzzy AHP and TOPSIS techniques. In other studies, Vijayakumar et al. [30] and Çalik [31] used Fuzzy AHP and Fuzzy TOPSIS to evaluate and select projects. Chatterjee [32] introduced VIKOR and ELECTRE for material selection, while Reddy et al. [33] proposed a model for choosing sustainable building materials using entropy-driven fuzzy TOPSIS. A hybrid model for prioritizing key parameters of a sustainable supply chain was proposed by Shirazi et al. [34]. Nevertheless, the techniques have not been widely used in the choice of coating to be used on bamboo when reinforcing concrete. The literature review indicates that Fuzzy TOPSIS model has been a common tool of MCDM that has been able to counter uncertainty and vagueness in the process of decision making [35].

This study is novel as it combines the concepts of entropy and fuzzy TOPSIS to rank and assess coating materials used in the bamboo reinforcement. These contributions are the building of a decision-making model on the selection of materials and the determination of the most suitable coating material to use in the bamboo reinforcement.

The proposed research applies the entropy approach to analyze objective criteria in terms of their overall importance and fuzzy TOPSIS to cope with the uncertainty of the material selection. The sensitivity analysis is used to confirm the strength of the results. The study allows developing a consistent framework of the strengthening of the bond and overall performance of bamboo reinforced concrete structures.

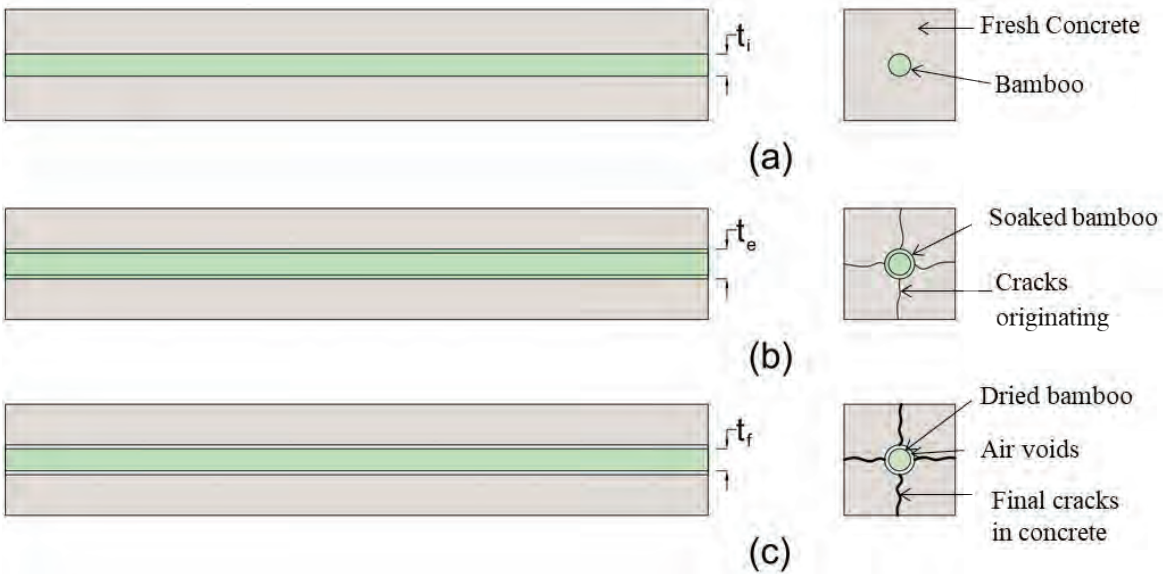


Figure 1. Behavior of untreated bamboo as reinforcement in concrete: (a) in fresh state, (b) in the curing process, and (c) in hardened concrete.

MATERIALS AND METHODS

Following an extensive literature review on the bond properties of bamboo-reinforced concrete, 13 studies were selected based on the availability of data relevant to the decision criteria of the proposed model. The optimal coating material identified in each study is chosen as a decision alternative for this research and is presented in Table 1. Four parameters that influence bond strength, including the compressive strength of concrete (fc), the tensile strength of bamboo (T), embedded length (la), and pullout load (P), were chosen as decision criteria. Each parameter is crucial in determining bond strength, as outlined below. Compressive strength of concrete (fc) makes the concrete stronger to hold the bamboo, facilitates the transfer of loads and minimizes cracking at the interface, which makes bonding stronger. The tensile strength of bamboo (T) is also significant since the higher the tensile capacity of the bamboo the higher the pullout forces that can be resisted resulting to better bond performance. Additionally, the embedded length (la) significantly affects bond strength, with longer lengths providing increased surface area for mechanical interlock, thereby improving resistance against

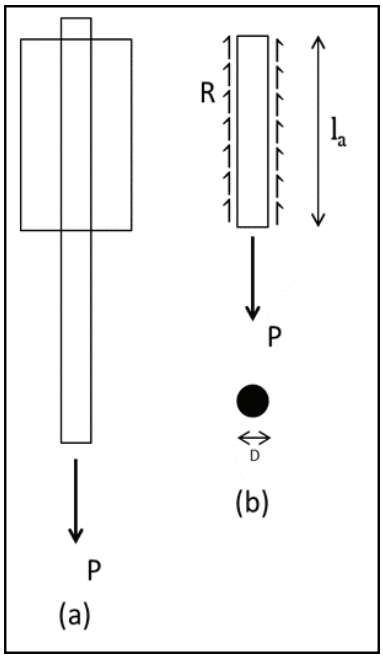


Figure 2. Schematic representation of (a) Pullout samples (b) Forces related to bond.

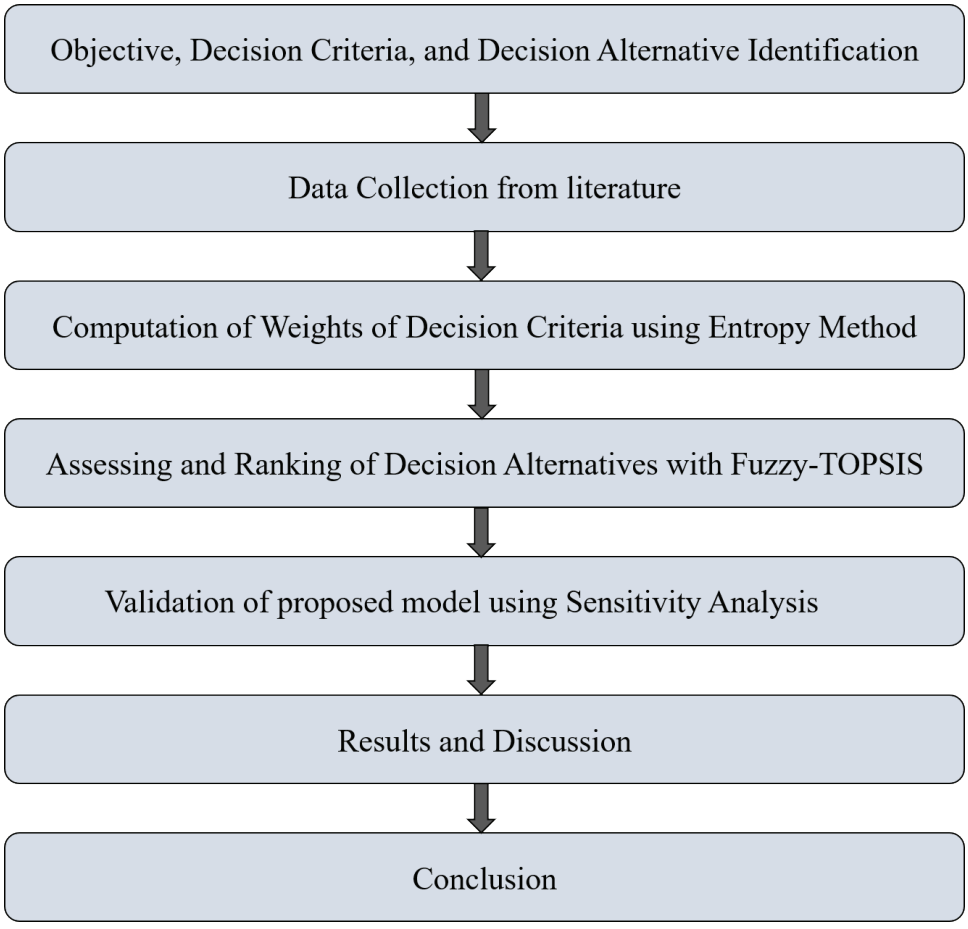


Figure 3. A flow chart for the selection of coating material using Entropy-fuzzy TOPSIS.

pullout. The pullout load (P), on the other hand, is the force applied during bond testing, higher loads require stronger bonds to prevent slippage, making it a key factor in assessing bond performance. The 28-day compressive strength of the concrete and the average ultimate tensile strength of bamboo were obtained from the studies. The embedded length varies depending on the specimen type and is expressed as a percentage to standardize the results across varying specimen sizes. Bond strength depends on the cross-sectional area of bamboo; the smaller the section, the stronger the bond is, and vice-versa. To overcome this, the pullout load is computed through equation 1, which is based on the equilibrium condition as shown in Figure 2.

$$P = \tau.p.la \quad (1)$$

Where, P = pullout force applied (kN); τ = bond strength (MPa); la = embedded length (mm) of a bamboo; and p = perimeter of bamboo (mm)

This study aims to develop a multi-criteria decision-making (MCDM) model, specifically utilizing the entropy-fuzzy TOPSIS model, to identify the optimal coating materials for bamboo reinforcement in concrete. The proposed model involves two stages: first, the calculation of decision criteria weights using Shannon's entropy method, and second, the evaluation and ranking of decision alternatives (coating material) using fuzzy TOPSIS. Figure 3 illustrates the flowchart of the proposed model.

Shannon Entropy Method

The Shannon Entropy Method is used for objective weight assignment in MCDM. It helps to determine the importance of different factors based on their entropy

Table 1. Data collected from existing literature

Alternative	Author	Concrete strength (MPa)	Tensile strength of Bamboo (MPa)	Embedded length (%)	Coating material	Chemical constituents	Bond strength (MPa)	Pullout load (kN)
A1	Ghavami [7]	20	119.02	33.33	Sikadur 32	Epoxy resin	2.75	24.75
A2	Sakaray et al. [18]	30	125	50	Algicoat RC-104	Polymer admixture	1.64	19.31
A3	Terai and Minami [23]	10	197	85	Synthetic Resin	Synthetic Resin	1.34	22.50
A4	Kute and Wakchaure [8]	22.52	133.6	100	Bituminous Paint	Bitumen	1.19	9.64
A5	Moroz et al. [21]	13.7	84.7	33.33	Spar varnish	Oil and resin	1.49	11.7
A6	Javadian et al. [3]	20	320	66.67	Moisture seal	Water-based epoxy	3.65	29.20
A7	Rahman et al. [35]	20	236.14	50	Bisphenol A	Bisphenol A	1.09	10.69
A8	Khatib and Nounu [20]	40	140	50	Linseed oil	Linseed oil	1.11	7.77
A9	Mali and Datta [15]	28	200	50	Bond tite	Epoxy adhesive	1.4	7.80
A10	Muhtar et al. [17]	31.31	126.68	66.67	Sikadur 752	Epoxy resins	2.25	27.00
A11	Rahim et al. [16]	32.95	169.27	50	Tack coat	Asphalt emulsions	2	9.42
A12	Tazowar et al. [19]	22.4	135	50	Araldite	Methacrylate adhesive	1.24	10.35
A13	Wang et al. [14]	41.1	85.6	50	SJK-61 epoxy mortar	Modified epoxy resin, curing agent, and functional aggregate in a mass ratio of 2:1:10.	4.7	9.23

values. Shannon [36] suggests that in decision analysis, the accuracy and reliability of outcomes are significantly influenced by the quantity and quality of data. Shannon's entropy measures the valuable information in a dataset and is used to determine criteria weights in MCDM methods like TOPSIS. Since the criteria used in making decisions are ambiguous and vague, the entropy method based on Shannon will be employed in Stage 1 to establish the objective weights. The advantage of Shannon entropy is that it is capable of capturing implicit information across various criteria to do a value dispersion analysis. In addition, studies of TOPSIS have established that there has been an increasing interest in the use of Shannon entropy [37]. The following steps outline Shannon's entropy methodology in detail:

Step 1: The initial decision matrix (D) is formed for m alternatives and n criteria.

$$D = \left(x_{ij}\right)_{mn} \tag{2}$$

Where x_{ij} shows the value of j^{th} criterion from the i^{th} alternative.

Step 2: Normalization of initial decision matrix (D)

$$Z_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}, \quad i = 1,2,3,\dots,m;\ j = 1,2,3,\dots,n \tag{3}$$

Step 3: The information entropy (e_j) of criterion j is given by

$$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} z_{ij} \ln z_{ij}\,, j = 1,2,3,\dots,n \tag{4}$$

Step 4: Compute the entropy weight (w_j) of criterion k

$$w_j = \frac{1-e_j}{\sum_{j=1}^{n}(1-e_j)}, j = 1,2,3,\dots,n \tag{5}$$

Fuzzy Topsis

Among the widely used multi-criteria decision-making (MCDM) techniques, TOPSIS is an effective and practical method for ranking and selecting alternatives by measuring their distances from ideal solutions. A TOPSIS model identifies the best solution by choosing the alternative nearest to the fuzzy positive ıdeal solution (FPIS) and the one furthest to the fuzzy negative ideal solution (FNIS). The FPIS represents the best alternatives by considering the highest performance values, while the FNIS identifies the worst alternatives by considering the lowest performance values. Since fuzzy numbers are commonly used to handle uncertainty and imprecision in decision-making, Fuzzy TOPSIS extends the traditional TOPSIS approach by incorporating fuzzy theory. This enhancement provides a more flexible evaluation of alternatives based on specific criteria, helping decision-makers assess trade-offs more effectively and make informed choices tailored to their priorities. The basic definitions of fuzzy numbers and fuzzy sets are given below [38–41]:

This study uses triangular fuzzy numbers (TFNs) as linguistic variables for the assessment and ranking of alternatives.

Definition 1: A membership function $\mu_{\tilde{A}}(x)$ often known as the grade of membership of x in $\tilde{A}$, characterizes a fuzzy set $\tilde{A}$ in a universe of discourse X. A triangular fuzzy membership function is shown in Figure 4.

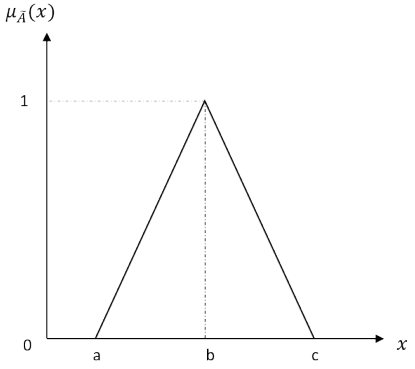


Figure 4. A fuzzy membership function in the form of a triangle.

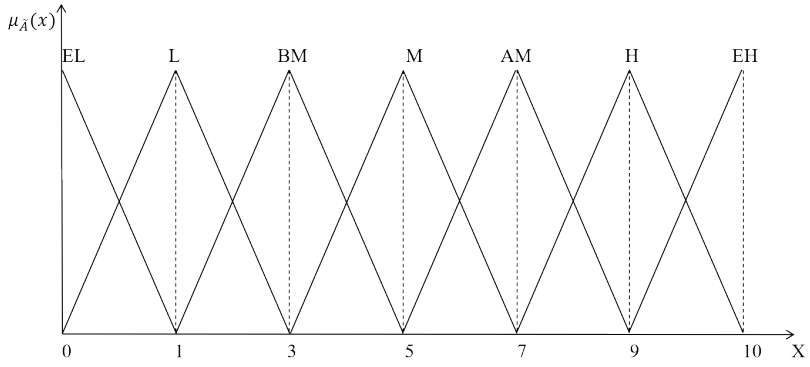


Figure 5. Membership functions of the linguistic variable for evaluating criteria.

Definition 2: A triangular fuzzy number is represented as a triplet $\tilde{A}$ = (a, b, c). The membership function $\mu_{\tilde{A}}(x)$ of triangular fuzzy number (TFN) can be calculated using equation 6.

$$\mu_{\tilde{A}}(x) = \begin{cases} 0 & x < 1, \\ \frac{x-a}{b-a} & a \le x \le b, \\ \frac{c-x}{c-b} & b \le x \le c, \\ 0 & x > c \end{cases} \quad (6)$$

Definition 3: Let $\tilde{p} = (p_1, p_2, p_3)$ and $\tilde{q} = (q_1, q_2, q_3)$ be two triangular fuzzy numbers. Then, the arithmetic operation between these two triangular fuzzy numbers is defined as follows:

$$\left\{ \begin{array}{c} \tilde{p} + \tilde{q} = (p_1, p_2, p_3) + (q_1, q_2, q_3) = (p_1 + q_1, p_2 + q_2, p_3 + q_3) \\ \tilde{p} - \tilde{q} = (p_1, p_2, p_3) - (q_1, q_2, q_3) = (p_1 - q_1, p_2 - q_2, p_3 - q_3) \\ \tilde{p} \times \tilde{q} = (p_1, p_2, p_3) \times (q_1, q_2, q_3) = (p_1 \cdot q_1, p_2 \cdot q_2, p_3 \cdot q_3) \\ \tilde{p}/\tilde{q} = (p_1, p_2, p_3)/(q_1, q_2, q_3) = (p_1/q_1, p_2/q_2, p_3/q_3) \\ k\tilde{p} = k(p_1, p_2, p_3) = (kp_1, kp_2, kp_3) \end{array} \right\} \quad (7)$$

This study employs triangular fuzzy numbers (TFNs) as linguistic variables to assess and rank alternatives. As shown in Figure 5, the criteria ratings from the fuzzy decision matrix are converted into TFNs. Following are the steps used in this study to implement the fuzzy-TOPIS model [38,42,43]:

Step 1: Construct a fuzzy decision matrix based on the ratings of m alternatives and n criteria using equation 8. Ratings for each alternative are assigned using linguistic variables represented by triangular fuzzy numbers (TFNs), as shown in Table 2.

$$\widetilde{D} = \begin{pmatrix} \tilde{k}_{11} & \tilde{k}_{12} & \ldots & \tilde{k}_{1n} \\ \tilde{k}_{21} & \tilde{k}_{22} & \ldots & \tilde{k}_{2n} \\ \ldots & \ldots & \tilde{k}_{ij} & \ldots \\ \tilde{k}_{m1} & \tilde{k}_{m2} & \ldots & \tilde{k}_{mn} \end{pmatrix} \quad (8)$$

Where, rows represent alternatives (i = 1, 2...., m) and the columns represent criteria (j = 1, 2...., n)

Step 2: Calculate the normalized fuzzy decision matrix ($\tilde{T}$) which can be represented by equation 9.

Table 2. Linguistic variables represented by triangular fuzzy numbers (TFNs)

Linguistic variable	Triangular fuzzy numbers (TFNs)
Extremely low importance (EL)	(0, 0, 1)
Low importance (L)	(0, 1, 3)
Below moderate importance (BM)	(1, 3, 5)
Moderate importance (M)	(3, 5, 7)
Above moderate importance (AM)	(5, 7, 9)
High importance (H)	(7, 9, 10)
Extremely high importance (EH)	(9, 10, 10)

$$\tilde{T} = \left[\tilde{t}_{ij}\right]_{m\times n} (i = 1,2,\ldots,m; \quad j = 1,2,\ldots,n) \quad (9)$$

The value of the beneficial criteria ($\tilde{t}_{ij}$), is calculated using the following formula:

$$\tilde{t}_{ij} = \left(\frac{p_{ij}}{r_j^+}, \frac{q_{ij}}{r_j^+}, \frac{r_{ij}}{r_j^+}\right) \quad (10)$$

$$\tilde{t}_{ij} = \left(\frac{p_j^-}{r_{ij}}, \frac{p_j^-}{q_{ij}}, \frac{p_j^-}{p_{ij}}\right) \quad (11)$$

Where $r_j^+ = \text{maximum } r_{ij}$ and $p_j^- = \text{minimum } p_{ij}$

Step 3: Determine the weighted normalized fuzzy decision matrix ($\tilde{V}$) The normalized fuzzy decision matrix and weight vector of evaluation criteria are used to compute the weighted triangular fuzzy decision matrix.

$$\tilde{V} = \left[\tilde{v}_{ij}\right]_{m\times n} (i = 1,2,\ldots,m; j = 1,2,\ldots,n) \quad (12)$$

$$\tilde{v}_{ij} = \tilde{t}_{ij} \times \tilde{w}_j \quad (13)$$

Step 4: Calculation of fuzzy positive ideal solutions (FPIS) and fuzzy negative ideal solutions (FNIS) can be accomplished using equations 14 and 15.

$$\begin{aligned} &A^G = \left\{\left(max\ \tilde{r}_{ij} \mid j \in J\right); \left(min\ \tilde{r}_{ij} \mid j \in J'\right)\right\}\ or \\ &A^G = \{\tilde{v}_1^+, \tilde{v}_2^+, \ldots\ldots, \tilde{v}_n^+\} \end{aligned} \quad (14)$$

$$\begin{aligned} &A^B = \left\{\left(min\ \tilde{r}_{ij} \mid j \in J\right); \left(max\ \tilde{r}_{ij} \mid j \in J'\right)\right\}\ or \\ &A^B = \{\tilde{v}_1^-, \tilde{v}_2^-, \ldots\ldots, \tilde{v}_n^-\} \end{aligned} \quad (15)$$

Where J = {j = 1, 2…, n | j associated with positive attribute} and J' = {j = 1, 2…, n | j associated with negative attribute}

Step 5: The separation measure of each alternative can be calculated using equations 16 and 17.

$$L_i^G = \sum_{j=1}^n d\left(\tilde{v}_{ij}; \tilde{v}_j^+\right)\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,n \quad (16)$$

$$L_i^B = \sum_{j=1}^n d\left(\tilde{v}_{ij}; \tilde{v}_j^-\right)\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,n \quad (17)$$

Where L_i^G and L_i^B are the distances of each alternative, forming the best and worst fuzzy solution, respectively. And d represents the distance between two fuzzy numbers and calculated by

$$d(\tilde{p}, \tilde{q}) = \sqrt{\frac{1}{3}\left[(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2\right]} \quad (18)$$

Step 6: Obtain the relative closeness to the ideal solution using equation 19.

$$C_i^+ = \frac{L_i^B}{L_i^G + L_i^B}\ \ i = 1,2,\ldots,m; 0 \le C_i^+ \le 1 \quad (19)$$

Table 3. Variation in criteria weights in percentage

Criteria	Change in weights (%)					
	C1	C2	C3	C4	C5	C6
Concrete strength	-6.4	+6.4	-13	+13	-19	+19
Tensile strength of bamboo	-6.4	+6.4	-13	+13	-19	+19
Embedded length	-6.4	+6.4	-13	+13	-19	+19
Pullout load	+10	-10	+20	-20	+30	-30

Step 7: Create a set of alternatives in descending order of preference based on relative closeness. The most preferred alternative is the one with the maximum (C_i^+), while the least preferred alternative is the one with the minimum (C_i^+).

Validation of the Proposed Model Using Sensitivity Analysis

A sensitivity analysis is conducted to confirm the stability of the proposed Entropy-Fuzzy TOPSIS model by systematically changing the weight of the criteria. The analysis is required to ensure that the outcomes are not too dependent on specific weight distributions and that results are consistent under different conditions. The objective is to assess how material rankings changed under various weighting scenarios, thereby verifying the model's reliability and robustness. This analysis enables an evaluation of the stability and consistency of the rankings produced by the Entropy-Fuzzy TOPSIS model under varying conditions [44,45]. In this study, the criteria weights were adjusted by ±10%, ±20%, and ±30%, with the remaining weights proportionally modified to maintain a total of 100%. Six conditions resulting from these adjustments are labeled C1, C2, C3, C4, C5, and C6, as shown in Table 3. These results are then compared to the proposed model, labeled C0.

RESULTS AND DISCUSSION

The findings of this study are categorized into three segments. In Section 2.1, the entropy weight method is employed to calculate the decision criteria weights. The performance and ranking of the coating materials are then determined using the fuzzy TOPSIS model, as outlined in Section 2.2. Validation of the entropy-fuzzy TOPSIS model is detailed in Section 2.3.

Weight Calculation Using Entropy Weight Method

The weights of the decision criteria are determined using the entropy weight method. The initial decision matrix for the coating material with all the chosen criteria is shown in Table 4.

A normalized decision matrix is created by normalizing the initial decision matrix, as shown in Table 5.

The information entropy for each decision criterion (e_j) and entropy weight for each criterion (w_j) are calculated and given in Table 6,

Table 4. The initial decision matrix for the coating material concerning decision criteria

Alternative	Concrete strength (MPa)	Tensile strength of Bamboo (MPa)	Embedded length (%)	Pullout load (kN)
A1	20	119.02	33.33	24.75
A2	30	125	50	19.31
A3	10	197	85	22.50
A4	22.52	133.6	100	9.64
A5	13.70	84.7	33.33	11.7
A6	20	320	66.67	29.20
A7	20	236.14	50	10.69
A8	40	140	50	7.77
A9	28	200	50	7.80
A10	31.31	126.68	66.67	27.00
A11	32.95	169.27	50	9.42
A12	22.4	135	50	10.35
A13	41.1	85.6	50	9.23

Table 5. Normalized decision matrix

Alternative	Concrete strength (MPa)	Tensile strength of Bamboo (MPa)	Embedded length (%)	Pullout load (kN)
A1	0.0602	0.0574	0.0453	0.1241
A2	0.0904	0.0603	0.0680	0.0969
A3	0.0301	0.0951	0.1156	0.1129
A4	0.0678	0.0645	0.1361	0.0484
A5	0.0413	0.0409	0.0453	0.0587
A6	0.0602	0.1544	0.0907	0.1465
A7	0.0602	0.1140	0.0680	0.0536
A8	0.1205	0.0676	0.0680	0.0390
A9	0.0843	0.0965	0.0680	0.0391
A10	0.0943	0.0611	0.0907	0.1354
A11	0.0993	0.0817	0.0680	0.0473
A12	0.0675	0.0652	0.0680	0.0519
A13	0.1238	0.0413	0.0680	0.0463

Table 6. Information entropy (e_j) and entropy weight (w_j) of the decision criteria.

Criteria	Concrete strength	Tensile strength of Bamboo	Embedded length	Pullout load
e_j	0.9749	0.9721	0.9812	0.9538
w_j	0.2128	0.2363	0.1594	0.3915

The entropy weights (w_j) for the selected decision criteria are shown in Figure 6. According to the outcomes, the most significant decision criterion in the evaluation of the performance of coating materials is the pullout load with the maximum weight of 0.3915. Bamboo tensile strength is second with a weight of 0.2363. Concrete strength is followed by weight of 0.2128 and embedded length has least weight of 0.1594.

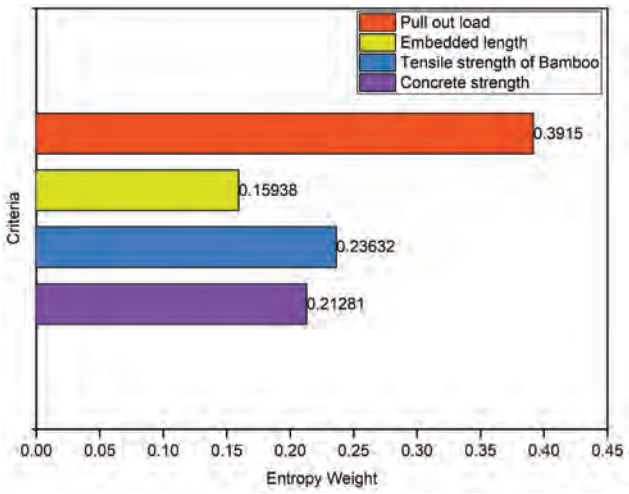


Figure 6. Entropy weight of decision criteria.

Assessment and Ranking of Coating Materials Using Fuzzy Topsis

The fuzzy decision matrix is obtained by converting the decision criteria of each coating material into TFNs to reflect the uncertainty and ambiguity in the data. The mapping of each criterion is determined by the location of the rating within the range of minimum and maximum values. The resulting fuzzy decision matrix, encompassing all decision criteria for the coating material, is shown in Table 7.

After normalizing the fuzzy decision matrix, the normalized fuzzy decision matrix is obtained. By multiplying normalized values by weights for each decision criterion in the weighted normalized fuzzy decision matrix, the weighted normalized fuzzy decision matrix is formed. Table 8 presents the weighted normalized fuzzy decision matrix for the coating material with the decision criteria.

In the next step, the distance of each alternative from the FPIS and FNIS is calculated. The closeness coefficients of each alternative are determined, and these coefficients are then used to rank them. Table 9 presents the entropy-fuzzy TOPSIS closeness coefficients along with the resulting ranking of alternatives.

The findings indicate that the coefficient values lie between 0.0533 and 0.6348 with the highest coefficient value of 0.6348 belonging to alternative A10. This alternative is an epoxy resin based material known as Sikadur-752, it is therefore considered the most suitable. Figure 7 shows

Table 7. The fuzzy decision matrix for the alternatives concerning decision criteria

Alternative	Concrete strength	Tensile strength of bamboo	Embedded length	Pullout load
A1	(1,3,5)	(0,1,3)	(0,0,1)	(7,9,10)
A2	(5,7,9)	(0,1,3)	(0,1,3)	(3,5,7)
A3	(0,0,1)	(3,5,7)	(7,9,10)	(5,7,9)
A4	(1,3,5)	(0,1,3)	(9,10,10)	(0,0,1)
A5	(0,0,1)	(0,0,1)	(0,0,1)	(0,1,3)
A6	(1,3,5)	(9,10,10)	(3,5,7)	(9,10,10)
A7	(1,3,5)	(5,7,9)	(0,1,3)	(0,0,1)
A8	(9,10,10)	(0,1,3)	(0,1,3)	(0,0,1)
A9	(5,7,9)	(3,5,7)	(0,1,3)	(0,0,1)
A10	(5,7,9)	(0,1,3)	(3,5,7)	(9,10,10)
A11	(7,9,10)	(1,3,5)	(0,1,3)	(0,0,1)
A12	(1,3,5)	(0,1,3)	(0,1,3)	(0,0,1)
A13	(9,10,10)	(0,0,1)	(0,1,3)	(0,0,1)

Table 8. The weighted normalized fuzzy decision matrix for alternatives based on decision criteria

Alternative	Concrete strength	Tensile strength of Bamboo	Embedded length	Pullout load
A1	(0.0213,0.0638,0.1062)	(0.0000,0.0236,0.0709)	(0.0000,0.0000,0.0159)	(0.2740,0.3523,0.3915)
A2	(0.1064,0.1490,0.1915)	(0.0000,0.0236,0.0709)	(0.0000,0.0159,0.0478)	(0.1174,0.1957,0.2740)
A3	(0.0000,0.0000,0.0213)	(0.0709,0.1182,0.1654)	(0.1116,0.1434,0.1594)	(0.1957,0.2740,0.3523)
A4	(0.0213,0.0638,0.1064)	(0.0000,0.0236,0.0709)	(0.0478,0.0797,0.1116)	(0.3523,0.3915,0.3915)
A5	(0.0000,0.0000,0.0213)	(0.0000,0.0000,0.0236)	(0.0000,0.0000,0.0159)	(0.0000,0.0391,0.1174)
A6	(0.0213,0.0638,0.1064)	(0.2127,0.2363,0.2363)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)
A7	(0.0213,0.0638,0.1064)	(0.1182,0.1654,0.2127)	(0.1434,0.1594,0.1594)	(0.0000,0.0000,0.0391)
A8	(0.1915,0.2128,0.2128)	(0.0000,0.0236,0.0709)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)
A9	(0.1064,0.1490,0.1915)	(0.0709,0.1182,0.1654)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)
A10	(0.1064,0.1490,0.1915)	(0.0000,0.0236,0.0709)	(0.0478,0.0797,0.1116)	(0.3523,0.3915,0.3915)
A11	(0.1490,0.1915,0.2128)	(0.0236,0.0709,0.1182)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)
A12	(0.0213,0.0638,0.1064)	(0.0000,0.0236,0.0709)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)
A13	(0.1915,0.2128,0.2128)	(0.0000,0.0000,0.0236)	(0.0000,0.0159,0.0478)	(0.0000,0.0000,0.0391)

the rankings of all the alternatives, according to their relative closeness coefficients.

Validation of the Proposed Entropy-Fuzzy Topsis Model

The proposed entropy-fuzzy TOPSIS model is validated through a sensitivity analysis. In this analysis, the weight of the most influential criterion, pullout load, which had the highest weight of 0.3915, is adjusted by ±10%, ±20%, and ±30%. The weights of the remaining criteria were proportionally adjusted to keep the total weight constant. Figure 8 presents the performance of the alternatives under different conditions in terms of relative closeness.

The analysis results are compared to the proposed model, revealing that material rankings remained consistent across all weight variation scenarios. It is interesting to note that alternative A9, Sikadur-752, always ranked first, regardless of the variations in weight distributions. These results indicate that epoxy resin-based material,

Table 9. Distance from FPIS & FNIS, relative closeness and ranking of coating material

Alternative	L_i^G	L_i^B	Closeness coefficient	Ranking
A1	0.5420	0.4215	0.4375	4
A2	0.5836	0.3540	0.3776	5
A3	0.4488	0.3992	0.4707	3
A4	0.4195	0.5055	0.5465	2
A5	0.8970	0.0505	0.0533	13
A6	0.6439	0.0832	0.1145	11
A7	0.5798	0.2116	0.2674	6
A8	0.6977	0.2194	0.2393	7
A9	0.6763	0.1649	0.1960	10
A10	0.3378	0.5872	0.6348	1
A11	0.6873	0.1990	0.2246	9
A12	0.8420	0.0832	0.0899	12
A13	0.7204	0.2194	0.2335	8

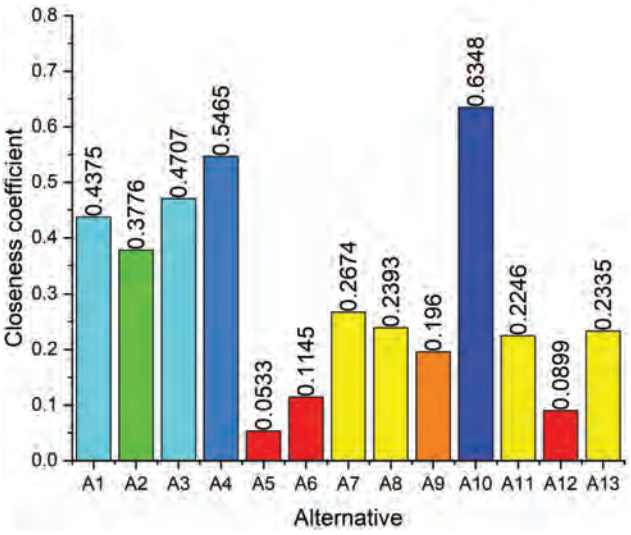


Figure 7. Ranking of coating materials based on their closeness coefficient.

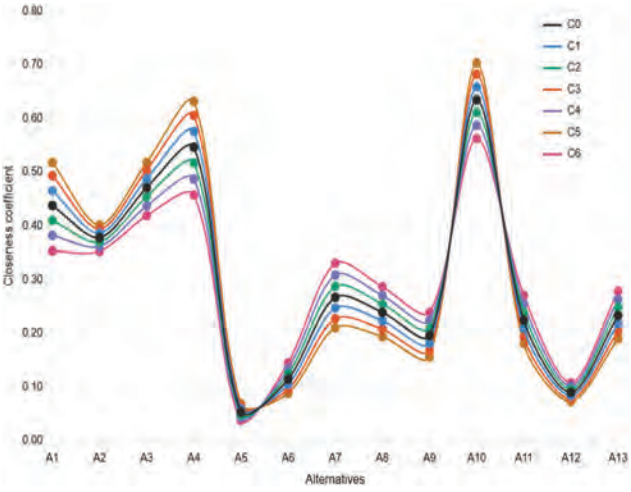


Figure 8. Performance of the alternatives under different conditions.

Sikadur-752 is the most suitable option. The sensitivity of rankings to different weights shows the strength of the Entropy-Fuzzy TOPSIS model that prioritizes materials using the criteria chosen. This consistency confirms the accuracy of the model in the decision making process since it is able to determine the best performing materials even when weighting changes are made subjectively.

CONCLUSION

The proposed study introduces a combined entropy-fuzzy TOPSIS model that is used to select an optimal coating material to be used in reinforcing bamboo in concrete. The model offers an effective, data-driven approach to the selection of a coating material by applying a comprehensive assessment of those materials in four crucial parameters, such as compressive strength of concrete, tensile strength of bamboo, embedded length, and pullout load.

The results determine an epoxy resin-based material, Sikadur-752 (alternative A10), as the best coating material, out of the 13 alternatives, with a relative closeness value of 0.6348. The strength of the selection model is tested using the sensitivity analysis, where the weight of the most influential criterion can be varied as high as 30 per cent, with corresponding changes of the other criteria. These findings confirm that this model is a trustworthy and effective approach to choosing the optimal coating materials that can be used to avoid delay and minimize costs. Epoxy resin Sikadur-752, with its superior bonding properties can be used as a bonding agent to bamboo reinforcement in concrete structures to significantly improve their performance.

The limitations of the proposed model are that lack of quantitative data on certain selection criteria and uncertainties about the economic feasibility, environmental impact, and long-term stability of coating materials. Future studies may fill in these gaps by adding more criteria, e.g., the water absorption properties of bamboo, the cost of bamboo, and the thickness and durability of the coating materials, thus allowing a more in-depth and reliable study.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] World Steel Association. Sustainability Indicators 2023 Report. Available at: https://worldsteel.org/wp-content/uploads/Sustainability-indicators-report-2023.pdf. Accessed on Apr 3, 2026.

[2] Mondal B, Maity D, Patra PK. Load and resistance factor design for bamboo reinforced concrete beam in ultimate flexural limit state. Struct Saf 2023;102:102323. [CrossRef]

[3] Javadian A, Wielopolski M, Smith IFC, Hebel DE. Bond-behavior study of newly developed bamboo-composite reinforcement in concrete. Constr Build Mater 2016;122:110–117. [CrossRef]

[4] Karthik S, Rao PRM, Awoyera PO. Strength properties of bamboo and steel reinforced concrete containing manufactured sand and mineral admixtures. J King Saud Univ Eng Sci 2017;29:400–406. [CrossRef]

[5] Puri V, Chakrabortty P, Anand S, Majumdar S. Bamboo reinforced prefabricated wall panels for low cost housing. J Build Eng 2017;9:52–59. [CrossRef]

[6] Kankam JA, Perry SH. Variability of bond strength between bamboo and concrete. ACI Mater J 1989;86:615–618. [CrossRef]

[7] Ghavami K. Bamboo as reinforcement in structural concrete elements. Cem Concr Compos 2005;27:637–649. [CrossRef]

[8] Kute SY, Wakchaure MR. Performance Evaluation for Enhancement of Some of the Engineering Properties of Bamboo as Reinforcement in Concrete. J Inst Eng Ser A 2014;94:235–242. [CrossRef]

[9] Glenn HE. Bamboo reinforcement in portland cement concrete. Texas, United States: Engineering Experiment Station; 1950.

[10] Fang HY, Fay SM. Mechanism of bamboo-water-concrete interaction. Proceedings of the International Conference on Materials of Construction for Developing Countries. Bangkok, Thailand; 1978. p. 37–48.

[11] Mulyati M, Arman A. The Evaluation of Bond Strength of Bamboo Reinforcement in Concrete. Mater Sci 2016;690–695. [CrossRef]

[12] Agarwal A, Nanda B, Maity D. Experimental investigation on chemically treated bamboo reinforced concrete beams and columns. Constr Build Mater 2014;71:610–617. [CrossRef]

[13] Ghavami K. Ultimate load behaviour of bamboo-reinforced lightweight concrete beams. Cem Concr Compos 1995;17:281–288. [CrossRef]

[14] Wang G, Wei Y, Chen S, Zhao K, Zhou Z. Bond performance between surface-modified bamboo bars and concrete under pull-out loading. J Build Eng 2023;79:107920. [CrossRef]

[15] Mali PR, Datta D. Experimental study on improving bamboo concrete bond strength. Adv Concr Constr 2019;7:191–201.

[16] Rahim NL, Ibrahim NM, Salehuddin S, Mohammed SA, Othman MZ. Investigation of bamboo as concrete reinforcement in the construction for low-cost housing industry. IOP Conf Ser Earth Environ Sci 2020;476. [CrossRef]

[17] Muhtar, Dewi SM, Wisnumurti, Munawir A. Enhancing bamboo reinforcement using a hose-clamp to increase bond-stress and slip resistance. J Build Eng 2019;26:100896. [CrossRef]

[18] Sakaray H, Togati NVVK, Reddy IVR. Investigattion on properties of bamboo as reinforcing material in concrete. Int J Eng Res Appl 2012;2:77–83.

[19] Tazowar M, Farhan A, Siddique A, Ahmed I. A novel approach for enhancing the bond performance of bamboo reinforced concrete by surface treatment and corrugation. Constr Build Mater 2023;409133728. [CrossRef]

[20] Khatib A, Nounu G. Corrugated bamboo as reinforcement in concrete. Proc Inst Civ Eng Struct Build 2017;170:311–318. [CrossRef]

[21] Moroz JG, Lissel SL, Hagel MD. Performance of bamboo reinforced concrete masonry shear walls. Constr Build Mater 2014;61:125–137. [CrossRef]

[22] Wairagade VR, Sonar IP. Bamboo concrete bond strength. Int J Eng Adv Technol 2019;9:747–752. [CrossRef]

[23] Terai M, Minami K. Research and Development on Bamboo Reinforced Concrete Structure. World Conf Earthq Eng 2012;15:1–10.

[24] Mondal B, Maity D, Patra PK. Bond Behavior between Bamboo and Normal-Strength Concrete: Experimental and Numerical Investigation. Pract Period Struct Des Constr 2022;27:1–11. [CrossRef]

[25] Chanpuypetch W, Niemsakul J, Atthirawong W, Supeekit T. An integrated AHP-TOPSIS approach for bamboo product evaluation and selection in rural communities. Decis Anal J 2024;12:100503. [CrossRef]

[26] Akyol SO, Baba A, Salman S. Selection of Material for Marine Environments Using Fuzzy TOPSIS Approach. J Inf Technol 2021;14:11–22. [CrossRef]

[27] Yildiz A, Karakoyun F, Parlak IE. The Most Suitable Mobile Rfid Reader Selection By Using Interval Type-2 Fuzzy Topsis Method. Sigma J Eng Nat Sci 2018;36:717–729.

[28] Rahim AAA, Musa SN, Ramesh S, Lim MK. Development of a fuzzy-TOPSIS multi-criteria decision-making model for material selection with the integration of safety, health and environment risk assessment. Proc Inst Mech Eng Part L J Mater Des Appl 2021;235:1532–1550. [CrossRef]

[29] Prasad RV, Rajesh R, Thirumalaikumarasamy D. Selection of coating material for magnesium alloy using Fuzzy AHP-TOPSIS. Sadhana Acad Proc Eng Sci 2020;45. [CrossRef]

[30] Vijayakumar SR, Suresh P, Sasikumar K, Pasupathi K, Yuvaraj T, Velmurugan D. Evaluation and selection of projects using hybrid MCDM technique under fuzzy environment based on financial factors. Mater Today Proc 2022;60:1347–1352. [CrossRef]

[31] Calik A. A Comparative Perspective in Sustainable Supplier Selection By Integrated Mcdm Techniques. Sigma J Eng Nat Sci 2020;38:835–852.

[32] Chatterjee AK. Sustainable construction and green buildings on the foundation of building ecology. Indian Concr J 2009;83:27–30.

[33] Reddy AS, Kumar PR, Raj PA. Entropy-based fuzzy TOPSIS framework for selection of a sustainable building material. Int J Constr Manag 2022;22:1194–1205. [CrossRef]

[34] Shirazi A. Prioritizing Key Parameters for Sustainable Supply Chain Development in the Construction Industry Using a Copeland Hybrid Method. Environ Energy Econ Res 2024;8.

[35] Rahman N, Shing LW, Simon L, Philipp M, Alireza J, Hebel Dirk E, et al. Enhanced bamboo composite with protective coating for structural concrete application. Energy Procedia 2017;143:167–172. [CrossRef]

[36] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423. [CrossRef]

[37] Sahu SK, Shekhar S, Khan A, Soni DL, Gangwar PK, Gupta M. A hybrid approach for the machinability analysis of Incoloy 825 using the entropy-MOORA method. High Temp Mater Process 2024;43. [CrossRef]

[38] Sodhi B, Tadinada P. A Simplified Description of Fuzzy TOPSIS. arXiv 2012;6–8.

[39] Tzeng GH, Huang JJ. Multiple Attribute Decision Making. 1st ed. New York: CRC Press; 2011.

[40] Mayyas A, Omar MA, Hayajneh MT. Eco-material selection using fuzzy TOPSIS method. Int J Sustain Eng 2016;9:292–304. [CrossRef]

[41] Khan A, Maity K. Application potential of combined fuzzy-TOPSIS approach in minimization of surface roughness, cutting force and tool wear during machining of CP-Ti grade II. Soft Comput 2019;23:6667–6678. [CrossRef]

[42] Rasi M, Abadi HMD. Proposing a framework for identifying the competitive advantage of small and medium-sized manufacturing organizations under uncertainty conditions. J Decis Oper Res 2022;7:515–529.

[43] Abadi HMD, Sadra A. Customer Clustering via Combined TOPSIS-Fuzzy -K-MEANS Method to Design an Efficient Customer Relationship System : A Real-life Case Study in the Copper Industry. Fuzzy Optim Model J 2024;5:13–26.

[44] Triantaphyllou E, Sánchez A. A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. Decis Sci 1997;28:151–194. [CrossRef]

[45] Kaymak U, van Nauta Lemke HR. A sensitivity analysis approach to introducing weight factors into decision functions in fuzzy multicriteria decision making. Fuzzy sets Syst. 1998;97:169–82. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1016–1028, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2022

Research Article

Genetic algorithm based smart irrigation system for agricultural fields using linear regression and IOT to ensure proper utilization and saving of natural water resources

Abhijit Dnyaneshwar JADHAV[1,*], Archana A. CHAUGULE[2], Harish U. TIWARI[2], Prashant G. AHIRE[3]

[1]School of CSIT, Symbiosis Skills and Professional University, Pune, 412101, India
[2]Pimpri Chinchwad College of Engineering and Research, Ravet, Pune, M.H., 412101, India
[3]Nutan Maharashtra Institute of Engineering and Technology, Talegaon Dabhade, M.H., 410507, India

ARTICLE INFO

Article history
Received: 29 October 2024
Revised: 23 December 2024
Accepted: 06 March 2025

Keywords:
Accuracy; Agricultural Irrigation; Internet of Things; Mean Squared Error (MSE); Water Predictions

ABSTRACT

Irrigation is crucial to agriculture. For available smart irrigations, accuracy is still a concern. Also, many of these systems are not affordable with practical usability, especially in rural areas of Maharashtra state. The proposed approach performs smart irrigation using Internet of Things (IoT) sensors, which collect the data about the water level in the agricultural field and provide the details of weather conditions, also uses Genetic Algorithm to optimize the input data, and Linear Regression model is used to predict the necessary water demands for the agricultural field. Here, the combination of Genetic Algorithm and linear regression is the useful and novel combination which helps to generate efficient results. By considering the level of soil moisture available in the crop field, the linear regression based model predicts the water needs and accordingly drives the motor pump for watering. The accuracy is found to be better as compared to existing approaches with 98.33% and the Mean Squared Error (MSE) is observed as 0.017, which is very less. The water demand predictions are given for various value ranges of soil moisture and temperature. For 278.93wfv soil moisture and 36°C temperature, the predicted water demand by model is 6.11mm against the actual water demand of 6.28mm. The finding justifies the usability of the system in real-time agricultural fields as given in Results section.

Cite this article as: Jadhav AD, Chaugule AA, Tiwari HU, Ahire PG. Genetic algorithm based smart irrigation system for agricultural fields using linear regression and IOT to ensure proper utilization and saving of natural water resources. Sigma J Eng Nat Sci 2026;44(2):1016–1028.

INTRODUCTION

The expansion of the agricultural sector is one of the best methods for guaranteeing social stability and economic development. In India a priority is given to the growth in this field in response to the significant economic, social, and geographical concerns that are highlighted by this industry [1]. The irrigation of agricultural fields is

*Corresponding author.
*E-mail address: abhijit.jadhav29@gmail.com

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

the most significant source for development. An efficient and suitable irrigation water supply considerably boosts agricultural productivity as water shortage has become a major worldwide problem [2]. There are two primary water sources for agriculture: irrigation and rainfall. But the precipitation isn't enough to supply the plant with the water it needs. On the other hand, excessive wetness hinders plants' capacity to absorb nutrients and increases their susceptibility to disease [3]. Open loop and closed-loop control are two sorts of irrigation control technologies [4, 5].

A farmer who uses an open-loop control system must manually choose when to irrigate. Based on his knowledge of the crop's perceived response, the operator chooses the quantity of irrigation, the amount of water used, and the rate of irrigation. This technique is widely used since it is simple to implement and doesn't need the use of sensors [6]. The consequence might be adverse water stress as some places receive excessive irrigation while others receive insufficient irrigation [7]. Significant water and energy losses are also experienced, which is in opposition to the worldwide goal of rationalizing and conserving water resources. Farmers who want to irrigate crops on fields that are stocked with vegetables must get up in the middle of the night due to the absence of power in rural areas. Furthermore, they were unable to go to the farms during COVID-19 to irrigate the plants, because of which many farmers have faced the heavy losses and also resulted in a food shortage across the nations [8].

Hence, automated irrigation systems powered by the Internet of Things (IoT) should replace manual irrigation. Lack of full end-to-end standards is one of the deployment-limiting concerns that currently affect the extent of IoT adoption [9]. As already said, contemporary situations have their limitations [10], we are exploiting the IoT and data science link to generate smart irrigation without the need for manual involvement. Therefore, the development of intelligent irrigation systems (closed-loop control) presents a useful alternative to traditional irrigation approaches. The intelligent system provides required amount of water to the agricultural field for plant growth and development by monitoring a variety of elements that have an impact on irrigation performance. The judgments are frequently made using the sensor data, which is then compared to the anticipated set points [11].

The disadvantages of the existing smart irrigation systems as well as the inconveniences of the manual irrigation systems are to be eliminated as a result of this research. Two components makeup the suggested smart irrigation system. First is a hardware sensor for gathering the input values of various parameters such as soil moisture, temperature etc. and second is processing of input data by using Genetic Algorithm (GA) and Linear Regression to predict the water demands. An effective adaptive method for resolving exploration and optimization problems is the IoT-based genetic algorithm [12]. In the second portion, a Linear Regression is used to estimate the probable water requirements of the agricultural field utilizing the optimum attribute values that have been gathered. As a result, this research project blends IoT with genetic algorithm optimization and machine learning. Here, the system's cost, timeliness, and accuracy are the main points of attention.

The next sections of this research paper discuss the literature review of existing methodologies, the recommended architecture of the research effort related to smart irrigation, the type of data to be used, the results and comparison with existing approaches and the conclusions of the study.

RELATED WORK

The thorough analysis of existing irrigation systems is done, which is as follows.

In an irrigation system review [13] has studied the combination of discrete forecasting and soil-, weather, and plant-based monitoring approaches in open fields. There are uncertainties in open-area agricultural-irrigation systems that need to be explored, unlike in environmentally controlled agriculture research. Also, in an smart irrigation using IoT presented by [14], it is concluded that two most important parameters for the creation of an intelligent irrigation system are the water level in the soil and the weather. It is also studied that, a variety of factors, including air temperature, soil temperature, air humidity, UV radiation, and many others, influence soil moisture.

An irrigation systems presented in [7] by Benzaouia Mohammed et al., in order to create an effective and long-lasting irrigation plan that intends to conserve water in the eastern part of Morocco. Closed-loop feedback control and a fuzzy logic-based irrigation strategy were proposed in this work. The length of irrigation for a particular plant type is determined by Mamdani's Fuzzification, Trapezoidal, and Triangular Membership functions. The risk of under-irrigation is eliminated by using the fuzzy control to keep soil moisture above the user-defined value. The findings demonstrate the system's effectiveness and dependability in both the summer and autumn. The work does not explain about automation of the system, also accuracy of the system can be enhanced with optimization approach for sensory value analysis. The system's timeliness is not stated by the researchers and comparison with existing systems is missing to prove the effectiveness of the system.

An irrigation system presented in [15] by Maona Li et al., to make irrigation scheduling easier, a fuzzy inference-based irrigation decision support system that incorporates a lucerne growth model, a soil water model, and pertinent software was created. The technique uses weather forecasting to anticipate the lucerne growth stage and variations in soil water. The fuzzy inference algorithm within the irrigation decision support system determines irrigation timing and amount of water taking the soil water and difference in lucerne height into account. The irrigation decision assistance system's effectiveness has been evaluated. The findings demonstrated NRMSE of 8.28% and 6.29%, respectively, the predictive models performed well in forecasting the development stage

of lucerne and soil water. The SW-based irrigation approach was contrasted with the irrigation decision support system's advantages and disadvantages. The disadvantage of the system is that, advancements should put an emphasis on crop yield, water conservation, and geographical variability as system doesn't provide this support.

Dr. Geetha S. et al. proposed the autonomous irrigation system in [16], which is designed to continuously monitor the moisture level of the soil. The system reacts appropriately by watering the soil precisely with proper amount of water and then cutting off the water supply once the required level of soil moisture is attained. There are two beneficial components. These are a moisture sensor and a motor or pump. Arduino boards are programmed using the Arduino IDE software. A soil moisture sensor is used to assess the moisture level of the soil. Water is supplied to plants via a motor or pump. Regardless of moisture level, the timer module distributes the water at specified intervals of time that are programmed in Arduino. The circuit's total current draw ranges from 200 to 220 mA. The main advantage of the system is that it promptly turns off the water pump when there is enough water and only activates it when it is required. This prevents water waste and ensures the most effective use of water. The method is limited in that just one feature is used to estimate a plant's water demands, and other significant parameters are not taken into account. This makes the system less accurate and increases the risk of false positives and drowning of plants.

An irrigation system proposed by V. Vinoth Kumar et al. presented in [17], is utilising the internet of things. In this system, data is collected at the user end regarding the status of the soil moisture, and if there is any deviation from the reference value, the user can send a command through the internet of things server to maintain the desired value of the field parameter for proper irrigation. The suggested IoT-based irrigation system is capable of measuring the field's temperature and moisture content and sending real-time data to the user using Wi-Fi and an IoT server. The system's drawback is that since the irrigation system is not automatically operated and technical field data are not provided to users who may not comprehend them, it cannot be used to all types of farms and gardens.

An irrigation system presented in [18] methodology, M. S. Goodchild et al. demonstrate the effective implementation of a PID (PI) controller for 32 strawberry plants cultivated in coir over a 94-day period and a potted poinsettia plant, utilizing a straightforward change to the integral function. The limited integral PI controller has been demonstrated to respond to the diurnal cycle in the potted poinsettia plant while keeping soil moisture to better than 1.5% with a 2-hour watering interval. Regardless of the input inaccuracy, the integral function is limited to only output values that are comparable to positive water supply volumes and to a maximum of half the daily water need. After "rainfall" and water supply disruption incidents, the PI controller was able to reestablish irrigation control. Additionally, none of the literature-described controller windup symptoms were triggered by these events. The improved PID controller described here has been demonstrated to give reliable and accurate irrigation management throughout a wide-range of solar radiation levels while responding promptly to changes in environmental conditions and the diurnal cycle. The strawberry growth experiment on coir substrate demonstrates how the limited integral PID controller can adapt water delivery to the demands of the plant from germinationto harvest while responding to changing solar radiation levels. The limitations of the system are its cost, unfriendliness for the farmers, lack of timeliness and also the accuracy of the system is not justified.

An irrigation system presented in [19], Gilberto Souza et. al, in order to limit the quantity of water irrigated, have offered an irrigation management based on a fuzzy inference system that employs the soil moisture level, as determined by capacitive soil moisture sensors, and the rain forecast from a PWS. Given the first findings, it is feasible to confirm the model's successful operation in this reduction, maintaining the soil moist at the plant's root zone. It is also possible to confirm that the system takes the soil moisture level into account when determining the irrigation prescription, and that when this level is high—for example, as a result of recent rains—it reduces the irrigation prescriptionto a low number. When compared to the FAO irrigation method, the results also show a 20% reduction in the amount of irrigated water, which allows us to predict that this reduction will remain the same or be similar after the crop cycle is over. Given the modest size on which the experiment was conducted, it would be important to assess the outcomes in a pilot under more realistic circumstances and with bigger dimensions. However, the initial outcomes in this direction are encouraging. The indirect automation of the soil moisture measuring process (using a capacitive sensor) enables the farmer to make an irrigation choice more quickly, increasing agricultural productivity and crop output. Still, full automation is required, also timeliness, accuracy and cost of the system is still the unsolved and non targeted issues in the irrigation system.

Using LabVIEW, Ravi Kumar et al. provided a contemporary design strategy for controlling and monitoring soil moisture in [20], which offers all the capabilities required to build any measurement or control irrigation in a noticeably shorter amount of time. A rain sensor is also part of the design, which is essential to limit unnecessary power use. According to the authors, farmers are not only able to produce crops with less water overall, but they are also able to increase crop yields and quality by regulating soil moisture more efficiently throughout crucial plant growth phases. The system's drawback is that it uses a small number of parameters and completely automates the watering system. Additionally, the quality attributes attained through the research work are not specifically mentioned in the results.

An irrigation system presented in [21] by Farid Touati et al. is a, wireless monitoring and feedback fuzzy control were

proposed as a way to create an irrigation system that promotes water conservation and improved irrigation management in extremely dry regions like Qatar. The created fuzzy controller effectively determined the timing and duration of irrigation for a particular crop by utilizing trapezoidal and triangular membership functions based on Mamdani fuzzification. With smooth oscillations, fuzzy control helped measure soil moisture above a preset value, preventing frequent system run-off and conserving water and energy. With a notable over-irrigation in winter, the system compensated for the quantity of water lost by evaporation as anticipated by the Penman-Monteith model in summer and winter. As a result, the technique makes it possible to estimate future watering demands. A wide-range ZigBee-GPRS based wireless system was also built in order to track system performance in real-time and build a database. Important environmental and climatic conditions, the timing and length of irrigation, as well as the quantity of water used, were all noted. This is much assisting with data analysis and system enhancement. The system is simple to use. An early cost-benefit study demonstrated that the system is financially viable. The system still should predict and automate the irrigation system with more accuracy and also time value of prediction and managing irrigation is also unsolved issue. Farmers in rural areas may not have sufficient technical knowledge to operate such complex irrigation design.

IoT Boards for Irrigation

As per the study done by GEORGE-CĂLIN SERIȚAN et al. [24], regarding various IoT boards, following comparison is done to ensure proper selection of IoT board for our proposed model (Table 1).

From the given Table 1, it is observed that An Arduino Nano is frequently chosen over a Raspberry Pi or ESP8266 for a basic irrigation system because it is easier to program, uses less power, and is made specifically for direct interaction with sensors and actuators. This makes it more effective for simple tasks like controlling water valves and reading soil moisture levels, especially when battery life is an issue. In contrast, a Raspberry Pi is better suited for complex applications with a full operating system and higher processing power, which may be excessive for a basic irrigation system. Also, the cost of the model is low when using the arduino board as compared to others with less maintenance.

From the discussion in survey and after studying the systems in [13-18] as well as few more studies provided by [19-22], the following problems are identified, which are commonly faced in smart irrigation systems.

- The input readings received from the sensors are passed to the irrigation controller without preprocessing and hence noisy readings are also used, which leads to erroneous watering requirements determination
- The accuracy is low and should be enhanced with more sophisticated irrigation model [23].
- The time constraints are missing, which leads to poor performance of the irrigation system. It would be incorrect to claim that a system's processing power rises in direct proportion to the quantity of resources used [24]. The amount of time and amount of water should be calculated while watering the crop fields.
- The complete automation is missing and manual intervention is still needed while setting the calculation of watering needs, which should have to be calculated automatically by the irrigation systems.
- The systems are not fault tolerant, Rollback and check pointing approaches provide intriguing ways to provide fault tolerance without a significant increase in cost and complexity. It includes practical features that reduce the amount of human programming work and increase availability [25].

One significant component influencing networked components performance is the distribution of workload among processing units (Fig. 2) [26]. To solve these problems, this research uses more optimized model elements to be added in designing the innovative irrigation system to overcome the mentioned challenges.

Table 1. Comparison of IoT boards

Sr. No.	Arduino nano	Raspberry Pi	ESP8266
1.	Microcontroller for simple control duties, sensor reading, and basic input/output.	More processing power and capabilities than Arduino are provided by this mini-computer like board with a complete operating system.	An inexpensive, Wi-Fi-capable microcontroller that is perfect for Internet of Things (IoT) applications.
2.	Large number of projects and tutorials, Easy to operate and better community support.	Ability to handle sophisticated programs, is adaptable, and comes with Ethernet and WiFi built in.	It is inexpensive, compact, and simple to integrate into Internet of Things applications.
3.	Less memory and computing power than the Raspberry Pi.	More costly and power-hungry than the ESP8266 or Arduino.	Less memory and computing power than the Raspberry Pi.
4.	· Applications: Basic automation, Small scale robotics, sensor projects	· Applications: IoT devices, Web servers, High level robotics and media centers	· Applications: Wireless sensors, Smart home devices and remote control

Objectives

Here, in this research, the researchers have identified the prioritized limitations and the objectives are defined as follows:

- To design and develop the smart irrigation system using machine learning and optimization technique
- To improve an accuracy of smart irrigation system using Genetic Algorithm
- To implement the smart irrigation system which will calculate the actual water needs to efficiently use the available water resources.

MATERIALS AND METHODS

The proposed model for smart irrigation comprises of important steps as shown in Figure 1, architecture of the system. In the first step the hardware circuitry of IoT using arduino collects the input values, which will be passed in the second step for optimization of the genetic algorithm and in the third phase the linear regression model uses these optimized input values to generate the watering requirements and driving the pump mechanism. The IoT plays very important role in connecting the various components of the smart irrigation system and driving the communication mechanism to pass information from one component to other component (Fig. 3) [27-29].

It takes care of collecting the input data as well as handling the output information generated by the module to drive the watering pump on and off mechanism in the proposed smart irrigation system. The mentioned three steps of the architecture are explained in the following sub sections.

Designing of Optimization Based Smart Irrigation Model Using IoT

The researchers encourage the use of Arduino based IoT circuit, which is user friendly and performance efficient hardware for irrigation system. The Figure 2 shows arduino board [30, 31] parts important for hardware connections to build smart irrigation system. Arduino boards may be used to transform analogue or digital input data from several sensors into an output, such as a motor, an LED, a link to the cloud, or a number of other tasks. By using the Arduino IDE to upload software to the board's CPU, we may configure the board to function in a certain manner; here we are configuring it for the settings to read

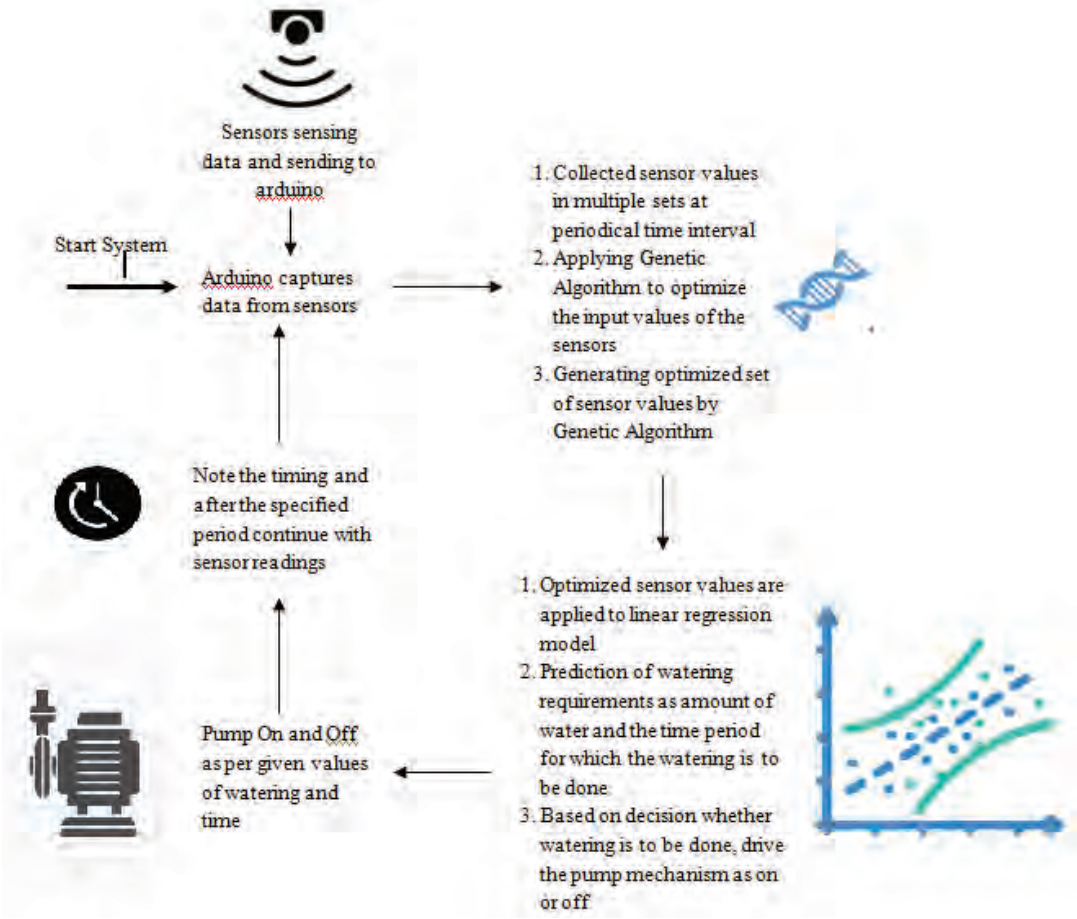


Figure 1. Proposed architecture of genetic algorithm based smart irrigation system.

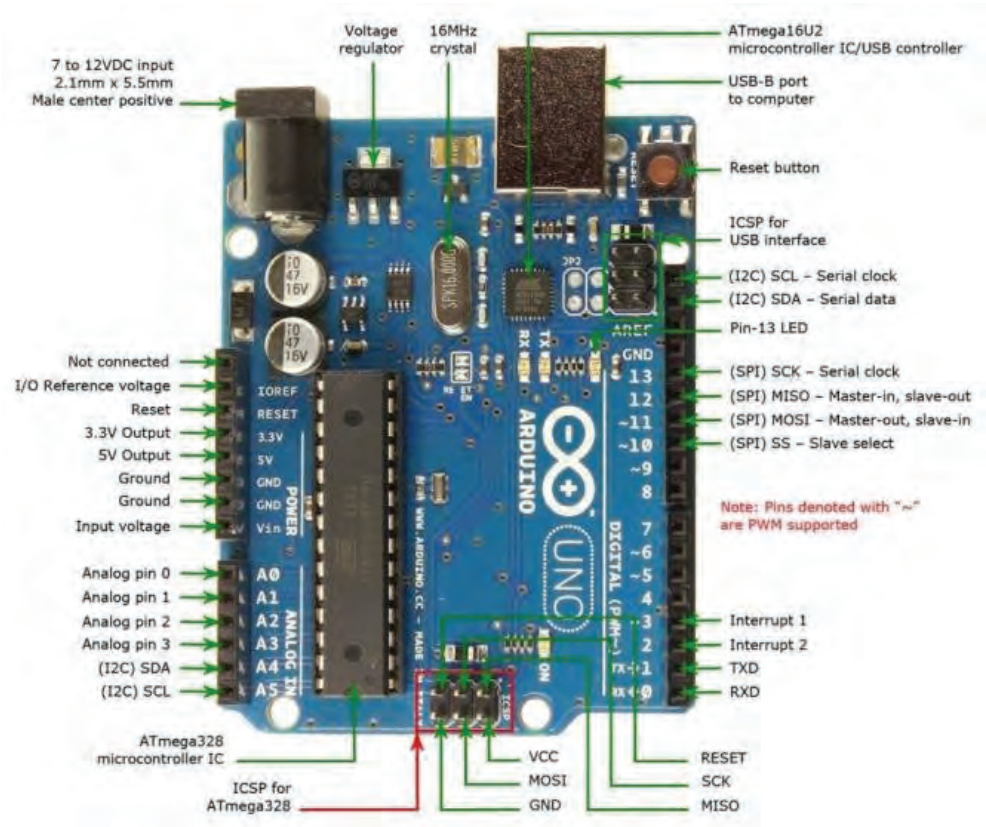


Figure 2. Arduino board image.

input from the temperature and moisture sensor for irrigation system. Unlike other previous programmable circuit boards, the Arduino does not require a separate piece of hardware called a programmer in order to load new code into the board. All you need is a USB cord. Once the hardware circuitry is ready, the sensors are reading the input features values such as soil moisture and the temperature value as well as values related to weather conditions.

Now, once the values are received from the sensors, these values are passed to the genetic algorithm, which optimizes the values as given in next section.

Genetic Algorithm Optimization

Utilizing the hardware-based system architecture, the method collects data on soil moisture, temperature, and weather conditions. The hardware consists of the required sensors for the Arduino circuit, which detects the data related to these attributes. The IoT sensors gather data from the agricultural field, which is then sent to genetic Algorithm (GA). Subsequent cycles of the GA will continue until the sensor input is properly optimized for the linear regression model processing. The GA is applied to optimize the sensor inputs, is given as follows.

Regarding application of Genetic Algorithm over other optimization approaches is to manage the time complexity while dealing with the accuracy. A Genetic Algorithm (GA) is often less computationally costly than brute force approaches when comparing the temporal complexity of optimization techniques; depending on implementation specifics, it usually falls into the "O(n log n)" or "O(n^2)" category. Its benefit is its capacity to investigate a wider range of possible solutions and maybe identify nearly ideal ones for challenging, non-differentiable issues [32].

The creation of an appropriate encoding or representation for the issue is a prerequisite to doing a GA. The ideal representation will vary depending on the demands and limitations of the problem that has to be solved. Because of its traceability and simplicity, [27,33] identified bit string format, also known as binary representation, as the most traditional method employed by GA researchers. Nevertheless, in certain circumstances, the binary representation could lead to a chromosome that is too big to manage or might not accurately depict the issue. However, GA is not just applicable to binary representation. [34,35], among others, have adopted real number representation for various applications where it was judged suitable to replace the usual binary representation. There are many additional instances of alternative encoding approaches used in a wide-range of GA applications. Here, we are representing the GA with the real-time sensor values to produce the optimize ones as that is our aim for using GA in this smart

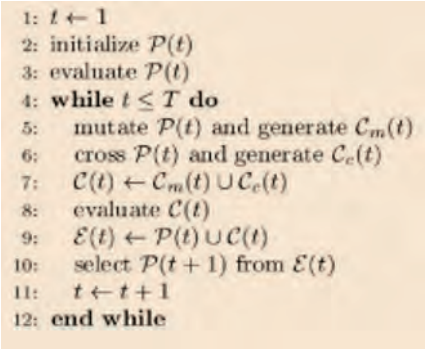


Figure 3. The reasoning behind using GA.

irrigation system. Following section discusses the application and working of GA in smart irrigation system.

Figure 3 illustrates the reasoning behind this genetic algorithm implementation. It includes the following variables: t, the generation counter; T, the maximum number of generations; P(t), the population at that generation; Cm(t) and Cc(t), the children populations obtained by the mutation and crossover operators, respectively; C(t), the children population; and E(t), the expanded population formed by the current population and their offspring.

The GA is implemented with following steps:

- **Initial Population**

The chromosome in proposed model is made up of non-repeated integer-valued genes with lengths equal to the number of sensor readings such as soil moisture and temperature values at given specified time. The input sensor readings of pair of soil moisture and temperature in a specified time interval represents gene. Since permutations are thought of as a natural representation for optimized readings, the population is made up of randomly produced permutations of set of paired values of soil moisture and temperature sensor readings.

- **Selection**

For the current models, the finest individual selection is employed. [36,37] refers to the best individual selection as "elitism," in which the elite member is not only chosen but also preserved and passes onto the following generation unaffected by crossover or mutation operators. The best individuals are chosen from a larger population in this instance's best individual selection. Both individuals from the existing population and children resulting from parental crossing and mutation makeup the larger population [38,39]. The Selection probability is Pc=0.6 to 0.8.

Hence, in our example the soil moisture and temperature sensor readings are selected based on their closeness with previous values and in relation with temperature readings the soil moisture readings are considered, as we know, the value of temperature drives the value of soil moisture. Appropriate and best values are selected for further iterations of GA.

- **Crossover**

Linear crossover is utilized for this proposed smart irrigation system iterations. In linear crossover the parents are selected for generating the offspring. After identifying parents the genes are selected at randomly among the parents. Now, as required in linear regression, we need to set the values for the α and β parameters.

Now to generate the offspring, the kth gene of the parents is modified using the following formula of equation (1):

$$O_{ik}: \alpha_i * P_{1k} + \beta_i * P_{2k} \quad (1)$$

Where,

i represent the number of offspring

P1, P2 represents the parents selected for the crossover.

After crossover the offspring are generated by the linear crossover as per formula given in equation (1).

- **Mutation**

The mutation operator utilized for all contiguous models is inversion mutation [40]. In inversion mutation, a randomly selected permutation section's order is reversed. For example, if the sequence is 1-2-3-4-5-6 and the randomly selected section is 3-4-5, then the mutated permutation is 1-2-5-4-3-6. Preserving both practicality and variety is the goal. With a 0.2 mutation rate, which was found satisfactory during preliminary experimentation, is used in the current models. The mutation rate is regarded as the likelihood of mutation of a particular genotype.

- **Termination**

For the current model, the number of generations or number of readings in a set has served as the termination criterion. The disadvantage, despite its simplicity and ease of implementation, is that evolution cycles continue needlessly until the very end if the optimal solution is discovered in the early generations. If the maximum number of generations is selected properly, this might be avoided. Termination criteria include either obtaining a certain value for the cost function that is desirable, hitting saturation, or after a specific number of iterations by choosing the generation number; this choice is based on experience and is often made big enough to guarantee convergence. It may be deemed adequate to double or triple the population size.

As given above the genetic algorithm at the end provides the optimized values of the soil moisture and temperature sensor readings. Now, these values are passed as input to the linear regression model to generate the watering needs. Now, we will discuss the third step of our implementation model.

Linear Regression and Smart Watering Mechanism of Irrigation System

By assuming a linear relationship between the independent and dependent variables, linear regression forecasts the relationship between two variables. In order to minimize the sum of squared differences between the anticipated and actual values, it searches for the ideal line. The

Table 2. Input attributes for linear regression model of smart irrigation system

Independent Variable	X1	X2	X3	X4	X5	X6	X7
Attributes/ Features	Soil Moisture (wfv)	Average temperature °C	Average wind speed (m/s)	Relative humidity (%)	Sunshine hours (h)	Pressure X5 (Pa)	solar radiation (W/m^2)

linear regression may have one dependent and one independent variable for developing a model, which is called as a simple linear regression. Other type is multi linear regression, which we are utilizing in this irrigation model, where, there are several independent variables and one dependent variable in this situation [41-45]. Following equation (2) is the multiple linear regression equation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots\ldots + \beta_n X_n \quad (2)$$

where:
Y is the dependent variable
$X_1, X_2, \ldots, X_n$ are the independent variables
β_0 is the intercept
$\beta_1, \beta_2, \ldots, \beta_n$ are the slopes

Y is the water demand in (mm), to be provided to the applicable crop field, which is a continuous value prediction by the linear regression module.

Linear regression calculates the amount of watering required by the crop field or agricultural field. When the amount of watering required is more than or equal to 1mm, then the linear regression model drives the watering pump mechanism to ON state and the watering is done till the calculated amount of water flows out of the watering pump. This is how the smart irrigation system using genetic algorithm and linear regression works.

RESULTS AND DISCUSSION

Dataset

The Del-Coco and et. al. [46] have found out that a lot of work has been done to combine ICT and IoT technologies to create smart architectures, or to focus on the physical layer. However, it takes a lot more work to reach maturity at the processing level and associated datasets. The key to utilizing recent machine learning techniques is the availability of trustworthy and carefully selected datasets. This could be the tipping point for a fully automated smart irrigation system, which should eventually incorporate application models into the training process. This will result in fully data-driven machine learning tactics from start to finish [46].

In order to ensure a representative sample by capturing variations in water application across the field, we have gathered data points from various locations across the irrigated area while taking into account factors like the type of irrigation system, the water source, and the topography of the field. We have also considered measurements such as soil moisture, temperature, and water flow rate parameters at each sampling point. For accurate analysis, it is important to conduct the survey of data values for these parameters at various locations in the irrigation field as well as conducting survey at various environmental conditions. Considering the place of irrigation, we have collected data by using random sampling method which leads to the selection of random locations in the irrigation field to collect data values of various irrigation parameters. Also, as data is collected season wise, during collection in each season the periodic sampling is executed to collect data samples at various time intervals in a day. As the samples are collected by us only, so we have taken care that, noisy samples should not be a part of the dataset, and also optimization of the input values to increase the model accuracy is done by using the Genetic Algorithm as a part of data preprocessing.

We started using IoT devices including sensors and arduino boards in different environments with several domestic plants in the mass collection process for the absolute need of information. This collection process was lasting for six months including winter, summer and rainy

Table 2. Sample data set with relevant attributes and Water demand

Soil Moisture (wfv)	Average temperature (°C)	Average wind speed (m/s)	Relative humidity (%)	Water demand (Y in mm)
278.93	36	2.32	19.62	6.28
318	34.5	2.41	16.56	6.02
.	.	.	.	.
.	.	.	.	.
777.85	15.3	3.35	18.52	4.68
865.29	14	1.78	19.33	4.46

season samples to ensure data samples with mostly all seasons for making system be useful throughout the year. Accompanying this growth are the following findings for few of the important attributes:

Attribute 1 - Soil Moisture: The data is generated by an analog sensor within a data interval of 0 to 1023. The smallest value is 278.93, and the maximum value is 865.29.

Attribute 2 – Temperature: These data were gathered with the help of a temperature sensor that displays the current temperature in Celsius. We can see that, during the collection months, the average temperature was 24.5°C, whereas the lowest value was 14°C, and the highest value was 36 °C.

Attribute 3 – Air Humidity: We collected humidity data using the same sensor that guaranteed the collection of temperature values. The results of our analysis pass show that the lowest value is 29.7%, and the greatest value is 84.1%.

Dependent Variable – Water demand: Based on the observations and communication with the expertise, the data for the water demands is collected in various environmental conditions and various possible combinations of the above mentioned input parameters. The maximum water needs identified are 6.65mm and minimum water demands is 2.94mm in different conditions. The model is trained and tested using the mentioned dataset only. The sample dataset with important attributes is as shown in Table 2.

Results for Smart Irrigation Using Linear Regression with GA and without GA

We have collected total 350 samples with depicting varying environmental conditions. The dataset is divided into 80% for training with 280 samples and 20% for testing with 70 samples. The results are explained as follows.

The Figure 4a to 4d shows the results for smart irrigation using linear regression without GA and with GA. All

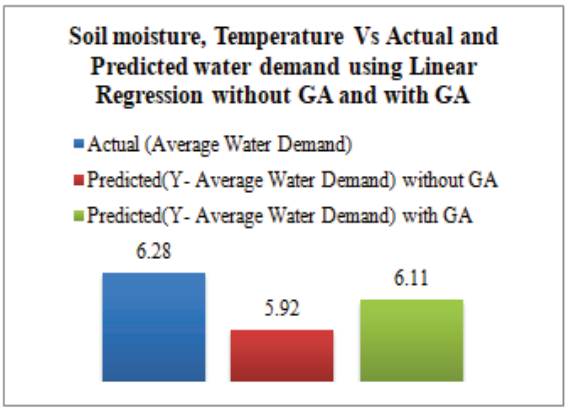


Figure 4a. Actual Vs Predicted water demand for given average soil moisture using Linear Regression without GA and with GA for Range of:
Soil moisture from 278.93 to 414.16 and Temperature from 30.5°C to 36°C

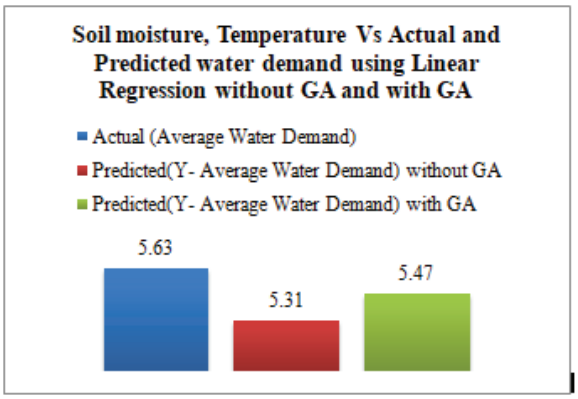


Figure 4b. Actual Vs Predicted water demand for given average soil moisture using Linear Regression without GA and with GA for Range of:
Soil moisture from 415.92 to 568.81 and Temperature from 25.6°C to 30.4°C

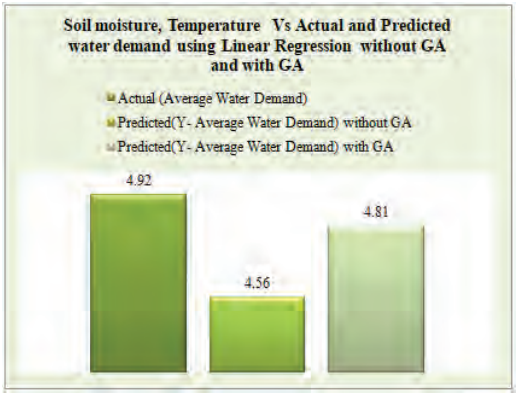


Figure 4c. Actual Vs Predicted water demand for given average soil moisture using Linear Regression without GA and with GA for Range of:
Soil moisture from 569.22 to 736.51 and Temperature from 20.5°C to 25.5°C

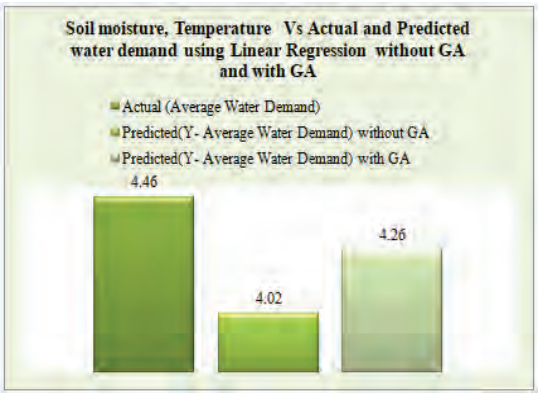


Figure 4d. Actual Vs Predicted water demand for given average soil moisture using Linear Regression without GA and with GA For Range of:
Soil moisture from 736.65 to 865.29 and Temperature from 14°C to 20.4°C

Figure 4. Graphs showing the actual water demands Vs predicted water demands without GA and with GA.

the graphs show the actual Vs predicted water demands using GA and without using GA for various soil moisture ranges and various temperature ranges.

As we can see, for low soil moisture range and high temperature range as shown in Figure 4a, the average water demand is 6.11mm as generated by our model using GA against the actual demand of 6.28mm. Also, for high soil moisture range and low temperature range as shown in Figure 4d, the average water demand is 4.26mm as generated by our model using GA against the actual demand of 4.46mm. The Figure 4b and Figure 4c shows the predicted water demands for medium soil moisture and temperature ranges as specified.

Internet of things along with optimization is found useful and effective to bring automation in agriculture [12], especially for irrigation system as stated in graphs.

The Figure 5 shows the comparison for MSE of our model with the other machine learning techniques. It is observed that MSE is very less as compared to the existing models of smart irrigation. The MSE of our model is only 0.017533.

This graph in Figure 5 shows the results for linear regression with genetic algorithm for smart irrigation are superior to other machine learning techniques used for smart irrigation without optimization technique. Here the GA helps to optimize the sensor reading values and generate nearly closer water demands to cope up with the accuracy issue of the smart irrigation system.

Figure 6 shows the accuracy comparison of our model with existing systems. We found that, our model of smart irrigation performs better as compared to existing systems in literature. The comparative study is as follows.

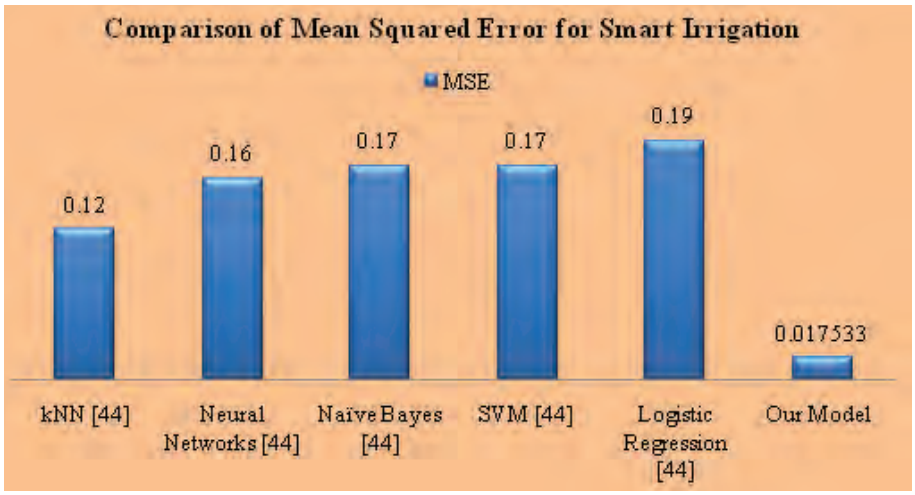


Figure 5. Comparison of our model with existing machine learning techniques for smart irrigation.

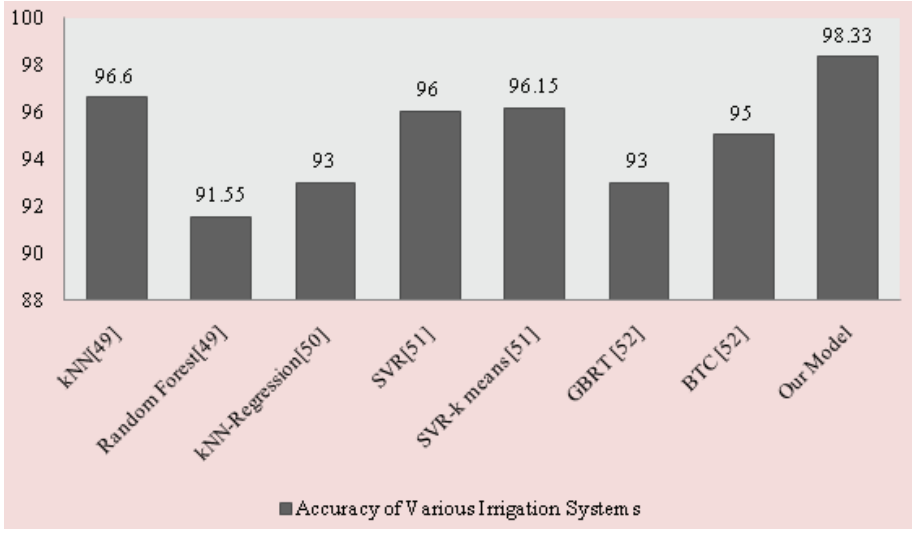


Figure 6. Accuracy Comparison of Our model with existing machine learning techniques for smart irrigation.

As per the Bavilla et al. [47], the accuracy of irrigation model using kNN is 96.60% and for Random forest it is 91.10%, with limited size dataset for training and same dataset is used for testing also. On the other hand, Akshay & Ramesh [48] have used kNN-Regression and observed the accuracy of 93%. In irrigation model by [49], the Goap, Amrendra et al. have used SVM for which the accuracy is 96%, also they have tested the system using combination of SVM-k-means, of which accuracy is found to be slightly increasing to 96.15%, which is better than SVM, kNN and Random Forest as studied above. In another research article, Goldstein, Anat et al. [50] have implemented the irrigation using Gradient Boosted Regression Trees (GBRT), for which the accuracy is observed as 93% and they have compared it with classifier model BRT i.e. Boosted Tree Classifier, which was found the best classifier with the accuracy of 95%.

After comparison, it is observed that accuracy of Smart Irrigation with linear regression model using genetic optimization is more than any of the standalone machine learning based irrigation systems. It is 98.33% and better than any of the other mentioned irrigation systems. This accuracy will help the real-time utilization of our irrigation system in rural areas of Maharashtra for watering to agricultural fields.

CONCLUSION

Herewith the researchers have implemented the smart irrigation system to calculate the watering needs for agricultural fields for various environmental conditions by using genetic algorithm optimization and linear regression. The soil moisture and other parameter values and readings generated by the IoT circuit which is developed using various sensors and the arduino board. The genetic algorithm optimizes the values of the input for linear regression and then linear regression is applied to generate the nearly accurate water demands. We have analyzed the results of smart irrigation using linear regression with and without genetic algorithm and found that the better accuracy is achieved with genetic algorithm optimization. The average MSE depicted in the results is 0.017 and accuracy is found to be 98.33 percent, which is better as compared to any of the existing smart irrigation systems without optimization algorithm. The future scope of the work considers the iterations for optimization in genetic algorithm with number of samples generated, which may lead to unnecessary processing of the input samples. With optimum number of iterations using more feasible approach, it can be avoided.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Ma'Mun SR, Loch A, Young MD. Robust irrigation system institutions: A global comparison. Glob Environ Change 2020;64:102128. [CrossRef]

[2] Gebbers R. Precision agriculture perspectives. European Conference on Precision Agriculture; Catalonia, Spain; 2013.

[3] Touati F, Al-Hitmi M, Benhmed K, Tabish R. A fuzzy logic based irrigation system enhanced with wireless data logging applied to the state of Qatar. Comput Electron Agric 2013;98:233–241. [CrossRef]

[4] Albiero D, Paulo RL, Santos JSG, Felix Junior JC, Melo R. Agriculture 4.0: A terminological introduction. 9th International proceedings on proceedings; 2010; pp. 1–2.

[5] Romero R, Muriel JL, García I, Peña DM. Research on automatic irrigation control: State of the art and recent results. Agric Water Manag 2012;114:59–66. [CrossRef]

[6] Heiman A, Miranowski J, Zilberman D, Alix J. The Increasing Role of Agribusiness in Agricultural Economics. J Agribusiness 2002;20:1S30.

[7] Awan S, Ahmed S, Ullah F, Nawaz A, Khan A, Uddin MI, et al. IoT with BlockChain: A Futuristic Approach in Agriculture and Food Supply Chain. Wireless Commun Mob Comput 2021;5580179. [CrossRef]

[8] Benzaouia M, Bouselham L, Hajji B, Dubois AM, Benslimane A, El Ouariachi M. Design and Performance Analysis of a Photovoltaic Water Pumping System based on DC-DC Boost Converter and BLDC Motor. 7th International Renewable and Sustainable Energy Conference; Agadir, Morocco; 2019; pp. 1–6. [CrossRef]

[9] Rathore NK, Pande S. IoE-Based Genetic Algorithms and Their Requisition. Computational Intelligent Security in Wireless Communications. Boca Raton: CRC Press; 2023. p. 25–50. [CrossRef]

[10] Jaiswal S, Ballal MS. Fuzzy inference based irrigation controller for agricultural demand side management. Comput Electron Agric 2020;175:105537. [CrossRef]
[11] Azar AT, Ammar HH, Silva GB, Razali MSAB. Optimal proportional integral derivative (PID) controller design for smart irrigation mobile robot with soil moisture sensor. International Conference on Advanced Machine Learning Technologies and Applications; Cairo, Egypt; 2019; pp. 349–359. [CrossRef]
[12] Khan F, Rathore N. Internet of things: A review article. J Cloud Comput 2018;5:20–25.
[13] Gamal Y, Soltan A, Said LA, Madian HAF. Radwan AG. Smart Irrigation Systems: Overview. IEEE Access 2023;13:66109–66121
[14] Vasanam SA, Jesmi J, Rasika AM, Lisha R. Smart Irrigation system using Internet of Things. 2024.
[15] Krishnan RS, Julie EG, Robinson YH, Raja S, Kumar R, Thong PH. Fuzzy Logic based Smart Irrigation System using Internet of Things. J Clean Prod 2020;252:119902. [CrossRef]
[16] D'Odorico P, Chiarelli DD, Rosa L, Bini A, Zilberman D, Rulli MC. The global value of water in agriculture. PNAS 2020;117:21985–21993. [CrossRef]
[17] Zerouali M, Zouirech S, El Ougli A, Tidhaf B, Zrouri H. Improvement of Conventional MPPT Techniques P&O and INC by Integration of Fuzzy Logic. 7th International Renewable and Sustainable Energy Conference; Agadir, Morocco; 2019; pp. 1–6. [CrossRef]
[18] Jamroen C, Komkum P, Fongkerd C, Krongpha W. An Intelligent Irrigation Scheduling System Using Low-Cost Wireless Sensor Network Toward Sustainable and Precision Agriculture. IEEE Access 2020;8:172756-172769. [CrossRef]
[19] Benzaouia M, Hajji B, Rabhi A, Mellit A, Benslimane A, Dubois AM. Energy Management Strategy for an Optimum Control of a Standalone Photovoltaic-Batteries Water Pumping System for Agriculture Applications. International Conference on Electronic Engineering and Renewable Energy; Singapore; 2020; pp. 855–868. [CrossRef]
[20] Capraro F, Patino D, Tosetti S, Schugurensky C. Neural network-based irrigation control for precision agriculture. 2008 IEEE International Conference on Networking, Sensing and Control; Sanya, China; 2008; pp. 357–362. [CrossRef]
[21] Li M, Sui R, Meng Y, Yan H. A real-time fuzzy decision support system for alfalfa irrigation. Comput Electron Agric 2019;163:104870. [CrossRef]
[22] Kumar V, Ramasamy R, Janarthanan S, Babu MV. Implementation of iot in smart irrigation system using arduino processor. Int J Civ Eng Technol 2017;8:1304–1314.
[23] Seritan GC, Enache BA, Vilciu I, Grigorescu SD, Mladenov V. Comparison Study of Top Development Boards in the Context of IOT. Rev Roum Sci Techn–Électrotechn et Énerg 2022;67:483–486.
[24] Rathore N. A review towards: load balancing techniques. I-manager's J Power Syst Eng 2016;4. [CrossRef]
[25] Rathore N. Checkpointing: fault tolerance mechanism. I-manager's J Cloud Comput 2017;4:28. [CrossRef]
[26] Rathore N, Singh PK. A comparative analysis of fuzzy based load balancing algorithms. I-manager's J Comput Sci 2017;5:23. [CrossRef]
[27] Delgodaa D, Saleemb SK, Malanoa H, Halgamuge MN. Root zone soil moisture prediction models based on system identification: formulation of the theory and validation using field and AQUACROP data. Agric Water Manag.
[28] Goodchild M, Kühn KD, Jenkins M, Burek AJ. A method for precision closed-loop irrigation using a modified pid control algorithm. Sens & Trans 2015;188:61.
[29] Rathore NK. Efficient agent based priority scheduling and loadbalancing using fuzzy logic in grid computing. System 2015;6:13–23.
[30] Geetha S, Phamila YA, Vaishnavi P, Charitha YLS, Jasayri T, Bhumika M, et al. Automatic watering system using soil moisture sensor and rtc timer with Arduino. Int J Adv Res Sci Commun-Technol 2022;2:2581–9429.
[31] Rathore N, Chana I. Report on hierarchal load balancing technique in grid environment. I-manager's J Inf Technol 2013;2:21. [CrossRef]
[32] Nopiah ZM, Khairir MI, Abdullah S, Baharin MN. Time complexity estimation and optimisation of the genetic algorithm clustering method. WSEAS Trans Math 2010;9:334–344.
[33] Rathore N. Ethical hacking and security against cyber crime. I-manager's J Inf Technol 2015;5:7. [CrossRef]
[34] Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. Multimed Tools Appl 2021;80:8091–8126. [CrossRef]
[35] Rathore N. Dynamic threshold based load balancing algorithms. Wirel Pers Commun 2016;91:151–85. [CrossRef]
[36] Afrouzy ZA, Nasseri SH, Mahdavi I. A genetic algorithm for supply chain configuration with new product development. Comput Ind Eng 2016;101:440–454. [CrossRef]
[37] Rathore N. Performance of hybrid load balancing algorithm in distributed web server system. Wirel Pers Commun 2018;101:1233–1246. [CrossRef]
[38] Alaoui A, Adamou-Mitiche ABH, Mitiche l. effective hybrid genetic algorithm for removing salt and pepper noise. Iet Image Process 2020;14:289–296. [CrossRef]
[39] Alvesa MJ, Almeidab M. Motga: a multiobjective tchebycheff based genetic algorithm for the multidimensional knapsack problem. Comput Oper Res 2017;34:3458–3470. [CrossRef]

[40] Azadeh A, Elahi S, Farahani MH, Nasirian B. A genetic algorithm-taguchi based approach to inventory routing problem of a single perishable product with transshipment. Comput Ind Eng 2017;104:124–133. [CrossRef]

[41] Chua M, Huber KE. Handbook for designing and conducting clinical and translational research. Cambridge: Academic Press; 2023. p. 145–152. [CrossRef]

[42] Akgun B, Oguducu SG. Streaming linear regression on spark mllib and moa. Proceedings of the 2015 ieee/acm international conference on advances in social networks analysis and mining 2015; 2015 August 25-28; Paris, France; 2015. pp. 1244–1247. [CrossRef]

[43] Lim HI. A linear regression approach to modeling software characteristics for classifying similar software. IEEE Xplore 2019;942–943. [CrossRef]

[44] Wu J, Liu C, Cui W, Zhang Y, Personalized collaborative filtering recommendation algorithm based on linear regression. IEEE Xplore 2019;139–142. [CrossRef]

[45] Tace Y, Tabaa M, Elfilali S, Leghris C, Bensag H, Renault E. Smart irrigation system based on iot and machine learning. Energy Rep 2022;8:1025–1036. [CrossRef]

[46] Del-Coco M, Leo M, Carcagnì P. Machine learning for smart irrigation in agriculture: how far along are we?. Information 2024;15:306. [CrossRef]

[47] Reddy BR, Rahiman MK. An smart irrigation system based on iot to enhance the accuracy using novel k-nearest neighbor compared with random forest algorithm to increase accuracy rates on agricultural fields. AIP Conf Proc 2024;3193:020023. [CrossRef]

[48] Akshay S, Ramesh TK. Efficient machine learning algorithm for smart irrigation. IEEE Xplore 2020;867–870. [CrossRef]

[49] Goap A, Sharma D, Shukla AK, Krishna CR. An iot based smart irrigation management system using machine learning and open source technologies. Comput Electron Agric 2018;155: 41–49. [CrossRef]

[50] Goldstein A, Fink L, Meitin A, Bohadana S, Lutenberg O, Ravid G. Applying machine learning on sensor data for irrigation recommendations: revealing the agronomist's tacit knowledge. Precis Agric 2018;19(3):421–444. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2006

Research Article

The effects of the feature reduction by machine learning algorithms and feature decomposition by fuzzification on the mathematical model success criterion

Filiz KANBAY[1,*], Melike GÖKSU[1]

[1]*Department of Mathematics, Yıldız Technical University, Istanbul, 34220, Türkiye*

ARTICLE INFO

Article history
Received: 01 July 2025
Revised: 24 October 2025
Accepted: 27 January 2026

Keywords:
Feature Construction; Feature Reduction; Fuzzy Logic; Machine Learning; Membership Function

ABSTRACT

In this study, the effects of reducing the number of features or classes and creating new features on the model success criteria were examined. For this purpose, the parameters that will maximize the model success of the models built with k-nearest neighbor algorithm, naive bayes algorithm, support vector machine, random forest algorithm, boosting, decision trees, artificial neural networks algorithms for the selected dataset were investigated and the model with the best performance was determined. In order to determine the best model, the Wilcoxon test was applied to all algorithm results evaluated on the same layers. To provide a more balanced evaluation, weighted classification report was generated and additionaly model accuracy values were obtained by applying SMOTE. All the results obtained were evaluated together and the best performed model was determined. Visualization of the data was plotted based on the two features with the highest impact on the model. Firstly the effects of supervised and unsupervised dimensionality reduction technique, such as principal component analysis, linear discriminant analysis, neighbourhood components analysis, non-negative matrix factorization, on the model performance were examined; then a method was proposed that calculates fuzzy set number and membership degree of each feature, and adds them to the dataset, and effects of using it together with dimensionality reduction methods on model success were examined. As a result of the analysis, it was observed that applying the proposed technique to the dataset before reducing it to two dimensions using the principal component analysis enhanced performance of all algorithms.

Cite this article as: Kanbay F, Göksu M. The effects of the feature reduction by machine learning algorithms and feature decomposition by fuzzification on the mathematical model success criterion. Sigma J Eng Nat Sci 2026;44(2):1029–1040.

INTRODUCTION

The importance of designing and analyzing artificial learning experiments has become even more evident as data mining methods have become increasingly widespread in many different fields such as robotics, medical diagnosis, biometrics and finance, thanks to the developments in the

***Corresponding author.**
*E-mail address: fkanbay@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

field of artificial learning in recent years [1]. In data mining, to make classification using a small number of features (a subset of features or combining features) allows us to better understand the process that produces this data, saves the cost required for measuring and obtaining an unnecessary input, and makes training the model easier by reducing the complexity of the model [1]. For this purpose, feature extraction and feature decomposition algorithms can be used. However, feature construction method which builds new features from the existing features in the dataset can be useful to maximize the model performance [2]. Vouk et al. [3] in their work give a summary of feature construction approaches and divide them into tree categories such as operator based feature construction, methods embedded into automat M.L. and feature interactions. Hummel in his work mentioned the different methods whose a common component is the use of additional values for involving uncertainty concepts in evidential reasoning. Moreover, fuzzy set theory presented by Zadeh [4] has been applied in practice in various disciplines such as data processing, decision support and others [5-17]. For example, the fuzzy set theory applied to multi attribute analysis by Bellman and Zadeh [18]. Bezdek developed fuzzy cluster method [19]. Fuzzy Multiple Attribute Decision Making Methods can be examined in [20]. Fuzzy soft set theoretic approach to decision making problems were studied [15,21]. Zadeh presented a mathematical basis the concept of linguistic variable whose values are words or sentences in natural or artificial language [22]. In [23], methods for measuring amount of information provided by a linguistic summary were discussed; the use of linguistic summaries as a tool for representing knowledge in data was suggested. After Yager many researcher studied on linguistic summary. Herrera and Martinez presented a model composed by a pair of values called 2 tuple (a linguistic term and a term) [14, 24]. Ranking fuzzy numbers is often a necessary step in many mathematical models [16]. There are several studies on the fuzzy versions of machine learning algorithms and classification methods: [9-10, 12, 25-34]. Der-Chiang Li and et al. mention that having too many features makes data analysis inefficient, while too few reduces model stability, and define a fuzzy logic-based transformation on the features using class labels [35]. Fuzzy sets are used in feature selection with the ability to handle uncertain and in complete data [36]. Unsupervised feature reduction algorithm is proposed by developing the fuzzy improved approximation set method for mixed data [37]. By defining a fuzzy learning ensemble system, unnecessary information is reduced and computational efficiency is increased [38]. Fuzzy rough approximation is used to extract the most important features from the low resolution image [39]. A new distance definition is made to measure the difference between nominal features and two feature algorithms based on this distance are proposed [40]. Dimensionality reduction techniques are used in many different areas [41-43]. The rest of the paper is structured into three parts. Firstly, the data description and the accuracy scores of some machine learning algorithms are presented. Secondly the number of features are reduced by using the feature reduction and feature decomposition methods. So the classification can be illustrated in the plane. Moreover the effects of reducing of the number of the classes are examined. In last part, in order to enhance accuracy of the model, as a novel approach, fuzzification method is utilized.

MATERIALS AND METHODS

Model Accuracy According to Machine Learning Algorithms

In this study, Python tools, including scikit-learn, Numpy, Pandas, Matplotlib, Seaborn, XGBoost were used within Jupyter Notebook to analyze a dataset consisting of 5 features with 403 samples and no missing data [44,45]. Dataset named 'User Knowledge Modeling' obtained from Ph.D. Thesis uploaded from the UCI Machine Learning Repository used for the empirical analysis of machine learning algorithms (https://archive.ics.uci.edu/) [46]. Kahraman et al. designed a model. The reason for designing this model is to disseminate adaptive smart instructional technology applications, as well as to enable knowledge space models to be developed without the support of software developers and to prevent students from getting lost in the hypermedia environment. This designed system has been tested. When the experimental analysis tool, which can make fourteen different pedagogical evaluations at the same time, was applied. The accuracy of following the pages related to an instructional goal of the students using the intelligent learning system was much higher than the users of the traditional teaching application [47]. It is the real dataset about the students' knowledge status about the subject of Electrical DC Machines. The inputs of the dataset (five features): Students degree of study time related to learning objects (STG), number of repetitions grade (SCG), user information performance (PEG) and degree of study time related to the student's prerequisite subjects (STR), the learning percentage (LPR) are classified as the knowledge levels of students (UNS) defined four different classes: high, medium, low and very low [48]. Mathematical models were built and the results were visualized using following machine learning algorithms with Python tools [44,45,49]. The model hyperparameters were determined using a grid search on dataset split into %80 train %20 test. Using these optimized hyperparameters the model's performance was then evaluated on the entire dataset via 5-fold cross validation. By using k nearest neighbour algorithm (KNN) that classifies a query point with the majority vote of its nearest neighbors in training dataset, the results obtained with different metrics (Minkowski distance, Euclidean distance and Manhattan distance) are visualized in Figure 1. So Figure 1 shows that the highest accuracy score of KNN is 88%.

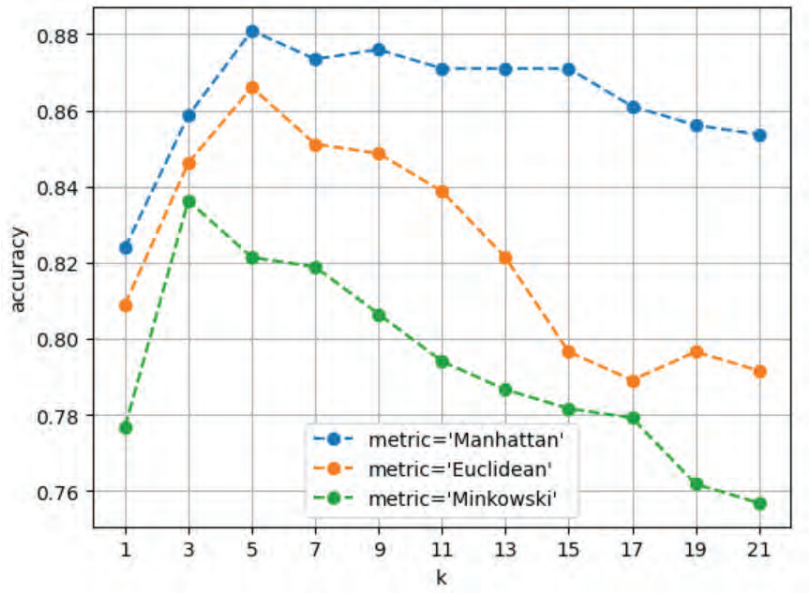


Figure 1. The accuracy score of KNN Model according to the number of neighborhoods.

In addition, the accuracy score of the Naive Bayes algorithm, which enables the classification of unlabeled data by using the conditional probability function for labeled data is 87.59%; the accuracy score of the Decision Trees algorithm is 93% with the best parameters {'criterion': 'gini', 'max_depth': 4, 'max_features': None, 'splitter': 'best'}. The accuracy score of the Random Forest algorithm is 93.3% with the best parameters { 'criterion': 'entropy', 'max_ depth': 4, 'max _features': None, 'n_estimators': 200 }. The accuracy score of Adaboost with the best parameters {'learning_ rate': 3, 'n_estimators': 400 } is 82.71%, the accuracy score of XGboost with the best parameters {'learning _ rate': 1, 'max_depth': 2, 'n_estimators': 300 } is 93.82%. The accuracy score of Support Vector Machines, which classifies the data points marked on the plane by leaving them at the maximum distance for both classes with a line to be drawn, with the best parameters { 'C': 100, kernel='linear' } is 96.02 %. The accuracy score of the Artificial Neural Networks (ANN) classifier is 93.7% with the best parameters {'activation': 'tanh', 'hidden_layer_sizes': (4,), 'learning_rate': 'constant', 'solver': 'lbfgs'}. The accuracy score of the Logistic Regresyon is 95.78% with the best parameters {'C': 100, 'class_weight': None, 'max_iter': 150, 'penalty': 'l1', 'solver': 'saga'}. The results obtained can be summarized in Figure 2.

Model success was measured using various algorithms on the dataset, with 'accuracy' selected as the model performance criterion. Among all models, the highest performance was achieved with Support Vector Machines (SVM),

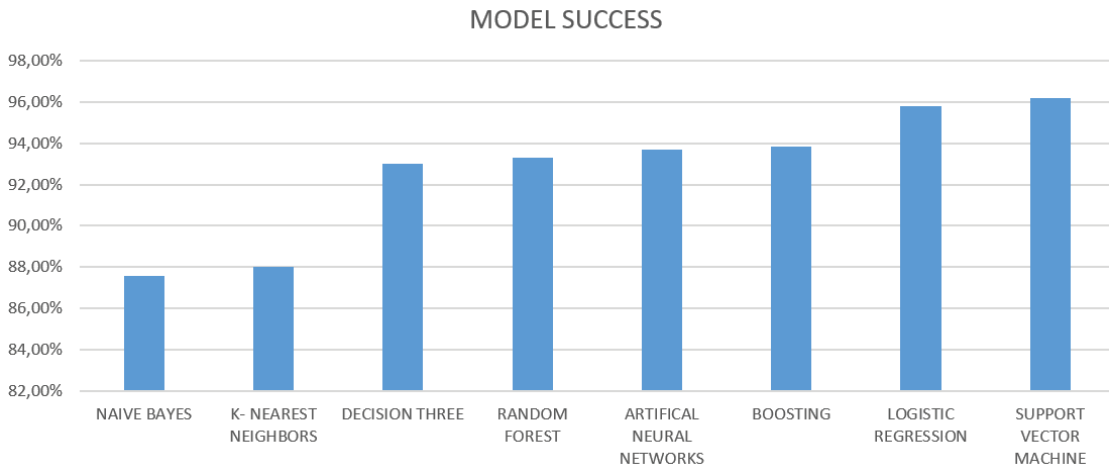


Figure 2. Model accuracy score according to classifiers.

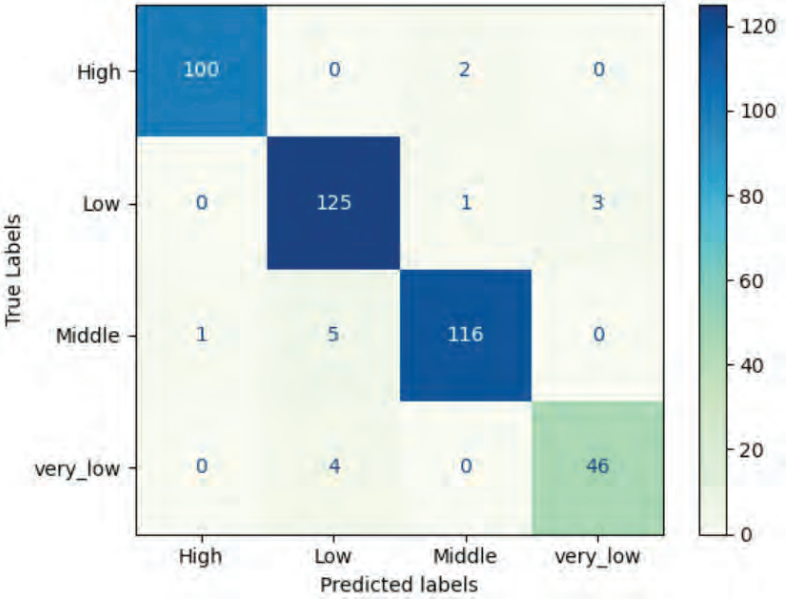


Figure 3. Support vector machine model confusion matrix.

as shown in Figure 2. Correct detection was made except for all but 11 out of a total of 403. The confusion matrix is shown in Figure 3.

Additionally, after stratified 30-fold cross validation (to increase the statistical power of the Wilcoxon test by obtaining by observations) the Wilcoxon test results comparing the sequential performance of the algorithms trained at each layer are presented as a Wilcoxon heat map in Figure 4. To determine the statistical significance of differences between two independent paired samples, p-values less than 0.05 and 0.01 were marked with * and ** respectively in Figure 4. These markings indicate the extent of difference between the performances of the two algorithms [50,51]. Since Naive Bayes and KNN perform worse than others, and while logistic regression and SVM perform better than others, they have more markings than others.

In order to provide a more balanced evaluation, the weighted classification metrics-computed using cross validation results from the previous stage and considering the class distribution- are given in Table 1. The last column of the table represents the accuracy obtained after applying the SMOTE technique [52].

In this regard, the algorithm with the highest model success was Support Vector Machines.

THE EFFECTS OF THE FEATURE SELECTION AND FEATURE EXTRACTION ON MODEL ACCURACY REDUCING THE NUMBER OF FEATURES

The random forest algorithm was applied identify the features that have little effect on the model. The importance ranking of the features were obtained using the random forest algorithm as the following form in Figure 5. Working with a dataset that includes variables of different importance allows for elaboration during the analyses.

In feature selection process, the number of features was reduced based on importance ranking obtained from the random forest algorithm, Using the most important three features the Support vector machines (SVM) model with the optimized parameters {'C':17, kernel='linear'} achieved 95.2% accuracy, when two most important features were used the SVM model with the optimized parameters {'C':5, kernel='rbf'} achieved 95%. So the model is visualized on the plane in Figure 6.

In the feature extraction process, when the number of features was reduced to three by using LDA, the accuracy score of the SVM model with the parameters {'C':17, kernel='poly'} was 96.0%, When the same operations were repeated using SVM with three and two features after applying NCA, the model accuracy scores with {'C':3, kernel='rbf'} were found to be 97.7%. After applying Principal Component Analysis (PCA), SVM model with three features and the parameters {'C':1000, kernel='linear'} achieved 78.1% accuracy; SVM model with two features and the parameters {'C':1000, kernel='rbf'} achieved 74.9% accuracy. After applying NMF, when the processes were repeated using SVM with three features and the parameters {'C':4, kernel='poly'}, the accuracy was 72.7% , and while the model for two features with the parameters {'C':3, kernel='rbf'} achieved 71.4% accuracy.

Reducing the Number of Classes

When the number of classes was reduced to two classes such as lower middle and upper middle, the SVM

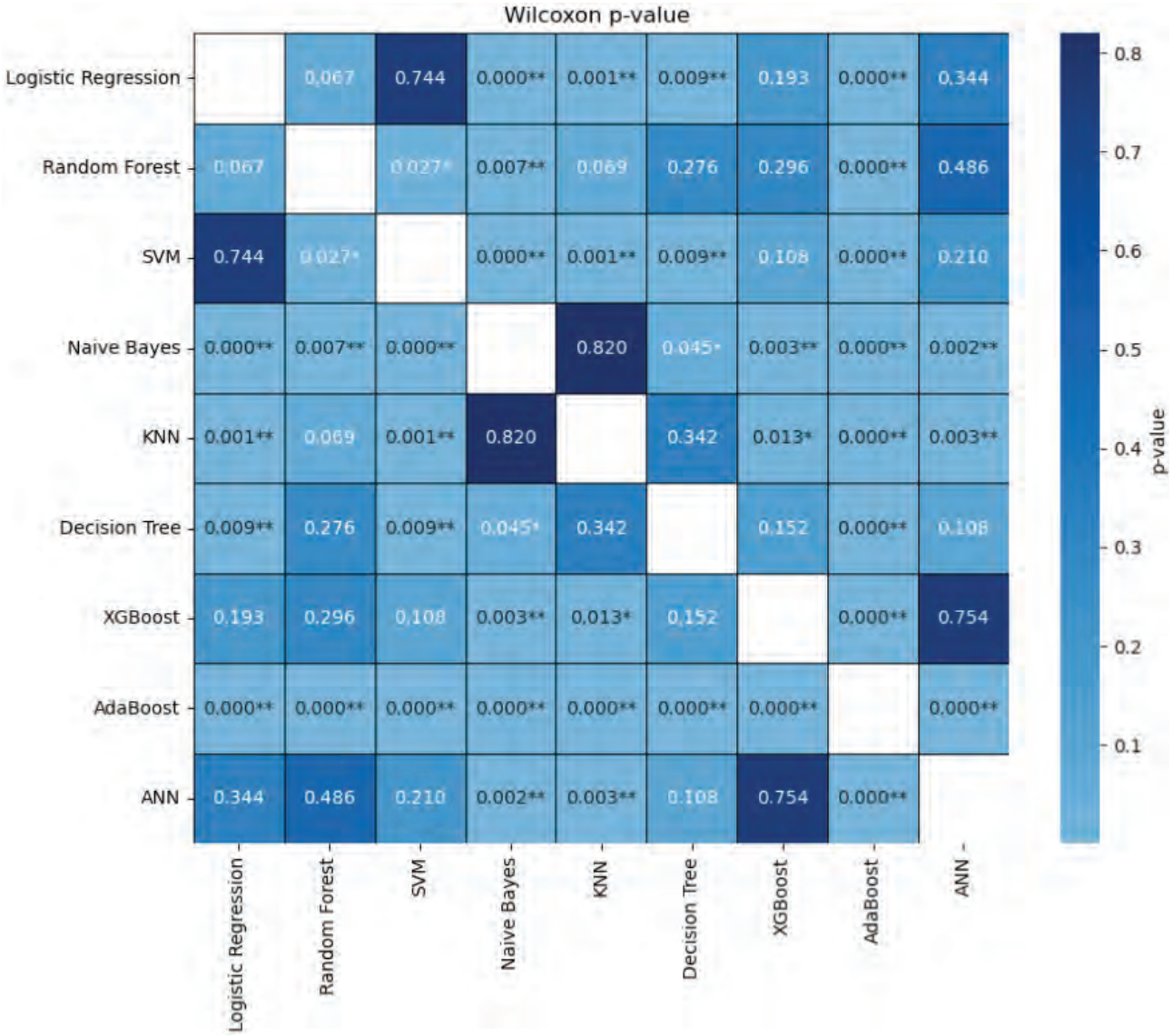


Figure 4. Wilcoxon p-value heatmap.

Table 1. Weighted classification metrics and model accuracy after applying SMOTE

Model	Weighted precision	Weighted recall	Weighted f1_score	Weighted accuracy	SMOTE accuracy
Logistic regression	0.9700	0.9630	0.9612	0.9584	0.9655
Random forest	0.9428	0.9377	0.9322	0.9213	0.9402
SVM	0.9726	0.9657	0.9646	0.9668	0.9582
Naive bayes	0.9176	0.8985	0.8943	0.9029	0.8886
KNN	0.9131	0.9076	0.9000	0.885	0.8952
Decision tree	0.9339	0.9258	0.9207	0.9118	0.9133
XGBoost	0.9498	0.9476	0.9424	0.9376	0.9479
AdaBoost	0.2796	0.4739	0.3391	0.4952	0.5406
ANN	0.9629	0.9527	0.9518	0.9559	0.9529

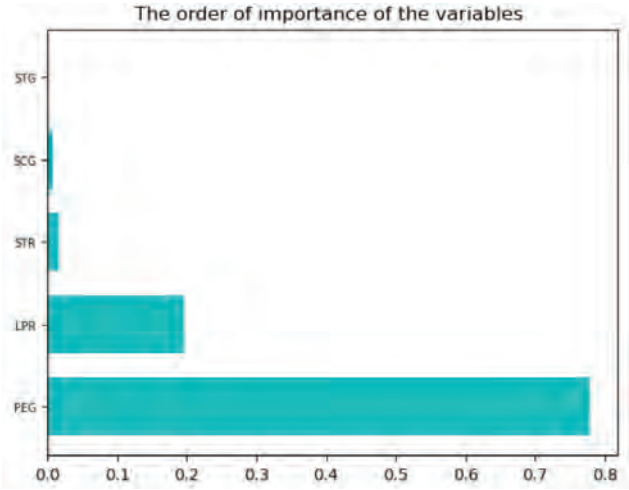


Figure 5. The order of importance of the variables determined by the random forest algorithm.

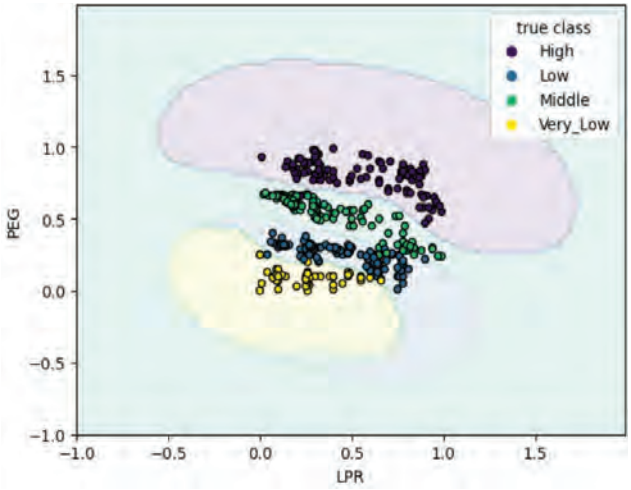


Figure 6. Visualization of the four-classes model on the data plane using two features.

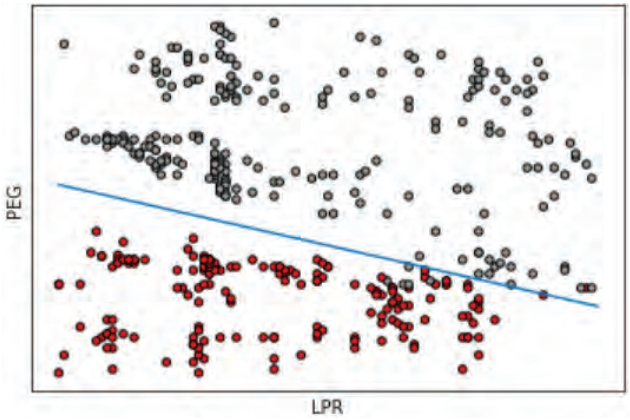


Figure 7. Visualization of the two-classes model on the data plane using SVM with a linear kernel.

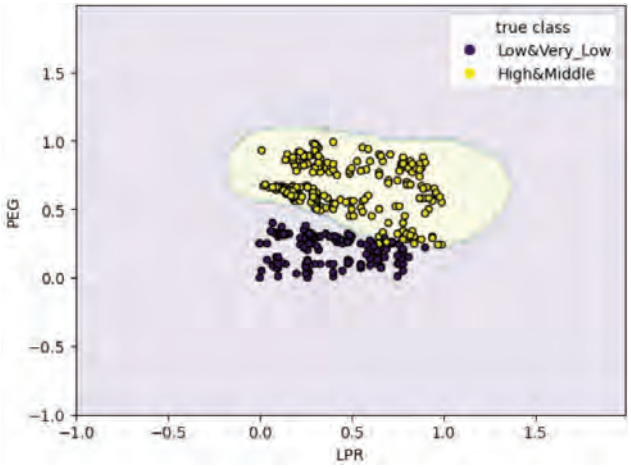


Figure 8. Visualization of the two-classes model on the data plane using SVM with an RBF kernel.

model with five features and the parameters {'C': 1000, kernel='linear'} yielded 98.7% accuracy. Moreover the number of features reduced by considering the order of importance in the random forest algorithm, the model with three features and the parameters {'C': 17, kernel='linear'} yielded an accuracy of 98.5%; while the model with two features and the parameters {'C': 4, kernel='linear'} achieved an accuracy of 98.5%. When the number of classes was reduced to two classes such as lower middle and upper middle and the number of features was reduced, model accuracy with {'C': 0.1, kernel='rbf'} for two classes and three features was 98.5%; for two classes and two features with {'C': 1000, kernel='rbf'} was 91.3%. Reducing the number of features to two means dividing the data into two regions on the plane, both linearly as seen in Figure 7 and curvilinearly as seen in Figure 8.

Feature Construction on Fuzzification

This section aims to develop a flexible modeling approach that uses statistical quantities of feature distributions, as opposed to classical fuzzy approaches. Similar data-driven approaches have been discussed in the literature under the term adaptive fuzzy partitioning, and it has been observed that they improve performance by better representing uncertainty in the data [53].

X_{mxn} denotes the dataset consisting of m observation with n features (here $m = 403$, $n = 5$). Y_{mx1} is a matrix consisting of labels. We denote the labels as a matrix Y with one column. So the i_{th} row of X_{mxn} represents the i_{th} unlabeled observation of the dataset as $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$. When the fuzzification process is applied on each p_{th} columns of the dataset: $x_p = (x_{1p}, x_{2p}, \ldots, x_{mp})$ which represents p_{th} feature, we get two new features such as membership degree

X_p *Fuzzy* and fuzzy class number X_p *FClass* for $p = 1, 2, .., n$. Here, the fuzzification process given by the following steps

For each p_{th} column of the dataset $X_p = (x_{1p}, x_{2p}, \ldots, x_{mp})$ for $p = 1, 2, .., n$.

Step 1. decide the boundaries of the fuzzy sets such as $[\beta]_{1p}$, $[\beta]_{2p}$, $[\beta]_{3p}$ according to the statistical quantities of the feature. ($[\beta]_{1p}$ is standart deviation of the variable X_p; $[\beta]_{2p}$ and $[\beta]_{3p}$ denote the mean value of the variable X_p)

Step 2. generate the fuzzy sets using following formulas for $i = 1, 2, .., m$

$$\mu_{1p} = \begin{cases} \frac{x_{ip}}{-[\beta]_{1p}} + 1, & 0 < x_{ip} \leq [\beta]_{1p} \\ 0, & [\beta]_{1p} < x_{ip} \leq 1 \end{cases} \quad (1)$$

$$\mu_{2p} = \begin{cases} \frac{x_{ip}}{[\beta]_{2p}}, & 0 < x_{ip} \leq [\beta]_{2p} \\ \frac{x_{ip}-1}{[\beta]_{2p}-1}, & [\beta]_{2p} < x_{ip} \leq 1 \end{cases} \quad (2)$$

$$\mu_{3p} = \begin{cases} 0, & 0 < x_{ip} \leq [\beta]_{3p} \\ \frac{x_{ip}-[\beta]_{3p}}{1-[\beta]_{3p}}, & [\beta]_{3p} < x_{ip} \leq 1 \end{cases} \quad (3)$$

Step 3. Calculate

$$X_p Fuzzy = \max_{t\in\{1,2,3\}}(\mu_{tp}) \quad (4)$$

and

$$X_p FClass = \arg\max_{i\in\{1,2,3\}}(\mu_{ip}) \quad (5)$$

Step 4. add two columns to the dataset X_{mxn} matrix (X_p *Fuzzy* to the dataset matrix as an entry $x_{i(n+p)}$); X_p *FClass* to the dataset matrix as an entry $x_{i(n+p+1)}$ in the i_{th} row, $(n+p)_{th}$; $(n+p+1)_{th}$ columns of the data matrix, $i = 1, 2, .., m$)

Step 5. apply steps 1 to 4 for every p value and obtain the new data matrix $X_{mx(3n)}$ (note: if it is necessary, the steps may be applied for only one feature column and obtained the new data matrix $X_{mx(n+2)}$ or the steps may be applied for only k feature columns so the new data matrix $X_{mx(n+2k)}$ will be obtained)

Step 6. build the machine learning model with the new data matrix and label matrix Y

From Step 2, when membership functions of the five features are calculated separately, for each feature Fuzzy sets are obtained as Figure 9-13. Here, the lines named low, medium and height were obtained with μ_{1p}, μ_{2p}, μ_{3p} formulas defined in step 2.

The steps are applied on each feature, we obtain following results: When the fuzzification is applied on STG, the model accuracy score with {'C': 100, kernel='linear'} is 94.8%. If STG is dropped from the new dataset, we get the accuracy score as 95.0% with {'C': 100, kernel='linear'}. When the fuzzification is applied on SCG, the model accuracy score with {'C': 17, kernel='linear'} is 95.55%. If SCG is dropped from the new dataset, we get the accuracy score as 93.81% with {'C': 150, kernel='linear'}. When the fuzzification is applied on STR, the model accuracy score with {'C': 100, kernel='linear'} is 94.79%. If STR is dropped from the

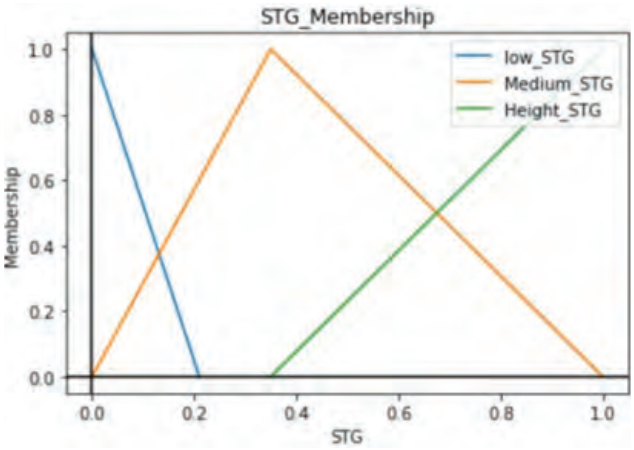


Figure 9. Fuzzy sets and Membership degrees of the feature STG.

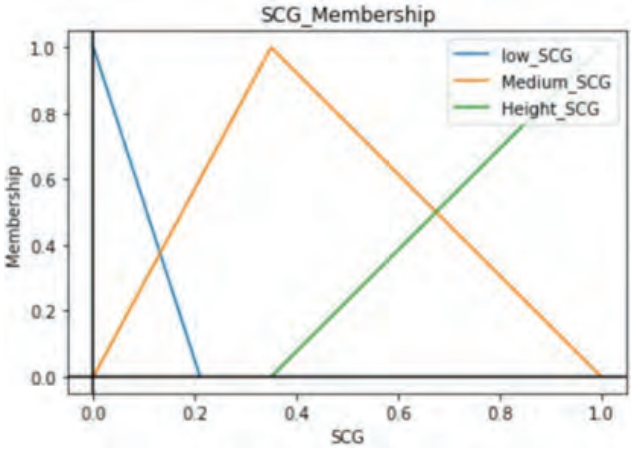


Figure 10. Fuzzy sets and Membership degrees of the feature SCG.

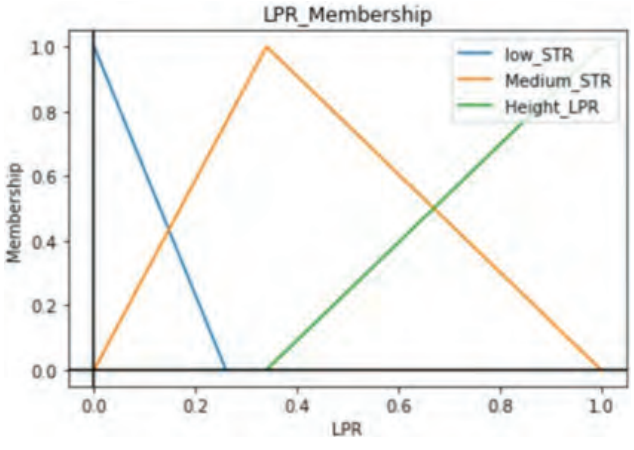


Figure 11. Fuzzy sets and Membership degrees of the feature LPR.

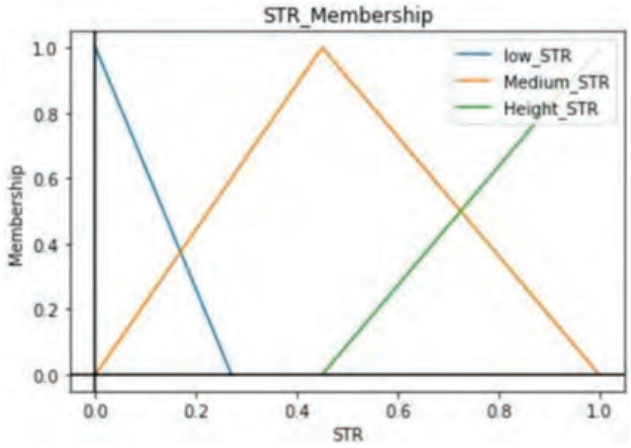


Figure 12. Fuzzy sets and Membership degrees of the feature STR.

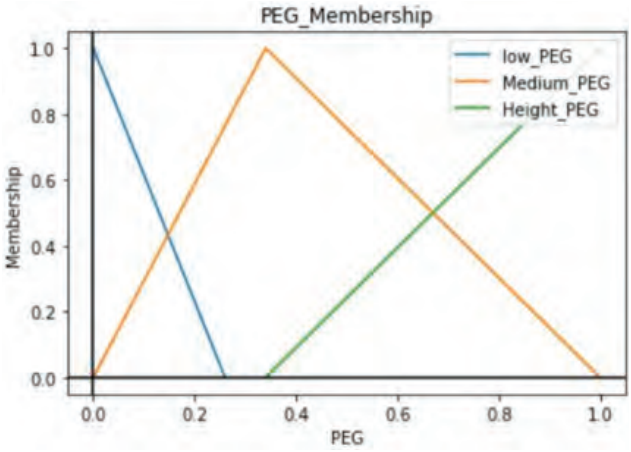


Figure 13. Fuzzy sets and Membership degrees of the feature PEG.

new dataset, we get the accuracy score as 95.05% with {'C': 1000, kernel='linear'}. When the fuzzification is applied on LPR, the model accuracy score with {'C': 4, kernel='linear'} is 95.05%. If LPR is dropped from the new dataset, we get the accuracy score as 95.05% with {'C': 5, kernel='linear'}. When the fuzzification is applied on PEG, the model accuracy score with {'C': 17, kernel='rbf'} is 96.78%. If PEG is dropped from the new dataset, we get the accuracy score as 85.60% with {'C': 17, kernel='rbf'}. When the fuzzification is applied on all features, adding fuzzy set labels and membership degrees to the dataset at the same time, The model accuracy score with {'C': 3, kernel='linear'} is 95.05%.

It was taken into consideration that the proposed approach, which aims to reduce uncertainty in the original feature space, could operate compatibly with unsupervised dimensionality reduction methods such as PCA, as PCA reduces inter-feature correlations and projects the data into a lower-dimensional space. By applying this approach, uncertainty in feature distributions is reduced, thereby enabling PCA to extract more informative components. Accordingly, the fuzzification process was applied prior to PCA.While the data is reduced to two dimensions using PCA, Table 2 shows the performance of all algorithms observed when the technique is applied or not applied before PCA.

To investigate the effects of applying fuzzy transformation before PCA on model performance, a data leakage-preventing processing algorithm, as shown in Figure 14, was implemented. All preprocessing steps were performed strictly within the inner loop of a stratified nested cross-validation framework to avoid any optimistic bias. The impact of the fuzzification module on classification performance was evaluated under two different conditions, with the module enabled and disabled, and visualized using a box plot. The results of evaluating the obtained F1 scores using a paired t-test, Hodge-Lehmann estimation, and Cohen's d test are given in Table 3.

Table 2. The effect of applying to the proposed technique to the data before reducing it to two dimensions applying PCA on the model's performance

Method	Accuracy 4D → 2D via PCA	Accuracy 4D → 2D via FUZY+PCA	Accuracy 3D → 2D via PCA	Accuracy 3D → 2D via FUZY+PCA
Logistic Regression	0.6604	0.7117	0.5974	0.7558
Random Forest	0.6476	0.7217	0.6580	0.8357
SVM	0.6604	0.7190	0.6075	0.7686
Naive Bayes	0.6622	0.6879	0.5902	0.7340
KNN	0.6503	0.7441	0.6113	0.8653
Decision Tree	0.6069	0.6666	0.6053	0.7641
XGBoost	0.5844	0.6582	0.5639	0.8276
AdaBoost	0.3849	0.3761	0.3895	0.3917
ANN	0.7049	0.7413	0.6282	0.7941

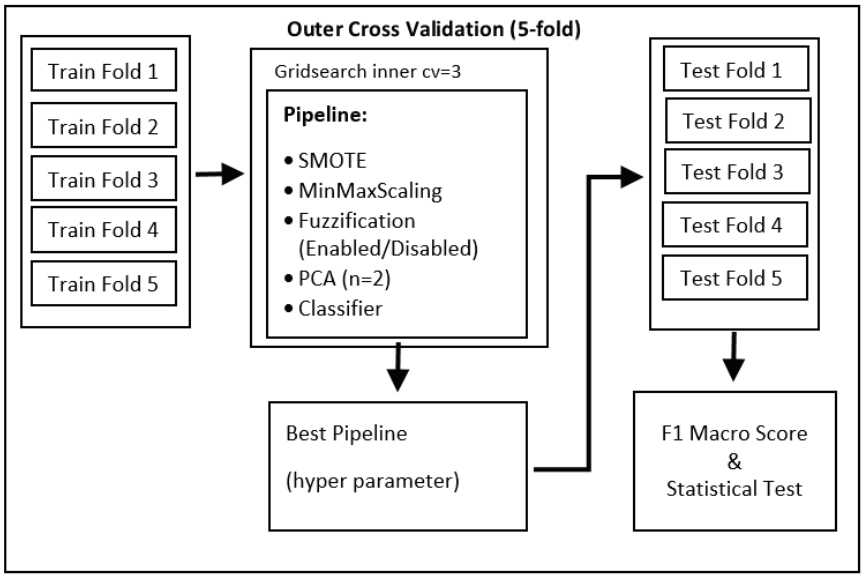


Figure 14. Proposed nested pipeline, showing how fuzzy preprocessing and PCA ordering are applied within inner and outer loops.

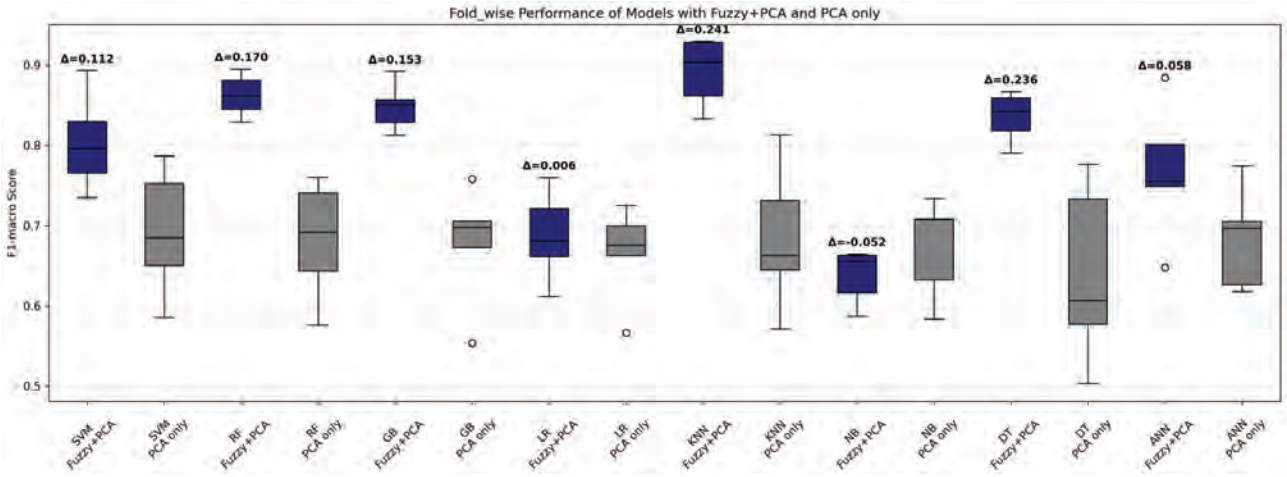


Figure 15. Effect of the fuzzification module on F1-score performance, illustrated using box plot.

Table 3. Statistical comparison of F1 scores with the fuzzification module enabled and disabled using a paired t-test, Hodges-Lehmann estimation, and Cohen's d

Model	Paired t_p	Hodges-Lehmann	Cohens d
SVM	0.021	0.112	1.658
Random forest	0.002	0.170	3.423
XGBoost	0.005	0.153	2.525
Logistic Regression	0.643	0.006	0.224
KNN	0.003	0.241	2.972
Naive bayes	0.423	-0.052	-0.398
DT	0.037	0.236	1.374
ANN	0.054	0.058	1.205

As shown in Figure 15, in our experiments, when the fuzzification module was applied prior to PCA, an increase in F1-scores was observed in all classifiers except NB. The consistency of these results was further confirmed by the paired t-test, Hodges-Lehmann estimation, and Cohen's analysis [54,55]. The outcomes are summarized in Table 3.

Table 3 shows that the Hodges-Lehmann median differences are positive, and Cohen's d values for the SVM, Random Forest, XGBoost, KNN and Decision Tree models, indicating that the fuzzification module significantly impacts the F1 scores. In contrast, no statistically significant effect is observed in the Logistic Regression and Naïve Bayes models.

RESULTS AND DISCUSSION

In this study, which introduces a novel fuzzification-based feature construction approach for ML models, the effects of feature reduction on the model accuracy score were examined using supervised and unsupervised methods. When the results are compared, it can be seen that reducing the number of features using supervised feature extraction methods has a greater positive impact on the model's accuracy score than reductions based on the importance ranking determined by random forest algorithm, and unsupervised feature extraction methods have negative effects on the model accuracy score, as they do not use class labels. However the proposed unsupervised method has positive effects on it. When the membership degrees and the set numbers are added as two new features to the dataset by creating the fuzzy sets based on the most important feature according to the random forest algorithm, the model yields a better accuracy score. It has been observed that unsupervised new feature construction approach based on fuzzification has a positive effect on the accuracy score even though unsupervised feature extraction methods have negative effects on the model accuracy score. Moreover, when the effect of applying the proposed technique to the data before dimensionality reduction was investigated, among all dimensionality reduction techniques a noticeable improvement in the performance of all algorithms was observed specifically when the data was reduced to two dimensions applying PCA. Applying fuzzification did not have a significant impact on training time or memory usage. Using a dataset of 403 samples, proposed method has shown its effectiveness, though the small dataset limits generalization, it can be generalized to other dataset. This results suggest that the applied technique can be recommended as a pre-processing technique and can be applied in areas such as image processing.

CONCLUSION

This study investigates the effects of reducing the number of features or classes and generating new features on model performance. Parameters that maximize model success were investigated for models created using various machine learning algorithms, including k-nearest neighbors, Naïve Bayes, support vector machines, random forest, boosting methods, decision trees, and artificial neural networks. The Wilcoxon test was applied to compare algorithm performances within the same layers and identify statistically significant differences. In addition, a weighted classification report was used to ensure a balanced evaluation, and model accuracy values were obtained using SMOTE.

The impact of both supervised and unsupervised dimensionality reduction techniques on model performance was analyzed. Furthermore, a novel fuzzification-based feature generation approach was proposed, which computes the number of fuzzy clusters, and membership degrees for each feature and incorporates them into the dataset. Experimental results demonstrate that applying the proposed method prior to dimensionality reduction, particularly before reducing data to two dimensions using PCA, generally leads to improvements in the performance of most evaluated algorithms.

REFERENCES

[1] Alpaydin E. Yapay öğrenme. 2. baskı. İstanbul: Boğaziçi Üniversitesi Yayınları; 2011.

[2] Vouk B, Guid M, Robnik Sikonja M. Feature construction using explanations of individual predictions. Eng Appl Artif Intell 2023;120:105823:1–27. [CrossRef]

[3] Hummel R, Manevitz LM. Combination calculi for uncertainty reasoning: representing uncertainty using distribution. Ann Math Artif Intell 1997;20:69–109. [CrossRef]

[4] Zadeh LA. Fuzzy sets. Inf Control 1965;8:338–353. [CrossRef]

[5] Zimmermann HJ. Fuzzy set theory and its application. Berlin/Heidelberg, Germany: Springer Science & Business Media; 2011.

[6] Zimmermann HJ. Fuzzy set theory. Wiley Interdiscip Rev Comput Stat 2010;2:317–332. [CrossRef]

[7] Shokori B, Shureshjoni RA, et al. A parametric method for ranking intuitionistic fuzzy numbers and its application to solve intuitionistic fuzzy network data envelopment analysis models. Complexity 2020;2020:6408613. [CrossRef]

[8] Celikyilmaz A, Turksen IB. Modelling uncertainty with fuzzy logic. In Kacprzyk J, (editor). Studies in Fuzziness and Soft Computing. Newyork: Springer Nature; 2009. p. 1–10. [CrossRef]

[9] Han J, Liu G. A novel approach based on fuzzy rough sets for web query system. In: Cao BY, et al., editors. Fuzzy İnformation and Engineering. Newyork: Springer Nature; 2010. p. 667–673. [CrossRef]

[10] Biswas N, Chakraborty C, Mullick SS, Das S. A parameter independent fuzzy weighted k nearest neighbor classifier. Pattern Recogn Lett 2018;101:80–87. [CrossRef]

[11] Nowicki RK, Starczewski JT. A new method for classification of imprecise data using fuzzy rough fuzzification. Inf Sci 2017;414:33–52. [CrossRef]
[12] Martino FD, Sessa S. A classification algorithm based on multidimensional fuzzy transforms. J Ambient Intell Humanized Comput 2022;13:2873–2885. [CrossRef]
[13] Hsu HM, Chen CT. Aggregation of fuzzy opinions under group decision makers. Fuzzy Sets Syst 1996;79:279–285. [CrossRef]
[14] Herrera F, Martinez LA. 2-tuple fuzzy linguistic representation model for computing with words. IEEE Trans Fuzzy Syst 2000;8:746–752. [CrossRef]
[15] Anil PN, Patil PG. Multi-parameter fuzzy soft set approach to decision making problems. Afr J Math Comput Sci Res 2018;11:61–71. [CrossRef]
[16] Brunelli M, Mezei J. How different are ranking methods for fuzzy numbers: A numerical study. Int J Approx Reason 2013;54:627–639. [CrossRef]
[17] Kanbay F, Vardar N. On a fuzzy application of the particulate matters estimation. Iran J Fuzzy Syst 2018;15:133–143.
[18] Belman RE, Zadeh LA. Decision making in a fuzzy environment. Manag Sci 1970;17:141–164. [CrossRef]
[19] Bezdek JC. Pattern recognition with fuzzy function algorithms. New York: Plenum Press; 1981. [CrossRef]
[20] Chen SJ, Hwang CL. Fuzzy Multiple Attribute Decision Making Methods. In: Beckmann M, Künzi HP, Fandel G, Trockel W, Dawid H, et al. (editors). Fuzzy Multiple Attribute Decision Making. Lecture Notes in Economics and Mathematical Systems. Heidelberg, Baden-Wurttemberg, Germany: Springer-Verlag; 1992. p. 289–486. [CrossRef]
[21] Roy AR, Maji PK. A fuzzy soft set theoretic approach to decision making problems. J Comput Appl Math 2007;203:412–418. [CrossRef]
[22] Zadeh LA. The concept of linguistic variable and its application to approximate reasoning I. Inf Sci 1975;8:199–249. [CrossRef]
[23] Yager RR. Database discovery using fuzzy sets. Int J Intell Syst 1996;11:691–712. [CrossRef]
[24] Martinez L, Rodriguez RM, Herrera F. The 2-tuple linguistic model, computing with words in decision making. In The 2-tuple Linguistic Model. New York: Springer Nature; 2015. p. 1–10. [CrossRef]
[25] Ezghari S, Zahi A, Zenkouar Z. A new nearest neighbor classification method based on fuzzy set theory and aggregation operators. Expert Syst Appl 2017;80:58–74. [CrossRef]
[26] Shang W, Huang H, Zhu H, Lin Y, Wang Z, Qu Y. An improved kNN algorithm fuzzy kNN. In: CIS Part I, LNAI 3801. Heidelberg, Baden-Wurttemberg, Germany: Springer-Verlag; 2005. p. 741–746. [CrossRef]
[27] Yuan Y, Shaw MJ. Introduction of fuzzy decision trees. Fuzzy Sets Syst 1995;69:125–139. [CrossRef]
[28] Bonissone P, Cadenas JM, et al. A fuzzy random forest. Int J Approx Reason 2010;51:729–747. [CrossRef]
[29] Lin CF, De Wang S. Fuzzy support vector machines. IEEE Trans Neural Netw 2002;13:464–471. [CrossRef]
[30] Wang XZ, Xing HJ, Li Y, Hua Q, Dong CR, Pedrycz W. A study on relationship between generalization abilities and fuzziness base classifiers in ensemble learning. IEEE Trans Fuzzy Syst 2014;23:1638–1654. [CrossRef]
[31] Kuncheva LI. Fuzzy classifier design. Vol. 49. Berlin/Heidelberg, Germany: Springer Science & Business Media; 2000. [CrossRef]
[32] Yang YL, Bai XY. A research on classification performance of fuzzy classifiers based on fuzzy set theory. Iran J Fuzzy Syst 2019;16:15–27.
[33] Yang MS, Chen CT. On strong consistency of the fuzzy generalized nearest neighbor rule. Fuzzy Sets Syst 1993;60:273–281. [CrossRef]
[34] Chen HL, Liu DY, Yang B, Wang SJ. An adaptive fuzzy k-nearest neighbor method based on parallel particle swarm optimization for bankruptcy prediction. In: Lecture Notes in Computer Science 6634 LNAI (Part 1). Berlin/Heidelberg, Germany: Springer-Verlag; 2011. p. 249–264. [CrossRef]
[35] Li DC, Liu CW, Hu SC. A fuzzy-based data transformation for feature extraction to increase classification performance with small medical datasets. Artif Intell Med 2011;52:45–52. [CrossRef]
[36] Theng D, Bhoyar KK. Feature selection techniques for machine learning: A survey of more than two decades of research. Knowl Inf Syst 2024;66:1575–1637. [CrossRef]
[37] Wen H, Zhao S, Liang M. Unsupervised attribute reduction algorithm for mixed data based on fuzzy optimal approximation set. Mathematics 2023;11:3452. [CrossRef]
[38] Gu X. Self-adaptive fuzzy learning ensemble systems with dimensionality compression from data streams. Inf Sci 2023;634:382–399. [CrossRef]
[39] Ismail M, Shang C, Yang J, et al. Supporting ANFIS interpolation for image super resolution with fuzzy rough feature selection. Appl Intell 2024;54:5373–5388. [CrossRef]
[40] Ma X, Hu H, Zhang Q, Xu Y. Feature selection for hybrid information systems based on β fuzzy covering and fuzzy evidence theory. Complex Intell Syst 2024;10:7531–7552. [CrossRef]
[41] Mohammed SJ, Abbas AK, Ahmad AA, Mohammed S, Sarı M, Uslu TH. Determining the parameters of data mining techniques and their effects on iron deficiency based prediction. Sigma J Eng Nat Sci 2025;43:655–664. [CrossRef]
[42] Akkiraz C, Türkmen Hİ, Güvensan MA, Can B. Battery-friendly tiny models for activity recognition on energy-constrained devices. Sigma J Eng Nat Sci 2023;41:716–728. [CrossRef]

[43] Yiğit ÖE, Öz E. Feature extraction for DNA capillary electrophoresis signals based on discrete wavelet transform combined with multi-scale permutation entropy. Sigma J Eng Nat Sci 2022;40:475–490. [CrossRef]

[44] Jake V. Python data science handbook: Essential tools for working with data. Sebastopol, California, USA: O'Reilly Media; 2017.

[45] Albon C. Machine learning with Python cookbook: Practical solutions from preprocessing to deep learning. 1st ed. Sebastopol, California, USA: O'Reilly Media; 2018.

[46] Kahraman H, Colak I, Sagiroglu S. User knowledge modeling [dataset]. Uttar Pradesh, India: UCI Machine Learning Repository; 2009.

[47] Kahraman HT. Designing and application of web-based adaptive intelligent education system (Unpublished doctorial thesis). Ankara: Gazi University; 2009.

[48] Kahraman HT, Sagiroglu S, Colak I. The development of intuitive knowledge classifier and the modeling of domain dependent data. Knowl-Based Syst Man Cybern 2013;37:283–295. [CrossRef]

[49] Sarker IH. Machine learning: Algorithms, real-world applications, and research directions. Sn Comput Sci 2021;2:160. [CrossRef]

[50] Dao PB. On Wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. Appl Energy 2022;318:1–15. [CrossRef]

[51] Bagkavos D, Patil PN. Improving the Wilcoxon signed rank test by a kernel smooth probability integral transformation. Stat Probab Lett 2021;171:1–10. [CrossRef]

[52] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 2002;16:321–357. [CrossRef]

[53] Ruiz E, Casillas J. Adaptive fuzzy partitions for evolving association rules in big data stream. Int J Approx Reason 2018;93:463–486. [CrossRef]

[54] Hodges Jr. L, Lehmann EL. Estimates of location based on rank tests. Ann Math Stat 1963;34:598–611. [CrossRef]

[55] Cohen J. Statistical power analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates; 1988.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.1944

Research Article

Synthesis of compounds containing Cyano Anchor Groups, production of dye-sensitive solar cells and invesigation of the antibacterial properties of these compounds

Fatma AYTAN KILIÇARSLAN[1], Melda ALTIKATOĞLU YAPAÖZ[1], İbrahim ERDEN[1,*]

[1]*Department of Chemistry, Yildiz Technical University, Istanbul, 34220, Türkiye*

ARTICLE INFO

Article history
Received: 22 November 2024
Revised: 15 January 2025
Accepted: 30 January 2025

Keywords:
Antibacterial Activity; Dye Sensitized Solar Cells (DSSCs); Fluorene; Triphenylamine

ABSTRACT

Global warming has become one of the most important factors threatening the world. The first place among the causes of global warming is taken by fossil fuels, which are mainly used for energy needs. The intensive consumption of fossil fuels causes the decrease of the world's fossil fuel reserves, environmental pollution and the greenhouse effect. Various renewable energy sources are shown as alternatives as a solution to these problems. Among all renewable energy forms, solar energy is the most popular. In this article, a dye-sensitive solar cell device of two compounds having a cyano anchor group containing fluorene and triphenyl amine as donor groups was prepared, and their antibacterial properties were investigated. The photovoltaic and electrochemical properties of these materials were evaluated, and it was investigated whether they could be used as photosensitizers in dye-sensitive solar cells. Photovoltaic cell efficiencies of the devices were in the range of 0.13-0.04% under 100 mW/cm^2 simulated AM 1.5 solar irradiance, and the highest open circuit voltage reached 0.45 V. When the photovoltaic cell efficiency values of the produced dye-sensitive solar cell devices were examined, it was seen that the L1 compound was higher than L2 due to the effect of the donor group. The antimicrobial properties of L1 and L2 compounds were examined, and the effect of donor groups was investigated. The conventional disk diffusion method was used to evaluate the antibacterial activity of the synthesized compounds against Escherichia coli and Staphylococcus aureus. As a result, it was seen that the produced compounds showed antibacterial activity, and the compound with high photovoltaic cell efficiency value had a high antibacterial properties.

Cite this article as: Aytan Kılıçarslan F, Altıkatoğlu Yapaöz M, Erden İ. Synthesis of compounds containing Cyano Anchor Groups, production of dye-sensitive solar cells and invesigation of the antibacterial properties of these compounds. Sigma J Eng Nat Sci 2026;44(2):1041–1047.

INTRODUCTION

Photovoltaic technology is known to be one of the most important techniques aiming to produce clean energy from sunlight [1]. Due to their low cost and simple fabrication methods, dye-sensitized solar cells (DSSCs), which convert sunlight into electricity, have been the subject of intense

*Corresponding author.
*E-mail address: ierden@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

research over the past several decades as a novel system in the field of photovoltaics. Other known advantages of DSSCs are flexibility, chemical versatility, colorful appearances, and the possibility of synthesizing a wide range of different organic and organometallic molecular structures [2-4].

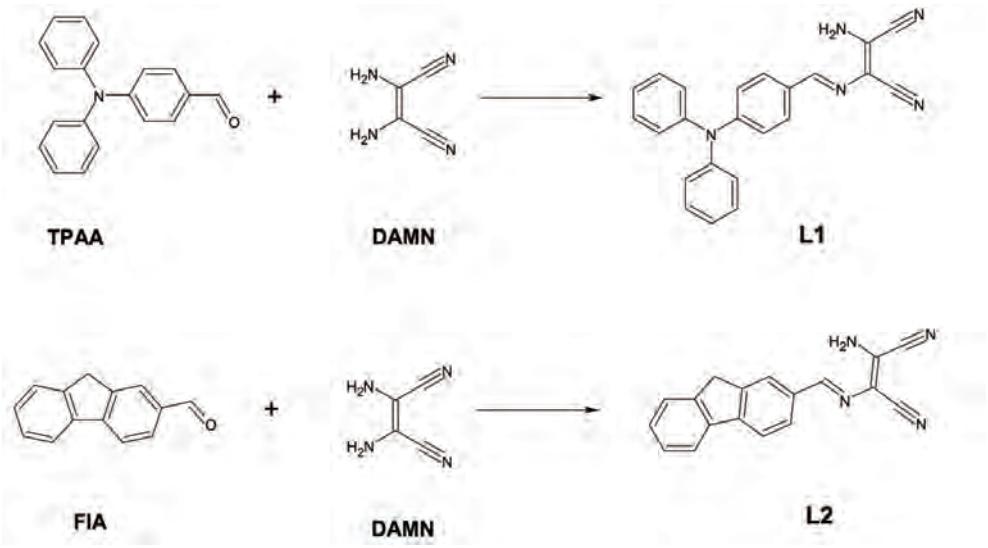
The device's photovoltaic and photocurrent levels affect how well solar energy is converted to electricity. Therefore, to enhance the photocurrent, the sensitizer's solar absorption spectra must be expanded from the visible to the near-infrared range, and its molar extinction coefficient must be raised. [5, 6].

In this case, novel chemical dyes with a red-shifted absorption band and increased molar absorptivity were created by attaching very prolonged conjugation units, such as triphenylamine and benzene, to the diaminomaleonitrile ligands. Within the organic photoelectric area, such as in organic light-emitting diodes (OLEDs) [7], organic field effect transistors (OFETs) [8], and perovskite solar cells (PSCs) [9], triphenylamine (TPA) and its derivatives, hole-transporting and emitting materials, are commonly utilized. Recent research revealed that because of its benefits, such as low excitation energy, less light damage, strong penetrability, etc., it has a promising future in the domains of biological probes and drug targeting [10]. Fluorene and its derivatives have been widely and successfully used in a variety of synthetic applications in recent years. For example, they are used in medical and optical applications. Additionally, fluorene derivatives show impressive spectroscopic and photophysical properties. Fluorine-based fluorophores are used significantly to improve dye absorption properties and fluorophore sensors [11, 12]. It is extensively employed in the production of several heterocyclic compounds.

2,3-Diaminomaleonitrile (DAMN), which is extensively employed in the production of several heterocyclic compounds, is a tetramer of hydrogen cyanide. DAMN derivatives have potential applications for the synthesis of heterocyclic compounds as well as the electron donor and/or due to the effects of the impact on the spectroscopic electron acceptor substituents, a wide variety of organic π-conjugated DAMN-based compounds and these compounds attention on the physicochemical properties, attractive and intense studies are in progress [13]. Because of their versatility in synthesis, mono-imitated DAMN-based compounds are also employed as intermediates in the synthesis of non-symmetrical ligands, pigments, and materials containing high nitrogen content [13,14]. It is commonly known that DAMN-based Schiff base is produced when DAMN reacts with aromatic aldehydes [15].

Azomethine groups (-RC=N-) are present in Schiff bases structurally, and they are created when amines and active carbonyl compounds condense. Schiff bases play a significant role as synthetic intermediates and are utilized in the thermostable optical material business, chemical sensors, aggregation-enhanced emission luminogen, jet-printing inks, and optically active materials [16]. Numerous scientists worldwide are studying the substantial anti-tumor, antibacterial, and antifungal action that various Schiff bases exhibit. According to reports, the nitrogen atom in the azomethine group may contribute to the hydrogen bond formation with the active centers of cell components, thereby interfering with regular cell functions [17].

In our study, we describe the synthesis and dye-sensitized solar cell performance and antimicrobial activity of organic electronic materials with D-π-A properties. We synthesized two dyes containing donor fluorene and triphenyl amine and containing a CN anchor group. It was determined whether these chemicals may be used as



Scheme 1. Synthesis scheme of **L1** and **L2** compounds.

photosensitizers in DSSCs by looking at their photovoltaic characteristics. The analysis results show that the L1 sensitizer used in the device has a strong donor group effect and increases the power conversion efficiency. When the antibacterial activities of these substances were investigated, it was determined that the synthesized substances showed the presence of antibacterial activity. Since it is a donor substance and has a high electron density, the highest antibacterial activity was observed in L1 substance.

MATERIALS AND METHODS

Materials

All of the chemicals were Fluka and Merck reagent grade. Unless specified otherwise, commercially available reagents were used without further purification, and solvents in processes were distilled from suitable drying agents before use. For column chromatography, Merck's silica gel 60 (0.04-0.063 mm) was used. The most important aspects of S. aureus and E. coli bacteria in the intestinal flora, which are quite common in nature and are found in dust, soil, on goods, human and animal skin, nasal mucosa, mouth, and nasopharynx flora, are chemotherapeutic. The rapid resistance too many of the substances, and therefore infections, are more common. E. coli is Gram negative; S. aureus is from Gram-positive bacteria group. The use of these two bacteria was deemed appropriate in our experimental study, since they are examples of both species and because they are widely used in antimicrobial studies in the literature. E. coli (ATCC25922) and S. aureus (ATCC 29213) strains were obtained from Yıldız Technical University, Faculty of Arts and Sciences, Chemistry Department, Biochemistry Laboratory Culture Collection.

SYNTHESIS

Synthesis of 2-amino-3-[(E)-{[4-(diphenylamino)phenyl] methylidene}amino]-3-aminomaleonitrile (L1) **[18]**

To the solution of 2,3-diaminomaleonitrile (0.1 g, 0.462 mmol) dissolved in 10 mL methanol, slowly add triphenylamine aldehyde (0.250 g, 0.462 mmol) dissolved in 10 mL methanol. The mixture is refluxed for 36 hours and then cooled at room temperature until a light yellow precipitate forms. The product is filtered, washed several times with cold methanol and acetonitrile and dried at ambient temperature. $C_{23}H_{17}N_5$ (L1), Yield: 0.24 g (72%). m.p. = 198 °C

Synthesis of 2-(((E)-(9H-fluoren-2-yl)methylene)amino)-3-aminomaleonitrile (L2) **[19]**

To the solution of 2,3-diaminomaleonitrile (0.139 g., 1.28 mmol) dissolved in 10 mL methanol, slowly add fluorene 2-aldehyde (0.250 g. 1.28 mmol) dissolved in 10 mL methanol. The mixture is refluxed for 24 hours and then cooled at room temperature until a light yellow precipitate forms. The product is filtered, washed several times with cold methanol and acetonitrile and dried at ambient temperature. $C_{18}H_{12}N_4$ (L2), Yield: 0.19 g (53%). m.p. = 195 °C

RESULTS AND DISCUSSION

Synthesis and Characterization

As shown in Scheme 1, L1 and L2 dyes were synthesized in good yields by the formation of Schiff base reactions. In the L1 and L2 FTIR spectra, the characteristic stretching bands of the amine (NH_2) group are seen at 3480-3350 cm^{-1}, and the stretching bands of the imine (C=N) group are seen at 1619 and 1590 cm^{-1}, respectively. These strong bands, which are indicative of the azomethine moiety of the majority of Schiff base compounds, are attributed to the

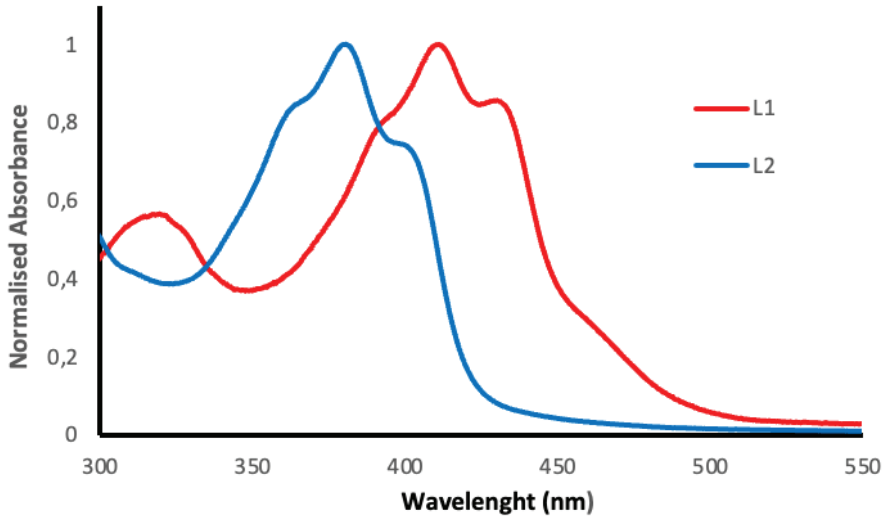


Figure 1. UV-vis spectra of L1 and L2 compounds in methanol.

ligands' C=N stretching frequencies. In the infrared spectra of L1 and L2, the absence of stretching vibrations at 1686 cm^{-1} belonging to the carbonyl group (C=O) in the starting compound TPAA and FIA structure indicates the formation of the Schiff base.

UV-vis absorption spectra of L1 and L2 compounds were taken by preparing $1x10^{-4}$ M in methanol. L1 and L2 showed two distinct bands that were different from the starting materials but similar to each other. In the absorption spectra of L1 and L2, the π-π* transitions with the highest energy levels were observed at peaks at 411 and 381 nm, respectively, as seen in Figure 1. Absorption peaks of strong charge transfer (CT) were not observed in the UV-vis spectrum of both substances.

Production and Characterization of DSSCs

In order to fabricate DSSCs, two transparent TiO_2 pastes were applied, as previously reported [20, 21], on the conductive side of the fluorine-doped tin oxide (FTO) substrate using Doctor Blade's method. The coated FTO substrates were immersed in an EtOH solution including 0.3 mM dye solution to create the dye-sensitized (L1 and L2) TiO_2 films (photoanodes). Using a platinum paste solution, the Pt counter electrodes were created on the FTO substrates. Surlyn film was used to seal the counter electrode and photoanode. Through a bored hole at the counter electrode, the redox electrolyte was injected between the electrodes. The current density-voltage (J-V) characteristics were investigated by using the AM 1.5 global one sun illumination (100 mW cm^{-2}) from a solar simulator.

The electrical performance of a solar cell is determined by the short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), current at the maximum power point (I_{mp}), voltage at the maximum power point (V_{mp}), maximum power (P_{max}), fill factor (FF), and photovoltaic cell efficiency (η). The solar cell efficiency and fill factor are calculated using the Eq. 1 and Eq. 2 equations, respectively [21].

$$\eta = \frac{P_{max}}{P_{in}} = \frac{I_{sc} \cdot V_{oc} \cdot FF}{P_{in}} \quad (1)$$

where the FF is most commonly determined from measurement of the I-V curve, is given by

$$FF = \frac{I_{mp} \cdot V_{mp}}{I_{mp} \cdot V_{oc}} \quad (2)$$

The DSSC parameters: J_{sc}, V_{oc}, and FF were extracted from the J-V curve to find the cell efficiency as illustrated in the Fig. 2. Furthermore, the effective area of the devices in our DSSC device was 0.52 cm^2, which is greater than that of the reported.

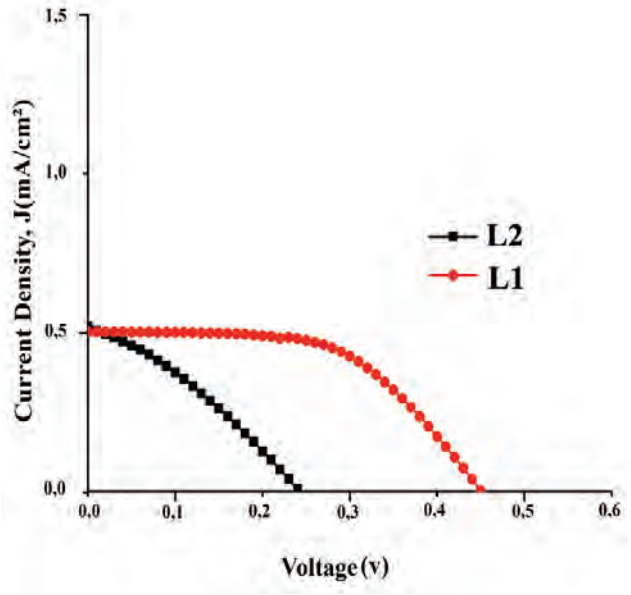


Figure 2. Current density-voltage curves with sensitizers L1 and L2 of illuminated DSSCs.

A clear association between the sensitizer's structure and the device's power conversion efficiency (PCE) can be seen in the current-voltage curves that were obtained in both light and dark conditions. The high values of the L1-based DSSC devices J_{sc}, V_{oc}, FF, and PCE from the generated devices contributed to a higher PCE value.

The higher PCE value obtained from the device fabricated with L1 sensitizer is a result of the TFA donor group having higher electron density compared to other L2 sensitizer derivatives and more electrons are presented to the TiO_2 surface. The values related to the results of the optimized DSSC device are presented in Table 1. The larger molar extinction coefficient and wider absorption in the visible region of L1 compound resulted in higher Jsc values compared to L2 compound. Weak bonding and electron transfer on semiconductor TiO_2 as well as the presence of anchor CN groups in the compound structures, are the causes of the low PCE value that was achieved. Moreover, the poor performance of L1-based DSSC might have been caused by electron recombination due to the aggregation of its molecules on the TiO_2 surface.

Table 1. Photovoltaic parameters for DSSCs sensitized with L1 and L2

Sample	Jsc (mA/cm²)	Voc (V)	FF	η (%)
L1	0,5	0,45	0,59	0,13
L2	0,5	0,25	0,32	0,04

There are few articles in the literature where the cyano group is used alone as an anchor/acceptor. In the sensitizer structure, the cyano group is generally located together with the carboxylic acid anchor group, and in this case, the life of the prepared devices is short due to the hydrolysis of the carboxylic acid. The L1 and L2 sensitizers used in the DSSC device were evaluated with the CN anchor/acceptor groups found in the literature. As a result, it was observed that higher Voc values were obtained in the produced devices, but lower Jsc and FF values were achieved. Low Jsc and FF values cause the power conversion efficiency of our device to decrease. Direct electron transfer from the donor group to the anchor group in the sensitizing structure leads to an increase in the performance of sensitized solar cells [21]. The existence of the π-conjugation system that provides the transfer of these electrons, as well as the electron density that will form from the donor group to the anchor group, is an important issue in terms of the performance of solar cells [22].

The functionalized dye (JY1) was synthesized by J. Cong et al. and included an electron-withdrawing nitro group along with CN and COOH anchor groups. The dye's power conversion efficiency was found to be η = 0.81 [23]. In a separate investigation, Badawy et al. synthesized two cyanoacetamide and four novel phenylacetonitrile dye derivatives and examined the effects of a thiophene spacer and several supplementary acceptors on the photophysical and electrochemical properties of the DSSC devices. Additionally, they assessed the DSSCs' total photo conversion efficiencies and photovoltaic characteristics for the sensitizers SFA5 and SFA6, obtaining the highest result reported in the literature [24]. In our next studies, we will synthesize new sensitizers by changing the acceptor and anchor groups in the structure to increase the efficiency of our dyes.

In the study conducted by S. M. Abdalhadi et al., three new 2,3-diaminomaleonitrile (DAMN) derivative dyes were prepared by simple Schiff base reaction in a one-pot reaction. The compounds were designed as a sensitizer in dye-synthesized solar cells (DSSCs) and were used in DSSCs. The data was found to show the best performance for SA3 dye with 0.38% efficiency at AM 1.5, followed by SA2 with 0.22% and the last dye was SA1 with 0.09%, 5.4% compared to the control cell (N719) [25].

Antimicrobial tests

E. coli (ATCC25922) and S. aureus (ATCC25923) were used as reference strains for Gram-negative and Gram-positive bacteria, respectively. Strains were grown on nutrient agar. The synthesized samples (L1 and L2) were tested for antibacterial activity by the standard disc diffusion method. Each bacterium was added to 100 µL of inoculum agar medium made by culturing for 18 to 24 hours and adjusting the turbidity to meet the 0.5-McFarland criterion. The E. coli broth culture was then inoculated onto nutrient agar plates for 24 hours. Sterile cotton swabs were used to swab each strain uniformly onto the separate plates.

Table 2. Antibacterial activity parameters of L1 and L2 samples

Bacteria	Inhibition zone (mm)	
	L1	*L2*
Echerichia coli	15	13
Staphylococcus aureus	19	17

Gel puncture was used to create 10 mm-diameter wells on nutrient agar plates. Ethanol was used as the solvent and 50 µL of sample solutions (1 mg/mL) were applied to each well in all plates with pipettes. Subsequently, the plates were incubated for 24 hours at 37 °C. The zone incubation of each well was measured in millimeters (mm), and the results were recorded.

When looking at the inhibition zone formed by sample L1, one of the Schiff bases used in the study, it was seen that the antibacterial effect was the highest. Alsantali et al. A new series of (4-(2,7-dichloro-9H-fluoren-4-yl)thiazol-yl) acetamide derivatives were synthesized, characterized, and evaluated for their antimicrobial activity. Additionally, the synthesized fluorene derivatives were found to show remarkable activity as antibacterial and antifungal agents when evaluated as antimicrobial agents. Although there is limited literature, these studies can be considered to open a new alternative field in the field of medicinal chemistry of fluorene derivatives [26].

By synthesizing new mono-, di- and tri-phosphonium ionic liquids and evaluating their antibacterial activity on Gram-positive and Gram-negative bacteria, Brunel et al. found that triphenylamine-phosphonium ionic liquid derivatives have a unique mechanism of action. They observed that although TPA-P compounds with structures with different chains, spacer lengths and charges were active on Gram-positive bacteria, this distinction was in favor of Gram-negative bacteria as the spacer length, and charge density increased and proved that TPA-P derivatives could constitute a useful biological tool [26].

CONCLUSION

In this study, compounds containing fluorene and triphenyl amine as donor groups and cyano as anchor groups were synthesized and used in the production of dye-sensitive solar cell devices. Photovoltaic cell efficiencies of the devices were in the range of 0.13-0.04% under 100 mW/cm^2 simulated AM 1.5 solar irradiance, and the highest open circuit voltage reached 0.45 V. It was observed that the L2<L1 efficiency of the produced dye-sensitive solar cell devices increased, respectively. The effective photovoltaic performance of sensitizer L1 in dye-sensitive solar cell devices appears to have a donor triphenylamine group in their structure and strongly present electrons to the TiO_2

surface. Our best result in this study was achieved with an L1-based dye-sensitive solar cell that exhibited a power conversion efficiency of up to 0.13%. It was observed that the antibacterial effect in the inhibition zone formed by the compounds L1 and L2 was the highest in the L1 compound and can be used in applications. Antibacterial activity and power return efficiency values of compounds L1 and L2 were measured and it was seen that compound L1 had higher values in both studies. We hope that in future studies this relationship will be used in some medical and biological applications.

ACKNOWLEDGEMENTS

This research has been supported by Yıldız Technical University Scientific Research Projects Coordination Department. Project Number: FBA-2021-4051.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Gong J, Sumathy K, Qiao Q, Zhou Z. Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends. Renew Sustain Energy Rev 2017;68:234–246. [CrossRef]

[2] Mozaffari S, Nateghi MR, Zarandi MB. An overview of the challenges in the commercialization of dye sensitized solar cells. Renew Sustain Energy Rev 2017;71:675–86. [CrossRef]

[3] Obotowo IN, Obot IB, Ekpe UJ. Organic sensitizers for dye-sensitized solar cell (DSSC): Properties from computation, progress and future perspectives. J Mol Struct 2016;1122:80–87. [CrossRef]

[4] Carella A, Borbone F, Centore R. Research progress on photosensitizers for DSSC. Front Chem 2018;6:113–127. [CrossRef]

[5] O'Regan B, Grätzel M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films. Nature 1991;353:737–740. [CrossRef]

[6] Grätzel M. Dye-sensitized solar cells. J Photochem Photobiol C Photochem Rev 2003;4:145–153. [CrossRef]

[7] Liu XC, Liang JF, You J, Ying L, Xiao Y, Wang SR, et al. Small molecular hole-transporting and emitting materials for hole-only green organic light-emitting devices. Dyes Pigm 2016;131:41–48. [CrossRef]

[8] Shirota Y, Kageyama H. Charge carrier transporting molecular materials and their applications in devices. Chem Rev 2007;107:953–1010. [CrossRef]

[9] Zhang F, Wang ZQ, Zhu HW, Pellet N, Luo JS, Yi CY, et al. Over 20% PCE perovskite solar cells with superior stability achieved by novel and low-cost hole-transporting materials. Nano Energy 2017;41:469–475. [CrossRef]

[10] Li W, Yang MD, Wang LK, Zhu WJ, Ye LN, Wu JY, et al. Aza-crown ether derivatives based on stilbene: Two-photon absorption and bioimaging. Dyes Pigm 2014;107:133–139. [CrossRef]

[11] Gualtier F, Teodori E, Bellucci C, Pesce E, Piacenza G. Structure-activity relationship studies in the field of calcium(II) antagonists: Effect of modifications at the tetrasubstituted carbon of verapamil-like compounds. J Med Chem 1985;28:1621–1628. [CrossRef]

[12] Alsantali RI, Hussein EM, Obaid RJ, Morad M, Altass HM, Alharbi A, et al. Bioactive fluorenes. Part II. Unprecedented biologically active thiazole derivatives based on 2,7-dichlorofluorene as competent DHFR inhibitors: Design, synthesis, and molecular docking approaches. Arab J Chem 2020;13:5451–5462. [CrossRef]

[13] Susdorf T, Bansal AK, Penzkofer A, Guo SL, Shi JM. Absorption and emission spectroscopic characterization of some azo dyes and a diamino-maleonitrile dye. Chem Phys 2007;333:49–56. [CrossRef]

[14] Khanmohammadi H, Arab V, Rezaeian K, Talei GR, Pass M, Shabani N. Diaminomaleonitrile-based azo receptors: Synthesis, DFT studies, and their antibacterial activities. J Mol Struct 2017;1129:169–178. [CrossRef]

[15] Kalhor M, Seyedzade Z. Rapid synthesis of 2-amino maleonitrile Schiff bases in aqueous media catalyzed by cerium(IV) ammonium nitrate (CAN) and a new method for the one-pot preparation of their dicyano imidazoles. Res Chem Intermed 2017;43:3349–3360. [CrossRef]

[16] Sheykhi-Estalkhjani A, Mahmoodi NO, Yahyazadeh A, Nadamani MP. Synthesis of new bis-benzylidene-hydrazides as a sensitive chromogenic sensor for naked-eye detection of CN^- and AcO^- ions. Tetrahedron 2018;74:4868–4874. [CrossRef]

[17] Horozić E, Suljagić J, Husejnagić D, Hasić N, Bratovčić A. Synthesis, characterization, and in vitro biological evaluation of the Schiff base derived from benzidine and 1,3-diphenyl-1,3-propanedione. J Eng Process Manag 2019;11:112–116. [CrossRef]

[18] Yang J, Shi R, Zhou P, Qiu Q, Li H. Asymmetric Schiff bases derived from diaminomaleonitrile and their metal complexes. J Mol Struct 2016;1106:242–258. [CrossRef]

[19] Sarveswari S, Beneto AJ, Siva A. Colorimetric sensing of cyanide and fluoride ions by diaminomalenonitrile based Schiff bases. Sens Actuators B Chem 2017;245:428–434. [CrossRef]

[20] Erden I, Kilicarslan FA, Karadogan B. Synthesis of phthalocyanines with donor–acceptor properties and their applications to dye-sensitized solar cells. Chem Pap 2019;1:1–6. [CrossRef]

[21] Kilicarslan FA, Erden I. Investigation of the effect of a nitro group as an anchor group in dye-sensitized solar cells. J Chem Res 2023;1:1–6. [CrossRef]

[22] Kilicarslan FA. Synthesis of sensitizer containing triphenylamine as donor group and investigation of solar cell applicability. Sigma J Eng Nat Sci 2025.

[23] Cong J, Yang X, Liu J, Zhao J, Hao Y, Wang Y, et al. Nitro group as a new anchoring group for organic dyes in dye-sensitized solar cells. Chem Commun 2012;48:6663–6665. [CrossRef]

[24] Badawy SA, Abdel-Latif E, Fadda AA, Elmorsy MR. Synthesis of innovative triphenylamine functionalized organic photosensitizers outperformed the benchmark dye N719 for high efficiency dye sensitized solar cells. Sci Rep 2022;12:12885. [CrossRef]

[25] Abdalhadi SM, Al-Baitai AY, Al-Zubaidi HA. Synthesis and Characterization of 2,3-Diaminomaleonitrile Derivatives by One-Pot Schiff Base Reaction and Their Application in Dye Synthesized Solar Cells. Indones J Chem 2021;21:443-451. [CrossRef]

[26] Brunel F, Lautard C, Giorgio C, Garzino F, Raimundo JM, Bolla JM, et al. Antibacterial activities of mono-, di-, and tri-substituted triphenylamine-based phosphonium ionic liquids. Bioorg Med Chem Lett 2018;28:926–929. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2024

Research Article

Evaluation of reliability and performance in multi-subsystem serial system: Analyzing K-out-of-n redundancy in complex series-parallel configuration using Copula methodology

Kabiru Garba IBRAHIM[1,3], Ibrahim YUSUF[2,*], Hamisu Ismail AYAGI[3]

[1]Deparment of Mathematics, Federal University, Dutse, 900242, Nigeria
[2]Department of Mathematical Sciences, Bayero University, Kano, 700233, Nigeria
[3]Department of Mathematics, Yusuf Maitama Sule University, Kano, 700282, Nigeria

ARTICLE INFO

Article history
Received: 08 October 2024
Revised: 30 December 2024
Accepted: 05 February 2025

Keywords:
Availability; Cost; Partial Failure; Reliability; Series-Parallel; Total Failure

ABSTRACT

This paper presents a novel in-depth cost and performance evaluation of a complex serial system comprising five subsystems arranged in a series-parallel configuration. The Subsystems B, C, and D each have one unit, whereas Subsystem A has n units with k-out-of-n redundancy. System E is made up of two units that function in active parallel. The high-performance system that supplies transportation, energy, and manufacturing, the problem of optimizing system dependability and cost effectiveness has been tackled in this study. To obtain a more reliable method than conventional methods and to address dependency in a complete failure, the general repair models and copula methodology are combined in a novel way. The state transition diagram was used to model the system. The system was modeled using a state transition diagram, which leads to a set of partial differential equations that were solved using Laplace transformation and an additional variable. The profit function, mean time to failure (MTTF), and availability of the system, which are the key reliable indicators, were produced by the methodology. Comparing with the usual repair model, our result showed that the copula -based repair method produced a 15% improvement in MTTF and availability, while maintaining a higher profitability index, which was increased by 12%. Our result also shows the trade-off between system profitability and repair costs, with emphasis on systems with optimal maintenance practices might cut off operation expenses by 20%. The results indicate that the copula method for reliability modeling offers engineers, maintenance managers, and system engineers valuable insights into performance and pertinent information. By offering quantitative assessments that can influence useful choices in system maintenance and resource allocation, and incorporating a dependency-aware repair mechanism, this alone surpasses the earlier results. In an attempt to increase cost-effectiveness in crucial applications and dependability, this research paves the way for feature research on sophisticated repair and optimization techniques.

Cite this article as: Ibrahim KG, Yusuf I, Ayagi HI. Evaluation of reliability and performance in multi-subsystem serial system: Analyzing K-out-of-n redundancy in complex series-parallel configuration using Copula methodology. Sigma J Eng Nat Sci 2026;44(2):1048–1068.

***Corresponding author.**
*E-mail address: iyusuf.mth@buk.edu.ng

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

The operational efficiency of the systems is assessing and optimizing in a range of domains, with mechanical and electrical systems from computer networks and software applications, requires reliability and performance analysis.

The goals stated where achieving through these two crucial aspects which guarantee the effectiveness and consistency of the systems operation. The system capacity is operated properly for a fixed amount of time, it's guaranteeing that the necessary functions is completes without failing, which referred to as consistency. Even though, the performance, refers to how well a system functions and includes elements like responsiveness, speed, and resource usage.

In system engineering, performance analysis and reliability are crucial elements that are often combined. The systems with high-performance are not necessarily dependable, while highly dependable systems are not always efficient. Stakeholders' and users' needs requires finding a stability between performance and dependability. Therefore, creating reliable, effective systems satisfy. While the expectations placed on its performance that guarantees the system performs at its best to satisfy, reliability guarantees that the system can be relied upon to function under certain conditions for a specific amount of time.

Many applications help find defects, make well-informed design choices, and eventually raise the general quality of systems through these evaluations. Possible flaws, forecast how the system will behave in different scenarios, Engineers can identify and make adjustments to increase the system's resilience and effectiveness by carrying out comprehensive reliability and performance evaluations. For example, these evaluations can aid in data flow optimization and downtime reduction in computer networks. Mechanical systems that require less upkeep and repair can result in the creation of their stronger parts. Medical devices, automotive systems, and aerospace are essential also maintaining dependable and effective procedures helps to avoid mishaps and save lives because failures can have disastrous results. Reliability is crucial in necessary applications including. Productivity and competitiveness in sectors like manufacturing and logistics is greatly boosted by high performance. Organizations may be significantly impacted financially by maintenance expenses, inefficiencies, and downtime. Therefore, system failures is related by lowering the expenses and inefficiencies, efficient reliability and performance management not only improves safety and user happiness but also helps organizations remain financially viable. Critical systems such as manufacturing processes, transportation networks, electricity grids, healthcare equipment, and communication systems depends on the efficient operation of modern society. Broader societal functions and economic stability are affected also these systems' malfunctions or inefficiency can cause serious disruptions but not only individual users.

A great deal of interest in the substantial effects of k-out-of-n systems and critical systems on quality, safety, and efficiency have revealed from a variety of disciplines and businesses in daily life by Scholars. Modeling, simulation, and empirical are the techniques applied in examine and improve system performance and reliability studied by these professionals. Creating systems was their main goal that exceed the strict requirements of their particular applications, guaranteeing efficiency, safety, and excellent results for stakeholders and users alike. In order to address the redundancy optimization of parallel k-out-of-n systems, for example, [1] presented a strong probabilistic and simulation-based methodology. Monte Carlo simulations and continuous-time Markov chains (CTMC) was their method use to improve system availability. The system running dependably to ensure the analysis creates the ideal number of operational elements needed. [2] Applied multivariate stochastic processes to analyze the reliability of complicated k-out-of-n systems with dependent components. Bayesian technique was used in deriving the system's credible intervals and reliability estimations. Sophisticated consecutive K-out-of-n: W CMO systems' signature reliability was examined in [3]. Universal generating function and structure-function framework were the tools used in determine the system's reliability function efficiently. [4] Proposed a method for calculating the time-dependent dependability of weighted and conventional k-out-of-n systems. This approach suggest a mathematical solution to assess dependability over time by considering the resistance deterioration, resistance correlation, and load non-stationarity. [5] In a series systems finding the solution to the redundancy optimization problem for k-out-of-n subsystems was concentrated. Optimizing the availability and configuration of the system used genetic algorithms and continuous-time Markov chains. The reliability of k-out-of-n:(G) systems under conflicting failure modes was examined in [6].

The dependability of the system was calculated using Monte Carlo simulations, and reliability functions for shock and cumulative shock models were derived using multivariate approaches [7]. Finite difference techniques was used to examine how two-parameter singly perturbed one-dimensional parabolic partial differential equations with a delay in the time variable can be solved. Creation and application of numerical methods was the main focuses of this research that guarantee accuracy and stability when managing the difficulties presented by time-delay effects and singular perturbations. Computational reliability and convergence features successfully captured by the strong gradients and boundary layers commonly associated with such differential equations through the use of a fitted operator approach proposed by the expands method [8]. Petri nets method used in investigating and evaluating the performance of complex manufacturing systems, were the function emphasizes in unpredictable situations [9] has out a thorough investigation of the ferrofluid flow over a flexible spinning disk kept at a constant temperature while taking magnetic

field-dependent viscosity into account. The resulting nonlinear coupled partial differential equations are converted into a system of coupled nonlinear ordinary differential equations using the well-known von Kármán transformation for rotating disks in order to enable numerical analysis. The modified equations are then numerically solved using the shooting method [10]. The economics and performance of a milk processing plant was assess by the petri net modeling technique [11]. Investigated the efficacy of a safety system that might encounter a sequence of cascading failures [12] examined the performance of a k-out-of-n redundant system under successive failures. In the framework of Itô calculus, the (4 + 1)-dimensional stochastic Fokas equation was examined [13] by some analytical methods [14] offer a reliable numerical method for resolving linear second-order Fredholm integro-differential equations with singular perturbations. Numerical method that comprises an exponentially fitted finite difference method for estimating the differential component of the equation in order to attain parameter-uniform accuracy was create by the authors. Moreover, the integral section is effectively approximated using the composite Simpson 1/3 rule, which guarantees better accuracy and stability of the numerical solution across a range of perturbation parameter values. Numerical method for solving linear second-order Fredholm integro-differential equations with singular perturbations is presented the stability by [14]. The problem of singular perturbations, were the work emphasis mainly in frequent result in abrupt gradients and boundary layer effects in the solution. The authors introduced a numerical technique that includes an exponentially fitted finite difference approach for estimating the differential component of the equation in order to attain parameter-uniform accuracy. Furthermore, Accuracy and stability of the numerical solution across a range of perturbation parameter values guarantees increased by the integral section is effectively approximated using the composite Simpson 1/3 rule [17] Examined the integration of the inverse Weibull distribution (IWD) with a dual hesitant fuzzy set (DHFS) theoretical method. In order to accurately simulate system failure rates at different levels—a crucial component of reliability analysis—the study suggested a methodology that makes use of both DHFS and IWD [18] created a Monte Carlo simulation method based on neuromorphic fuzzy sets to improve medical diagnostic decision-making in a fuzzy setting [19] successfully optimized transportation solutions in a Fermatean fuzzy environment by proposing an algorithm designed to handle three kinds of Fermatean fuzzy transportation problems [20] worked on improving solar photovoltaic (PV) system dependability and performance. In their study, so many tactics and technological advancements intended to increase the longevity of PV systems were examined, including the important related variables. Further, degradation events that can be much reduced the dependability of solar PV systems over time, such as environmental effects and material wear. To evaluate the performance of active parallel system that has two unit, where specialized repair machines manage corrective maintenance, where specialized repair machines manage corrective maintenance [21], formulate a compounded reliability model. The research develops mathematical models that predict reliability and performance indicators across a range of operational scenarios while considering repair rate, maintenance efficiency and failure rate. The models help in optimizing maintenance schedules by offering vital information about how various approaches in maintenance impact, system performance and dependability. The research also considered alternative design and operational improvements for repair machines. Using redundancy techniques is a crucial part of improving the dependability of such systems that would increase system capability and efficiency; contribute to more dependable and high-performance parallel systems utilized in industrial and engineering applications. The k out of n redundancy strategy stands out among these. Due to its practical significance and efficacy in improving system dependability, the k-out-of-n redundancy scheme, the significance of its practicality in which sub-systems are operationally deemed if at least k of n components are operating, is consequently the focus of research. For instance, the Copula approach to analyze each maintenance parameter [22] creates a model to evaluate the efficiency of the manufacturing system stage [23] conducted an in-depth analysis of the performance measures of a computer network system, which is structured as a series-parallel system incorporating four integrated subsystems. The system under study comprises two load balancers (LB), five web servers (WB), and three database replica servers (DBRS). Each of these components plays a critical role in ensuring the seamless operation and reliability of the network by employing the k-out-of-n redundancy approach to model the subsystems' reliability. By integrating these subsystems into a cohesive series-parallel architecture, some important performance metrics like system reliability, availability, and mean time to failure (MTTF) are developed and analyzed numerically [24] conducted a comprehensive study focusing on the reliability and performance measures of a complex system, which is organized into two subsystems arranged in a series configuration and supported by a switching device. For every sub-system if redundancy policy is implemented on them, the system's design aimed to guarantee excellence reliability and performance. Considering 2 out of 5 G policy govern subsystems 2. Availability, mean time to failure (MTTF) where accessed, system dependability and mean time to repair (MTTR). Their study shows how the switching device's functioning and the subsystems' redundancy rules affect the system's overall performance.

[25] delve into the performance modeling and assessment of a complex, repairable system composed of two subsystems arranged in series. Multiple identical units are included in the subsystems which comprise the units operating with a minimum of k out of n with an operation of k out of n:G units. Subsystem-1 in this system architecture

is made up of n units that follow the k-out-of-n: G policy. Subsystem-2 is made up of m units that follow the r-out-of-m: G policy. Numerous performance measurements, such as mean time to repair (MTTR), mean time to failure (MTTF), availability, and system dependability [26] examines several reliability metrics for a complex system made up of two subsystems with controllers connected in series is covered in their work, which is an intriguing option for certain design issues. Subsystem-2 has m units that work under the r-out-of-m: G; policy, whereas subsystem-1 has n units that work under the K-out-of-n: G; policy. It is expected that the failure rates of both subsystems would follow an exponential distribution. There are two sorts of distributions that can be repaired: general and Gumbel-Hougaard family copula distributions [27] present the dependability characteristics of a computer network system comprising distributed database servers, load balancers, and a centralized server set up as a series parallel system with three subsystems are discussed in their study. Subsystem 1: A load balancer (LB) is a device that channels requests to servers and receives requests or responses from servers. Subsystem 3 is the centralized distributed database server (CDDS), which handles all of the application data in DDS. Subsystem 2 is made up of three identical distributed database servers (DDS) I, II, and III that are connected in parallel and operate under the 2-out-of-3: G policy, which manages and stores some of the application data [28] initially used a copula technique to analyze and improve the performance of honeynet systems. The performance of any honeynet system can be categorized according to its profitability, dependability, and availability. Therefore, the purpose of this work was to examine how well a multi-state honeynet system performed in terms of availability, dependability, and predicted profit. Two kinds of fixes are examined in their paper. Type II repairs, often referred to as copula repairs, are used to recover from a complete or lethal failure to a perfect state, whereas Type I repairs, also known as general repairs, are used to recover from a partial or non-lethal failure [29] This paper looks at the reliability properties of a parallel system. The parallel system being looked at needs two of its three active units to be working for it to work. This study's primary goal is to measure and analyze the impact of both offline and online preventive maintenance. There are two methods for performing preventative maintenance on the systems: online and offline. Online preventative maintenance is carried out following the failure of each system's first unit. After the second unit of each system fails, offline preventative maintenance is done. There are two types of failures that might happen: partial and full. Both systems can fail and be fixed in an exponential way [30] increase RAMD, their study sought to improve the dependability, reliability, maintainability, availability, and metrics such as MTBF and MTTF of textile manufacturing systems. The textile system under study is a serial system with five subsystems: the weaving section (subsystem A), the dry clean section (subsystem B), the cross-cut section (subsystem C), the side seam section (subsystem D), and the cleaning section (subsystem E). The main unit, heated standby unit, and cold standby unit make up each subsystem. The system regulating the differential difference equation is assembled from the state-to-state transition diagram using the Markovian birth-death method for design and prediction.

The method and model utilized for simpler two subsystems are insufficient to handle array of issues brought up by their increased complexity. The intricate failure behavior and performance characteristics, which are difficult to predict and assess through conventional methods, may arise from the interconnections and interactions among different subsystems.

This gap serves as the impetus for the present study. The objective is to examine the reliability and performance improvements in serial systems that incorporate various subsystems through a k-out-of-n redundancy approach. In this work, dependency-aware copula modelling is used to analyze the influence of redundant components on systems dependability and efficiency. Through a detailed evaluation of subsystems interactions and backup structures, the study contributes guidance for designing more robust series systems.

This study combines k-out-of-n redundancy with copula-based and general repair techniques to provide a comprehensive framework for reliability modeling and performance evaluation of complex serial systems. The study employs numerical trials to derive key reliability indicators such as profit functions, availability, mean time to failure, and overall reliability, thereby assisting engineers, maintenance managers, and system designers in making informed decisions.

This work is organized in to the following sections: The notation used to analyze the suggested model is introduced in Section 2. The system's state is described in detail in Section 3. The performance and reliability models, along with considerations of particular scenarios, are presented in Section 4. In Section 5, these models are validated numerically, and in Section 6, the results are discussed. The findings presented in Section 7 serve as the study's final conclusion.

NOTATIONS

r: representing time variable.

s: representing variable of transformation of Laplace

ν_1: rate of failure representation of subsystem A

ν_2: rate of failure representation of subsystem B

ν_3: rate of failure representation of subsystem C

ν_4: rate of failure representation of subsystem D

ν_5: rate of failure representation of subsystem E

$T(x)/T_0(x)$: rate of repair representation for diminish performance/complete breakdown of subsystem A

$T(y)/T_0(y)$: rate of repair representation for diminish performance/complete breakdown failure of subsystem E

$T_0(m)$: rate of repair representation for complete breakdown of subsystem C

$T_0(z)$: rate of repair representation for complete breakdown of subsystem B

$T_0(n)$: rate of repair representation for complete breakdown of subsystem D

$Q_i(r)$: stand for chance of the equipment staying in any state at instants for i = 0 to 13

$\bar{Q}_0(s)$: representation for transformation of Laplace with probability $N(q)$

$Q_k(x,r)$: representation for chance of the system staying in any state with service duration is (x,r) with service variable x and time r

$Q_k(y,r)$: representation for chance of the equipment sojourning in any state for i = 1…….13, the with service duration is (y, r) with service y and time q

$E_p(r)$: profit anticipation profit in [0, r)

A_1, A_2: income and cost of service cost per unit time, respectively.

$T_0(x)$: representation of joint probability according to Gumbel-Hougaard family Copula definition is given as (failed state S_i to good state S_0)

SYSTEM DESCRIPTION

A k-out-of-n system is a common redundancy strategy in reliability engineering in which overall operation is maintained as long as at least k out of the n available components remain functional. The k-out-of-n architectures significantly improve system dependability by tolerating partial component loss while avoiding the cost and complexity associated with full duplication of all components. Specifically, the system continue to work whenever at least k out of the total number of n components are working. This type of redundancy introduces fault tolerance by allowing a limited number of component failures. Thus, the system can withstand up to n-k failed components without collapsing into total failure. This type of system can seen in some applications such as electric power systems, telecommunication networks, etc. This type of redundancy is particularly valuable in engineering applications where complete component availability is not essential for acceptable performance. Systems like communication networks, electrical power infrastructures, and safety-critical installations often implement k-out-of-n designs to attain an effective equilibrium among reliability, efficiency, and resource utilization. Permitting controlled failure without complete system shutdown enhances the economic viability of these configurations.

This study analyzes a series-configured system consisting of five interconnected subsystems, each of which is essential to the overall operation of the system. Because the subsystems are arranged in series, the failure of any single subsystem will result in a total system failure.

a. Subsystem A utilizes a k-out-of-n: G configuration comprising identical units. The subsystem continues to operate effectively as long as k units are active, ensuring built-in fault tolerance and improved reliability via redundancy.
b. Subsystem B is comprised of a singular unit, and its operational status is wholly contingent upon the performance of that unit.
c. Subsystem C is one unit subsystem whose failure lead to total system collapse. \
d. Subsystem D contain two units in parallel redundancy.
e. Subsystem E, similar to Subsystems B and C, comprises a single unit and does not incorporate redundancy. For the overall system to maintain functionality, it is essential that all five subsystems operate effectively. Any failure, whether in the redundant k-out-of-n structure of Subsystem A, the parallel configuration of Subsystem D, or the single-unit subsystems, leads to a total system failure.

S_0: Initial state, the subsystems and the system are in perfect state. The system is up and running.

S_1: One unit has failed in subsystem A, other units are up. The system is up and running.

S_2: The second unit in subsystem A failed, other units are up. The system is up and running.

S_3: The third unit has failed in subsystem A, other units are up. The system is up and running.

S_4: The first unit in subsystem E failed, other unit is up. The system is up and running in reduced capacity.

S_5: Previously the first unit in subsystem A failed, the first unit in subsystem E, others unit in subsystem A and one unit in subsystem E are up. The system is up and running in reduced capacity.

S_6: Previously the first unit in subsystem E failed, the first unit in subsystem A, others unit in subsystem A and one unit in subsystem E are up. The system is up and running in reduced capacity.

S_7: Previously the first unit in subsystem E and first unit in subsystem A failed, the second unit in subsystem A failed, others unit in subsystem A and one unit in subsystem E are up. The system is up and running in reduced capacity.

S_8: Previously the first unit in subsystem B and first and second unit in subsystem A failed, the third units in subsystem A failed, others unit in subsystem A and one unit in subsystem E are up. The system is up and running in reduced capacity.

S_9: All units in subsystem A failed. The system is down.

S_{10}: All units in subsystem E failed. The system is down.

S_{11}: Subsystem B failed. The system is down.

S_{12}: Subsystem C failed. The system is down.

S_{13}: Subsystem D failed. The system is down.

METHODOLOGY

It is a multi-step process to develop the partial differentiation equations for reliability and performance models for

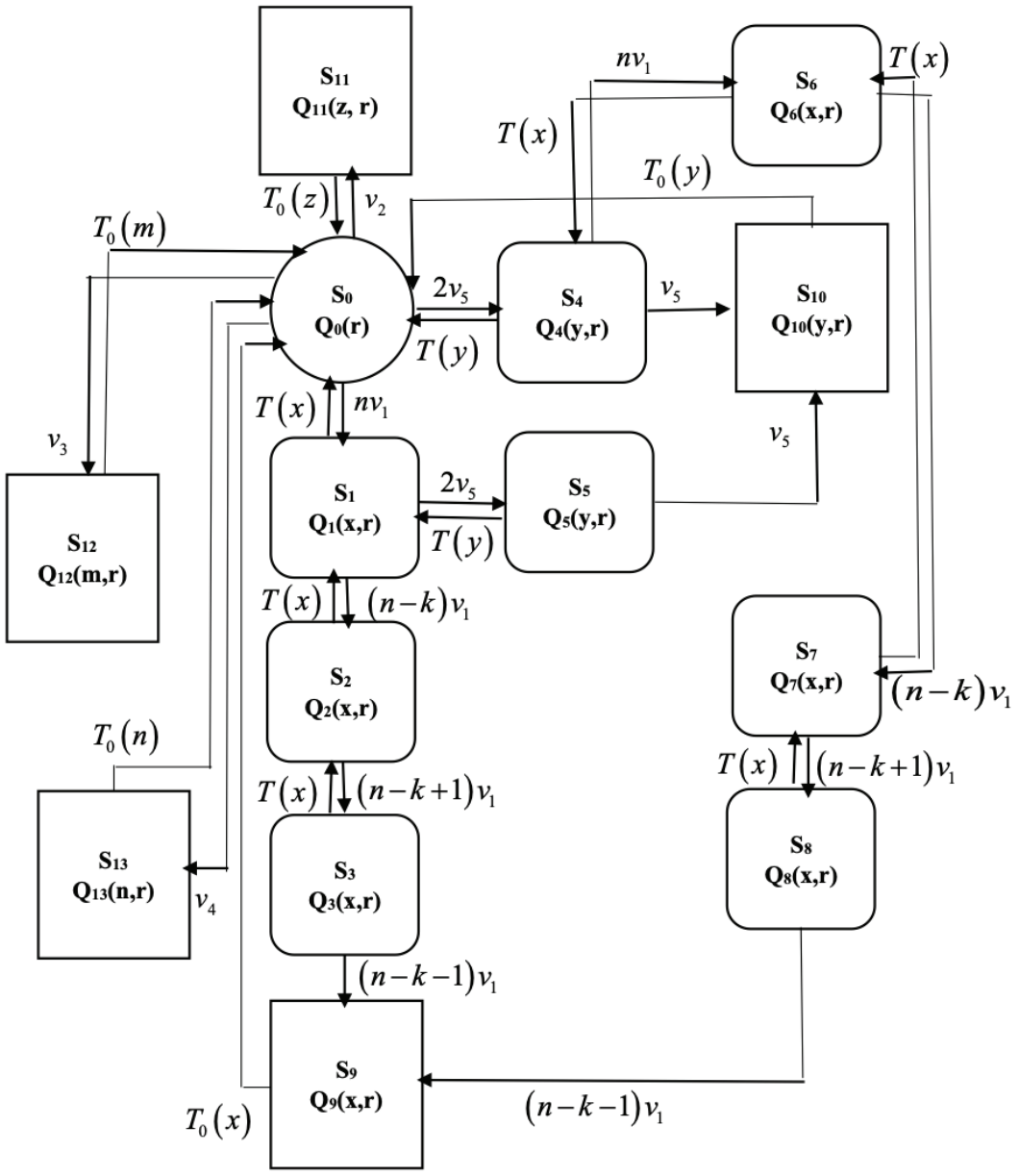


Figure 1. Transition diagram of the system.

the system using Figure 1 and solving them with Laplace transformation and the supplementary variable technique.

1. Clearly states all essential variables and parameters employed in this study. These include the structural configuration of the system, transition rates, transition probabilities, repair intensities, and unit failure rates.
2. Formulating reliability and performance models for the system based on the transition diagram in Figure 1 and resolving them through Laplace transformation and the supplementary variable technique. .
3. Develop Partial Differential Equations (PDEs) utilizing the transition diagram to compute transition rates, which indicate the likelihood of state changes. Develop partial differential equations that define the dynamics of state probabilities over time, articulating these rates as functions of the designated variables and parameters.
4. Utilize the Laplace Transformation: Transform the partial differential equations into the frequency domain through the application of the Laplace transformation. Transforming time-dependent derivatives into algebraic terms that incorporate the Laplace variable simplifies the equations.
5. Solve the Equations After Transformation: To determine state probabilities or other reliability and performance indicators, simplify and analyze the modified

equations. In this step, it is often necessary to rearrange algebraic expressions and solve for unknown variables.

6. The equations that have been transformed require resolution: To determine state probabilities or other indicators of reliability and performance, streamline and engage with the modified equations. This phase often requires addressing unknown variables and performing algebraic manipulations.

FORMULATION OF RELIABILITY MODELS AND THEIR SOLUTIONS

Laplace transforms and the additional variable technique were employed to develop dependability models for system modeling and analysis. The transition diagram facilitated the derivation of differential equations through a probabilistic method. The steady-state probabilities were then found by solving these differential equations using initial and boundary conditions. The creation of reliability models is based on these probability. The following partial differential equations were obtained from Figure 1:

$$\left\{\frac{\partial}{\partial r}+nv_1+v_2+v_3+v_4+v_5\right\}Q_0(r)=\int_0^\infty T(x)Q_1(x,r)dx + \int_0^\infty T(y)Q_4(y,r)dy+\int_0^\infty T_o(x)Q_9(x,r)dx + \int_0^\infty T_o(y)Q_{10}(y,r)dy+\int_0^\infty T_o(z)Q_{11}(z,r)dz + \int_0^\infty T_o(m)Q_{12}(m,r)dm+\int_0^\infty T_o(n)Q_{13}(n,r)dn \quad (1)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k)v_1+2v_5+T(x)\right\}Q_1(x,r)=0 \quad (2)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k+1)v_1+T(x)\right\}Q_2(x,r)=0 \quad (3)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k-1)v_1+T(x)\right\}Q_3(x,r)=0 \quad (4)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial y}+nv_1+v_5+T(y)\right\}Q_4(y,r)=0 \quad (5)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial y}+v_5+T(y)\right\}Q_5(y,r)=0 \quad (6)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k)v_1+T(x)\right\}Q_6(x,r)=0 \quad (7)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k+1)v_1+T(x)\right\}Q_7(x,r)=0 \quad (8)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+(n-k-1)v_1+T(x)\right\}Q_8(x,r)=0 \quad (9)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial x}+T_0(x)\right\}Q_9(x,r)=0 \quad (10)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial y}+T_0(y)\right\}Q_{10}(y,r)=0 \quad (11)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial z}+T_0(z)\right\}Q_{11}(z,r)=0 \quad (12)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial m}+T_0(m)\right\}Q_{12}(m,r)=0 \quad (13)$$

$$\left\{\frac{\partial}{\partial r}+\frac{\partial}{\partial n}+T_0(n)\right\}Q_{13}(n,r)=0 \quad (14)$$

Boundary Conditions

$$Q_1(x,r)=nv_1Q_0(r) \quad (15)$$

$$Q_2(x,r)=n(n-k)v_1^2Q_0(r) \quad (16)$$

$$Q_3(x,r)=n(n-k)(n-k+1)v_1^3Q_0(r) \quad (17)$$

$$Q_4(y,r)=2v_5Q_0(r) \quad (18)$$

$$Q_5(y,r)=2nv_1v_5Q_0(r) \quad (19)$$

$$Q_6(x,r)=2nv_1v_5Q_0(r) \quad (20)$$

$$Q_7(x,r)=2n(n-k)v_1^2v_5Q_0(r) \quad (21)$$

$$Q_8(x,r)=2n(n-k)(n-k+1)v_1^3v_5Q_0(r) \quad (22)$$

$$Q_9(x,r)=2n(n-k)(n-k+1)(n-k-1)v_1^4v_5Q_0(r) \quad (23)$$

$$Q_{10}(y,r)=2v_5^2(1+nv_1)Q_0(r) \tag{24}$$

$$Q_{11}(z,r)=v_2Q_0(r) \tag{25}$$

$$Q_{12}(m,r)=v_3Q_0(r) \tag{26}$$

$$Q_{13}(n,r)=v_4Q_0(r) \tag{27}$$

Solution

Laplace transforms

$$\begin{aligned}\{s+nv_1+v_2+v_3+v_4+v_5\}\overline{Q}_0(s)&=\int_0^\infty T(x)\overline{Q}_1(x,s)dx\\&+\int_0^\infty T(y)\overline{Q}_4(y,s)dy+\int_0^\infty T_o(x)\overline{Q}_9(x,s)dx\\&+\int_0^\infty T_o(y)\overline{Q}_{10}(y,s)dy+\int_0^\infty T_o(z)\overline{Q}_{11}(z,s)dz\\&+\int_0^\infty T_o(m)\overline{Q}_{12}(m,s)dm+\int_0^\infty T_o(n)\overline{Q}_{13}(n,s)dn\end{aligned} \tag{28}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k)v_1+2v_5+T(x)\right\}\overline{Q}_1(x,s)=0 \tag{29}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k+1)v_1+T(x)\right\}\overline{Q}_2(x,s)=0 \tag{30}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k-1)v_1+T(x)\right\}\overline{Q}_3(x,s)=0 \tag{31}$$

$$\left\{s+\frac{\partial}{\partial y}+nv_1+v_5+T(y)\right\}\overline{Q}_4(y,s)=0 \tag{32}$$

$$\left\{s+\frac{\partial}{\partial y}+v_5+T(y)\right\}\overline{Q}_5(y,s)=0 \tag{33}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k)v_1+T(x)\right\}\overline{Q}_6(x,s)=0 \tag{34}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k+1)v_1+T(x)\right\}\overline{Q}_7(x,s)=0 \tag{35}$$

$$\left\{s+\frac{\partial}{\partial x}+(n-k-1)v_1+T(x)\right\}\overline{Q}_8(x,s)=0 \tag{36}$$

$$\left\{s+\frac{\partial}{\partial x}+T_0(x)\right\}\overline{Q}_9(x,s)=0 \tag{37}$$

$$\left\{s+\frac{\partial}{\partial y}+T_0(y)\right\}\overline{Q}_{10}(y,s)=0 \tag{38}$$

$$\left\{s+\frac{\partial}{\partial z}+T_0(z)\right\}\overline{Q}_{11}(z,s)=0 \tag{39}$$

$$\left\{s+\frac{\partial}{\partial m}+T_0(m)\right\}\overline{Q}_{12}(m,s)=0 \tag{40}$$

$$\left\{s+\frac{\partial}{\partial n}+T_0(n)\right\}\overline{Q}_{13}(n,s)=0 \tag{41}$$

Boundary Conditions

$$\overline{Q}_1(x,s)=nv_1\overline{Q}_0(s) \tag{42}$$

$$\overline{Q}_2(x,s)=n(n-k)v_1^2\overline{Q}_0(s) \tag{43}$$

$$\overline{Q}_3(x,s)=n(n-k)(n-k+1)v_1^3\overline{Q}_0(s) \tag{44}$$

$$\overline{Q}_4(y,s)=2v_5\overline{Q}_0(s) \tag{45}$$

$$\overline{Q}_5(y,s)=2nv_1v_5\overline{Q}_0(s) \tag{46}$$

$$\overline{Q}_6(x,s)=2nv_1v_5\overline{Q}_0(s) \tag{47}$$

$$\overline{Q}_7(x,s)=2n(n-k)v_1^2v_5\overline{Q}_0(s) \tag{48}$$

$$\overline{Q}_8(x,s)=2n(n-k)(n-k+1)v_1^3v_5\overline{Q}_0(s) \tag{49}$$

$$\overline{Q}_9(x,s)=2n(n-k)(n-k+1)(n-k-1)v_1^4v_5\overline{Q}_0(s) \tag{50}$$

$$\overline{Q}_{10}(y,s)=2v_5^2(1+nv_1)\overline{Q}_0(s) \tag{51}$$

$$\overline{Q}_{11}(z,s)=v_2\overline{Q}_0(s) \tag{52}$$

$$\overline{Q}_{12}(m,s)=v_3\overline{Q}_0(s) \tag{53}$$

$$\overline{Q}_{13}(n,s)=v_4\overline{Q}_0(s) \tag{54}$$

$$\overline{Q}_0(s)=\frac{1}{H(s)} \tag{55}$$

$$\overline{Q}_1(s)=\frac{nv_1}{H(s)}\left\{\frac{1-\overline{S}_T(s+(n-k)v_1+2v_2)}{s+(n-k)v_1+2v_2}\right\} \tag{56}$$

$$\overline{Q}_2(s)=\frac{n(n-k)v_1^2}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+(n-k+1)v_1\right)}{s+(n-k+1)v_1}\right\} \quad (57)$$

$$\overline{Q}_3(s)=\frac{n(n-k)(n-k+1)v_1^3}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+(n-k-1)v_1\right)}{s+(n-k-1)v_1}\right\} \quad (58)$$

$$\overline{Q}_4(s)=\frac{2v_5}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+nv_1+v_5\right)}{s+nv_1+v_5}\right\} \quad (59)$$

$$\overline{Q}_5(s)=\frac{2nv_1v_5}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+v_5\right)}{s+v_5}\right\} \quad (60)$$

$$\overline{Q}_6(s)=\frac{2nv_1v_5}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+(n-k)v_1\right)}{s+(n-k)v_1}\right\} \quad (61)$$

$$\overline{Q}_7(s)=\frac{2n(n-k)v_1^2v_5}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+(n-k+1)v_1\right)}{s+(n-k+1)v_1}\right\} \quad (62)$$

$$\overline{Q}_8(s)=\frac{2n(n-k)(n-k+1)v_1^3v_5}{H(s)}\left\{\frac{1-\overline{S}_T\left(s+(n-k-1)v_1\right)}{s+(n-k-1)v_1}\right\} \quad (63)$$

$$\overline{Q}_9(s)=\frac{2n(n-k)(n-k-1)(n-k+1)v_1^4v_5}{H(s)}\left\{\frac{1-\overline{S}_T(s)}{s}\right\} \quad (64)$$

$$\overline{Q}_{10}(s)=\frac{2v_5^2(1+nv_1)}{H(s)}\left\{\frac{1-\overline{S}_T(s)}{s}\right\} \quad (65)$$

$$\overline{Q}_{11}(s)=\frac{v_2}{H(s)}\left\{\frac{1-\overline{S}_T(s)}{s}\right\} \quad (66)$$

$$\overline{Q}_{12}(s)=\frac{v_3}{H(s)}\left\{\frac{1-\overline{S}_T(s)}{s}\right\} \quad (67)$$

$$\overline{Q}_{13}(s)=\frac{v_4}{H(s)}\left\{\frac{1-\overline{S}_T(s)}{s}\right\} \quad (68)$$

Where

$$H(s)=\left(s+nv_1+v_2+v_3+v_4+2v_5-\Delta_0\right) \quad (69)$$

$$\Delta_0=\begin{Bmatrix} nv_1\overline{S}_T\left(s+(n-k)v_1+2v_5\right)+2v_5\overline{S}_T\left(s+nv_1+v_5\right)+ \\ 2n(n-k)(n-k+1)(n-k-1)v_1^4v_5\overline{S}_{T_0}(s)+ \\ 2v_5^2(1+nv_1)\overline{S}_{T_0}(s)+v_2\overline{S}_{T_0}(s)+v_3\overline{S}_{T_0}(s)+v_4\overline{S}_{T_0}(s) \end{Bmatrix} \quad (70)$$

$$\overline{Q}_{up}(s)=\overline{Q}_0(s)+\overline{Q}_1(s)+\overline{Q}_2(s)+\overline{Q}_3(s)+\overline{Q}_4(s)+\overline{Q}_5(s)+\overline{Q}_6(s)+\overline{Q}_7(s)+\overline{Q}_8(s) \quad (71)$$

$$\overline{Q}_{up}(s)=\frac{1}{H(s)}\begin{pmatrix} 1+nv_1\left\{\frac{1-\overline{S}_T\left(s+(n-k)v_1+2v_2\right)}{s+(n-k)v_1+2v_2}\right\}+\left\{\frac{1-\overline{S}_T\left(s+(n-k+1)v_1\right)}{s+(n-k+1)v_1}\right\} \\ +n(n-k)(n-k+1)v_1^3\left\{\frac{1-\overline{S}_T\left(s+(n-k-1)v_1\right)}{s+(n-k-1)v_1}\right\} \\ +2v_5\left\{\frac{1-\overline{S}_T\left(s+nv_1+v_5\right)}{s+nv_1+v_5}\right\}+\left\{\frac{1-\overline{S}_T\left(s+v_5\right)}{s+v_5}\right\} \\ +\left\{\frac{1-\overline{S}_T\left(s+(n-k)v_1\right)}{s+(n-k)v_1}\right\}+\left\{\frac{1-\overline{S}_T\left(s+(n-k+1)v_1\right)}{s+(n-k+1)v_1}\right\} \\ +\left\{\frac{1-\overline{S}_T\left(s+(n-k-1)v_1\right)}{s+(n-k-1)v_1}\right\} \end{pmatrix} \quad (72)$$

MODELING RELIABILITY CHARACTERISTICS AND OPTIMIZING PERFORMANCE

Availability Analysis Using Copula Repair

In the availability model for both general and copula repairs, the following scenarios are taken into account for consistency.

Fixing $S_{c_0}(s)=\bar{S}_{exp[x^\theta+\{log\,\varphi(x)\}^\theta]^{1/\theta}}(s)=\frac{exp[x^\theta+\{log\,\varphi(x)\}^\theta]^{1/\theta}}{s+exp[x^\theta+\{log\,\varphi(x)\}^\theta]^{1/\theta}}$, $\bar{S}_c(s)=\frac{c}{s+c}$, and taking failure rates are set at various values in equation (65), the availability equations for n = 15 and k = 5,9, and 11 respectively for copua repair using the inverse Laplace transform for the following cases:

Using failure rates as: v_1= 0.0101, v_2 = 0.0202, v_3 = 0.0303, v_4 = 0.0404, v_5 = 0.0505

Case 1 for n=15, k=5

$$\overline{Q}_{up}(t)=\begin{pmatrix} -0.003378e^{-1.05050t}+0.135117e^{-1.14140t} \\ +0.001299e^{-1.11110t}-0.000111e^{-1.09090t} \\ +0.036585e^{-2.89418t}-0.068227e^{-1.36240t} \\ -0.125105e^{-1.15176t}+1.023819e^{-0.02026t} \end{pmatrix} \quad (73)$$

Case 2 for n=15, k=9:

$$\overline{Q}_{up}(t)=\begin{pmatrix} -0.000968e^{-1.05050t}+0.000757e^{-1.07070t} \\ +0.082685e^{-1.10100t}+0.036832e^{-2.89409t} \\ -0.056733e^{-1.35515t}-0.077578e^{-1.12678t} \\ -1.015005e^{-0.01589t} \end{pmatrix} \quad (74)$$

Case 3 for n=15, k=11:

$$\overline{Q}_{up}(t)=\begin{pmatrix} 0.001636e^{-1.05050t}+0.074962e^{-1.08080t} \\ +0.036947e^{-2.89405t}-0.051843e^{-1.34693t} \\ -0.072161e^{-1.11349t}+1.010480e^{-0.01361t} \\ -0.000020e^{-1.03030t} \end{pmatrix} \quad (75)$$

Availability Analysis Using General Repair

In a similar way, we obtain expressions for system availability for general repair by substituting each of the cases considered in equation (72) and using Laplace transformation for a fixed value of n = 15 and k = 5,9, and 11 respectively for general repair as:

Case 1 for n=15, k=5:

$$\overline{Q}_{up}(t)=\begin{pmatrix}-0.006226e^{-1.05050t}+0.076913e^{-1.14140t}\\-0.016692e^{-1.43883t}-0.054757e^{-1.15953t}\\+0.027110e^{-1.02875t}+0.972568e^{-0.01927t}\\+0.001204e^{-1.11110t}-0.000120e^{-1.09090t}\end{pmatrix} \quad (76)$$

Case 2 for n=15, k=9:

$$\overline{Q}_{up}(t)=\begin{pmatrix}-0.001107e^{-1.05050t}-0.004748e^{-1.42816t}\\-0.024274e^{-1.14306t}+0.013049e^{-1.01963t}\\+0.965055e^{-0.01513t}+0.051346e^{-1.10100t}\\+0.000679e^{-1.07070t}\end{pmatrix} \quad (77)$$

Case 3 for n=15, k=11:

$$\overline{Q}_{up}(t)=\begin{pmatrix}-0.044862e^{-1.08080t}+0.001418e^{-1.05050t}\\+0.000415e^{-1.42346t}-0.014844e^{-1.13542t}\\+0.006976e^{-1.01395t}+0.961197e^{-0.01295t}\\-0.000025e^{-1.03030t}\end{pmatrix} \quad (78)$$

Table 1. Availability of the system for n=15 and k=5, 9, 11 under copula and general repair

Time	Availability					
	Copula repair			General repair		
	n=15, k=5	n=15, k=9	n=15, k=11	n=15, k=5	n=15, k=9	n=15, k=11
0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1	0.9906	0.9886	0.9876	0.9653	0.9633	0.9623
2	0.9797	0.9805	0.9809	0.9401	0.9408	0.9413
3	0.9626	0.9671	0.9695	0.9195	0.9238	0.9262
4	0.9439	0.9523	0.9568	0.9009	0.9089	0.9132
5	0.9251	0.9374	0.9439	0.8834	0.8949	0.9010
6	0.9065	0.9226	0.9312	0.8664	0.8813	0.8893
7	0.8884	0.9081	0.9186	0.8498	0.8680	0.9778
8	0.8705	0.8937	0.9062	0.8336	0.8550	0.8665
9	0.8531	0.8796	0.8939	0.8176	0.8421	0.8553
10	0.8360	0.8658	0.8818	0.8020	0.8295	0.8443

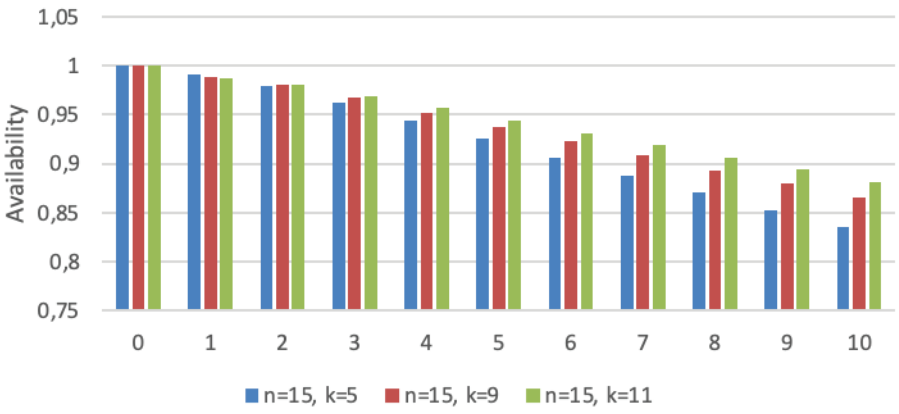


Figure 2. Graph of availability against time for n=15 and k=5, 9, 11 using copula repair.

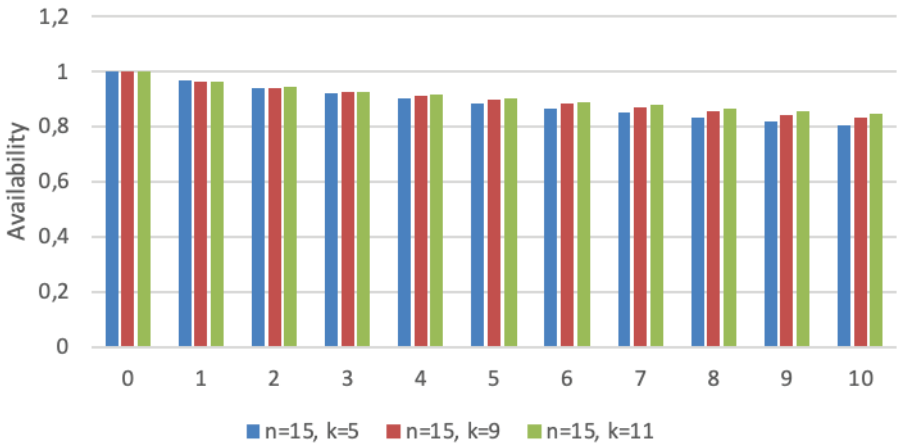


Figure 3. Graph of availability against time for n=15 and k=5, 9, 11 using general repair.

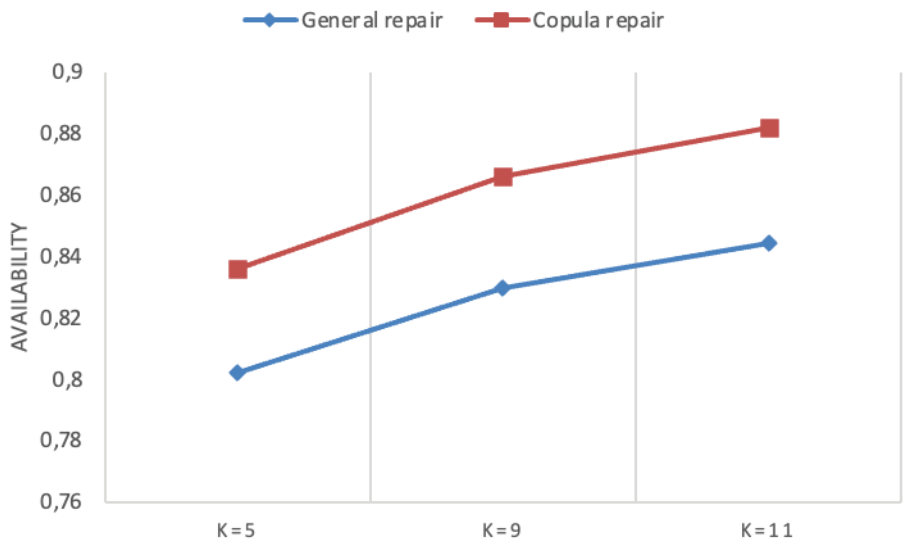


Figure 4. Availability comparison between Copula and General repair.

Reliability Analysis

In this situation, we set all system repairs to zero in equation (72), and applying inverse Laplace transform to give system's reliability. We obtain the following reliability expressions for three cases

Case 1 for n=15, k=5:

$$R(t) = \begin{pmatrix} -0.659587e^{-0.34340t} + 0.072522e^{-0.1111t} \\ + 0.007412e^{-0.09090t} + 0.063125e^{-0.10100t} \\ +0.052241e^{-0.05050t} + 0.714285e^{-0.20200t} \\ + 0.750000e^{-0.14140t} \end{pmatrix} \quad (79)$$

Case 2 for n=15, k=9:

$$R(t) = \begin{pmatrix} -0.485141e^{-0.34340t} + 0.0625000e^{-0.10100t} \\ + 0.037067e^{-0.07070t} + 0.054681e^{-0.05050t} \\ +0.714285e^{-0.20200t} + 0.054107e^{-0.06060t} \end{pmatrix} \quad (80)$$

Case 3 for n=15, k=11:

$$R(t) = \begin{pmatrix} -0.418044e^{-0.34340t} + 0.001086e^{-0.03030t} \\ + 0.576923e^{-0.08080t} + 0.050500e^{-0.04040t} \\ +0.075248e^{-0.05050t} + 0.714285e^{-0.20200t} \end{pmatrix} \quad (81)$$

Table 2. Reliability of the System for fixed n and k=5, 9, and 11

T	n=15, k=5	n=15, k=9	n=15, k=11
0	1.0000	1.0000	1.0000
1	0.9452	0.9419	0.9403
2	0.8733	0.8729	0.8729
3	0.7940	0.8002	0.8035
4	0.7138	0.7279	0.7355
5	0.6365	0.6586	0.6709
6	0.5641	0.5939	0.6107
7	0.4978	0.5344	0.5553
8	0.4380	0.4802	0.5050
9	0.3845	0.4314	0.4594
10	0.3372	0.3875	0.4183

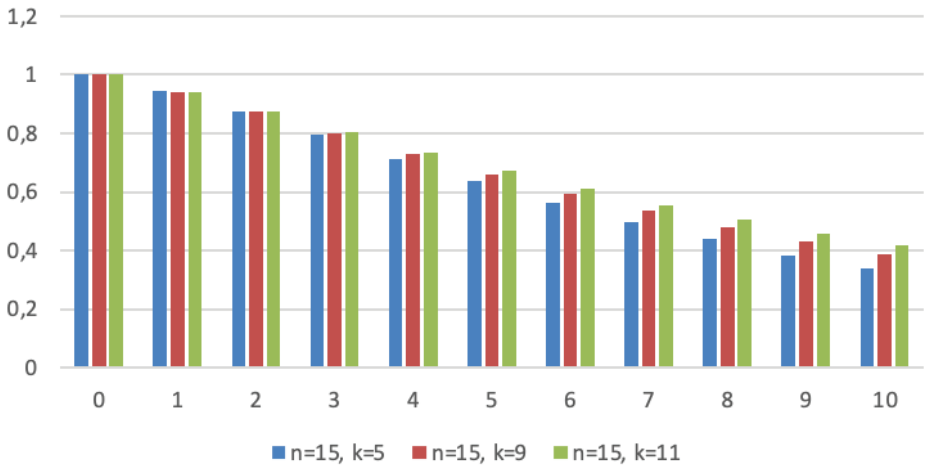


Figure 5. Graph of reliability against time for n=15 and k=5, 9, 11.

Mean Time To Failure Analisis

Supposing all repairs vanished to zero in (72). When *s* tends to zero, the MTTF can be evaluated using limit as follows:

$$MTTF = \lim_{s\to 0}\overline{Q}_{up}(s) = \frac{1}{\Delta(v)}\left(\begin{array}{l} 1+\dfrac{nv_1}{(n-k)v_1+2v_2}+\dfrac{n(n-k)v_1^2}{(n-k+1)v_1} \\ +\dfrac{n(n-k)(n-k+1)v_1^3}{(n-k-1)v_1}+\dfrac{2v_5}{nv_1+v_5} \\ +\dfrac{2nv_1v_5}{v_5}+\dfrac{2nv_1v_5}{(n-k)v_1}+\dfrac{2n(n-k)v_1^2v_5}{(n-k+1)v_1} \\ +\dfrac{2n(n-k)(n-k+1)v_1^3v_5}{(n-k-1)v_1} \end{array}\right) \quad (82)$$

$$\Delta(v) = nv_1 + v_2 + v_3 + v_4 + 2v_5$$

Cost Analysis Using Copula repair

The expression below is the expected profit for $t \geq 0$ is

$$E_p(t) = A_1\int_0^t Q_{up}(t)dt - A_2t \quad (83)$$

Case 1 for n=15, k=5:

$$E_p(t) = A_1\left\{\begin{array}{l} 0.003215e^{-1.05050t} - 0.118379e^{-1.14140t} \\ -0.012641e^{-2.89418t} + 0.050776e^{-1.36248t} \\ +0.108620e^{-1.15176t} - 50.521432e^{-0.02026t} \\ -0.001169e^{-1.11110t} + 0.000102e^{-1.09090t} \\ +50.49160769 \end{array}\right\} - A_2t \quad (84)$$

Case 2 for n=15, k=9:

$$E_p(t) = A_1\left\{\begin{array}{l} 0.000922e^{-1.05050t} - 0.000707e^{-1.07070t} \\ -0.012726e^{-2.89409t} + 0.041977e^{-1.35155t} \\ +0.068849e^{-1.12678t} - 63.839506e^{-0.01589t} \\ -0.075099e^{-1.10100t} + 63.81629150 \end{array}\right\} - A_2t \quad (85)$$

Table 3. MTTF of the System against Failure rate n=15 and k=5

Failure Rate	MTTF for V_1	MTTF for V_2	MTTF for V_3	MTTF for V_4	MTTF for V_5
0.0001	15.7910	9.6604	9.3600	9.6719	11.4034
0.0002	15.2521	9.6523	9.3571	9.6688	11.3954
0.0003	14.7878	9.6443	9.3541	9.6656	11.3873
0.0004	14.3843	9.6363	9.3511	9.6624	11.3793
0.0005	14.0307	9.6284	9.3481	9.6592	11.3713
0.0006	13.7187	9.6205	9.3451	9.6560	11.3633
0.0007	13.4415	9.6126	9.3421	9.6528	11.3553
0.0008	13.1938	9.6048	9.3392	9.6497	11.3474
0.0009	12.9710	9.5970	9.3362	9.6465	11.3394

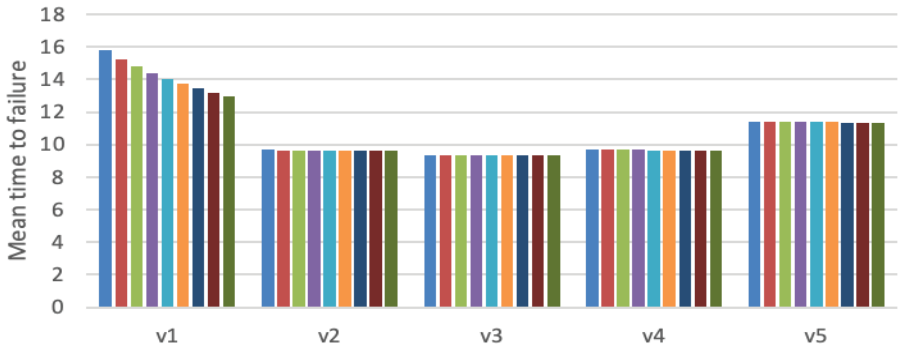


Figure 6. Graph of MTTF against failure rate for k=5.

Table 4. MTTF of the System against Failure rate n=15 and k=9

Failure Rate	MTTF for V_1	MTTF for V_2	MTTF for V_3	MTTF for V_4	MTTF for V_5
0.0001	16.3146	13.0047	11.0045	11.3712	13.1143
0.0002	15.7764	12.9795	11.0010	11.3675	13.1056
0.0003	15.3157	12.9545	10.9975	11.3637	13.0970
0.0004	14.9184	12.9296	10.9940	11.3600	13.0884
0.0005	14.5732	12.9049	10.9905	11.3562	13.0798
0.0006	14.2715	12.8804	10.9870	11.3525	13.0712
0.0007	14.0062	12.8560	10.9835	11.3488	13.0627
0.0008	13.7717	12.8319	10.9800	11.3450	13.0541
0.0009	13.5633	12.8078	10.9765	11.3413	13.0456

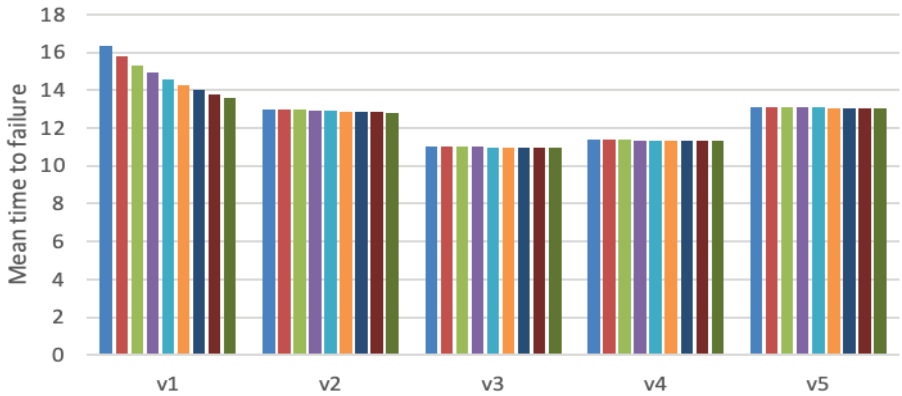


Figure 7. Graph of MTTF against failure rate for k=9.

Table 5. MTTF of the System against Failure rate n=15 and k=5

Failure Rate	MTTF for V_1	MTTF for V_2	MTTF for V_3	MTTF for V_4	MTTF for V_5
0.0001	16.9678	17.1922	12.5660	12.9848	14.6144
0.0002	16.4268	17.1344	12.5620	12.9805	14.6056
0.0003	15.9649	17.0772	12.5580	12.9762	14.5967
0.0004	15.5679	17.0205	12.5540	12.9719	14.5879
0.0005	15.2244	16.9644	12.5500	12.9676	14.5791
0.0006	14.9256	16.9089	12.5460	12.9634	14.5703
0.0007	14.6643	16.8539	12.5420	12.9591	14.5615
0.0008	14.4348	16.7995	12.5380	12.9548	14.5528
0.0009	14.2324	16.7456	12.5340	12.9506	14.5440

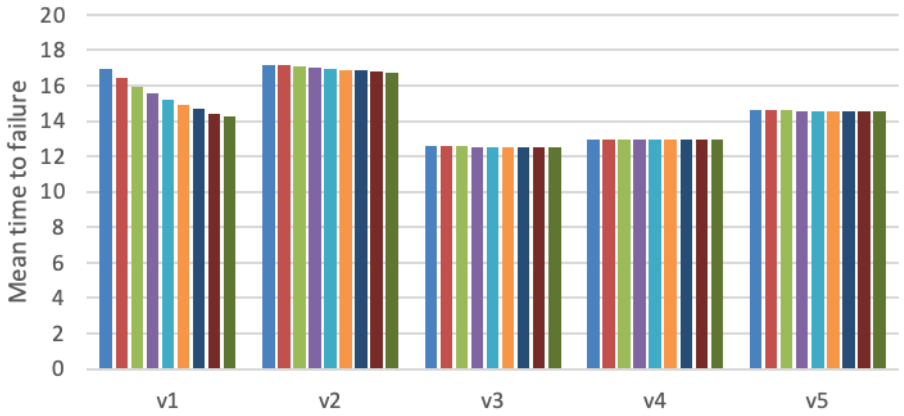


Figure 8. Graph of MTTF against failure rate for k=11.

Case 3 foe n=15, k=11:

$$E_p(t) = A_1 \begin{Bmatrix} 0.000020e^{-1.03030t} - 0.001557e^{-1.05050t} \\ -0.012766e^{-2.89405t} + 0.038490e^{-1.34693t} \\ +0.064806e^{-1.11349t} - 74.225387e^{-0.01361t} \\ -0.069357e^{-1.08080t} + 74.20575210 \end{Bmatrix} - A_2 t \quad (86)$$

Cost Analysis Using General Repair

Case 1: For n=15, k=5:

$$E_p(t) = A_1 \begin{Bmatrix} 0.005927e^{-1.05050t} - 0.067385e^{-1.14140t} \\ +0.011601e^{-1.43883t} + 0.047223e^{-1.15953t} \\ -0.026352e^{-1.02875t} - 50.461648e^{-0.01927t} \\ -0.001083e^{-1.11110t} + 0.000110e^{-1.09090t} \\ +50.49160769 \end{Bmatrix} - A_2 t \quad (87)$$

Case 2: For n=15, k=9:

$$E_p(t) = A_1 \begin{Bmatrix} 0.001054e^{-1.05050t} + 0.003325e^{-1.42816t} \\ +0.021236e^{-1.14306t} - 0.012798e^{-1.01963t} \\ -63.781838e^{-0.01513t} - 0.046635e^{-1.10100t} \\ -0.000634e^{-1.07070t} + 63.81629148 \end{Bmatrix} - A_2 t \quad (88)$$

Case 3: For n=15, k=11:

$$E_p(t) = A_1 \begin{Bmatrix} -0.041508e^{-1.08080t} - 0.001349e^{-1.05050t} \\ -0.000292e^{-1.142346t} + 0.013074e^{-1.13542t} \\ -0.006880e^{-1.01395t} - 74.168820e^{-0.01295t} \\ +0.000025e^{-1.03030t} + 74.20575204 \end{Bmatrix} - A_2 t \quad (89)$$

RESULTS AND DISCUSSION

Table 1and Figure 2 and Figure 3 present the results of availability over time for n=15 and K5,9 and 11units for copula and general repair. The analysis compared the effectiveness of two different repair and general repair. It is evident from the Tables and figures that the system's

Table 6. Profit of the system for n=15 and k=5 under copula and general repair

	Profit for Copula repair for k=5					Profit General repair for k=5				
Time	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**
0	0	0	0	0	0	0	0	0	0	0
1	0.8931	0.7931	0.6931	0.5931	0.4931	0.8815	0.7815	0.6815	0.5815	0.4815
2	1.7791	1.5791	1.3791	1.1791	0.9791	1.7337	1.5337	1.3337	1.1337	0.9337
3	2.6506	2.3506	2.0506	1.7506	1.4506	2.5633	2.2633	1.9633	1.6633	1.3633
4	3.5039	3.1039	2.7039	2.3039	1.9039	3.3734	2.9734	2.5734	2.1734	1.7734
5	4.3384	3.8384	3.3384	2.8384	2.3384	4.1656	3.6656	3.1656	2.6656	2.1656
6	5.1543	4.5543	3.9543	3.3543	2.7543	4.9405	4.3405	3.7405	3.1405	2.5405
7	5.9518	5.2518	4.5518	3.8518	3.1518	5.6986	4.9986	4.2986	3.5986	2.8986
8	6.7312	5.9312	5.1312	4.3312	3.5312	6.4403	5.6403	4.8403	4.0403	3.2403
9	7.4931	6.5931	5.6931	4.7931	3.8931	7.1659	6.2659	5.3659	4.4659	3.5659
10	8.2376	7.2376	6.2376	5.2376	4.2376	7.8758	6.8758	5.8758	4.8758	3.8758

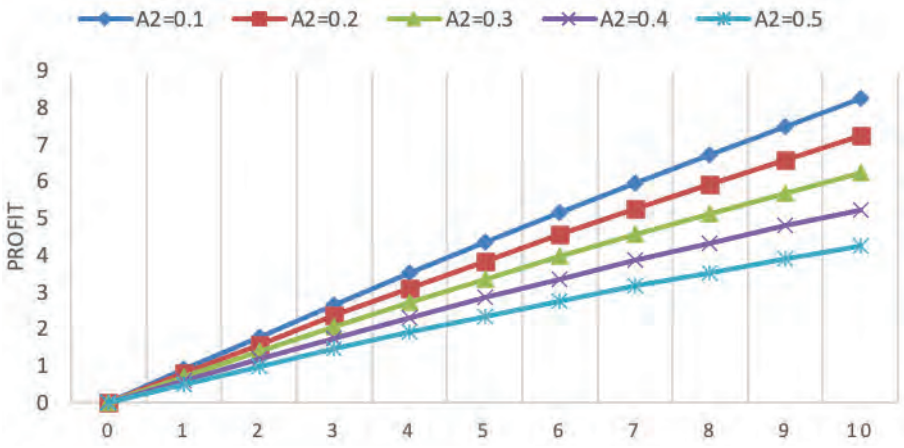


Figure 9. Graph of Profit against time for n=15 and k=5 using Copula repair.

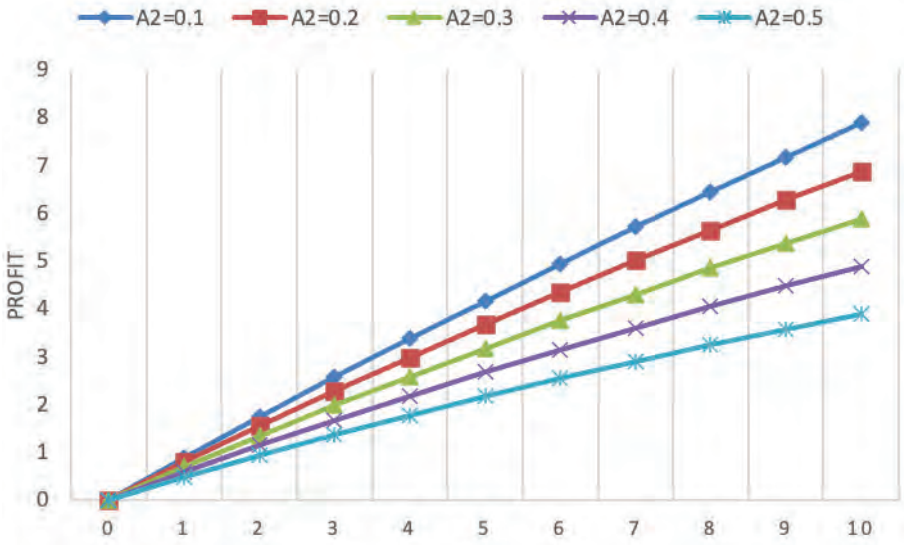


Figure 10. Graph of Profit against time for n=15 and k=5 for General repair.

Table 7. Profit of the system for n=15 and k=9 under copula and general repair

	Profit for Copula repair for k=9					Profit General repair for k=9				
Time	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**
0	0	0	0	0	0	0	0	0	0	0
1	0.8913	0.7713	0.6913	0.5913	0.4913	0.8798	0.7798	0.6798	0.5798	0.4798
2	1.7765	1.5765	1.3765	1.1765	0.9765	1.7312	1.5312	1.3312	1.1312	0.9312
3	2.6506	2.3506	2.0506	1.7506	1.4506	2.5633	2.2633	1.9633	1.6033	1.3633
4	3.5104	3.1104	2.7104	2.3104	1.9104	3.3796	2.9796	2.5796	2.1796	1.7796
5	4.3553	3.8553	3.3553	2.8553	2.3553	4.1815	3.6815	3.1815	2.6815	2.1815
6	5.1853	4.5853	3.9853	3.3853	2.7853	4.9696	4.3696	3.7696	3.1696	2.5696
7	6.0007	5.3007	4.6007	3.9007	3.2007	5.7443	5.0443	4.3443	3.6443	2.9443
8	6.8016	6.0016	5.2016	4.4016	3.6016	6.5059	5.7059	4.9059	4.1059	3.3059
9	7.5883	6.6883	5.7883	4.8883	3.9883	7.2545	6.3545	5.4545	4.5545	3.6545
10	8.3610	7.3610	6.3610	5.3610	4.3610	7.9903	6.9903	5.9903	4.9903	3.9903

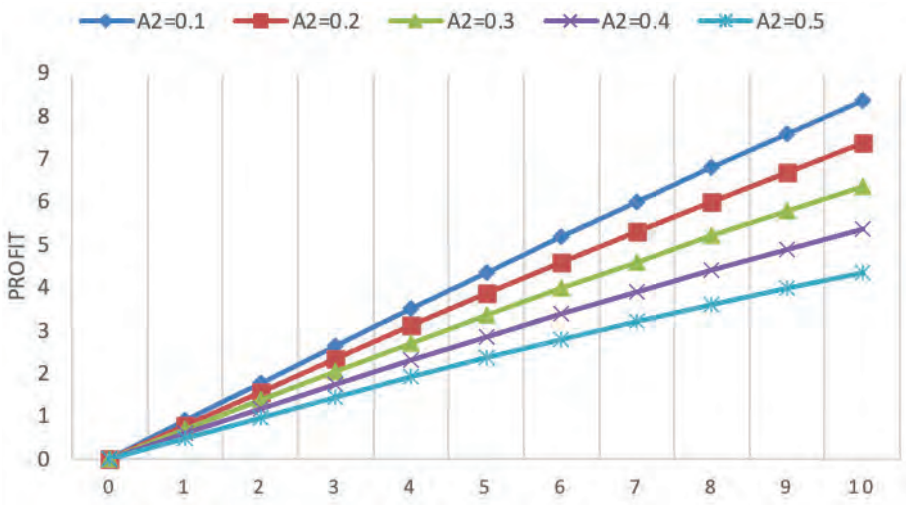


Figure 11. Graph of Profit against time for n=15 and k=9 for Copula repair.

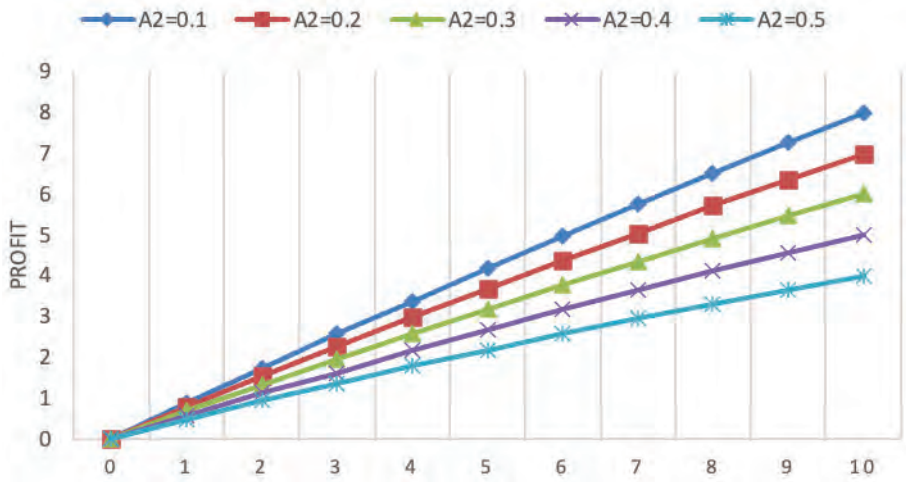


Figure 12. Graph of Profit against time for n=15 and k=9 for General repair

Table 8. Profit of the system for n=15 and k=11 under copula and general repair

	Profit for Copula repair for k=11					Profit General repair for k=11				
Time	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**	**A_2 =0.1**	**A_2 =0.2**	**A_2 =0.3**	**A_2 =0.4**	**A_2 =0.5**
0	0	0	0	0	0	0	0	0	0	0
1	0.8905	0.7905	0.6905	0.5905	0.4905	0.8790	0.7790	0.6790	0.5790	0.4790
2	1.7754	1.5754	1.3754	1.1754	0.9754	1.7301	1.5301	1.3301	1.1301	0.9301
3	2.6509	2.3509	2.0509	1.7509	1.4509	2.5636	2.2636	1.9636	1.6636	1.3636
4	3.5141	3.1141	2.7141	2.3141	1.9141	3.3832	2.9832	2.5832	2.1832	1.7832
5	4.3646	3.8646	3.3646	2.8646	2.3646	4.1903	3.6903	3.1903	2.6903	2.1903
6	5.2022	4.6022	4.0022	3.4022	2.8022	4.9855	4.3855	3.7855	3.1855	2.5855
7	6.0271	5.3271	4.6271	3.9271	3.2271	5.7691	5.0691	4.3691	3.6691	2.9691
8	6.8395	6.0395	5.2395	4.4395	3.6395	6.5413	5.7413	4.9413	4.1413	3.3413
9	7.6396	6.7396	5.8396	4.9396	4.0396	7.3022	6.4022	5.5022	4.6022	3.7022
10	8.4275	7.4275	6.4275	5.4275	4.4275	8.0521	7.0521	6.0521	5.0521	4.0521

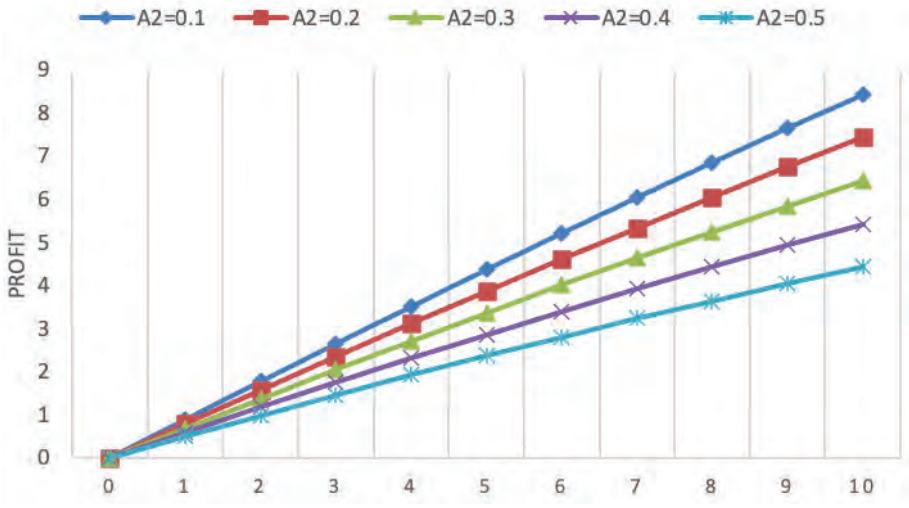


Figure 13. Graph of profit against time for n=15 and k=11 for copula repair

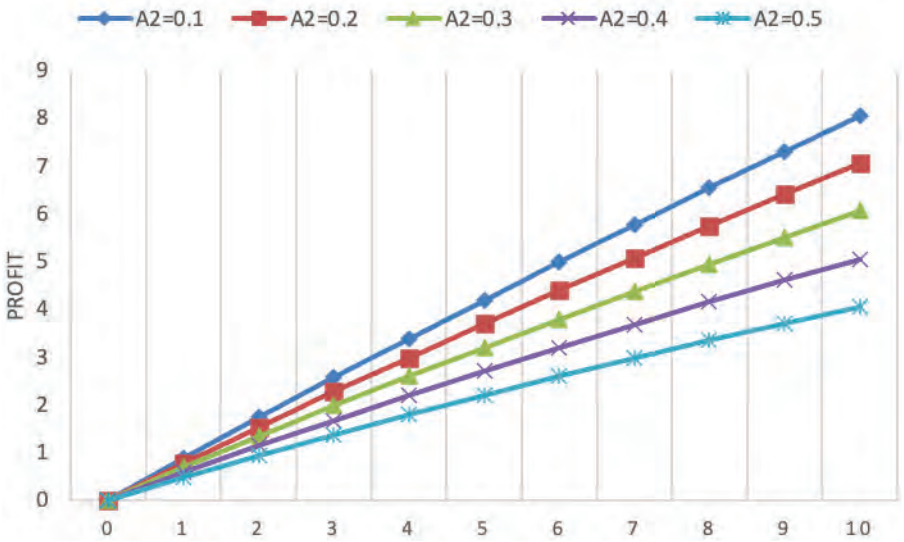


Figure 14. Graph of profit against time for n=15 and k=11 for general repair.

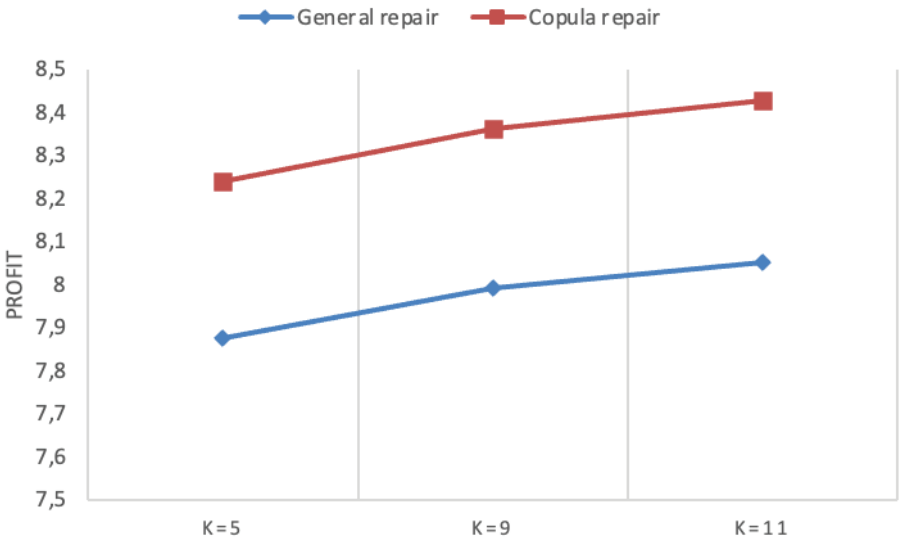


Figure 15. Maximum profit comparison between copula and general repair.

availability decreases as time progresses, irrespective of the repair strategy used or the value of k. However, the rate of decline in availability varies depending on the number of units required for operation. For both repair strategies, system availability is consistently higher when k=11 compared to k=5 and k=9. This suggests that an increased number of operational units, specifically k=11, improves system resilience and sustains higher availability over time. The configuration with k=11 demonstrates a greater tolerance to panel failures, as it can operate effectively even with up to four panels out of commission. Configurations with k5 or k=9 exhibit reduced resistance to panel failures, resulting in a more significant decline in availability.

The availability outcomes of the Copula repair and general repair procedures differ, as shown in the table and figures. The table and Figure 2, Figure 3 and Figure 4 make it evident that, for all values Copula repair method consistently produce better availability than that of general repair technique. The Copula repair method performs better because it can be more accurately simulate the dependability and interactions between units, which results in more effective maintenance and repair schedules. As a result, the system has greater sustained availability and is more dependable. By utilizing both the greater redundancy of higher k value and enhance efficiency of the Copula repair approach, this combination provides the highest and the most sustained availability (Fig. 5).

The influence of varying failure rates v1, v2, v3, v4, and v5 on the mean time to failure (MTTF) of subsystems A, B, C, D, and E is detailed in the findings presented in Table 3 and Figure 6 for k=5, Table 4 and Figure 7 for k=9, and Table 5 and Figure 8 for k=11. The tables and figures illustrate a distinct pattern: irrespective of the value of k, the mean time to Failure (MTTF) decreases as the failure rates v1, v2, v3, v4, and v5 increase. For a specified k, each table presents numerical data demonstrating the relationship between MTTF and increasing failure rates. Table 3 and Figure 6 focus on k=5, illustrating how an increased failure rate in any of the subsystems a through E leads to a reduced MTTF. The observed trend of diminishing MTTF alongside increasing failure rates is evident in Table 4 and Figure 7 for k=9, as well as in Table 5 and Figure 8 for k=11. Regardless of the subsystem failure rate, the data presented in the tables and figures clearly indicate that the MTTF is consistently higher when k=11 in comparison to k=5 and k=9. This indicates that an increased k enhances the overall MTTF by bolstering the system's resilience against failures in particular subsystems. The graphical representations (Figures 6, 7, and 8) effectively illustrate the decreasing trend of MTTF in relation to increasing failure rates for each k, thereby reinforcing the numerical data presented. The enhanced performance regarding MTTF for k=11 is clearly illustrated by these figures, where the bars for k=11 consistently exceed those for k=5 and k=9. Consequently, increasing k to 11 significantly enhances the MTTF, indicating a higher level of system robustness and reliability across various failure rates. Considering the diverse service costs (A2) that range from 0.1 to 0.5, the profit over time results are presented in Table 6 and Figure 9 for copula repair, as well as Figure 10 for general repair when k=5. Additionally, Table 7 and Figure 11 illustrate copula repair and Figure 12 depicts general repair when k=9. Finally, Table 8 and Figure 13 showcase copula repair, while Figure 14 represents general repair when k=11. It is evident from the tables and figures that profit increases over time for both copula and general repair procedures, irrespective of the value of k or the service cost (A2). This steady increasing trend shows that, under both repair procedures, the systems become more profitable over time. Lower service costs (A2) result in higher profits; for both repair procedures, the greatest earnings are shown at

A2=0.1. Reducing service expenses increases net profit by decreasing total maintenance and repair costs. The data presented in the tables and figures clearly demonstrate that, across all values of k and service costs, the copula repair method consistently generates higher profits compared to the general repair technique. Consequently, the copula repair technique, when paired with a low service cost (A2=0.1), yields the highest profit. This outcome highlights the importance of enhancing revenue through the reduction of service costs and the optimization of repair methods. The research presented in Figure 15 indicates that profit increases over time for both Copula and generic repair procedures. When service expenses are minimized, the Copula repair technique consistently yields greater profitability. This highlights the importance of efficient repair methods and cost management in enhancing the profitability of the systems. An important trade-off was revealed through the cost function analysis: lower expenses enhance profit, whereas higher service costs diminish overall system profitability. The results indicate that the copula-based repair policy outperforms the general repair approach regarding availability, mean time to failure, and reliability, offering valuable strategies for improving system performance. For organizations aiming to achieve an optimal balance between cost and performance, minimize failure rates, and improve operational processes, these findings carry significant implications. Industries can achieve sustainable and profitable system operations by leveraging the study's insights to make informed decisions regarding resource allocation, service pricing, and maintenance policies. The discussion effectively underscores the overarching trend in the data, indicating that system availability, reliability, and profit diminish over time, regardless of the repair approach or k-value employed. This provides the analysis with a solid and comprehensible foundation. The observation of greater k-values (k=11) leading to enhanced system resilience, along with increased availability, dependability, and profit, is quite insightful. This illustrates the way redundancy enhances system resilience and reliability, aligning with established principles of reliability engineering. Unlike continuous-time Markov chains (CTMCs), which often depend on memoryless assumptions (exponential distributions) for transition rates, the copula-based approach facilitates the modeling of more intricate dependencies and marginal distributions. This highlights a key advantage of your results and findings with the copula-based repair policy compared to methods such as continuous-time Markov chain, Monte Carlo simulation, Bayesian bootstrap method, generating function and structure-function approach, genetic algorithm, and multivariate analysis. This is particularly beneficial when dealing with systems that exhibit non-exponential characteristics, such as repair durations or failure occurrences that are not independent. The copula-based repair policy demonstrated superior performance in availability, dependability, and mean time to failure (MTTF) when contrasted with conventional methods. While comprehensive estimating methodologies are offered by approaches like the Bayesian bootstrap or Monte Carlo simulations, these techniques may not be specifically tailored to address complex dependency structures. While effective, generating function and structure-function techniques may not sufficiently capture component dependency patterns. Our copula-based approach enables a more realistic modeling of repair strategies and system behaviors, effectively quantifying and incorporating interdependencies.

CONCLUSION

This study used Copula and general repair technique, reliability modeling, and performance evaluation of complicated serial systems incorporating k out of n redundancy. The expression of reliability measurements such as mean time to failure, profit function, availability, and reliability are developed and verified numerically. The effect of growth of any manufacturing industry are metrics etc. It was found through the cost function analysis that lower system profit resulted from greater service costs and vice versa. The copula distribution repair policy improved system performance more than general repair according to mean time to failure, reliability, availability, and cost analysis. The research makes a contribution by modeling and assessing the system performance using a copula and general repair techniques. Businesses can utilize this data to reduce failure and improve their operational procedure. In order to increase profit while preserving system performance, managers might use this knowledge to inform decisions about resources allocation and service pricing. To improve our comprehension of system performance under various configurations or settings, future work could consider sophisticated techniques such as genetic algorithms, particle swarm optimization.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest with the regard to the manuscript

REFERENCES

[1] Oszczypała M, Konwerski J, Ziółkowski J, Małachowski J. Reliability analysis and redundancy optimization of k-out-of-n systems with random variable k using continuous time Markov chain and Monte Carlo simulation. Reliab Eng Syst Saf 2024;242:109780. [CrossRef]

[2] Oszczypała M, Ziółkowski J, Małachowski J, Oszczypała M, Ziółkowski J, Małachowski J. Redundancy allocation problem in repairable k-out-of-n systems with cold, warm, and hot standby: A genetic algorithm for availability optimization. Appl Soft Comput 2024;165:112041. [CrossRef]

[3] Naaz S, Khanna R, Rawat R. Signature reliability analysis of complex consecutive k-out-of-n:W CMO system via UGF and SFA. Proc Inst Mech Eng O J Risk Reliab 2024;238:1037-1052. [CrossRef]
[4] Wang C. Time-dependent reliability of (weighted) k-out-of-n systems with identical component deterioration. J Infrastruct Preserv Resil 2021;2:3. [CrossRef]
[5] Oszczypała M, Ziółkowski J, Małachowski J, Oszczypała M, Ziółkowski J, Małachowski J. Redundancy allocation problem in repairable k-out-of-n systems with cold, warm, and hot standby: A genetic algorithm for availability optimization. Appl Soft Comput 2024;165:112041. [CrossRef]
[6] Bian L, Wang G, Liu P. Reliability analysis for k-out-of-n(G) systems subject to dependent competing failure processes. Comput Ind Eng 2023;177:109084.
[7] Negero NT, Duressa GF, Rathour L, Mishra VN. A novel fitted numerical scheme for singularly perturbed delay parabolic problems with two small parameters. Partial Differ Equ Appl Math 2023;8:100546. [CrossRef]
[8] Kumar A, Kumar V, Modgil V, Kumar A, Sharma A. Performance analysis of complex manufacturing system using petri nets modeling method. J Phys Conf Ser 2021;1950:012061. [CrossRef]
[9] Meghashree GR, Asha CS, Prasad SK, Nargund AL, Rathour L, Singh V, et al. Study of magnetic field dependent viscosity and temperature on Jenkins model fluid flow over a rotating disk. Ann Math Comput Sci 2024;25:33-53. [CrossRef]
[10] Kumar N, Tewari PC, Sachdeva A. Performance assessment of milk processing system using petri nets. SN Appl Sci 2020;2:1488. [CrossRef]
[11] Xie L, Lundteigen MA, Liu Y. Performance assessment of K-out-of-N safety instrumented systems subject to cascading failures. ISA Trans 2021;118:35-43. [CrossRef]
[12] Xie L, Lundteigen MA, Liu Y. Performance analysis of safety instrumented systems against cascading failures during prolonged demands. Reliab Eng Syst Saf 2021;216:107975. [CrossRef]
[13] Mohammed WW, Cesarano C. The soliton solutions for the (4 + 1)-dimensional stochastic Fokas equation. Math Methods Appl Sci 2023;46:7589-7597. [CrossRef]
[14] Hogeme MS, Woldaregay MM, Rathour L, Mishra VN. A stable numerical method for singularly perturbed Fredholm integro differential equation using exponentially fitted difference method. J Comput Appl Math 2024;441:115709. [CrossRef]
[15] Sharma MK, Chaudhary S, Rathour L, Mishra VN. Modified genetic algorithm with novel crossover and mutation operator in travelling salesman problem. Sigma J Eng Nat Sci 2024;42:1876-1883. [CrossRef]
[16] Fuzhang W, Imtiaz A, Hijaz A, Alsulami MD, Alimgeer KS, Cesarano C, et al. Meshless method based on RBFs for solving three-dimensional multi-term time fractional PDEs arising in engineering phenomenons. J King Saud Univ Sci 2021;33. [CrossRef]
[17] Rathour L, Singh V, Yadav H, Sharma MK, Mishra VN. A dual hesitant fuzzy set theoretic approach in fuzzy reliability analysis of a fuzzy system. Inf Sci Lett 2024;13:433-440. [CrossRef]
[18] Sharma MK, Dhiman N, Kumar S, Rathour L, Mishra VN. Neutrosophic Monte Carlo simulation approach for decision making in medical diagnostic process under uncertain environment. Int J Neutrosophic Sci 2023;22:8-16. [CrossRef]
[19] Sharma MK, Bhargava SAK, Kumar S, Rathour L, Mishra LN, Pandey S. A fermatean fuzzy ranking function in optimization of intuitionistic fuzzy transportation problems. Adv Math Models Appl 2022;7:191-204.
[20] Alimi OA, Meyer EL, Olayiwola OI. Solar photovoltaic modules' performance reliability and degradation analysis—A review. Energies 2022;15:5964. [CrossRef]
[21] Lawan AM, Ali UA, Lado A, Isa MS, Maihulla AS, Yusuf I. Reliability and performance analysis of two unit active parallel system attended by two repairable machines. Int J Oper Res 2022;19:27-40.
[22] Maihulla AS, Yusuf I, Khoo MBC. Reliability and performance analysis on stages to control manufacturing system using Copula technique. Life Cycle Reliab Saf Eng 2023;12:197-205. [CrossRef]
[23] Poonia P. Performance assessment of a multi-state computer network system in series configuration using copula repair. Int J Reliab Saf 2021;15:68-88. [CrossRef]
[24] Singh VV, Poonia PK, Abdullahi AH. Performance analysis of a complex repairable system with two subsystems in series configuration with an imperfect switch. J Math Comput Sci 2020;10:359-383.
[25] Singh VV, Abubakkar MI, Ibrahim KH, Yusuf I. Performance assessment of the complex repairable system with n-identical units under k-out-of-n: G scheme and copula linguistic repair approach. Int J Qual Reliab Manag 2021;39:367-386. [CrossRef]
[26] Singh VV, Poonia PK. Stochastic analysis of k-out-of-n: G type of repairable system in combination of subsystems with controllers and multi repair approach. J Optim Ind Eng 2022;15:121-130.
[27] Yusuf I, Lado AI, Singh VV, Ali UA, Sufi NA. Performance analysis of multi computer system consisting of three subsystems in series configuration using copula repair policy. SN Comput Sci 2020;1:241. [CrossRef]

[28] Isa MS, Wu J, Yusuf I. Performance estimation of honeynet system for network security enhancement via copula linguistic. Sigma J Eng Nat Sci 2024;42:1169-1182. [CrossRef]

[29] Ayagi HI, Wan Z, Yusuf I, Sanusi A. Performance estimation of parallel system under online and offline preventive maintenance. Sigma J Eng Nat Sci 2024;42:1-13.

[30] Iggi AM, Yusuf I. RAMD analysis of mixed standby serial manufacturing system. Sigma J Eng Nat Sci 2024;42:1-17.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2026

Research Article

A comprehensive approach to decision making in uncertain environments using picture fuzzy hypersoft topology

Muhammad IMRAN HARL[1], Muhammad SAEED[1,*], Muhammad HARIS SAEED[2]

[1]*Department of Mathematics, University of Management and Technology, Lahore, Punjab, 54000 Pakistan*
[2]*Department of Chemistry, University of Management and Technology, Lahore, Punjab, 54000 Pakistan*

ARTICLE INFO

Article history
Received: 24 July 2024
Revised: 02 December 2024
Accepted: 08 January 2025

Keywords:
Bipolar Fuzzy Set; Decision Making; Fuzzy Set; Fuzzy Topology; Hypersoft Set; Soft Set

ABSTRACT

In this work we introduced picture fuzzy hypersoft topological spaces (PFHSTS) as an extension of intuitionistic fuzzy hypersoft topology to address vague, incomplete, and inconsistent information. PFHSTS is a novel concept for analyzing complex systems in uncertain environments by combining the strengths of picture fuzzy sets with the parameterization features of hypersoft sets. Some basic PFHSTS concepts such as closure, interior, and exterior are defined, and their properties are examined with the help of illustrative examples. Along with the theoretical development, a group decision-making model based on the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) is developed in the PFHST setting. This model is designed to handle decision problems involving imprecise and multi-dimensional information. A numerical example in the field of small firms and industries is included to demonstrate the practical implementation of the proposed method. In this study, a unified framework is proposed to extend existing fuzzy topological models by combining the concepts of picture fuzzy sets and hypersoft sets. The results show that PFHSTS deals uncertainty better and can help in making decisions in complex situations.

Cite this article as: Imran Harl M, Saeed M, Haris Saeed M. A comprehensive approach to decision making in uncertain environments using picture fuzzy hypersoft topology. Sigma J Eng Nat Sci 2026;44(2):1069–1081.

INTRODUCTION

The idea of "combinatorial innovation" has grown to be progressively more popular in the topical era. This idiom describes the narrative and amazing combination of existing opinions, methods, and technologies. Combinatorial innovation has grown to be a trendy technique be- cause it allows innovators to generate innovative goods and solutions by utilizing information and technologies that are currently in existence. Making decisions entails choosing the para- mount course of action from an assortment of options, and it is a crucial progression. It is an essential ability that people and organizations exploit to get from side to side in a multiplicity of situations and accomplish their goals. The procedure of forming decisions involves a num- ber of indispensable elements and crucial phases. These comprise decisive the issue at hand, obtaining

***Corresponding author.**
*E-mail address: muhammad.saeed@umt.edu.pk

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

important data, locating and assessing prospective alternatives, coming to a knowl- edgeable choice, carrying out the selected course of action, and assessing the consequences. Multi-criteria decision-making (MCDM) is an approach to handling thorny choice issues with several challenging elements. While evaluating options, MCDM takes numerous criteria or goals into a description. It seeks to suggest an organized and systematic approach to managing such multifarious decision situations, where a number of rudiments must be painstaking. It is important to remember that established mathematical methods have addressed a wide assortment of problems and have played a significant role in the development of mathematics. As mathematics advances and tackles gradually more multifaceted issues, the latest thoughts and methodologies are developed to be fruitful for and extend the adventurous approaches. In classical set theory, a component determines whether it is a member of a set based on a well-defined measure. In real-world situations, the link of an element in a set is frequently not binary or well-defined. Fuzzy set (FS) which were presented by Zadeh [1] in 1965 offer a distinguished basis for dealing with and modelling ambiguity and has a major impact on the field of mathematics. It has opened up new avenues for argument and assumption in situations where mathematical models were insufficient or inept. An extension of FS, the intuitionistic fuzzy set (IFS) was produced in 1986 by Krassimir Atanassov [2]. While IFS also describes uncertainty or indeterminacy, FS deals with delusion and biased membership. In the existent globe, IFS cannot handle paradoxical information despite having many victorious applications. For instance, there are four categories for voting outcomes: ”refrain,” ”vote against,” ”vote for,” and ”refuse to vote.”

To tackle this kind of problem, Cuong [3] presented a picture fuzzy set (PFS) in 2013. The picture fuzzy set is composed of the positive membership function, neutral membership function, and negative membership function. Correlation coefficients of PFS and their uses in clustering analysis were first presented by Sing [4] in 2014. This paper develops correlation measures specifically intended for PFS in order to investigate their use in clustering analysis. On the basis of the PFS domain, Son [5] and Thong [6] presented innovative fuzzy computations in 2015. These structures were used for temporal measurements and climate predive analysis applications. By adopting PFS-based fuzzy computations, the goal is to increase the precision and dependability of measurements and predictions in various domains. Picture fuzzy separation measures, generalised picture distance measures, and picture association measures were defined by Son [7, 8] in 2016. Under the PFS paradigm, grouping investigation concerns are addressed with these measures. The objective is to provide efficient clustering analysis techniques that can manage picture fuzzy data. An innovative fuzzy derivation framework on PFS was proposed by Son [9] in 2017. To handle complex data, Thong [10,11] used a revolutionary photo fuzzy clustering algorithm in 2016. A picture fuzzy aggregation operators approach was introduced by Wei [12] in 2017 and used for multi-attribute decision-making (MADM) ranking enterprise resource planning (ERP) structures.

A decision-making method based on the picture fuzzy weighted cross-entropy was studied in 2016 by Wei [13]. The suggested approach is used to rate options during decision-making. Soft set (SS), which deals with uncertainty, was introduced in 1999 by Molodtsov [14]. The fuzzy soft set (FSS), which is an extension of both FS and SS, was introduced by Maji [15] in 2001. Unlike a SS, which has clear parameters for each element, a FSS associates each element with a FS. Every element is assigned a membership degree by the FS, signifying its level of importance or membership in the FSS. The strength of membership is indicated by these membership degrees, which can vary from 0 to 1. The notion of intuitionistic fuzzy soft set (IFSS), which integrates the ideas of IFS and SS, was first introduced by Maji [16] in 2001. Instead of a crisp parameter or a FS in a IFSS, each element is associated with a FSS. The picture fuzzy soft set (PFSS), first presented by Yang [17] in 2015. In 2018, Smarandache [18] introduced the hypersoft concept set (HSS), which extends the notion of SS to handle situations where the characteristics of a group of parameters contain additional sub-attributes. He also gave the concept of fuzzy hypersoft sets (FHSS)and intuitionistic fuzzy hypersoft sets (IFHSS). In 2020, Saeed [19] contributed to the development of HSS*s* by introducing several fundamental concepts and operators. In 2021, Jafar [20] proposed FHSS and developed aggregation operators for these sets. They were aimed to extend the concept of HSS by incorporating aspects of fuzzy set theory and provide suitable aggregations methods for the FHSS hybrid. In 2021, Saeed et al. [21, 22] introduced the concept of complex multi-fuzzy hypersoft (CMFHSS) sets as a highly complex solution to multiple criteria decision-making (MCDM) problems where they utilized entropy and similarity measures for the evaluation of efficiency of SWMS (Solid Waste Management Systems). In 2021, Jafar [23] introduced the concept of matrix theory in interval-valued fuzzy hypersoft sets (IVFHSS) and proposed a decision making solution based on the designed mathematical structure. They applied this algorithm in the context of staff selection. In 2021, Saeed [24,25] suggested complex neutrosophic hypersoft (CNHSS) mappings, which combine the concepts of complex neutrosophic sets (NSS) and HSS. These mappings were utilized in diagnosing infectious disorders and hepatitis.

In 2021, Saqlain [26,27] expanded the concept of hypersoft sets (HSS) to include neutrosophic hypersoft sets (NHSS). He then used the TOPSIS method within this new framework, applying similarity measures. He looked at both single-valued and multi-valued NHSS and introduced tangent similarity measures for single-valued NHSS. In 2022, Rahman [28] used NHSS-FHSS in decision-making, providing examples to show how NHSS can be effectively used in real-world situations. In 2023, Saeed [29] introduced picture

fuzzy hypersoft sets (PFHSS), which included basic operations and were used in multi-attribute decision-making. Also in 2022, Mahmood [30] presented the T-bipolar soft set, which offers a better way to represent both positive and negative aspects of information compared to earlier bipolar soft set models. This method is especially useful when dealing with conflicting or dual perspectives in decision-making. In 2024, Majed Albaity [31] introduced a Complex hesitant fuzzy rough set for choosing the best data sources for integration in data science. That same year, Ubaid Ur Rehman [32] proposed bipolar fuzzy Aczel-Alsina prioritized aggregation operators, which help in selecting the best platforms for energy trading while effectively managing priorities and uncertainty. In 1968,Chang [33], developed fuzzy topology (FT) by modifying classical topology concepts to handle fuzziness and uncertainty. His work plays role for later studies and applications in mathematics and other areas. In 1995, Coker [34] extended the Chang's ideas by introducing intuitionistic fuzzy topology (IFT), In the context of IFT, the concepts of continuity, compactness, and connectedness have been studied and generalized to accommodate the IFS framework. In 2011, Cagman [35] defined the soft topology (ST), which is a generalization of the traditional topology that incorporates the notion of SS. In 2011, Tanay [36] developed the foundations of fuzzy soft topological spaces (FSTS) and investigated their properties and relationships with traditional topology and fuzzy set theory. In 2013, Osmanoglu [37] introduced intuitionistic fuzzy soft topology (IFSTS), also he further investigated the structure, properties, and applications of IFSTS. In 2022, Musa

[38] introduced hypersoft topology (HSTS), which is hybrid structure of HSS and ST. In 2023, Yolcu proposed the concept of fuzzy hypersoft topology (FHST) [39] and intuitionistic fuzzy hypersoft topology (IFHST) [40], also he investigated basic operations and results relating to these notions.

In 1981, Hwang [41] developed the Technique for Order of Preference by Similarity to ideal solution (TOPSIS) as a MCDM analysis. Boran [42] developed the TOPSIS technique based on intuitionistic fuzzy sets for multi-criteria decision-making issues. Furthermore, Selim [43] briefly described the fuzzy soft TOPSIS method as a multi-criteria decision-making technique. Based on the TOPSIS method, they suggested a group decision-making process for use in a soft, fuzzy environment. Many scholars in [44,45] also studied the TOPSIS method for resolving decision-making issues in various fuzzy environments.

MOTIVATION

Topological structures (TSs) on various existing theories of HSS have many applications in fields such as pattern recognition, decision-making, medical diagnosis, and image processing. These TSs provide a framework for analyzing and understanding the relationships and prop- erties of HSS in these domains.However, it is important to note that these TSs may face challenges when dealing with scenarios where each element of the universal set has inconsistent information.PFHSS handles inconsistent information in HSS environment so introducing topology on PFHSS is indeed crucial.

i. To examine the connectedness, compactness, separation, and other topological characteristics of PFHSS by introducing topology to it.This analysis assists in a more systematic and thorough understanding of the PFHSS's structure and behaviour.
ii. Topology on PFHSS allows for a more precise way to represent and analyze vague or imprecise information. It helps in challenging situations how PFHSS components connect and influence each other, making it easier to choose wisely best options.
iii. Weak topological structures can handle uncertainty but in absence of union, which is important in modeling and decision-making. PFHST improve this by including not just membership and non-membership degrees, but also neutrality and refusal degrees. These extra features offer a more dependable approach for sub attributive Decisions and other tasks that needed for choosing accurate information.

The key purpose to propose the initiative of PFHST on PFHSS, and to deal with some basic operations such as PFHS-interior, PFHS-closure and PFHS-exterior. Furthermore, we employ the PFHT in a TOPSIS-based group decision-making approach in a PFHSS environment.

ORGANIZATION OF THE PAPER

The structure of the rest of the paper is outlined as follows:

Section 2: This section provides a brief review of some fundamental concepts which are related to PFHST.

Section 3: The concept of PFHST is introduced on PFHSS, and several basic definitions are presented in this section.

Section 4: This section presents an algorithm that is based on an extension of the TOPSIS approach and a group decision-making mechanism. A numerical example is also presented to show how the algorithm is used.

Section 5: To illustrate the efficacy of the suggested structure in contrast to the current structures, a comparative analysis is carried out.

Section 6: It summarizes the findings and contributions of the research. Additionally, recommendation for future work are provided, suggesting potential avenues for further exploration and improvement in the discussed areas.

PRELIMINARIES

Definition 2.1. ([2]) An IFS on universal set $U = \{u_1,u_2,\ldots\ldots\ldots..u_n\}$ is defined as:

$$Z = \{m(u_k), n(u_k) \mid u_k \in U\}$$

Where $m_Z(u_k)$: $U \to [0,1]$ indicates the degree of membership for $u_k$in Z, $n_Z(u_k)$: $U \to [0,1]$ indicates the degree of non membership for u_k in Z, *s.t* $0 \le m(u_k) + n(u_k) \le 1$. $h(u_k) = 1 - m(u_k) - n(u_k)$ indicates the degree of hesitancy for u_k in Z.

Definition 2.2. ([3]) An PFS on universal set $U = \{u_1, u_2, \ldots\ldots\ldots u_n\}$ is defined as:

Where $m_V(u_k)$: $U \to [0,1]$ indicates the degree of membership for u_k in V, $t_V(u_k)$: $U \to [0,1]$ indicates the neutral degree for u_k in V, $n_V(u_k)$: $U \to [0,1]$ indicates the degree of non membership for u_k in V, *s.t* $0 \le m(u_k) + t(u_k) + n(u_k) \le 1$. $r(u_k) = 1 - m(u_k) - t(u_k) - n(u_k)$ indicates the refusal degree for u_k in Z.

Definition 2.3. (([14]) $\boldsymbol{P}$(U) be the power set of U and T be a attributes set. When D: T → $\boldsymbol{P}$(U) Then, the pair (D, T) is a SS over U.

Definition 2.4. ([17]) A function D: T → $\boldsymbol{P}$(PF(U), where $\boldsymbol{P}$(PF(U)), is power set of PFS over $\boldsymbol{U}$ then pair (D, T) is called PFSS.

Definition 2.5. ([18]). Suppose $\boldsymbol{P}$(U) the power set of U. Let T = $\{g_1, g_2, \ldots\ldots\ldots g_n\}$ be n sets of distinct parameters, each of whose associated attribute values are S_1, S_2,…… S_n. Suppose $S = S_1 \times S_2 \times \ldots\ldots\ldots \times S_n$., with $S_p \cap S_q = \emptyset$, and p, q ∈ {1, 2....., n}. The pair (F, S), where F: $S \to \boldsymbol{P}(U)$ is called a HSS over U.

Definition 2.6. ([18]). Suppose $\boldsymbol{P}$(PF(U)), is power set of PFS over $\boldsymbol{U}$. Let T = $\{g_1, g_2, \ldots\ldots\ldots g_n\}$ be n sets of distinct parameters, each of whose associated attribute values are S_1, S_2,……S_n. Suppose $S = S_1 \times S_2 \times \ldots\ldots\ldots \times S_n$., with $S_p \cap S_q = \emptyset$,, and p, q ∈ {1, 2....., n}. The pair (F, S), where F: $S \to \boldsymbol{P}$(PF(U)), is called a PHSS over U.

Definition 2.7. ([38]) If Γ_H is a collection of HSS over $\boldsymbol{U}$, then Γ_H is a HSST on $\boldsymbol{U}$ if:

(i) (Φ, Ψ, S), (Ψ, Φ, S) belong to Γ_H.

(ii) The intersection of any two HSSs in Γ_H belongs to Γ_H.

(iii) The union of any number of HSSs in Γ_H belongs to Γ_H.

Then, (U, Γ_H, S) is called a hypersoft topological space HSTS over U.

PICTURE FUZZY HYPERSOFT TOPOLOGY

Here, we proposed the innovative idea of picture fuzzy hypersoft topology (PFHST) which is hybrid structure of PFHSS and HST. The basic operation with relavent examples are discussed in this section.

Definition 3.1. A PFHSS (F, S) over U is said to be a relative null PFHSS denoted by (Φ, S) if for all $c_k \in S$, $\Phi(c_k) = < 0, 0, 1 >$.

Definition 3.2. A PFHSS (F, S) over U is said to be a relative whole PFHSS denoted by (Ψ, S) if for all $c_k \in S$, $\Psi(c_k) = < 0, 0, 1 >$.

Definition 3.3. Let Γ_{PFH} be a collection of picture fuzzy hypersoft sets over U, then Γ_{PFH} is a PFHTS on U if:

(i) (Φ, S), (Ψ, S) belong to Γ_{PFH}.

(ii) The intersection of any two picture fuzzy hypersoft sets in Γ_{PFH} belongs to Γ_{PFH}.

(iii) The union of any number of picture fuzzy hypersoft sets in Γ_{PFH} belongs to Γ_{PFH}.

The triplet (U, Γ_{PFH}, S) is called picture fuzzy hypersoft topological space over U.

Definition 3.4. The members of Γ_{PFH} are said to be picture fuzzy hypersoft open sets in U.

Example 3.5. Suppose $U = \{u_1, u_2, u_3, u_4\}$ be universe. Let $T = \{g_1, g_2, g_3\}$ be three distinct parameters, each of whose associated attribute values are S_1, S_2, S_3.

$$S_1 = \{b_{11}, b_{12}, b_{13}, b_{14}\}, S_2 = \{b_{21}, b_{22}, b_{23}, b_{24}\}, \quad S_3 = \{b_{31}, b_{32}, b_{33}, b_{34}\}$$

then

$$S = \{c_1 = (b_{11}, b_{21}, b_{32}, b_{44}), c_2 = (b_{14}, b_{24}, b_{33}, b_{43}), \quad c_3 = (b_{12}, b_{23}, b_{31}, b_{41})\}.$$

Consider the following picture fuzzy hypersoft sets

$$(\Phi, S) = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.0, 0.0, 1.0\rangle}{u_1}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_2}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_3}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.0, 0.0, 1.0\rangle}{u_1}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_2}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_3}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.0, 0.0, 1.0\rangle}{u_1}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_2}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_3}, \frac{\langle 0.0, 0.0, 1.0\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\},$$

$$(\Psi, S) = \left\{ \begin{array}{l} \left(\left[\frac{\langle 1.0, 0.0, 0.0\rangle}{u_1}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_2}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_3}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 1.0, 0.0, 0.0\rangle}{u_1}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_2}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_3}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 1.0, 0.0, 0.0\rangle}{u_1}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_2}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_3}, \frac{\langle 1.0, 0.0, 0.0\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\},$$

$$(\mathrm{F}, S)^1 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3, 0.1, 0.3\rangle}{u_1}, \frac{\langle 0.4, 0.2, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.4, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.1, 0.3\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.3, 0.1, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.3\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.2, 0.3, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right). \end{array} \right\},$$

$$(\mathrm{F}, S)^2 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.5, 0.2, 0.1\rangle}{u_1}, \frac{\langle 0.5, 0.3, 0.1\rangle}{u_2}, \frac{\langle 0.3, 0.5, 0.1\rangle}{u_3}, \frac{\langle 0.4, 0.3, 0.2\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.5, 0.2, 0.2\rangle}{u_1}, \frac{\langle 0.6, 0.2, 0.1\rangle}{u_2}, \frac{\langle 0.3, 0.4, 0.1\rangle}{u_3}, \frac{\langle 0.5, 0.2, 0.2\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.4, 0.4, 0.1\rangle}{u_1}, \frac{\langle 0.6, 0.2, 0.1\rangle}{u_2}, \frac{\langle 0.5, 0.3, 0.2\rangle}{u_3}, \frac{\langle 0.5, 0.3, 0.1\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$(\mathrm{F}, S)^3 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.5, 0.1, 0.1\rangle}{u_1}, \frac{\langle 0.5, 0.2, 0.1\rangle}{u_2}, \frac{\langle 0.3, 0.4, 0.1\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.2\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.5, 0.1, 0.2\rangle}{u_1}, \frac{\langle 0.6, 0.1, 0.1\rangle}{u_2}, \frac{\langle 0.3, 0.3, 0.1\rangle}{u_3}, \frac{\langle 0.5, 0.1, 0.2\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.4, 0.3, 0.1\rangle}{u_1}, \frac{\langle 0.6, 0.1, 0.1\rangle}{u_2}, \frac{\langle 0.5, 0.2, 0.2\rangle}{u_3}, \frac{\langle 0.5, 0.1, 0.1\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$(\mathrm{F}, S)^4 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3, 0.2, 0.3\rangle}{u_1}, \frac{\langle 0.4, 0.3, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.5, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.3, 0.3\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.3, 0.1, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.3\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.2, 0.3, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

It can be seen that $\Gamma_{PFH} = \{(\Phi, S)(\Psi, S), (\mathrm{F}, S)^1, (\mathrm{F}, S)^2, (\mathrm{F}, S)^3, (\mathrm{F}, S)^4\}$ is PFHST over U.

Definition 3.7. The $\Gamma_{PFH} = \{(\Phi, S), (\Psi, S)\}$ is a PFHSS on U, it is called picture fuzzy hypersoft indiscrete topology on U. This type of topology is denoted by $[\Gamma_{PFH}]^i$.

Definition 3.8. If Γ_{PFH} is a collection of all picture fuzzy hypersoft subsets over U called picture fuzzy hypersoft discrete topology on U. It is denoted by $[\Gamma_{PFH}]^d$.

Definition 3.9. A PFHSS (F, S) over U is said to be a picture fuzzy hypersoft closed set over U if (F, S)^C belongs to Γ_{PFH}.

Proposition 3.10. Consider two PFHST (U, $[\Gamma_{PFH}]^1$, S) and (U, $[\Gamma_{PFH}]^2$, S) then the followings are valid

i. Are comparable if $[\Gamma_{PFH}]^1 \subseteq [\Gamma_{PFH}]^2$
ii. $[\Gamma_{PFH}]^1 \subseteq [\Gamma_{PFH}]^2$, $[\Gamma_{PFH}]^1$ is coarser than $[\Gamma_{PFH}]^2$, $[\Gamma_{PFH}]^2$ is stronger than $[\Gamma_{PFH}]^1$.

Proposition 3.11. The intersection of two (U, $[\Gamma_{PFH}]^1$, S) and (U, $[\Gamma_{PFH}]^2$, S) is also PFHSTS.

Proof:

(i) $\because$ $[\Gamma_{PFH}]^1$ and $[\Gamma_{PFH}]^2$ be two PFHSTs then (Φ, D) and (Ψ, D) belong to $[\Gamma_{PFH}]^1 \cap [\Gamma_{PFH}]^2$

(ii) Suppose $(\Psi, S)_t \in [\Gamma_{PFH}]^1 \cap [\Gamma_{PFH}]^2$ then $(\Psi, S)_k \in [\Gamma_{PFH}]^1$ this implies $\cap^n_{k=1} (\Psi, S) \in [\Gamma_{PFH}]^1$ also $(\Psi, S)_k \in [\Gamma_{PFH}]^2$ this implies $\cap^n_{k=1} (\Psi, S) \in [\Gamma_{PFH}]^1 \cap [\Gamma_{PFH}]^2$.

(iii) Suppose $(\Psi, S)_t \in [\Gamma_{PFH}]^1 \cap [\Gamma_{PFH}]^2$ then$(\Psi, S)_k \in [\Gamma_{PFH}]^1$ this implies $\cup^n_{k=1} (\Psi, S) \in [\Gamma_{PFH}]^1$ also $(\Psi, S)_k \in [\Gamma_{PFH}]^2$ this implies $\cup^n_{k=1} (\Psi, S) \in [\Gamma_{PFH}]^1 \cap [\Gamma_{PFH}]^2$.

Remarks 3.11. The unoin of two (U, $[\Gamma_{PFH}]^1$, S) and (U, $[\Gamma_{PFH}]^2$, S) need not to be PFHSTS.

Theorem 3.11. Consider a PFHSTS topological space (U, $[\Gamma_{PFH}]^1$, S) over U, then

(i) (Φ, S), (Ψ, S) both are PFHS closed.
(ii) The arbitrary intersection of PFHS closed sets is PFHS closed.
(iii) The finite union of PFHS closed sets is closed.

Proof:

(i) Since $(\Phi, S)^C = (\Psi, S)$ and $(\Psi, S)^C = (\Phi, S)$ so both are PFHSS closed.
(ii) $(\Psi, S)_k$ is closed PFHSS so $(((\Psi, S)_k)^c)^c \in \Gamma_{PFH}$ thus we have $\cup ((\Psi, S)_k)^c \in \Gamma_{PFH}$ therefore $\cup((\Psi, S)_k)^c)^c) = \cap (\Psi, S)_k$
(iii) For each $(\Psi, S)_k$ is closed therefore $((\Psi, S)_k)^c \in \Gamma_{PFH}$ this implies $\cap (\Psi, S)_k \in \Gamma_{PFH}$ hence $\cap (((\Psi, S)_k)^c)^c) = \cup (\Psi, S)_k$ is closed.

Definition 3.11. Suppose that (U, $[\Gamma_{PFH}]$, S) is a PFHSTS over U then closure of (U, $[\Gamma_{PFH}]$, S) is denoted by $[\Gamma_{PFH}]^{clr}$ and defined as the intersection of all picture fuzzy hypersoft closed supersets of U. It is seen that closure of PFHSS is the smallest subset of PFHSS collection.

Example 3.12. It is clear from example 3.5

$$\Gamma_{PFH} = \{(\Phi, S)^c, (\Psi, S)^c, [(\mathrm{F}, S)^1]^c, [(\mathrm{F}, S)^2]^c, [(\mathrm{F}, S)^3]^c, [(\mathrm{F}, S)^4]^c\}$$

$$[(\mathrm{F}, S)^1]^c = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.1, 0.1, 0.3\rangle}{u_1}, \frac{\langle 0.3, 0.2, 0.4\rangle}{u_2}, \frac{\langle 0.3, 0.4, 0.1\rangle}{u_3}, \frac{\langle 0.3, 0.1, 0.2\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.4, 0.1, 0.3}{u_1}, \frac{\langle 0.3, 0.1, 0.5\rangle}{u_2}, \frac{\langle 0.5, 0.3, 0.1\rangle}{u_3}, \frac{\langle 0.3, 0.1, 0.4\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.4, 0.3, 0.2\rangle}{u_1}, \frac{\langle 0.3, 0.1, 0.5\rangle}{u_2}, \frac{\langle 0.6, 0.2, 0.1\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$[(\mathrm{F}, S)^2]^c = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.1, 0.2, 0.5\rangle}{u_1}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_2}, \frac{\langle 0.1, 0.5, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.3, 0.4\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.2, 0.2, 0.5\rangle}{u_1}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_2}, \frac{\langle 0.1, 0.4, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.2\ 0.5\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.1, 0.4, 0.4\rangle}{u_1}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_2}, \frac{\langle 0.2, 0.3, 0.5\rangle}{u_3}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$[(\mathrm{F}, S)^3]^c = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.1, 0.1, 0.5\rangle}{u_1}, \frac{\langle 0.1, 0.2, 0.5\rangle}{u_2}, \frac{\langle 0.1, 0.4, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.1, 0.4\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.2, 0.1, 0.5\rangle}{u_1}, \frac{\langle 0.1, 0.1, 0.6\rangle}{u_2}, \frac{\langle 0.1, 0.3, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.1, 0.5\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.1, 0.3, 0.4\rangle}{u_1}, \frac{\langle 0.1, 0.1, 0.6\rangle}{u_2}, \frac{\langle 0.2, 0.2, 0.5\rangle}{u_3}, \frac{\langle 0.1, 0.1, 0.5\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$[(\mathrm{F}, S)^4]^c = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3, 0.2, 0.3\rangle}{u_1}, \frac{\langle 0.3, 0.3, 0.4\rangle}{u_2}, \frac{\langle 0.3, 0.5, 0.1\rangle}{u_3}, \frac{\langle 0.3, 0.3, 0.2\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.4, 0.1, 0.3\rangle}{u_1}, \frac{\langle 0.3, 0.1, 0.5\rangle}{u_2}, \frac{\langle 0.5, 0.3, 0.1\rangle}{u_3}, \frac{\langle 0.3, 0.1, 0.4\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.4, 0.3, 0.2\rangle}{u_1}, \frac{\langle 0.3, 0.1, 0.5\rangle}{u_2}, \frac{\langle 0.6, 0.2, 0.1\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

Suppose $[(\mathrm{F}, S)^5]$ is PFHSS over U

$$(\mathrm{F}, S)^5 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3, 0.1, 0.3\rangle}{u_1}, \frac{\langle 0.4, 0.2, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.4, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.1, 0.3\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.3, 0.1, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.3\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.2, 0.3, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

According to the definition $[(\mathrm{F}, S)^5]^{clr} = [(\mathrm{F}, S)^3]^c$

Theorem 3.13. If (F, S)^1 and (F, S)^2 are two PFHSSs then

i. (F, S)$\subseteq$ (F, S)^{clr}
ii. if (F, S)$^1 \subseteq$ (F, S)^2 then $[(\mathrm{F}, S)^1]^{clr} \subseteq [(\mathrm{F}, S)^2]^{clr}$
iii. $[(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr} = [(\mathrm{F}, S)^1]^{clr} \cup [(\mathrm{F}, S)^2]^{clr}$
iv. $[(\mathrm{F}, S)^{clr}]^{clr} = (\mathrm{F}, S)$

Proof

(i) is obvious.
(ii) Since (F, S)$^1 \subseteq$ (F, S)^2 from (i) (F, S)$^1 \subseteq [(\mathrm{F}, S)^1]^{clr}$ and (F, S)$^2 \subseteq [(\mathrm{F}, S)^2]^{clr}$ (F, S)$^1 \subseteq [(\mathrm{F}, S)^2]^{clr}$ (F, S)$^1 \subseteq [(\mathrm{F}, S)^1]^{clr}$ Hence $[(\mathrm{F}, S)^1]^{clr} \subseteq [(\mathrm{F}, S)^2]^{clr}$
(iii) Since (F, S)$^1 \subseteq$ (F, S)$^1 \cup$(F, S)^2,(F, S)$^2 \subseteq$ (F, S)$^1 \cup$ (F, S)^2 and (F, S) $\subseteq$ (F, S)^{clr} then ((S, $[(\mathrm{F}, S)^1]^{clr} \subseteq [(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr}$ and $[(\mathrm{F}, S)^2]^{clr} \subseteq [(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr}$, $[(\mathrm{F}, S)^1]^{clr} \cup [(\mathrm{F}, S)^2]^{clr} \subseteq [(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr}$ also $[(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr} \subseteq [(\mathrm{F}, S)^1]^{clr} \cup [(\mathrm{F}, S)^2]^{clr}$ Hence $[(\mathrm{F}, S)^1 \cup (\mathrm{F}, S)^2]^{clr} = [(\mathrm{F}, S)^1]^{clr} \cup [(\mathrm{F}, S)^2]^{clr}$
(iv) From (i), (F, S) $\subseteq$ (F, S)^{clr} then $[(\mathrm{F}, S)]^{clr} \subseteq [(\mathrm{F}, S)^{clr}]^{clr}$, $[(\mathrm{F}, S)^{clr}]^{clr} =$ (F, S) $\subseteq$ (F, S)^{clr}, then $[(\mathrm{F}, S)^{clr}]^{clr} \subseteq$ (F, S)^{clr} Consequently, we have $[(\mathrm{F}, S)^{clr}]^{clr} =$ (F, S)

Definition 3.14. Consider a PFHSTS topological space (U, $[\Gamma_{PFH}]$, S) over U, then interior of PFHSS is denoted by $[\Gamma_{PFH}]^{int}$ and defined as the union of all PFHS open subsets of U.

Example 3.15. It is clear from example 3.5 Suppose (F, S)^6 be PFHSS

$$(\mathrm{F}, S)^6 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3, 0.1, 0.3\rangle}{u_1}, \frac{\langle 0.4, 0.2, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.4, 0.3\rangle}{u_3}, \frac{\langle 0.2, 0.1, 0.3\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.3, 0.1, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.3, 0.5\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.3\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.2, 0.3, 0.4\rangle}{u_1}, \frac{\langle 0.5, 0.1, 0.3\rangle}{u_2}, \frac{\langle 0.1, 0.2, 0.6\rangle}{u_3}, \frac{\langle 0.4, 0.1, 0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

$$[(\mathrm{F}, S)^6]^{int} = (\mathrm{F}, S)^3$$

Definition 3.16. Consider a PFHSTS topological space (U, $[\Gamma_{PFH}]$, S) over U, then interior of PFHSS is denoted by $[\Gamma_{PFH}]^{ext}$ defind as $[\Gamma_{PFH}]^{ext} = [(\Gamma_{PFH})^c]^{int}$.

Example 3.17. Using example 3.5 Suppose $(F, S)^7$ is PFHS subset over U.

$$(F,S)^7 = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3,0.1,0.3\rangle}{u_1}, \frac{\langle 0.4,0.2,0.3\rangle}{u_2}, \frac{\langle 0.1,0.4,0.3\rangle}{u_3}, \frac{\langle 0.2,0.1,0.3\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.3,0.1,0.4\rangle}{u_1}, \frac{\langle 0.5,0.1,0.3\rangle}{u_2}, \frac{\langle 0.1,0.3,0.5\rangle}{u_3}, \frac{\langle 0.4,0.1,0.3\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.2,0.3,0.4\rangle}{u_1}, \frac{\langle 0.5,0.1,0.3\rangle}{u_2}, \frac{\langle 0.1,0.2,0.6\rangle}{u_3}, \frac{\langle 0.4,0.1,0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

Then

$$[(F,S)^7]^c = \left\{ \begin{array}{l} \left(\left[\frac{\langle 0.3,0.1,0.3\rangle}{u_1}, \frac{\langle 0.3,0.2,0.4\rangle}{u_2}, \frac{\langle 0.3,0.4,0.1\rangle}{u_3}, \frac{\langle 0.3,0.1,0.2\rangle}{u_4} \right], \{c_1\} \right), \\ \left(\left[\frac{\langle 0.4,0.1,0.3\rangle}{u_1}, \frac{\langle 0.3,0.1,0.5\rangle}{u_2}, \frac{\langle 0.5,0.3,0.1\rangle}{u_3}, \frac{\langle 0.3,0.1,0.4\rangle}{u_4} \right], \{c_2\} \right), \\ \left(\left[\frac{\langle 0.4,0.3,0.2\rangle}{u_1}, \frac{\langle 0.3,0.1,0.5\rangle}{u_2}, \frac{\langle 0.6,0.2,0.1\rangle}{u_3}, \frac{\langle 0.4,0.1,0.4\rangle}{u_4} \right], \{c_3\} \right) \end{array} \right\}$$

This implies $[(F, S)^7]^{ext} = (F, S)^2$

Theorem 3.18. If $(F, S)^1$and $(F, S)^2$ belongs to ΓPFH, then

(i) $(F, S)^{int} \subseteq (F, S)$

(ii) If $(F, S)^1 \subseteq (F, S)^2$ then $((F, S)^1)^{int} \subseteq ((F, S)^2)^{int}$

(iii) $((F, S)^1 \cap (F, S)^2)^{int} = (F, S)^1)^{int} \cap (F, S)^1)^{int}$

(iv) $(F, S)^{int})^{int} = (F, S)$

Proof (i) is obvious.

(ii) Given $(F, S)^1 \subseteq (F, S)^1$ From (i) $((F, S)^1)^{int} \subseteq (F, S)^1$ and $((F, S)^2)^{int} \subseteq (F, S)^2$. $\Longrightarrow ((F, S)^1)^{int} \subseteq (F, S)^1 \subseteq (F, S)^2 \Longrightarrow ((F, S)^1)^{int} \subseteq (F, S)^2$ but $((F, S)^2)^{int} \subseteq (F, S)^2$. Hence $((F, S)^1)^{int} \subseteq ((F, S)^2)^{int}$

(iii) $\because (F, S)^1 \cap (F, S)^2 \subseteq (F, S)^1$ and $(F, S)^1 \cap (F, S)^2 \subseteq (F, S)^2$, from (i) $((F, S)^1)^{int} \subseteq (F, S)^1$ implies $(F, S)^1 \cap (F, S)^2 \subseteq ((F, S)^1)^{int}$ and $(F, S)^1 \cap (F, S)^2 \subseteq ((F, S)^2)^{int}$. $(F, S)^1 \cap (F, S)^1 \subseteq (F, S)^1)^{int} \cap (F, S)^1)^{int}$ $(F, S)^1)^{int} \cap (F, S)^1)^{int} \subseteq (F, S)^1 \cap (F, S)^2$. Also intersection of interior of $(F, S)^1$ and interior of $(F, S)^1$ is subset of intersection of $(F, S)^1$ and $(F, S)^1$. Therefore intersection of interior of $(F, S)^1$ and interior of $(F, S)^1$ is open subset of intersection of $(F, S)^1$ and $(F, S)^1$. Hence $((F, S)^1 \cap (F, S)^2)^{int} = (F, S)^1)^{int} \cap (F, S)^1)^{int}$

(iv) From (i), it follows $(F, S)^{int})^{int} \subseteq (F, S)^{int}$. For any open set (F, S) s.t $(F, S) \subseteq (F, S)^{int}$., $(F, S) = (F, S)^{int}$. $\subseteq (F, S)^{int})^{int}$, so $(F, S)^{int} \subseteq (F, S)^{int})^{int}$ Consequently, we have $(F, S)^{int} = (F, S)$

APPLICATION OF PICTURE FUZZY HYPERSOFT TOPOLOGY IN DECISION MAKING

In this part, we apply the picture PFHST under the specified conditions to the group decision-making (GDM) process in the PFHSS environment. We provided the TOPSIS technique concept and its integration into the suggested PFHTS.

Proposed Algorithm with TOPSIS Method

Consider about a MADM procedure that includes a specific collection of alternatives U = $\{u_1, u_2, u_3 \ldots\ldots\ldots u_n\}$. Each alternative is assessed using a unique set of attributes denoted by T = $\{g_1, g_2, g_3, \ldots\ldots\ldots g_m\}$ each of which has a sub attribute value. Suppose $(SDM)^1$, $(SDM)^2$, $(SDM)^3$, ..., $(SDM)^n$ be n different experts. Each expert has assessed the provided options and provided their assessments using linguistic phrases like Superb, wonderful, etc. This study takes into account all of the linguistic factors and their related weights from the list that is condensed in Table 1. Then, we summarise the remaining phases of the suggested approach as follows in order to get the best alternative(s) from the provided alternative.

ALGORITHM

Step 01: Establish a weighted PFH parameter matrix

$$L_w = [\sigma_{ij}]_{m\times n} = \begin{Bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \ldots\ldots\ldots & \sigma_{1m}, \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \ldots\ldots\ldots & \sigma_{2m}, \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \ldots\ldots\ldots & \sigma_{3m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \ldots\ldots\ldots & \sigma_{nm} \end{Bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\text{(i)}$$

Step 02: Create the weighted normalized PFH parameter matrix as follows:

$$Z_H = \begin{Bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \ldots\ldots\ldots & \delta_{1m}, \\ \delta_{21} & \delta_{22} & \delta_{23} & \ldots\ldots\ldots & \delta_{2m}, \\ \delta_{31} & \delta_{32} & \delta_{33} & \ldots\ldots\ldots & \delta_{3m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \delta_{n1} & \delta_{n2} & \delta_{n3} & \ldots\ldots\ldots & \delta_{nm} \end{Bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\text{(ii)}$$

Where

$$\delta_{ij} = \frac{\sigma_{ij}}{\sqrt{\sum_{i=1}^{n} (\sigma_{ij})^2}}$$

Step 03: Compute the weight vector $\Upsilon_i = \{\Upsilon_1, \Upsilon_2, \Upsilon_3, \ldots\ldots\ldots\Upsilon_n\}$, Υ_i are obtained as

$$\Upsilon_i = \frac{\lambda_j}{\sum_{j=1}^{m} \lambda_j}$$

where

$$\lambda_j = \frac{\sum_{i=1}^{n} \delta_{ij}}{j}, \quad \ldots\ldots\ldots\ldots\ldots\text{(iii)}$$

Step 04: Determine PFH aggregate matrix $(SDM)^{agrt}$ as follows:

$$(SDM)^{agrt} = \frac{(SDM)^1 + (SDM)^2 + (SDM)^3 + \ldots\ldots\ldots + (SDM)^m}{m} = [\, v_{ij}]_{m\times n} \quad \ldots\ldots\ldots\text{(iv)}$$

Step 05: Construct the weighted PFH decision matrix

$$\chi = [\mu_{ij}]_{m\times n} = \begin{pmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \ldots\ldots\ldots & \mu_{1m}, \\ \mu_{21} & \mu_{22} & \mu_{23} & \ldots\ldots\ldots & \mu_{2m}, \\ \mu_{31} & \mu_{32} & \mu_{33} & \ldots\ldots\ldots & \mu_{3m} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mu_{m1} & \mu_{m2} & \mu_{m3} & \ldots\ldots\ldots & \mu_{mm} \end{pmatrix} \quad \ldots\ldots\ldots\ldots\ldots\text{(v)}$$

where $\mu_{ij} = \Upsilon_i \times v_{ij} = \{[m_{c_i}(u_j)], [t_{c_i}(u_j)], [n_{c_i}(u_j)]\}$

Step 06: Obtain PFH positive ideal solution (PFHV+) and PFH negative ideal solution (PFHV−), where

$$(PFHV^{+}) = \{max[m_{c_i}(u_j)], min[t_{c_i}(u_j)], min[n_{c_i}(u_j)]\} = \{m^{+}_{c_i}(u_j), t^{+}_{c_i}(u_j), n^{+}_{c_i}(u_j)\} \quad \text{.....(vii)}$$

$$(PFHV^{-}) = \{min[m_{c_i}(u_j)], min[t_{c_i}(u_j)], max[n_{c_i}(u_j)]\} = \{m^{-}_{c_i}(u_j), t^{-}_{c_i}(u_j), n^{-}_{c_i}(u_j)\} \quad \text{.....(viii)}$$

Step 07: Compute the PFH separation measurements KR_j^{+} and KR_j^{-} p for all j = 1, 2, 3, ...m, defined as follows:

$$KR_j^{+} = \{\Sigma^{n}_{j=1} [[m_{c_i}(u_j) - m^{+}_{c_i}(u_j)]^2 + [t_{c_i}(u_j) - t^{+}_{c_i}(u_j)]^2 + [n_{c_i}(u_j) - n^{+}_{c_i}(u_j)]^2]\}^{1/2} \quad \text{.....(ix)}$$

$$KR_j^{-} = \{\Sigma^{n}_{j=1} [[m_{c_i}(u_j) - m^{-}_{c_i}(u_j)]^2 + [t_{c_i}(u_j) - t^{-}_{c_i}(u_j)]^2 + [n_{c_i}(u_j) - n^{-}_{c_i}(u_j)]^2]\}^{1/2} \quad \text{.....(x)}$$

Step 08: Obtain the PFH closeness coefficient C_j of each alternatives.

$$C_j = \frac{KR_j^{-}}{KR_j^{+} + KR_j^{-}} \in [0, 1] \quad \text{.....(xi)}$$

Step 07: Based on the PFH closeness coefficient, rank the alternatives in decreasing (or increasing) order and choose the optimal object from the alternatives. The flowchart of the algorithm is presented in Figure 1.

Numerical Example: Small firms and industries, usually referred to as microbusinesses or microindustries, are crucial to the economies of various countries. These companies are frequently identified by their diminutive size, constrained human resources, and little capital expenditures. The flexibility and ability of PFHST to handle uncertainty and imprecise information make it a valuable tool in decision-making, quality control, supply chain management, risk assessment, product development, customer relationship management, resource optimization, and market analysis. Its application can help small industries enhance their decision-making processes, improve operational efficiency, and achieve competitive advantage in the market. Consider a multi-national organization looking to develop an industry in an under-developed country. Now, there are a number of industrial sectors that the organization can expand in. So, the selection of the optimal sector for investment is quite the task as it is influenced by numerous factors. With that, consider U = {u_1, u_2, u_3, u_4, u_5} be five industrial sectors. In order to make optimal decisions, the organization assembles a team of six experts {$(SDM)^1$, $(SDM)^2$, $(SDM)^3$, $(SDM)^4$, $(SDM)^5$, $(SDM)^6$}. For the analysis of these industries, a selection of parameters is made beforehand on which they have to present their expert opinion on (T_i = {g_1, g_2, g_3, g_4, g_5}). Here, T_i stands for "revenue and profitability", "operational efficiency", "product quality", "technological

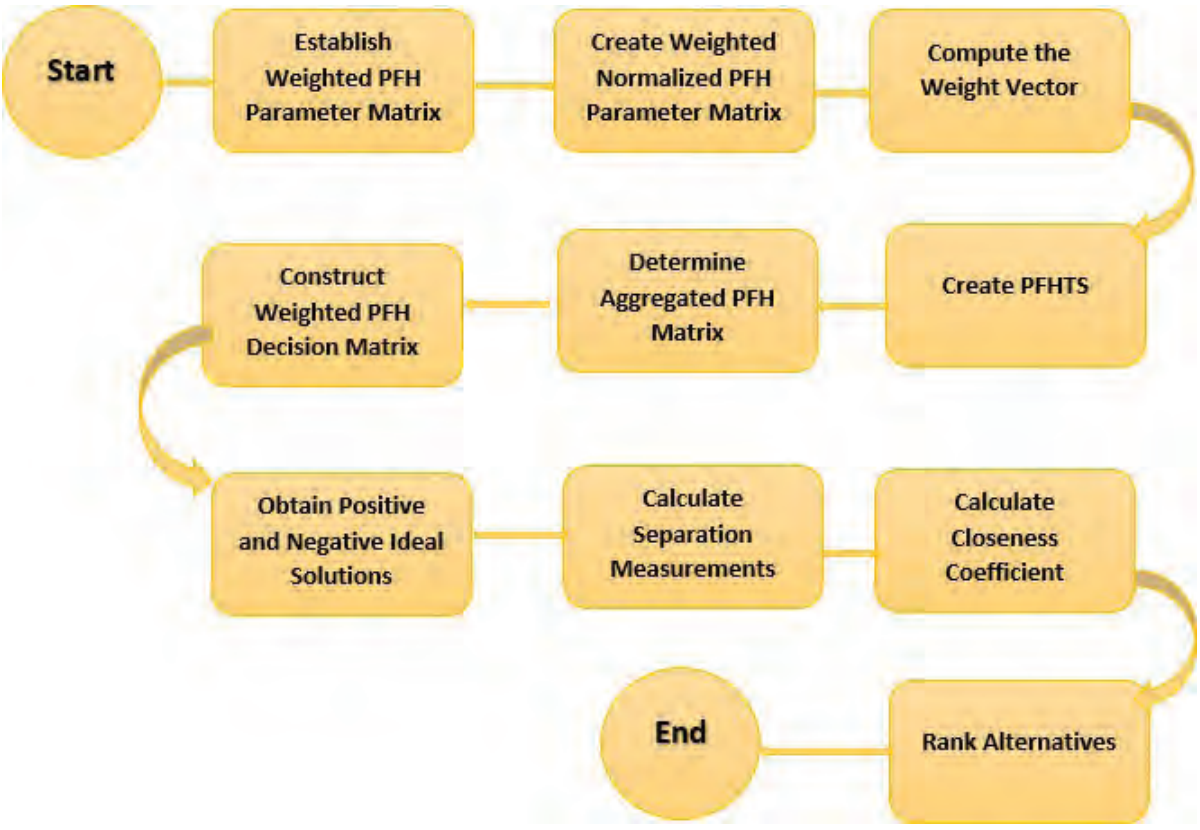


Figure 1. Flow chart of algorithm.

adoption", and "taxation and costs" be the attribute values. The corresponding attribute values of each of the attributes is given by the set $\{S_1, S_2, S_3, S_4, S_5\}$, where:

$S_1 = \{b_{11}$ = revenue growth, b_{12} = profit margins, b_{13} = return on investment, b_{14} = cash flow$\}$,

$S_2 = \{b_{21}$ = production efficiency, b_{22} = inventory turnover, b_{23}= lead times, b_{24} = cost management $\}$,

$S_3 = \{b_{31}$ = service quality, b_{32} = innovation, b_{33} = differentiation$\}$,

$S_4 = \{b_{41}$ = adoption rate, b_{42} = automation levels, b_{43} = digital marketing, b_{44} = IT infrastructure $\}$,

$S_5 = \{b_{51}$ = taxes, b_{52} = debt-to-equity ratio, b_{53} = ability to attract investments b_{54} = secure loans $\}$. There are seven hundred sixty eight possible outcomes but due to computational barriers and better explanation of the algorithm, the five outcomes illustrated below are addressed:

$$S = \begin{Bmatrix} c_1 = (b_{11}, b_{12}, b_{13}, b_{14} b_{15}), \\ c_2 = (b_{21}, b_{22}, b_{23}, b_{24}, b_{25}), \\ c_3 = (b_{31}, b_{32}, b_{33}, b_{34}, b_{35}), \\ c_4 = (b_{41}, b_{42}, b_{43}, b_{44}, b_{45}) \\ c_5 = (b_{51}, b_{52}, b_{53}, b_{54}, b_{55}) \end{Bmatrix}$$

The expert data is processed as illustrated below:

Step 1: On the basis of the weighted PFH parameter matrix is constructed as follows:

$$L_w = \begin{Bmatrix} 0.95\ 0.75\ 0.95\ 0.55\ 0.35 \\ 0.75\ 0.55\ 0.55\ 0.35\ 0.95 \\ 0.35\ 0.20\ 0.95\ 0.75\ 0.55 \\ 0.55\ 0.95\ 0.20\ 0.55\ 0.75 \\ 0.95\ 0.75\ 0.35\ 0.95\ 0.20 \\ 0.20\ 0.95\ 0.75\ 0.20\ 0.95 \end{Bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (xii)$$

Table 1. Linguistic terms to identify the alternatives

Linguistic terms	Weights
Superb	0.95
Wonderful	0.75
Nice	0.55
Not good	0.35
Unacceptable	0.20

Step 2: The weighted normalized PFH parameter matrix ZH is calculated by using (ii).

$$Z_H = \begin{Bmatrix} 0.56\ 0.41\ 0.56\ 0.37\ 0.21 \\ 0.45\ 0.33\ 0.33\ 0.23\ 0.56 \\ 0.21\ 0.12\ 0.56\ 0.50\ 0.33 \\ 0.33\ 0.56\ 0.12\ 0.37\ 0.45 \\ 0.56\ 0.45\ 0.21\ 0.64\ 0.12 \\ 0.12\ 0.56\ 0.45\ 0.13\ 0.56 \end{Bmatrix}$$

Step 3: By using (iii), the weight vector of the given attributes are computed as

$\Upsilon = \{0.197, 0.215, 0.197, 0.198, 0.197\}$.

Step 4: For each decision-maker $(SDM)^i$, i = 1 to 6 and their corresponding PFH decision matrices, we get a PFHSTS on U as

$$(F,S)^1=(SDM)^1=\left\{\begin{array}{l}\left(\left[\frac{\langle 0.4,0.1,0.5\rangle}{u_1},\frac{\langle 0.5,0.2,0.1\rangle}{u_2},\frac{\langle 0.8,0.1,0.1\rangle}{u_3},\frac{\langle 0.6,0.1,0.3\rangle}{u_4},\frac{\langle 0.7,0.2,0.1\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 0.6,0.2,0.1\rangle}{u_1},\frac{\langle 0.5,0.3,0.1\rangle}{u_2},\frac{\langle 0.4,0.3,0.2\rangle}{u_3},\frac{\langle 0.4,0.1,0.3\rangle}{u_4},\frac{\langle 0.4,0.2,0.2\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 0.3,0.3,0.4\rangle}{u_1},\frac{\langle 0.7,0.1,0.2\rangle}{u_2},\frac{\langle 0.2,0.2,0.4\rangle}{u_3},\frac{\langle 0.5,0.2,0.2\rangle}{u_4},\frac{\langle 0.8,0.1,0.1\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 0.3,0.1,0.4\rangle}{u_1},\frac{\langle 0.6,0.2,0.1\rangle}{u_2},\frac{\langle 0.4,0.1,0.2\rangle}{u_3},\frac{\langle 0.6,0.2,0.1\rangle}{u_4},\frac{\langle 0.9,0.1,0.0\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 0.8,0.1,0.1\rangle}{u_1},\frac{\langle 0.4,0.3,0.2\rangle}{u_2},\frac{\langle 0.7,0.2,0.1\rangle}{u_3},\frac{\langle 0.4,0.1,0.4\rangle}{u_4},\frac{\langle 0.5,0.2,0.3\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

$$(F,S)^2=(SDM)^2=\left\{\begin{array}{l}\left(\left[\frac{\langle 0.8,0.1,0.1\rangle}{u_1},\frac{\langle 0.5,0.3,0.1\rangle}{u_2},\frac{\langle 0.5,0.3,0.1\rangle}{u_3},\frac{\langle 0.3,0.4,0.2\rangle}{u_4},\frac{\langle 0.6,0.2,0.1\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 0.4,0.1,0.2\rangle}{u_1},\frac{\langle 0.4,0.2,0.1\rangle}{u_2},\frac{\langle 0.5,0.2,0.1\rangle}{u_3},\frac{\langle 0.5,0.2,0.2\rangle}{u_4},\frac{\langle 0.4,0.3,0.1\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 0.4,0.4,0.1\rangle}{u_1},\frac{\langle 0.6,0.2,0.1\rangle}{u_2},\frac{\langle 0.5,0.3,0.2\rangle}{u_3},\frac{\langle 0.5,0.3,0.1\rangle}{u_4},\frac{\langle 0.4,0.2,0.4\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 0.5,0.2,0.1\rangle}{u_1},\frac{\langle 0.4,0.3,0.3\rangle}{u_2},\frac{\langle 0.5,0.2,0.2\rangle}{u_3},\frac{\langle 0.7,0.2,0.1\rangle}{u_4},\frac{\langle 0.6,0.1,0.3\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 0.4,0.2,0.2\rangle}{u_1},\frac{\langle 0.5,0.2,0.3\rangle}{u_2},\frac{\langle 0.5,0.1,0.3\rangle}{u_3},\frac{\langle 0.6,0.1,0.2\rangle}{u_4},\frac{\langle 0.3,0.3,0.3\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

$$(F,S)^3=(SDM)^3=\left\{\begin{array}{l}\left(\left[\frac{\langle 0.4,0.1,0.5\rangle}{u_1},\frac{\langle 0.5,0.2,0.1\rangle}{u_2},\frac{\langle 0.5,0.1,0.1\rangle}{u_3},\frac{\langle 0.3,0.1,0.3\rangle}{u_4},\frac{\langle 0.6,0.2,0.1\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 0.4,0.1,0.2\rangle}{u_1},\frac{\langle 0.4,0.2,0.1\rangle}{u_2},\frac{\langle 0.4,0.2,0.2\rangle}{u_3},\frac{\langle 0.4,0.1,0.3\rangle}{u_4},\frac{\langle 0.4,0.2,0.2\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 0.3,0.3,0.4\rangle}{u_1},\frac{\langle 0.6,0.1,0.2\rangle}{u_2},\frac{\langle 0.2,0.2,0.4\rangle}{u_3},\frac{\langle 0.5,0.2,0.2\rangle}{u_4},\frac{\langle 0.4,0.1,0.4\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 0.3,0.1,0.4\rangle}{u_1},\frac{\langle 0.4,0.2,0.3\rangle}{u_2},\frac{\langle 0.4,0.1,0.2\rangle}{u_3},\frac{\langle 0.6,0.2,0.1\rangle}{u_4},\frac{\langle 0.6,0.1,0.3\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 0.4,0.1,0.2\rangle}{u_1},\frac{\langle 0.4,0.2,0.3\rangle}{u_2},\frac{\langle 0.5,0.1,0.3\rangle}{u_3},\frac{\langle 0.4,0.1,0.4\rangle}{u_4},\frac{\langle 0.3,0.2,0.3\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

$$(F,S)^4=(SDM)^4=\left\{\begin{array}{l}\left(\left[\frac{\langle 0.8,0.1,0.1\rangle}{u_1},\frac{\langle 0.5,0.2,0.1\rangle}{u_2},\frac{\langle 0.8,0.1,0.1\rangle}{u_3},\frac{\langle 0.6,0.1,0.2\rangle}{u_4},\frac{\langle 0.7,0.2,0.1\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 0.6,0.1,0.1\rangle}{u_1},\frac{\langle 0.5,0.2,0.1\rangle}{u_2},\frac{\langle 0.5,0.2,0.1\rangle}{u_3},\frac{\langle 0.5,0.1,0.2\rangle}{u4},\frac{\langle 0.4,0.2,0.1\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 0.4,0.3,0.1\rangle}{u_1},\frac{\langle 0.7,0.1,0.1\rangle}{u_2},\frac{\langle 0.5,0.2,0.2\rangle}{u_3},\frac{\langle 0.5,0.2,0.1\rangle}{u4},\frac{\langle 0.8,0.1,0.1\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 0.5,0.1,0.1\rangle}{u_1},\frac{\langle 0.6,0.2,0.1\rangle}{u_2},\frac{\langle 0.5,0.1,0.2\rangle}{u_3},\frac{\langle 0.7,0.2,0.1\rangle}{u4},\frac{\langle 0.9,0.1,0.0\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 0.8,0.1,0.1\rangle}{u_1},\frac{\langle 0.5,0.2,0.2\rangle}{u_2},\frac{\langle 0.7,0.1,0.1\rangle}{u_3},\frac{\langle 0.6,0.1,0.2\rangle}{u4},\frac{\langle 0.5,0.2,0.3\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

$$(F,S)^5=(SDM)^5=\left\{\begin{array}{l}\left(\left[\frac{\langle 1.0,0.0,0.0\rangle}{u1},\frac{\langle 1.0,0.0,0.0\rangle}{u2},\frac{\langle 1.0,0.0,0.0\rangle}{u_3},\frac{\langle 1.0,0.0,0.0\rangle}{u_4},\frac{\langle 1.0,0.0,0.0\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 1.0,0.0,0.0\rangle}{u1},\frac{\langle 1.0,0.0,0.0\rangle}{u2},\frac{\langle 1.0,0.0,0.0\rangle}{u_3},\frac{\langle 1.0,0.0,0.0\rangle}{u_4},\frac{\langle 1.0,0.0,0.0\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 1.0,0.0,0.0\rangle}{u1},\frac{\langle 1.0,0.0,0.0\rangle}{u2},\frac{\langle 1.0,0.0,0.0\rangle}{u_3},\frac{\langle 1.0,0.0,0.0\rangle}{u_4},\frac{\langle 1.0,0.0,0.0\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 1.0,0.0,0.0\rangle}{u1},\frac{\langle 1.0,0.0,0.0\rangle}{u2},\frac{\langle 1.0,0.0,0.0\rangle}{u_3},\frac{\langle 1.0,0.0,0.0\rangle}{u_4},\frac{\langle 1.0,0.0,0.0\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 1.0,0.0,0.0\rangle}{u1},\frac{\langle 1.0,0.0,0.0\rangle}{u2},\frac{\langle 1.0,0.0,0.0\rangle}{u_3},\frac{\langle 1.0,0.0,0.0\rangle}{u_4},\frac{\langle 1.0,0.0,0.0\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

$$(F,S)^6=(SDM)^6=\left\{\begin{array}{l}\left(\left[\frac{\langle 0.0,0.0,1.0\rangle}{u1},\frac{\langle 0.0,0.0,1.0\rangle}{u2},\frac{\langle 0.0,0.0,1.0\rangle}{u_3},\frac{\langle 0.0,0.0,1.0\rangle}{u_4},\frac{\langle 0.0,0.0,1.0\rangle}{u_5}\right],\{c_1\}\right),\\ \left(\left[\frac{\langle 0.0,0.0,1.0\rangle}{u1},\frac{\langle 0.0,0.0,1.0\rangle}{u2},\frac{\langle 0.0,0.0,1.0\rangle}{u_3},\frac{\langle 0.0,0.0,1.0\rangle}{u_4},\frac{\langle 0.0,0.0,1.0\rangle}{u_5}\right],\{c_2\}\right),\\ \left(\left[\frac{\langle 0.0,0.0,1.0\rangle}{u1},\frac{\langle 0.0,0.0,1.0\rangle}{u2},\frac{\langle 0.0,0.0,1.0\rangle}{u_3},\frac{\langle 0.0,0.0,1.0\rangle}{u_4},\frac{\langle 0.0,0.0,1.0\rangle}{u_5}\right],\{c_3\}\right),\\ \left(\left[\frac{\langle 0.0,0.0,1.0\rangle}{u1},\frac{\langle 0.0,0.0,1.0\rangle}{u2},\frac{\langle 0.0,0.0,1.0\rangle}{u_3},\frac{\langle 0.0,0.0,1.0\rangle}{u_4},\frac{\langle 0.0,0.0,1.0\rangle}{u_5}\right],\{c_4\}\right),\\ \left(\left[\frac{\langle 0.0,0.0,1.0\rangle}{u1},\frac{\langle 0.0,0.0,1.0\rangle}{u2},\frac{\langle 0.0,0.0,1.0\rangle}{u_3},\frac{\langle 0.0,0.0,1.0\rangle}{u_4},\frac{\langle 0.0,0.0,1.0\rangle}{u_5}\right],\{c_5\}\right)\end{array}\right\}$$

Thus, the collection $\{(SDM)^1, (SDM)^2, (SDM)^3, (SDM)^4, (SDM)^5, (SDM)^6\}$ gives a PFHT on U. The PFH decision matrices is shown in Table 2.

Step 5: The aggregated PFH matrix (SDM) is obtained by using (iv) and summarized in Table 3.

Step 6: The weighted PFH decision matrix is obtained by using (vi) and written in Table 4.

Step 7: From the weighted matrix J and utilizing equations (vi), (vii), we obtain ideal solutions P F HV^+ and P F HV^- are

Table 2. PFH decision matrices

	c_1	c_2	c_3	c_4	c_5
$(SDM)^1$					
u_1	0.4, 0.1, 0.5	0.6, 0.2, 0.1	0.3, 0.3, 0.4	0.3, 0.1, 0.4	0.8, 0.1, 0.1
u_2	0.5, 0.2, 0.1	0.5, 0.3, 0.1	0.7, 0.1, 0.2	0.6, 0.2, 0.1	0.4, 0.3, 0.2
u_3	0.8, 0.1, 0.1	0.4, 0.3, 0.2	0.2, 0.2, 0.4	0.4, 0.1, 0.2	0.7, 0.2, 0.1
u_4	0.6, 0.1, 0.3	0.4, 0.1, 0.3	0.5, 0.2, 0.2	0.6, 0.2, 0.1	0.4, 0.1, 0.4
u_5	0.7, 0.2, 0.1	0.4, 0.2, 0.2	0.8, 0.1, 0.1	0.9, 0.1, 0.0	0.5, 0.2, 0.3
$(SDM)^2$					
u_1	0.8, 0.1, 0.1	0.4, 0.1, 0.2	0.4, 0.4, 0.1	0.5, 0.2, 0.1	0.4, 0.2, 0.2
u_2	0.5, 0.3, 0.1	0.4, 0.2, 0.1	0.6, 0.2, 0.1	0.4, 0.3, 0.3	0.5, 0.2, 0.3
u_3	0.5, 0.3, 0.1	0.5, 0.2, 0.1	0.5, 0.3, 0.2	0.5, 0.2, 0.2	0.5, 0.1, 0.3
u_4	0.3, 0.4, 0.2	0.5, 0.2, 0.2	0.5, 0.3, 0.1	0.7, 0.2, 0.1	0.6, 0.1, 0.2
u_5	0.6, 0.2, 0.1	0.4, 0.3, 0.1	0.4, 0.2, 0.4	0.6, 0.1, 0.3	0.3, 0.3, 0.3
$(SDM)^3$					
u_1	0.4, 0.1, 0.5	0.4, 0.1, 0.2	0.3, 0.3, 0.4	0.3, 0.1, 0.4	0.4, 0.1, 0.2
u_2	0.5, 0.2, 0.1	0.4, 0.2, 0.1	0.6, 0.1, 0.2	0.4, 0.2, 0.3	0.4, 0.2, 0.3
u_3	0.5, 0.1, 0.1	0.4, 0.2, 0.2	0.2, 0.2, 0.4	0.4, 0.1, 0.2	0.5, 0.1, 0.3
u_4	0.3, 0.1, 0.3	0.4, 0.1, 0.3	0.5, 0.2, 0.2	0.6, 0.2, 0.1	0.4, 0.1, 0.4
u_5	0.6, 0.2, 0.1	0.4, 0.2, 0.2	0.4, 0.1, 0.4	0.6, 0.1, 0.3	0.3, 0.2, 0.3
$(SDM)^4$					
u_1	0.8, 0.1, 0.1	0.6, 0.1, 0.1	0.4, 0.3, 0.1	0.5, 0.1, 0.1	0.8, 0.1, 0.1
u_2	0.5, 0.2, 0.1	0.5, 0.2, 0.1	0.7, 0.1, 0.1	0.6, 0.2, 0.1	0.5, 0.2, 0.2
u_3	0.8, 0.1, 0.1	0.5, 0.2, 0.1	0.5, 0.2, 0.2	0.5, 0.1, 0.2	0.7, 0.1, 0.1
u_4	0.6, 0.1, 0.2	0.5, 0.1, 0.2	0.5, 0.2, 0.1	0.7, 0.2, 0.1	0.6, 0.1, 0.2
u_5	0.7, 0.2, 0.1	0.4, 0.2, 0.1	0.8, 0.1, 0.1	0.9, 0.1, 0.0	0.5, 0.2, 0.3
$(SDM)^5$					
u_1	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0
u_2	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0
u_3	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0
u_4	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0
u_5	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0	1.0, 0.0, 0.0
$(SDM)^6$					
u_1	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0
u_2	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0
u_3	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0
u_4	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0
u_5	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0	0.0, 0.0, 1.0

Table 3. The aggregated PFH matrix

$(SDM)^{agr}$	c_1	c_2	c_3	c_4	c_5
u_1	0.57, 0.07, 0.37	0.50, 0.08, 0.27	0.40, 0.38, 0.33	0.43,0.08,0.33	0.57,0.08,0.27
u_2	0.50, 0.15, 0.23	0.47, 0.15, 0.23	0.60, 0.08, 0.27	0.50,0.15,0.30	0.47,0.15,0.33
u_3	0.60, 0.10, 0.23	0.47, 0.15, 0.27	0.40, 0.15, 0.37	0.47,0.08,0.13	0.57,0.08,0.30
u_4	0.47, 0.12, 0.33	0.47, 0.08, 0.33	0.50, 0.15, 0.27	0.60,0.13,0.23	0.50,0.07,0.37
u_5	0.60, 0.17, 0.23	0.43, 0.15, 0.27	0.57, 0.08, 0.33	0.63,0.08,0.28	0.45,0.13,0.35

Table 4. The weighted PFH decision matrix

	c_1	c_2	c_3	c_4	c_5
u_1	0.112,0.013,0.072	0.098,0.015,0.053	0.078,0.074,0.065	0.085,0.016,0.065	0.112,0.016,0.053
u_2	0.107,0.032,0.049	0.101,0.032,0.049	0.129,0.017,0.058	0.101,0.032,0.065	0.101,0.032,0.071
u_3	0.118,0.020,0.045	0.092,0.029,0.053	0.078,0.029,0.072	0.091,0.016,0.026	0.112,0.016,0.059
u_4	0.093,0.023,0.065	0.093,0.015,0.065	0.099,0.029,0.053	0.119,0.026,0.046	0.099,0.014,0.073
u_5	0.118,0.033,0.045	0.084,0.029,0.053	0.112,0.015,0.065	0.124,0.016,0.055	0.087,0.026,0.069

$PFHV^+$ = {(0.112, 0.013, 0.053), (0.129, 0.017, 0.049), (0.118, 0.020, 0.045), (0.119, 0.015, 0.053), (0.124, 0.015, 0.045)}

$PFHV^-$ = {(0.078, 0.017, 0.071), (0.101, 0.017, 0.058), (0.078, 0.016, 0.072), (0.093, 0.014, 0.073), (0.084, 0.015, 0.069)}

Step 8: For each j = 1, 2, 3, the PFH separation measurements KR+ p and KR− p are calculated by using equations (ix), (x) as

KR^+_j = {0.138, 0.193, 0.183, 0.181, 0.198}

KR^-_j = {0.076, 0.055, 0.075, 0.052, 0.067}

Step 9: Using equation (xi), compute the picture fuzzy hypersoft closeness coefficients Cp, $c_1 = 0.355$, $c_2 = 0.222$, $c_3 = 0.290$, $c_4 = 0.223$, $c_5 = 0.253$

Step 10: Based on the ratings of Cp, we can obtain the ordering of the given alternatives as $c_1 > c_3 > c_5 > c_4 > c_2$ Which corresponds to the alternatives ratings as: $u_1 > u_3 > u_5 > u_4 > u_2$

Using the recently defined TOPSIS to evaluate alternatives successfully creates new opportunities for improved decision making. TOPSIS offers an organised and methodical approach to assess and prioritise options according to how close they are to the perfect answer. Through the simultaneous consideration of several criteria and goals, TOPSIS gives decision makers a thorough grasp of the advantages and disadvantages of the various options. This strategy gets around the drawbacks of conventional decision-making techniques, which frequently concentrate on a single criterion or neglect to take the complexity of real-world situations into account. Through the integration of multiple aspects and their respective relevance, TOPSIS application enables decision makers to make well-informed choices.

Comparative Analysis

We will compare the suggested structure with the current hypersoft topological structure in this section.

i. HSTS [38] deals information with attributes and sub attributes. It faced challenges in ranking.
ii. The FHSTS [39] deals data attributes and sub attributes in an uncertain way.
iii. The IFHSTS [40], which takes into account both the membership and non-membership degrees. Both the existence and absence of attributes and sub attributes IFHSTS can be handled well.
iv. The characteristics and subattributes of the PFHSTS (proposed approach) with positive, neutral, and negative membership functions overcome the limitations of existing structure of topology.

Our proposed structure has several advantages over present hypersoft topologies.

Structural Comparison

PFHSTS has been shown to be versatile with structural comparisons being based on critical evaluating characteristics such as DM (membership degree), NMD

Table 5. A comprehensive comparison of the proposed structure with hybrid fuzzy structures reported in literature

Structures	DoM	DoNM	DoNTM	SAAF	MAAF	DFPT
FS [1]	✓	×	×	×	×	×
IFS [2]	✓	✓	×	×	×	×
SS [14]	×	×	×	✓	×	×
FT [33]	✓	✓	×	×	×	×
IFT [34]	✓	✓	×	×	×	×
ST [35]	×	×	×	✓	×	×
FSTS [36]	✓	×	×	✓	×	×
IFSTS [37]	✓	✓	×	✓	×	×
HSTS [38]	×	×	×	✓	✓	✓
FHST [39]	✓	×	×	✓	✓	✓
IFHST [40]	✓	✓	×	✓	✓	✓
PFHST (pro- posed)	✓	✓	✓	✓	✓	✓

Table 6. Sensitivity analysis with existing structures

Set Structure	c_1	c_2	c_3	c_4	c_5	Ranking
FHTS	0.134	0.267	0.237	0.212	0.250	$c_2 \geq c_5 \geq c_3 \geq c_4 \geq c_1$
IFHTS	0.247	0.391	0.234	0.297	0.278	$c_2 \geq c_4 \geq c_5 \geq c_3 \geq c_1$
PFHTS (proposed)	0.355	0.290	0.253	0.223	0.222	$c_1 \geq c_3 \geq c_5 \geq c_4 \geq c_2$

(non-membership degree), AD (neutral degree), SAAF (single-argument approximate function), MAAF (multiargument approximate function) and DFPT (deep focus on parametric tuples). A summary of these comparisons is presented in Table 5.

Sensitivity Analysis

A sensitivity analysis of the developed PFHST TOPSIS method, in comparison with some existing structures from the Hypersoft set, is presented in Table 6. PFHSTS is versatile as shown by structural comparisons in terms of critical evaluating characteristics such as DM (membership degree), NMD (non-membership degree), AD (neutral degree), SAAF (single-argument approximate function), MAAF (multiargument approximate function), and DFPT (deep focus on parametric tuples). Table 5 summarizes these comparisons..

CONCLUSION

The picture fuzzy hypersoft set is a novel concept that surpasses all other accessible fuzzy soft set models. This most recent model is more precise and reasonable, making it capable of solving numerous problems more efficiently and sensibly. In this paper, we introduce certain essential aspects of picture fuzzy hypersoft topology. This framework is constructed based on the concept of union and intersection peculiar to picture fuzzy hypersoft sets. In this topological framework, the paper also gives the basic definitions concerning the interior, closure, boundary and exterior. Based on these definitions, several theorems are established. Furthermore, an algorithm is proposed alongside an application demonstrating its relevance in a group decision-making method. The model is also presented as an extension of the Technique for Order of Preference by Similarity to Ideal Solution approach. The efficiency and applicability of the suggested algorithm is explained through a numerical example. In the work to come, we shall examine the algebraic properties of picture fuzzy hypersoft topology and examine how it could be used in decision-making, medical diagnosis, clustering analysis, pattern recognition, and information science. The picture fuzzy hypersoft topology also presents new opportunities to be explored in the advanced fuzzy set theories. In the next work, we would like to expand this framework to include the idea of bipolarity and complex fuzzy sets. The use of bipolar picture fuzzy hypersoft sets will allow modeling both positive and negative information as well as the degrees of neutral and refusal, which will be more effective in modeling the uncertainty in a complex decision-making situation. Also, the incorporation of the concept of complex fuzzy sets, where membership functions include phase and amplitude, can further be used to increase the potential of the model in solving problems in dynamic environments like signal processing, quantum computing, and advanced pattern recognition. These extensions will greatly increase the applicability and usefulness of picture fuzzy hypersoft topology in solving complex and multifaceted problems in decision-making and more.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Zadeh L. Fuzzy sets. Inform Control 1965;8:338–353. [CrossRef]

[2] Atanassov KT. More on intuitionistic fuzzy sets. Fuzzy Sets Syst 1989;33:37–45. [CrossRef]

[3] Cuong BC, Kreinovich V. Picture fuzzy sets. J Comput Sci Cybern 2014;30:409–420.

[4] Singh P. Correlation coefficients for picture fuzzy sets. J Intell Fuzzy Syst 2015;28:591–604. [CrossRef]

[5] Son LH. DPFCM: A novel distributed picture fuzzy clustering method on picture fuzzy sets. Expert Syst Appl 2015;42:51–66. [CrossRef]

[6] Thong PH, Son LH. A new approach to multi-variable fuzzy forecasting using picture fuzzy clustering and picture fuzzy rule interpolation method. In: Knowledge and Systems Engineering (KSE 2014), Springer; 2015. p. 679–690. [CrossRef]

[7] Son LH. Generalized picture distance measure and applications to picture fuzzy clustering. Appl Soft Comput 2016;46:284–295. [CrossRef]

[8] Son LH. Measuring analogousness in picture fuzzy sets: From picture distance measures to picture association measures. Fuzzy Optim Decis Mak 2017;16:359–378. [CrossRef]

[9] Son LH, Van Viet P, Van Hai P. Picture inference system: A new fuzzy inference system on picture fuzzy set. Appl Intell 2017;46:652–669. [CrossRef]

[10] Thong PH, Son LH. Picture fuzzy clustering for complex data. Eng Appl Artif Intell 2016;56:121–130. [CrossRef]

[11] Thong PH, Son LH. A novel automatic picture fuzzy clustering method based on particle swarm optimization and picture composite cardinality. Knowl Based Syst 2016;109:48–60. [CrossRef]

[12] Wei G. Picture fuzzy aggregation operators and their application to multiple attribute decision making. J Intell Fuzzy Syst 2017;33:713–724. [CrossRef]

[13] Wei G. Picture fuzzy cross-entropy for multiple attribute decision making problems. J Bus Econ Manag 2016;17:491–502. [CrossRef]

[14] Molodtsov D. Soft set theory—first results. Comput Math Appl 1999;37:19–31. [CrossRef]

[15] Maji P, Biswas R, Roy A. Fuzzy soft set theory. J Fuzzy Math 2001;9:589–602.

[16] Maji PK. More on intuitionistic fuzzy soft sets. In: Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC 2009), New York: Springer; 2009. p. 231–240. [CrossRef]

[17] Yang Y, Liang C, Ji S, Liu T. Adjustable soft discernibility matrix based on picture fuzzy soft sets and its applications in decision making. J Intell Fuzzy Syst 2015;29:1711–1722. [CrossRef]

[18] Smarandache F. Extension of soft set to hypersoft set, and then to plithogenic hypersoft set. Neutrosophic Sets Syst 2018;22:168–170.

[19] Saeed M, Ahsan M, Siddique MK, Ahmad MR. A study of the fundamentals of hypersoft set theory. Int J Scientif Eng Res 2020;1:230.

[20] Jafar MN, Saeed M. Aggregation operators of fuzzy hypersoft sets. Turk J Fuzzy Syst 2021;11:1–17. [CrossRef]

[21] Saeed M, Ahsan M, Abdeljawad T. A development of complex multi-fuzzy hypersoft set with application in MCDM based on entropy and similarity measure. IEEE Access 2021;9:60026–60042. [CrossRef]

[22] Saeed M, Ahsan M, Saeed MH, Mehmood A, El-Morsy S. Assessment of solid waste management strategies using an efficient complex fuzzy hypersoft set algorithm based on entropy and similarity measures. IEEE Access 2021;9:150700–150714. [CrossRef]

[23] Jafar MN, Saeed M, Haseeb M, Habib A. Matrix theory for intuitionistic fuzzy hypersoft sets and its application in multi-attributive decision-making problems. Theory Appl Hypersoft Set 2021;65. [CrossRef]

[24] Saeed M, Ahsan M, Saeed MH, Mehmood A, Abdeljawad T. An application of neutrosophic hypersoft mapping to diagnose hepatitis and propose appropriate treatment. IEEE Access 2021;9:70455–70471. [CrossRef]

[25] Saeed M, Ahsan M, Mehmood A, Saeed MH, Asad J. Infectious diseases diagnosis and treatment suggestions using complex neutrosophic hypersoft mapping. IEEE Access 2021;9:146730–146744. [CrossRef]

[26] Saqlain M, Riaz M, Saleem MA, Yang MS. Distance and similarity measures for neutrosophic hypersoft set (NHSS) with construction of NHSS-TOPSIS and applications. IEEE Access 2021;9:30803–30816. [CrossRef]

[27] Saqlain M, Xin XL. Interval valued, m-polar and m-polar interval valued neutrosophic hypersoft sets. Infinite Study 2020.

[28] Rahman AU, Saeed M, Khalifa H. Decision making application based on parameterization of fuzzy hypersoft set with fuzzy setting. Ital J Pure Appl Math 2022.

[29] Saeed M, Harl MI. Fundamentals of picture fuzzy hypersoft set with application. Neutrosophic Sets Syst 2023;53:24.

[30] Mahmood T, Hussain K, Ahmmad J, Rehman UU, Aslam M. A novel approach toward TOPSIS method based on lattice ordered t-bipolar soft sets and their applications. IEEE Access 2022;10:69727–69740. [CrossRef]

[31] Albaity M, Rehman UU, Mahmood T. Data source selection for integration in data sciences via complex hesitant fuzzy rough multi-attribute decision-making method. IEEE Access 2024;12:110146–110159. [CrossRef]

[32] Rehman UU, Labassi F, Alsuraiheed T, Mahmood T, Khan MA. Selection of energy trading platform for peer-to-peer (P2P) energy trading by using a multi-attribute decision-making approach based on bipolar fuzzy Aczel-Alsina prioritized aggregation operators. IEEE Access 2024;12:80847–80858. [CrossRef]

[33] Chang CL. Fuzzy topological spaces. J Math Anal Appl 1968;24:182–190. [CrossRef]
[34] Coker D. An introduction to intuitionistic fuzzy topological spaces. Fuzzy Sets Syst 1997;88:81–89. [CrossRef]
[35] Cagman N, Karatas S, Enginoglu S. Soft topology. Comput Math Appl 2011;62:351–358. [CrossRef]
[36] Tanay B, Kandemir MB. Topological structure of fuzzy soft sets. Comput Math Appl 2011;61:2952–2957. [CrossRef]
[37] Osmanoglu I, Tokat D. On intuitionistic fuzzy soft topology. Gen Math Notes 2013;19:59–70.
[38] Musa SY, Asaad BA. Hypersoft topological spaces. Neutrosophic Sets Syst 2022;49:397–415. [CrossRef]
[39] Yolcu A, Ozturk TY. On fuzzy hypersoft topological spaces. Caucasian J Sci 2022;9:1–19. [CrossRef]
[40] Yolcu A. Intuitionistic fuzzy hypersoft topology and its applications to multi-criteria decision-making. Sigma 2023;41:106–118. [CrossRef]
[41] Hwang CL, Yoon K. Methods for multiple attribute decision making. In: Multiple Attribute Decision Making: Methods and Applications. New York: Springer; p. 58–191. [CrossRef]
[42] Boran FE, Genc S, Kurt M, Akay D. A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. Expert Syst Appl 2009;36:11363–11368. [CrossRef]
[43] Eraslan S, Karaaslan F. A group decision making method based on TOPSIS under fuzzy soft environment. J New Theory 2015;3:30–40.
[44] Garg H, Arora R, Arora R. TOPSIS method based on correlation coefficient for solving decision-making problems with intuitionistic fuzzy soft set information. AIMS Math 2020;5:2944–2966. [CrossRef]
[45] Keikha A, Garg H, Mishmast Nehi H. An approach based on combining Choquet integral and TOPSIS methods to uncertain MAGDM problems. Soft Comput 2021;25:7181–7195. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1082–1093, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2028

Research Article

New ranking for picture fuzzy sets and novel way to address picutre fuzzy transportation problem

K. HEMALATHA[1], VENKATESWARLU. B[2,*]

[1]Department of Mathematics, Prathyusha Engineering College, Tamil Nadu, 602025, India
[2]Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Tamil Nadu, 632014, India

ARTICLE INFO

Article history
Received: 19 July 2024
Revised: 13 October 2024
Accepted: 30 December 2024

Keywords:
Initial Basic Feasible Solution; Intuitionistic Fuzzy Set; Transportation Problem; Picture Fuzzy Set; Picture Fuzzy Transportation Problem

ABSTRACT

A significant issue that is being extensively researched in the arena of operation research involves the issue of transportation. The primary objective of this hurdle is to figure out the lowest possible total transportation costs for commodities to provide the items from the origin and satisfy customer demands at destinations. In practical scenarios, the decision-maker might not be certain of the exact figures of the coefficients applicable to the transport issue. Transportation problems with ambiguous and uncertain information can be analyzed using a picture fuzzy set, an extension of an intuitionistic fuzzy set, specifically when extra-linguistic assessment elements are needed. This research combines picture fuzzy figures to represent each unit's availability, demand, and expense. In this study, picture fuzzy numbers are transformed into precise values using a proposed novel ranking way. Additionally, we developed an innovative approach for determining an initial basic feasible solution to the picture fuzzy transportation problem. To illustrate the proposed technique, we considered three types of numerical examples. The findings of the suggested algorithm were compared with existing tactics in the literature to establish its efficacy.

Cite this article as: Hemalatha K, Venkateswarlu B. New ranking for picture fuzzy sets and novel way to address picutre fuzzy transportation problem. Sigma J Eng Nat Sci 2026;44(2):1082–1093.

INTRODUCTION

The ideal planning tool across many fields, such as finance, technology, economics, and enterprise, is linear programming. In every firm, decision-making is crucial. Depending on the circumstances and the information available, different decisions must be made. Individuals had to adjust their decision-making in earlier times to fit the situation. Uncertainty exists at present in every decision we make. As a result, making decisions is difficult. Zadeh proposed a set termed fuzziness to tackle uncertainty [1].

Transportation experts have difficulties at the start of the twenty-first century as a result of the complexity that is growing. To provide prompt, secure, and dependable transportation when minimizing the negative effects on the environment and societies, transportation experts must strive to meet certain desired outcomes. Capacity limitations, toxic emissions, a lack of reliability, poor safety records, and

*Corresponding author.
*E-mail address: venkatesh.reddy@vit.ac.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

energy waste are some of the obstacles that transportation specialists must deal with. Taking into account these realities, transport systems are typically complicated systems with numerous components and various stakeholders, each with a variety of goals that are frequently at odds with one another [2]. Transporting raw materials or completed goods from a single manufacturer to another while attempting to keep the overall cost of shipping to a minimum is known as a transportation problem (TP). Hitchcock first came up with the fundamental transportation issue. The goal of the issue of transit is to establish the ideal quantity of a good to be delivered from multiple sources (origins) to several consumer points (destinations), such that the total expenditure associated with transport is lowered or the total gain from mobility is maximized [3]. Expense details are necessary for a minimization TP, and in this instance, the goal is to reduce overall costs. A maximization TP, on the other hand, uses revenue, turnover, or gain information; in this circumstance, the approach aims to maximize the overall revenue. Whenever precise data is offered, there exist traditional approaches to tackle those transportation issues. However, data might not have been known with precision in real-world transportation circumstances [4].

The uncertainty of any number of variables, which cannot be precisely determined in practical situations, can arise from a variety of factors, including the imprecise tool of measurement, perished and missing data, as well as computation mistakes. In these cases, standard techniques to resolve TP are no longer appropriate. Fuzzy mathematics has become necessary to solve this kind of problem [5]. To solve this issue, the issue's details are provided with fuzzy figures, which represent the uncertainties in the information being sought.

The concept of fuzzy sets originated by Zadeh [6] to deal with ambiguous data mathematically when making decisions, and it has been utilized effectively in a wide range of fields. Following the groundbreaking work by Bellman and Zadeh [7], the idea of fuzzy sets was applied in the optimization sector. However, hesitations are owing to a variety of reasons, including the need for greater interaction, inaccurate data, an understanding of markets, a lack of customer perception, etc. Similar to the costs, there are still many unknowns and hesitations due to many factors such as variations in gasoline prices, clogged roads, meteorological conditions, etc. In such cases, the decision-makers are unable to accurately predict the expense of transportation. As a result, Atanassov and Stoeva [8] presented the intuitionistic fuzzy collection (IFS), which is to be more reliable than the framework of fuzzy sets developed by Zadeh, for managing these ambiguous inputs. After that, a sizable number of effective solutions to TPs and IFTPs were reported in the research literature, and the scope of applications progressively expanded [9,10].

The neutrality degree is not one of the key concepts of IFS. The idea of neutrality grade is frequently observed under conditions where we are faced with human viewpoints containing more responses of the form: yes, abstain, no, or rejection [11]. For example, human voting, choosing features, and so forth. In this vein, Cuong and Kreinovich [12] created the picture fuzzy setting, which is an outgrowth of the IFS by including the ideas of the positive, neutral, and negative membership ratings of an aspect. Cuong and Kreinovich [13] examined a few picture fuzzy set (PFS) characteristics and proposed PFS distance measurements. Phong et al. [14] investigated a few combinations of picture fuzzy connections. In their study of the fundamental fuzzy reasoning operators, Cuong and Pham [15] examined denials, interjections, disjunctions, and their effects on picture sets that are fuzzy. They also developed the basic functions for fuzzy inference procedures in picture fuzzy structures. On picture fuzzy systems, Cuong et al. [16] demonstrated the features associated with an involutive picture negator and several related De Morgan uncertain triples. A picture uncertain inference framework created using a membership plot has been presented by Viet et al. [17]. PFS correlation factors were investigated by Singh [18]. For picture fuzzy sets, Cuong et al. [19] explored the categorization of attainable t-norms as well as t-conorms of picture operators. Son [20] suggested a novel distance metric between PFSs and implemented it in clustering with fuzzy data. Son [21] extended fundamental metrics for distance in PFSs and investigated a few of their characteristics. The fuzzy inference was suggested for PFSs by Son et al., [22]. By using a new separation metric as the basis for decision-making, Peng and Dai [23] established the algorithm for PFS. Wei [24] provided various methods for determining the similarities of PFS.

Garg [25] investigated several uses of picture-uncertain aggregating methods in multicriteria decision-making processes. The concept of complex PFS, resulting in a generalized version of complex IFS and pythagorean fuzzy set, was put forth by Akram et al. [26], who also created a decision-making paradigm. Harmonic operators involving TrPFNs were utilized by Shit et al. [27] to demonstrate their importance in multi-criteria decision-making issues. The extension technique was used by Dutta and Ganju [11] to conduct picture fuzzy arithmetic using illustrations. Employing trapezoidal picture fuzzy figures, Akram et al. [28] looked into the shortest path issues. Geetha and Selvakumari [29] found low-cost solutions to picture-fuzzy transport issues. Mehmood and Bashir [30] proposed a method for resolving the full-picture fuzzy transportation issue. An innovative hybrid framework built on PFS and linear assignment was proposed by Gündodu et al. [31] and used for the first time in a real-world setting to address a challenge related to the growth of public transit. By using a novel approach, Akram et al. [32] were able to tackle linear programming issues in the context of picture fuzzy sets. The concept of picture fuzzy entropy along with its utilization was created by Kumar et al. [33] using a combination of the picture fuzzy technique and partial value data.

PFSs were successfully employed in collaboration, administration, and other real-life decision-making and transportation problems that involve uncertainty. Kirişci [34] created an algorithm designed in generalized pythagorean fuzzy sets. This algorithm was provided to assist with solving multi-attribute issues related to decision-making. For a pythagorean fuzzy transit problem of three types, Hemalatha and Venkateswarlu [35] devised a novel mean square approach to determine the initial basic feasible solution.

As previously discussed, limited studies have been done on the picture fuzzy transportation problem. According to the literature review, there are no parallel fresh approaches for both ranking and initial basic feasible solution (IBFS) to resolve TP in a picture uncertain setting. This research gap led the authors to develop a novel ranking and IBFS technique that can optimize the transportation problem in a Picture uncertain context and its optimal value without any mathematical tools. This study aims to address transportation issues where supply, demand, and transportation prices are PFN. The primary goal of this research is to reduce total transport expenses in a picture-fuzzy context. The following is the paper's key contribution:

(i) Created a new ranking function to convert fuzzily-defined picture integers into precise values.

(ii) Developed an algorithm to handle the picture fuzzy transport problem's initial basic feasible answer. So, using the allocation of the suggested algorithm, we continue to the modified distribution (MODI) strategy for the ideal answer.

(iii) Investigated six numerical instances of random values for three different types of picture fuzzy transportation problem (PFTP) to demonstrate the validity and effectiveness of the suggested approach. With thorough comparisons to certain previous studies, its advantages and viability are also stated.

The paper is structured as follows: The preliminary material is presented in Section 2. The mathematical description of the picture-fuzzy transport issue is described in Section 3 in depth. Section 4 of the proposal contains a description of the algorithm. Numerical examples are provided in Section 5. Results and discussion are presented in Section 6, and the work is concluded in Section 7.

PRELIMINAIES

In this part, we define certain fundamental concepts related to the current studies.

Fuzzy Set [36]

Any of the fuzzy subsets of the universal collection $\tilde{z}'$ can be expressed as a collection of ordered pairs, as shown below:

$$A''=\left\{\left\{(x)'_0,\mu_{A''}(x'_0)\right)\middle|x'_0\in\tilde{z}'\right\}$$

where $\mu_{A''}(x'_0)$ symbolizes an association functionality, which accepts values in the range [0,1].

Intuıtionistic Fuzzy Set [37]

Any IFS in the universal collection $\tilde{z}'$ can be described with the following syntax:

$$\tilde{A}''=\{\langle x'_0,\mu_{\tilde{A}''}(x'_0),\nu_{\tilde{A}''}(x'_0)\rangle|x'_0\in\tilde{z}'\},$$

where $\mu_{\tilde{A}''}$, $\nu_{\tilde{A}''}$ clarifies the role of belonging and not belonging, these two functions accept inputs in the range [0,1] with the specified condition. $0\le\mu_{\tilde{A}''}(x'_0),\nu_{\tilde{A}''}(x'_0)\le 1$.

Picture Fuzzy Set [38]

Any PFS in the Universal $\tilde{z}'$ collection can be defined as

$$\tilde{A}''=\left\{\left\{\langle x\rangle'_0,\mu_{\tilde{A}''}(x'_0),\eta_{\tilde{A}''}(x'_0),\nu_{\tilde{A}''}(x'_0)\right\rangle\middle|x'_0\in\tilde{z}'\right\},$$

where $\mu_{\tilde{A}''}(x'_0)$, $\eta_{\tilde{A}''}(x'_0)$, $\nu_{\tilde{A}''}(x'_0)\in[0,1]$; $\mu_{\tilde{A}''},\eta_{\tilde{A}''}$, and $\nu_{\tilde{A}''}$ indicate the levels of belonging, neutral, and not belonging x'_0 in $\tilde{A}''$, respectively. The following requirements must be met: $0\le\mu_{\tilde{A}''}(x'_0)+\eta_{\tilde{A}''}(x'_0)+\nu_{\tilde{A}''}(x'_0)\le 1$.

Proposed Ranking Function

Let $P_F=(\mu_{\tilde{A}''}(x'_0),\eta_{\tilde{A}''}(x'_0),\nu_{\tilde{A}''}(x'_0))$ be a picture fuzzy number. The ranking function is a defuzzification tool of picture fuzzy numbers to crisp numbers. It is used to compare fuzzy numbers. A novel ranking technique based on a collection of picture fuzzy numbers is outlined as follows:

$$R(P_F)=|\mu_{\tilde{A}''}(x'_0)-[\eta_{\tilde{A}''}(x'_0)+\nu_{\tilde{A}''}(x'_0)]| \quad (1)$$

Based on the new rank function we can analyze two picture fuzzy numbers. If $\sigma=\left(\mu_1^{P_F},\eta_1^{P_F}\nu_1^{P_F}\right)$ and $\varphi=\left(\mu_2^{P_F},\eta_2^{P_F},\nu_2^{P_F}\right)$ are two picture fuzzy numbers, then the relation between those are given by,

Case (1) $\sigma>\varphi$ iff $R(\sigma)>R(\varphi)$

Case (2) $\sigma<\varphi$ iff $R(\sigma)<R(\varphi)$

Case (3) $\sigma=\varphi$ iff $R(\sigma)=R(\varphi)$

MATHEMATICAL MODEL FOR PICTURE FUZZY TRANSPORTATON PROBLEM

Assume there are n destinations and m sources as shown in Table 1. A picture fuzzy TP can be mathematically expressed in the following equations (2-4):

$$\text{Minimize } \bar{z}^{P_F}=\sum_{i=1}^{m}\sum_{j=1}^{n}\bar{c}_{ij}^{P_F}.x_{ij} \quad (2)$$

Subject to constraints,

$$\sum_{j=1}^{n}x_{ij}=\bar{a}_i^{P_F},i\ =1\ to\ m, \quad (3)$$

$$\sum_{i=1}^{m}x_{ij}=\bar{b}_j^{P_F},j\ =\ 1\ to\ n, \quad (4)$$

$x_{ij}\ge 0$ for each i, j

where,

Table 1. Picture fuzzy trnsportation problem

Sources	Destinations				
		$\mathcal{D}_1$	$\mathcal{D}_2$	....	$\mathcal{D}_n$	Supply
	s_1	$c_{11}^{P_F}$	$c_{12}^{P_F}$	....	$c_{1n}^{P_F}$	$a_1^{P_F}$
	s_2	$c_{21}^{P_F}$	$c_{22}^{P_F}$	....	$c_{2n}^{P_F}$	$a_2^{P_F}$
	.....	.....	.....	.....	.....	.....
	s_m	$c_{m1}^{P_F}$	$c_{m2}^{P_F}$	....	$c_{mn}^{P_F}$	$a_m^{P_F}$
	Demand	$b_1^{P_F}$	$b_2^{P_F}$	....	$b_n^{P_F}$	

$\bar{c}_{ij}^{P_F}$- picture fuzzy expense of moving one unit of a given good supplier *i* to recipient *j*,

x_{ij}- transferred quantity from input *i* to terminal *j*,

$\bar{a}_i^{P_F}$- picture fuzzy units of supply to be carried between *n* places,

$\bar{b}_j^{P_F}$- picture fuzzy number of demand units needed at endpoints.

Proposed Alogrithm for Solving Picture Fuzzy Transportation Problem

The steps of the proposed algorithm are described below and the flowchart is shown in Figure 1.

Step 1: Select Transportation Problem under the Picture fuzzy setting.

Step 2: Refine picture fuzzy values into crisp quantities by applying the advised ranking function that has been created in equation (1).

Step 3: After turning them into crisp, determine whether the issue at hand is balanced or not.

(i) Go to step 5 if the given problem is balanced.

(ii) Go to step 4 if the given problem is unbalanced.

Step 4: To balance the aggregate demand and availability, add a cost-free dummy row or column.

Step 5: Divide each cost of a row by a number of columns, then add all of the divided values for each row.

Step 6: Divide each cost of a column by a number of rows, then add all of the divided values for each column.

Step 7: Select one penalty with the greatest value between the rows and columns. If the maximum value is the same for more than one option, pick either one.

Step 8: Choose the least cost value in the relevant row or column of the maximum penalty.

Step 9: For that specific cell, allocate the least value among the supply and need.

Step 10: When there is no longer any demand or supply for that specific row or column, remove the entire row or column.

Step 11: Repeat steps 5–10 until all allocations have been met.

Step 12: After determining IBFS, use the MODI approach to locate the ideal solution.

Numerical Illustrations

Example 5.1.[29] Consider the four places that require the delivery of four dairy-based goods. As shown in the matrix that appears below Table 2, the unit cost of shipping appears as picture-fuzzy values. We have to determine a transport strategy that will cost the least overall.

Step 1: Select Transportation Problem under the Picture fuzzy setting.

Step 2: Using equation (1) convert the picture fuzzy into crisp which is shown in Table 3.

Since the chosen problem is balanced. By using steps 5 and 6, divide each cost of a row by a number of columns,

Table 2. Transportation problem in picture fuzzy environment

Dairy products	L_1	L_2	L_3	L_4	Availability
Milk Powder	(0.8,0.1,0.1)	(0.4,0.3,0.3)	(0.5,0.3,0.2)	(0.7,0.1,0.1)	(0.6,0.3,0.1)
Ghee	(0.2,0.6,0.1)	(0.5,0.3,0.1)	(0.8,0.1,0)	(0.4,0.5,0.1)	(0.8,0.1,0.1)
Butter	(0.6,0.3,0.1)	(0.7,0.3,0)	(0.3,0.4,0.1)	(0.6,0.2,0.2)	(0.5,0.3,0.1)
Cheese	(0.4,0.4,0.2)	(0.8,0.2,0)	(0.4,0.6,0)	(0.7,0.1,0.1)	(0.4,0.4,0.1)
Requirement	(0.6,0.2,0.1)	(0.7,0.2,0.1)	(0.4,0.3,0.2)	(0.6,0.4,0)	

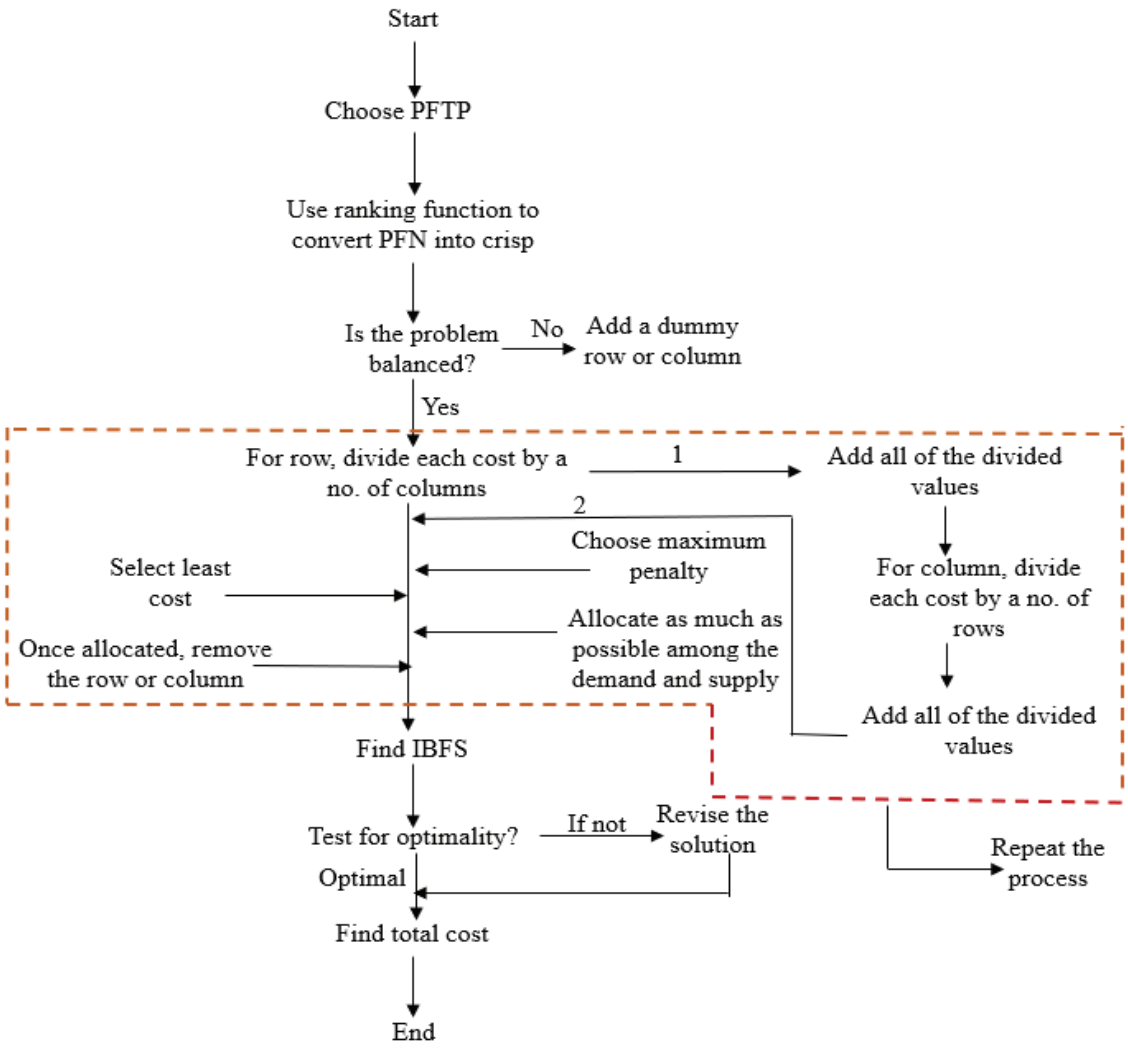


Figure 1. Flowchart.

Table 3. Defuzzified values

Dairy products	L_1	L_2	L_3	L_4	Availability
Milk Powder	0.6	0.2	0	0.6	0.2
Ghee	0.5	0.1	0.7	0.2	0.6
Butter	0.2	0.4	0.2	0.4	0.1
Cheese	0.2	0.6	0.2	0.6	0.1
Requirement	0.3	0.4	0.1	0.2	

Table 4. Row and column penalties

Dairy products	L_1	L_2	L_3	L_4	Availability	Row Penalty
Milk Powder	0.6	0.2	0	0.6	0.2	0.35
Ghee	0.5	0.1	0.7	0.2	0.6	0.375
Butter	0.2	0.4	0.2	0.4	0.1	0.3
Cheese	0.2	0.6	0.2	0.6	0.1	0.4
Requirement	0.3	0.4	0.1	0.2		
Column Penalty	0.375	0.325	0.275	0.45		

then add all of the divided values for each row. Similarly, divide each cost of a column by a number of rows, then add all of the divided values for each column which is shown in Table 4.

Step 7: Select one with the greatest value between the rows and columns. If the maximum value is the same for more than one option, pick either one.

Step 8: Choose the least cost value in the relevant row or column of the maximum penalty.

Step 9: For that specific cell, allocate the least value among supply and need which is shown in Table 5.

Step 10: When there is no longer any demand or supply for that specific row or column, remove the entire row or column. Table 6 shows the adjusted matrix with the second allocation.

Table 5. First allocation

Dairy products	L_1	L_2	L_3	L_4	Availability	Row penalty
Milk Powder	0.6	0.2	0	0.6	0.2	0.35
Ghee	0.5	0.1	0.7	0.2 [0.2]	0.6	0.375
Butter	0.2	0.4	0.2	0.4	0.1	0.3
Cheese	0.2	0.6	0.2	0.6	0.1	0.4
Requirement	0.3	0.4	0.1	0.2		
Column Penalty	0.375	0.325	0.275	0.45		

Table 6. Second allocation

Dairy products	L_1	L_2	L_3	Availability	Row penalty
Milk Powder	0.6	0.2	0	0.2	0.267
Ghee	0.5	0.1 [0.4]	0.7	0.4	0.427
Butter	0.2	0.4	0.2	0.1	0.267
Cheese	0.2	0.6	0.2	0.1	0.334
Requirement	0.3	0.4	0.1		
Column Penalty	0.375	0.325	0.275		

Table 7. Third allocation

Dairy products	L_1	L_3	Availability	Row penalty
Milk Powder	0.6	0	0.2	0.3
Butter	0.2 [0.1]	0.2	0.1	0.2
Cheese	0.2	0.2	0.1	0.2
Requirement	0.3	0.1		
Column Penalty	0.334	0.134		

Table 8. Fourth allocation

Dairy products	L_1	L_3	Availability	Row penalty
Milk powder	0.6	0	0.2	0.3
Cheese	0.2 [0.1]	0.2	0.1	0.2
Requirement	0.2	0.1		
Column penalty	0.4	0.1		

Table 9. Final allocations

Dairy products	L_1	L_3	Availability
Milk Powder	$0.6^{0.1}$	$0^{0.1}$	0.2
Requirement	0.1	0.1	

Step 11: Repeat steps 5–10 until all allocations have been met.

By step 12, utilizing the suggested algorithm, Table 3-9 displays the defuzzified quantities and full allocations of demand and availability. The suffix entries of each Table from 5-9 display allocations. We used the MODI approach after getting the IBFS, and the overall minimal transit cost was 0.18.

Example 5.2. With the capacity of production facilities S, T, and U must supply goods to three separate sites, P1, P2, and P3. Table 10 displays the mobility problem matrix together with demand and supply data as picture fuzzy numbers. The total least transportation expense must be determined.

For the complete transportation issue, which involves supply and demand, use the prescribed rank function to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 0.08.

Example 5.3. Three plants producing fertilizers have been established across the nation, and their products will be delivered to the states of J, K, and L, respectively. The cost of transportation must be kept to a minimum. Table 11 shows the cost of transportation for the items in picture-fuzzy settings while supply and demand information are presented in precise form.

For the complete transportation issue, which involves supply and demand, use the prescribed rank function

Table 10. Input data for PFTP

Production shops	P_1	P_2	P_3	Availability
S	(0.2,0.6,0.1)	(0.3,0.1,0.5)	(0.5,0.2,0.3)	(0.6,0.1,0.2)
T	(0.7,0.2,0.1)	(0.4,0.3,0.2)	(0.8,0,0.1)	(0.5,0.4,0)
U	(0.4,0.3,0)	(0.1,0.8,0.1)	(0.6,0.2,0.2)	(0.7,0.1,0.1)
Requirement	(0.5,0.1,0.1)	(0.3,0.6,0.1)	(0.9,0,0)	

Table 11. Input data for PFTP

Factories	J	K	L	Availability
X_1	(0.8,0.1,0.1)	(0.5,0.2,0.1)	(0.9,0,0)	26
X_2	(0.4,0.6,0)	(0.3,0.4,0.2)	(0.2,0.7,0)	42
X_3	(0.7,0,0.3)	(0.6,0.1,0.1)	(0.5,0.3,0)	33
Requirement	19	72	58	

Table 12. Input data for PFTP

Factories	*q*	*r*	*s*	*t*	Availability
F_1	(0.5,0.3,0)	(0.3,0.4,0.2)	(0.8,0.1,0.1)	(0.6,0.2,0.1)	82
F_2	(0.2,0.4,0.2)	(0.8,0.1,0)	(0.7,0.5,0.1)	(0.6,0,0)	58
F_3	(0.7,0.1,0.1)	(0.6,0.2,0.2)	(0.1,0.8,0.1)	(0.3,0.1,0.5)	30
F_4	(0.4,0.3,0.3)	(0.9,0.1,0)	(0.3,0.5,0)	(0.9,0,0)	36
Requirement	83	42	25	56	

to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 22.5.

Example 5.4. Four cement plants and four local distributors are owned by a corporation. From the following Table 12, we must determine the minimum price for transportation.

For the complete transportation issue, which involves supply and demand, use the prescribed rank function to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 59.6.

Example 5.5. Four plants are owned by a fan manufacturing company. The corporation has three separate warehouse locations. The next Table 13 provided picture-fuzzy values that represented the warehouse demand and the maximum capabilities of the plants. For it, we need to discover minimal transit costs.

For the complete transportation issue, which involves supply and demand, use the prescribed rank function to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 0.11.

Example 5.6. The issue of figuring out the minimal cost plan for supplying dealers with the appropriate quantity of cars is one that an automobile dealer must solve. The pertinent facts are shown in Table 14 below, where supply, as well as demand, were depicted as picture fuzzy numerals. The cost of transportation must be kept to a minimum.

For the complete transportation issue, which involves supply and demand, use the prescribed rank function to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 78.8.

Example 5.7. A petroleum business must deliver fuel to four distinct depots from its four refineries. The price of delivering one unit of fuel coming from each refinery to each depot is shown in Table 15. For it, we need to discover minimal transit costs.

Table 13. Input data for PFTP

Plants	W_1	W_2	W_3	Availability
U	(0.3,0.4,0.1)	(0.5,0.3,0.1)	(0.8,0.2,0)	(0.7,0.2,0.1)
V	(0.5,0.2,0.1)	(0.6,0.1,0.1)	(0.9,0,0)	(0.3,0.6,0.1)
W	(0.7,0.2,0.1)	(0.1,0.8,0.1)	(0.3,0.1,0.5)	(0.6,0.1,0.2)
X	(0.5,0.2,0.2)	(0.7,0.3,0)	(0.4,0.4,0.1)	(0.3,0.4,0.2)
Requirement	(0.6,0.1,0.3)	(0.8,0.3,0.1)	(0.5,0.3,0.1)	

Table 14. Input data for PFTP

Plant	*a*	*b*	*c*	*d*	Availability
A	25	28	40	34	(0.8,0.3,0)
B	66	54	43	52	(0.8,0.1,0.1)
C	36	30	56	48	(0.9,0,0)
Requirement	(0.5,0.2,0.2)	(0.6,0.1,0.1)	(0.9,0.1,0)	(0.9,0.1,0.1)	

Table 15. Input data for PFTP

Refinery	*p*	*q*	*r*	*s*	Availability
l	40	25	46	34	(0.5,0.2,0.1)
m	32	38	42	50	(0.5,0.3,0)
n	44	60	22	56	(0.7,0.1,0.1)
o	28	45	33	40	(0.8,0.1,0)
Requirement	(0.9,0.1,0)	(0.7,0.3,0.1)	(0.7,0.1,0.2)	(0.4,0.4,0.2)	

Table 16. Comparative analysis

	Numerical Examples						
	1	**2**	**3**	**4**	**5**	**6**	**7**
VAM	0.47	0.08	27.1	62.6	0.11	79.2	46.6
Existing [35]	0.41	0.08	27.1	59.6	0.13	78.2	46.3
Proposed	0.18	0.08	22.5	59.6	0.11	78.2	44.1
Optimum	0.18	0.08	22.5	52.1	0.11	78.2	44.1

For the complete transportation issue, which involves supply and demand, use the prescribed rank function to transform picture fuzziness into accurate numbers. Following that, applying the offered algorithm, we were able to determine the total minimal cost of transportation, which is 44.1.

RESULTS AND DISCUSSION

Data derived from transit issues are frequently ambiguous in many real-world circumstances. Fuzzy mathematics is a way to quantify such ambiguity to deal with issues like these. In this paper, we suggested a method for choosing the initial basic feasible answer to the transportation problem with a picture fuzzy. The capacity, consumer demand, and cost of the item's transport are all represented as picture numbers that are fuzzier in our algorithm. For capturing ambiguous, inconsistent, and imperfect information, picture fuzzy collections are preferable. In contrast to intuitionistic fuzzy sets, picture fuzzy sets take refusal membership grades into account. In Table 16 below, the disparity between the proposed method and the current one is displayed. The created PFTP framework, except for the first quantitative instance, produces results that are satisfactory and nearly identical to those of the comparing technique, based on comparison results. This indicates that the new algorithm and ranking function offer a better answer than the one that already exists. We took six mathematical instances representing random PFTP values into consideration and solved them using the proposed method as well as the existing approaches to demonstrate the efficacy of the proposed approach. On the other hand, the problems that we randomly selected incorporate three different forms of balanced and imbalanced PFTPs. Executives in the fields of transportation and supply chain management could potentially be guided in making judgments because the process is straightforward regarding the steps required for the computation. Figures 2 and 3 shows the comparison chart of proposed with existing and traditional approach.

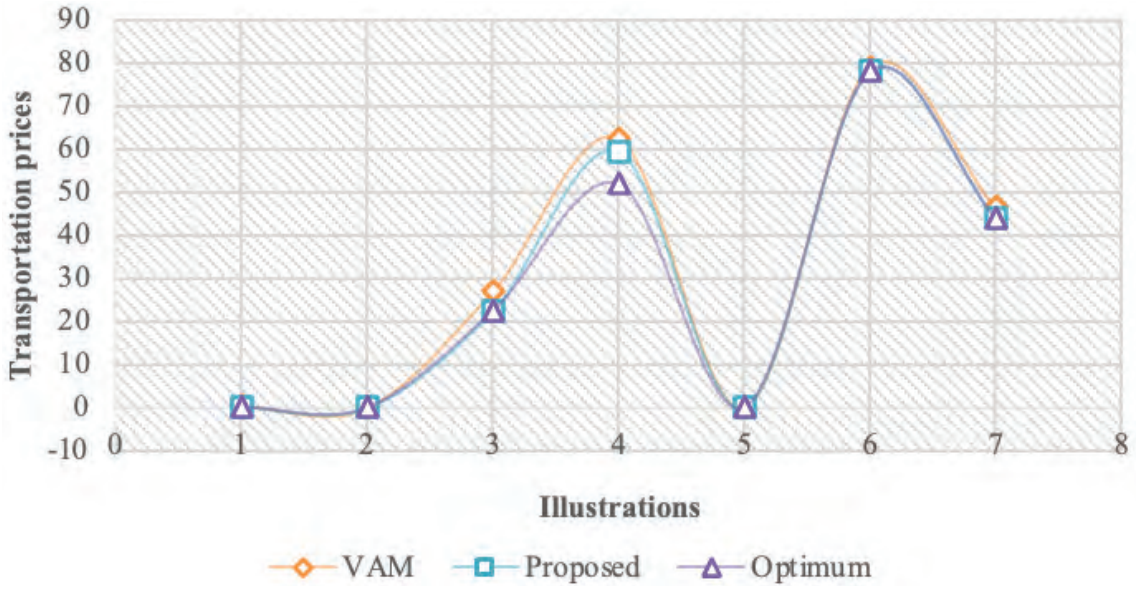


Figure 2. Comparison of proposed with traditional method.

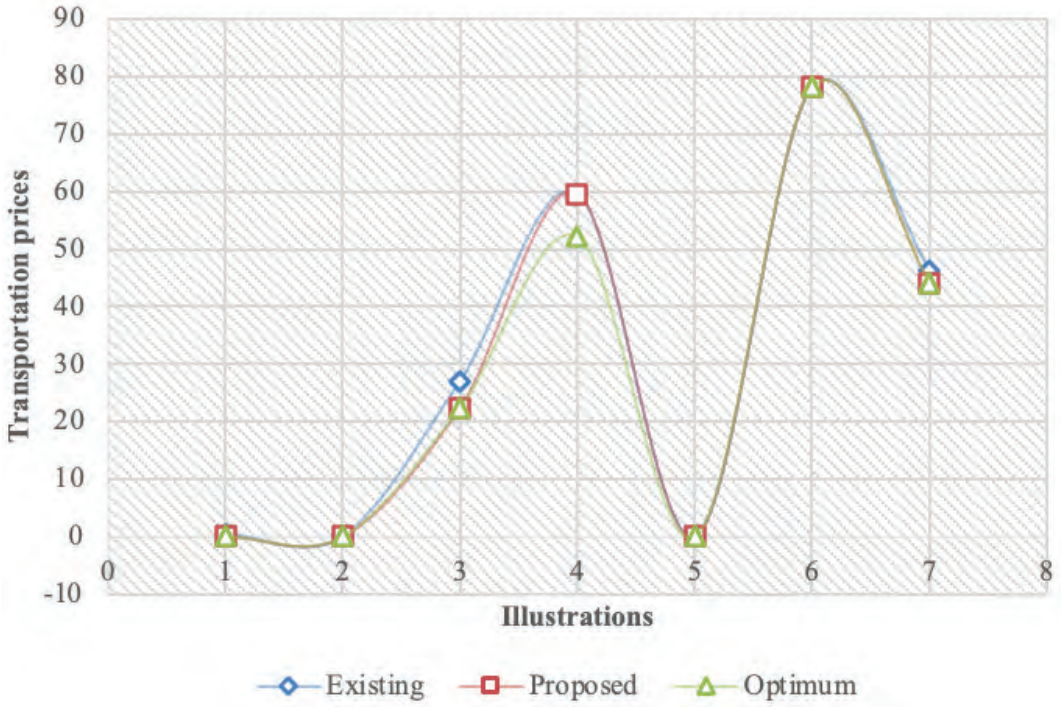


Figure 3. Comparison of proposed with existing method.

CONCLUSION

In this study, a specific family of uncertain transportation problems, the picture fuzzy transportation problem is optimized. Many published works have investigated a wide range of transport challenges of varying complexity. The characteristics of transportation can change over a particular period due to globalization in the current technology era as well as other inevitable factors. Instead of employing the traditional technique to resolve transit concerns, we investigated picture imprecise data and devised an easy alternating solution. In practical situations, a ranking approach is required to handle a transportation challenge in a fuzzy environment. In the current study, we first developed a rank function for transforming an uncertain picture figure into a crisp figure, followed by an algorithm for determining the first fundamental feasible solution to the picture fuzzy transportation problem. The effectiveness of the proposed approach has been supported by numerical evidence obtained through a thorough comparison of the proposed method to other existing approaches. Demonstrating the approach with numerical examples revealed no weaknesses. Thus, the suggested work's objective has been achieved. The suggested technique will be used to address the transshipment issue, in which sources may pass via intermediate locations before getting to their final destination. Among the applications of the transportation dilemma, the proposed strategy resolves issues with allocation and physical distribution. The proposed ranking approach cannot handle complex picture fuzzy sets. In the future, we plan to expand our inquiry into assignment and fractional transit concerns in a picture-fuzzy scenario.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Kalyani S, Nagarani S. A fully fuzzy transportation problem with hexagonal fuzzy number. AIP Conf Proc 2020;1. [CrossRef]

[2] Maity G, Mardanya D, Roy SK, Weber GW. A new approach for solving dual-hesitant fuzzy transportation problem with restrictions. Springer 2019;44. [CrossRef]

[3] Kumar PS. Algorithms for solving the optimization problems using fuzzy and intuitionistic fuzzy set. Int J Syst Assur Eng Manag 2020;11:189–222. [CrossRef]

[4] Kour D, Mukherjee S, Basu K. Solving intuitionistic fuzzy transportation problem using linear programming. Int J Syst Assur Eng Manag 2017;8:1090–101. [CrossRef]

[5] Hedid M, Zitouni R. Solving the four index fully fuzzy transportation problem. Oper Res 2020;11:199–215.

[6] Zadeh LA. Fuzzy sets. Inf Control 1965;8:338–353. [CrossRef]

[7] Bellman RE, Zadeh LA. Decision-making in a fuzzy environment. Manag Sci 1970;17:B–141. [CrossRef]

[8] Atanassov KT. Intuitionistic fuzzy sets. Fuzzy Sets Syst 1986;20:87-96. [CrossRef]

[9] Das KN, Das R, Acharjya DP. Least-looping stepping-stone-based ASM approach for transportation and triangular intuitionistic fuzzy transportation problems. Complex Intell Syst 2021;7:2885–94. [CrossRef]

[10] Mishra LN, Raiz M, Mishra VN. Tauberian theorems for weighted means of double sequences in intuitionistic fuzzy normed spaces. Yugosl J Oper Res 2022;32:377–388. [CrossRef]

[11] Dutta P, Ganju S. Some aspects of picture fuzzy set. Trans A Razmadze Math Inst 2018;172:164–175. [CrossRef]

[12] Cuong BC, Kreinovich V. Picture fuzzy sets-a new concept for computational intelligence problems. IEEE Xplore 2013;1–6. [CrossRef]

[13] Cuong BC, Kreinovich V. Picture fuzzy sets. J Comput Sci Cybern 2014;30:409–420.

[14] Phong PH, Hieu DT, Ngan RT, Them PT. Some compositions of picture fuzzy relations. Appl Inform Tech Res 2014;1:19–20.

[15] Cuong BC, Pham VH. Some fuzzy logic operators for picture fuzzy sets. IEEE Xplore 2015;1:132–137. [CrossRef]

[16] Cuong BC, Ngan RT, Hai BD. An involutive picture fuzzy negator on picture fuzzy sets and some De Morgan triples. IEEE Xplore 2015;1:126–131. [CrossRef]

[17] Van VP, Thi MCH, Pham VH. Some extensions of membership graphs for picture inference systems. IEEE Xplore 2015;1:192–197. [CrossRef]

[18] Singh P. Correlation coefficients for picture fuzzy sets. J Intell Fuzzy Syst 2015;28:591–604. [CrossRef]

[19] Cuong BC, Kreinovitch V, Ngan RT. A classification of representable t-norm operators for picture fuzzy sets. IEEE Xplore 2016;1:19–24. [CrossRef]

[20] Son LH. Generalized picture distance measure and applications to picture fuzzy clustering. Appl Soft Comput 2016;46:284–295. [CrossRef]

[21] Son LH. Measuring analogousness in picture fuzzy sets: from picture distance measures to picture association measures. Fuzzy Optim Decis Mak 2017;16:359–378. [CrossRef]

[22] Son LH, Viet PV, Hai PV. Picture inference system: a new fuzzy inference system on picture fuzzy set. Appl Intell 2017;46:652–669. [CrossRef]

[23] Wei G. Some similarity measures for picture fuzzy sets and their applications. Iran J Fuzzy Syst 15 2018;15:77–89.

[24] Kumar S, Arya V, Kumar S, Dahiya A. A new picture fuzzy entropy and its application based on combined picture fuzzy methodology with partial weight information. Int J Fuzzy Syst 2022;24:3208–3225. [CrossRef]

[25] Kirişci M. Generalized Pythagorean fuzzy sets and new decision-making method. Sigma J Eng Nat Sci 2022;40:806–813. [CrossRef]

[26] Hemalatha K, Venkateswarlu B. Pythagorean fuzzy transportation problem: New way of ranking for Pythagorean fuzzy sets and mean square approach. Heliyon 2023;9. [CrossRef]

[27] Akram M, Bashir A, Garg H. Decision-making model under complex picture fuzzy Hamacher aggregation operators. Comput Appl Math 2020;39:1–38. [CrossRef]

[28] Peng X, Dai J. Algorithm for picture fuzzy multiple attribute decision-making based on new distance measure. Int J Uncertain Quantif 2017;7. [CrossRef]

[29] Garg H. Some picture fuzzy aggregation operators and their applications to multicriteria decision-making. Arab J Sci Eng 2017;42:5275–5290. [CrossRef]

[30] Shit C, Ghorai G, Xin Q, Gulzar M. Harmonic aggregation operator with trapezoidal picture fuzzy numbers and its application in a multiple-attribute decision-making problem. Symmetry 2022;14:135. [CrossRef]

[31] Akram M, Ullah I, Allahviranloo. A new method to solve linear programming problems in the environment of picture fuzzy sets. Iran J Fuzzy Syst 2022;19: 29–49.

[32] Akram M, Habib A, Alcantud JCR. An optimization study based on Dijkstra algorithm for a network with trapezoidal picture fuzzy numbers. Neural Comput Appl 2021;33:1329–1342. [CrossRef]

[33] Gundogdu FK, Duleba S, Moslem S, Aydin S. Evaluating public transport service quality using picture fuzzy analytic hierarchy process and linear assignment model. Appl Soft Comput 2021;100:106920. [CrossRef]

[34] Mehmood MA, Bashir S. Extended Transportation Models Based on Picture Fuzzy Sets. Math Probl Eng 2022.

[35] Geetha SS, Selvakumari K. A picture fuzzy approach to solving transportation problem. Eur J Mol Clin Med 2020;7:4982–4990.

[36] Yolcu A. Intuitionistic fuzzy hypersoft topology and ıts applications to multi-criteria decision-making. Sigma J Eng Nat Sci 2023;41:106–118. [CrossRef]

[37] Singh S, Ganie AH. Applications of a picture fuzzy correlation coefficient in pattern analysis and decision-making. Granul Comput 2022:1–15. [CrossRef]

[38] Srinivasan R, Karthikeyan N, Jayaraja A. A Proposed Technique to Resolve Transportation Problem by Trapezoidal Fuzzy Numbers. Indian J Sci Technol 2021;14:1642–1646. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1094–1110, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2030

Research Article

Bayesian computational methods for estimation of q-Weibull distribution

Ibrahim SADOK[1,*]

[1]*Department of Mathematics and Computer Science, Faculty of Exact Sciences, University of Bechar, Bechar, 08000, Algeria, Laboratory of Mathematics, Djillali Liabes University of Sidi Bel-Abbès, P. O. Box 89, 22000, Sidi Bel-Abbès, Algeria*

ARTICLE INFO

Article history
Received: 17 October 2024
Revised: 02 January 2025
Accepted: 13 February 2025

Keywords:
Bayesian Inference; Metropolis-Hastings Algorithm; MLE; Parameter Estimation; Reliability Analysis; Q-Weibull Distribution

ABSTRACT

The q-Weibull distribution is a flexible probability distribution that is commonly used in reliability analysis, survival analysis, and extreme value modelling. Obtaining accurate parameter estimates is essential for these applications. This paper investigates several estimation techniques for the q-Weibull distribution, with special emphasis on the Metropolis–Hastings algorithm. Through an extensive simulation study, we evaluate the performance of these methods across different sample sizes. In addition, we apply them to a real dataset to illustrate their practical utility and to highlight situations where the Metropolis–Hastings approach proves particularly advantageous.

Cite this article as: Sadok I. Bayesian computational methods for estimation of q-Weibull distribution. Sigma J Eng Nat Sci 2026;44(2):1094–1110.

INTRODUCTION

The Weibull distribution has served as a workhorse in reliability analysis across many disciplines. In medicine, the Weibull distribution is often used to estimate survival times for cancer patients, helping researchers better understand treatment effectiveness and expected life spans [1]. In engineering, it plays an important role in predicting the fatigue life of mechanical components that are exposed to repeated or cyclic stresses [2]. It is also widely applied in the wind energy field, where it helps describe extreme wind speeds for turbine design and risk evaluation [3]. In finance, analysts use it to model stock price variations and capture rare but significant market events [4]. Moreover, in image processing, the Weibull distribution can be useful for modeling different types of noise patterns [5].

The classical Weibull distribution has the following probability density function (pdf):

$$f(x|k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad (1)$$

where $k > 0$ denotes the shape parameter, and $\lambda > 0$ represents the scale parameter. This formulation describes the distribution using two parameters, which together provide a high degree of flexibility in modeling different types of data, particularly in applications involving failure times and survival analysis.

Equation (1) can be interpreted as a generalization of the exponential distribution, since it reduces to the exponential case when the shape parameter $k = 1$. One of the main reasons for the widespread use of the Weibull distribution is its ability to capture the behavior of many

*Corresponding author.
*E-mail address: ibrahim.sadok@univ-bechar.dz

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

real-world systems. In practice, such systems often exhibit a relatively high probability of failure at early stages, followed by a decreasing failure rate over time [6]. For this reason, the Weibull distribution has become a standard tool in reliability engineering and survival analysis.

However, its broad applicability, the classical Weibull distribution may not adequately describe data with heavy tails or complex, non-extensive behavior. To overcome these limitations, several generalized forms have been proposed. Among them, a notable extension is based on the q-exponential function, leading to the so-called q-Weibull distribution [7–9].

The q-exponential function, which generalizes the classical exponential function, is defined:

$$e_q(x) = [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (2)$$

where $x \in \mathbb{R}$ and q determines the level of non-extensive behavior in the model. When $q = 1$, the q-exponential function reduces to the traditional exponential function [10]. The q-exponential definition incorporates a cut-off criterion that effectively precludes negative or complex values, a critical attribute when associating the function with probabilities. Notably, for select parameter values, the q-exponential displays a transition between an exponential profile and a power-law regime [11, 12]. More specifically, when $a > 0$ and $q > 1$, $e_q(2ax)$ exhibits power-law behavior for large values of x, leading to heavy-tailed distributions..

The q-Weibull distribution provides greater flexibility and adaptability, particularly in datasets characterized by power-law behaviour or non-extensive statistics. This generalized form has been applied to diverse areas, such as dielectric breakdown modelling in electronic devices [13] and financial market volatility analysis [14]. Recent applications in environmental sciences further demonstrate the relevance of distribution in capturing extreme weather events and modelling climate change and flood data .

In this article, we introduce a Bayesian approach for parameter estimation in the context of the q-Weibull distribution. The estimation procedure relies on Markov Chain Monte Carlo (MCMC) techniques, where Gibbs sampling is used to iteratively update the parameters of the distribution. In recent years, advances in Markov Chain Monte Carlo (MCMC) methods, including adaptive algorithms [15, 16] and variational inference techniques [17], have significantly improved both convergence properties and computational efficiency. In this study, particular attention is given to the Metropolis–Hastings algorithm [18] and two of its widely used variants: the Independent Metropolis–Hastings (IMH) and the Random Walk Metropolis (RWM) algorithms [19].

Bayesian estimation methods, as considered in this work, offer several advantages in practical applications such as engineering reliability and survival analysis. In particular, their ability to incorporate prior information and to explicitly account for uncertainty makes them especially valuable when dealing with complex real-world data.

This study provides a comparative analysis of maximum likelihood estimation (MLE) and Bayesian approaches (specifically IMH and RWM) for estimating the parameters of the q-Weibull distribution. Although the q-Weibull model is recognized for its flexibility in capturing diverse data behaviors, the issue of reliable parameter estimation has not yet been thoroughly investigated. Through an extensive simulation study covering different sample sizes, the performance of MLE, IMH, and RWM is examined. The results indicate that the RWM approach provides improved performance compared with standard MCMC implementations when estimating posterior distributions. In addition, the study illustrates the usefulness of Bayesian methods through applications to real datasets, emphasizing their relevance in fields such as engineering and the natural sciences.

The practical importance of this work lies in the flexibility of the q-Weibull distribution for modelling a wide range of phenomena. In engineering, it is commonly used in reliability analysis, for instance, to model material fatigue under varying stress conditions. In the natural sciences, the distribution is well suited for survival modelling and extreme-value analysis, where classical models often struggle to capture tail behaviour. Moreover, the comparative assessment of estimation techniques (particularly the advantages of Bayesian methods such as IMH and RWM) offers useful guidance for researchers and practitioners working with complex and heterogeneous datasets.

The subsequent sections of this paper are structured as follows. Section 2 introduces the q-Weibull model and delineates its characteristic statistics. Section 3 presents the frequentist analysis of the q-Weibull distribution, where the parameters are estimated using the MLE and the corresponding confidence intervals are derived. Section 4 then introduces the Bayesian framework developed for parameter estimation. We elucidate the resampling procedures employing both IMH and RWM algorithms. Section 5 presents a comprehensive simulation study. Section 6 demonstrates the practical application of the proposed methods through the analysis of two real-world datasets. Finally, Section 7 summarizes the main findings of the study.

BACKGROUND

The q-Weibull model is derived by replacing the exponential function in the classical Weibull model (represented by equation (1)) with a q-exponential. A detailed explanation of this substitution can be found in [13]. The resulting expression for the q-Weibull pdf is given by

$$f_q(x|q, k, \lambda) = (2 - q)\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e_q\left(-\left(\frac{x}{\lambda}\right)^k\right), \quad (3)$$

where $k > 0$ and $q < 2$ are shape parameters, and $\lambda > 0$ is a scale parameter.

By substituting the q-exponential function $e_q(x)$ into equation (3), we can rewrite the pdf of the q-Weibull distribution as follows:

$$f_q(x|q,k,\lambda) = (2-q)\frac{k}{\lambda^k}x^{k-1}\left[1-(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{1}{1-q}}$$

where the domain of the q-Weibull distribution is given by the following piecewise function:

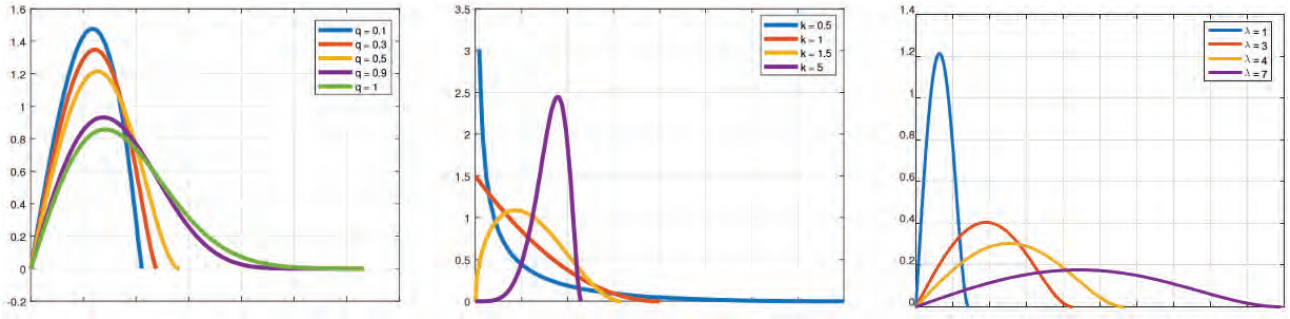
$$x \in \begin{cases}[0,+\infty) & 1 < q < 2 \\ \left[0,\lambda(1-q)^{-1/k}\right] & q < 1\end{cases}$$

As the parameter $q \to 1$, the function f_q transforms into the Weibull pdf (1). When k is set to 1, it corresponds to the q-exponential pdf (2). Moreover, when both q and k tend towards 1, it transitions into the exponential distribution.

Consider a random sample, denoted as $x = (x_1, x_2, \ldots, x_n)$, drawn from the q-Weibull distribution. The likelihood function for (q, k, λ) given x, takes the form

$$L(q,k,\lambda|x) = \prod_{i=1}^{n} f_q(x|q,k,\lambda) = \prod_{i=1}^{n}(2-q)\frac{k}{\lambda^k}x^{k-1}\left[1-(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{1}{1-q}} \quad (4)$$

Figure 1 illustrates density curves for the q-Weibull distribution, highlighting the effects of varying parameters q, k, and λ. In the first set of curves, $k = 2$ and $\lambda = 1$ are held constant, while q varies from 0.1 to 1. This clearly shows how the parameter q affects the tail behaviour: smaller values of q lead to heavier tails, indicating greater dispersion in the data. In the second set, $q = 0.5$ and $\lambda = 1$ are fixed while



(a) (b) (c)

Figure 1. q-Weibull probability density function with reference parameters $q = 0.5$, $k = 2$, $\lambda = 1$ with changing: (a) parameter q, (b) shape parameter k, (c) scale parameter λ.

Table 1. Statistical characteristics of the q-Weibull Distribution for $q < 1$.

Characteristics	Functional form
pdf	$(2-q)\frac{k}{\lambda^k}x^{k-1}\left[1-(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{1}{1-q}}$
cdf	$1-\left[1-(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{2-q}{1-q}};\ 0 \le x \le \lambda(1-q)^{-1/k}$
hrf	$\frac{k(2-q)x^{k-1}}{\lambda^k\left[1-(1-q)\left(\frac{x}{\lambda}\right)^k\right]};\ \lvert x\rvert < \lambda(1-q)^{-1/k}$
s^{th} moment	$\frac{\lambda^s(2-q)}{(1-q)^{1+s/k}}B\left(\frac{s}{k}+1,\frac{2-q}{1-q}\right)$
Mean	$\frac{\lambda(2-q)}{(1-q)^{1+1/k}}B\left(\frac{1}{k}+1,\frac{2-q}{1-q}\right)$
Variance	$\frac{\lambda^2(2-q)}{(1-q)^{1+2/k}}\left[B\left(\frac{2}{k}+1,\frac{2-q}{1-q}\right)-\frac{2-q}{1-q}B^2\left(\frac{1}{k}+1,\frac{2-q}{1-q}\right)\right]$
Mode	$\lambda\left[\frac{k-1}{k+(k-1)(1-q)}\right]^{\frac{1}{k}};\ k>0$

Table 2. Statistical Characteristics of the q-Weibull Distribution for $1 < q < 2$.

Characteristics	Functional form
pdf	$(2-q)\frac{k}{\lambda^k}x^{k-1}\left[1+(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{1}{1-q}}$
cdf	$1-\left[1+(1-q)\left(\frac{x}{\lambda}\right)^k\right]^{\frac{2-q}{1-q}};\ x>0$
hrf	$\frac{k(2-q)x^{k-1}}{\lambda^k\left[1+(1-q)\left(\frac{x}{\lambda}\right)^k\right]}$
s^{th} moment	$\frac{\lambda^s(2-q)}{(1-q)^{1+s/k}}B\left(\frac{s}{k}+1,\frac{1}{q-1}-\frac{s}{k}-1\right)$
Mean	$\frac{\lambda(2-q)}{(1-q)^{1+1/k}}B\left(\frac{1}{k}+1,\frac{1}{q-1}-\frac{1}{k}-1\right)$
Variance	$\frac{\lambda^2(2-q)}{(1-q)^{1+2/k}}\left[B\left(\frac{2}{k}+1,\frac{1}{q-1}-\frac{2}{k}-1\right)+\frac{2-q}{1-q}B^2\left(\frac{1}{k}+1,\frac{1}{q-1}-\frac{1}{k}-1\right)\right]$
Mode	$\lambda\left[\frac{k-1}{k-(k-1)(1-q)}\right]^{\frac{1}{k}};\ k>0$

k ranges from 0.5 to 5. Here, we see that k mainly controls the shape and the height of the peak, with larger values producing sharper and more pronounced peaks. In the third set of curves, q = 0.5 and k = 1 are kept constant, and λ varies from 1 to 7. This highlights its role as a scale parameter, shifting the distribution along the horizontal axis. The significance of studying the q-Weibull distribution lies in its ability to generate diverse density curves by varying parameters q, k, and λ, which influence tail behaviour, shape, and scale, respectively. This flexibility enables the modelling of various data types, from light-tailed to heavy-tailed distributions. These characterisations allows for a clear visualization of how changes in each parameter impact the distribution's behaviour, making these relationships more intuitive and accessible.

Tables 1 and 2 present a detailed summary of the main statistical properties of the q-Weibull distribution for the cases $q < 1$ and $1 < q < 2$, respectively. Each table lists the probability density function (pdf), cumulative distribution function (cdf), hazard rate function (hrf), the s^{th} moment, mean, variance, and mode, along with their corresponding mathematical formulas. Additionally, the tables reference the beta function $B(x,y)$, defined as:

$$B(x,y)=\int_0^1 t^{x-1}(1-t)^{y-1}dt$$

This information collectively offers a detailed insight into the behaviour and properties of the q-Weibull distribution for cases $q < 1$ and $1 < q < 2$.

FREQUENTIST ANALYSIS OF THE q-WEIBULL DISTRIBUTION

Maximum Likelihood Estimates of the q-Weibull Distribution Parameters

In engineering applications, the primary challenge lies in the estimation of the q-Weibull parameters. Among the available methods for estimating q-Weibull distribution parameters, the most commonly used one is the MLE [11]. However, it is important to note that the first derivatives of the log-likelihood function with respect to the parameters are non-linear, making it extremely challenging to derive analytical solutions.

The logarithm of the likelihood function for this sample, which quantifies the probability of observing these values in the given distribution, can be expressed as follows:

$$\ln L(q,k,\lambda|x)=n\ln(2-q)+n\ln k-kn\ln\lambda+(k-1)\sum_{i=1}^{n}\ln(x_i)+\frac{1}{1-q}\sum_{i=1}^{n}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right]. \quad (6)$$

The partial derivatives of the log-likelihood function concerning the parameters q, k, and λ are as follows:

$$\frac{\partial \ln L(q,k,\lambda|x)}{\partial q} = \frac{n}{2-q} + \frac{1}{(1-q)}\sum_{i=1}^{n}\frac{x_i^k}{\lambda^k-(1-q)x_i^k} + \frac{1}{(1-q)^2}\sum_{i=1}^{n}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right],$$

$$\frac{\partial \ln L(q,k,\lambda|x)}{\partial k} = \frac{n}{k} - n\ln\lambda + \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\frac{x_i^k\ln\left(\frac{x_i}{\lambda}\right)}{\lambda^k-(1-q)x_i^k},$$

$$\frac{\partial \ln L(q,k,\lambda|x)}{\partial \lambda} = -\frac{k}{\lambda}\left(n - \sum_{i=1}^{n}\frac{x_i^k}{\lambda^k-(1-q)x_i^k}\right).$$

Because of the non-linearity of these derivatives, obtaining closed-form solutions for the MLEs, namely $\hat{q}$, $\hat{k}$, and $\hat{\lambda}$, is a formidable task. To tackle this challenge, a constrained optimization problem is formulated, seeking to maximize the natural logarithm of the likelihood function ln$L$ given by equation (6) while satisfying the conditions $q < 2$, $k > 0$, and $\lambda > 0$.

The q-Weibull model indeed poses a disadvantage due to the complexity of parameter estimation. Given the intricacies of the calculation process and the substantial computational requirements for each method, we have resorted to the Newton–Raphson method to obtain the optimal parameters. This approach reduces the challenges of estimating the q-Weibull parameters, leading to more accurate results in engineering applications.

Confidence Intervals for the q-Weibull Parameters

In this section, we focus on constructing confidence intervals (CIs) for both the q-Weibull parameters and the reliability parameters, based on the asymptotic normality property of maximum likelihood estimators (MLEs). To achieve this, we frequently turn to the Fisher information matrix, a valuable tool in establishing CIs for parameters when capitalizing on the asymptotic normality of the MLEs.

In this context, let us consider the parameter of interest, referred to as Θ. Under specified regularity conditions, it can be observed that the MLEs $\hat{\Theta}$approximate a normal distribution $\mathcal{N}(\Theta, I^{-1}(\Theta))$ as the sample size $n \to \infty$. Here, $I^{-1}(\Theta)$ represents the inverse matrix derived from the Fisher information matrix. Notably, the components of the Fisher information matrix are directly linked to the expected values of the second-order derivatives of ln$L$, as exemplified in equation (6). This matrix is expressed as

$$I(q,k,\lambda) = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q^2} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q\,\partial k} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q\,\partial \lambda} \\ \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k\,\partial q} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k^2} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k\,\partial \lambda} \\ \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda\,\partial q} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda\,\partial k} & \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda^2} \end{bmatrix}.$$

A previous study conducted in [20-28] utilized this approximation to determine the confidence intervals for the q-Weibull parameters. In this context, we will perform the derivation of the expected values. When calculating the expected values becomes a challenging task, an alternative approach is to estimate them by using the negations of the second-order derivatives of ln$L$ evaluated at the MLEs [5, 27]. The second-order derivatives of the natural logarithm of the likelihood function are as follows

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q^2} = -\frac{n}{(2-q)^2} - \frac{1}{1-q}\sum_{i=1}^{n}\frac{x_i^{2k}}{(\lambda^k-(1-q)x_i^k)^2} + \frac{2}{(1-q)^2}\sum_{i=1}^{n}\frac{x_i^k}{\lambda^k-(1-q)x_i^k} + \frac{2}{(1-q)^3}\sum_{i=1}^{n}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right],$$

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q\,\partial k} = \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k\,\partial q} = \sum_{i=1}^{n}\frac{x_i^{2k}\ln\left(\frac{x_i}{\lambda}\right)}{(\lambda^k-(1-q)x_i^k)^2},$$

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial q\,\partial \lambda} = \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda\,\partial q} = -\frac{k}{\lambda}\sum_{i=1}^{n}\frac{x_i^{2k}}{(\lambda^k-(1-q)x_i^k)^2},$$

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k^2} = -\frac{n}{k^2} - \lambda^k\sum_{i=1}^{n}\frac{x_i^k\ln^2\left(\frac{x_i}{\lambda}\right)}{(\lambda^k-(1-q)x_i^k)^2},$$

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial k\,\partial \lambda} = \frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda\,\partial k} = -\frac{1}{\lambda}\left(n - \sum_{i=1}^{n}\frac{x_i^k}{\lambda^k-(1-q)x_i^k}\right) + k\lambda^{k-1}\sum_{i=1}^{n}\frac{x_i^k\ln(x_i/\lambda)}{(\lambda^k-(1-q)x_i^k)^2},$$

$$\frac{\partial^2 \ln L(q,k,\lambda|x)}{\partial \lambda^2} = \frac{k}{\lambda^2}\left(n - \sum_{i=1}^{n}\frac{x_i^k}{\lambda^k-(1-q)x_i^k}\right) - k^2\lambda^{k-2}\sum_{i=1}^{n}\frac{x_i^k}{(\lambda^k-(1-q)x_i^k)^2}.$$

The covariance matrix for the MLEs of the q-Weibull distribution parameters can be estimated by taking the inverse of the observed information matrix $I(q, k, \lambda)$. This covariance matrix is also equivalent to the inverse of the Fisher information matrix, which is

$$Cov(q,k,\lambda) = I^{-1}(q,k,\lambda) = \begin{bmatrix} var_{11} & cov_{12} & cov_{13} \\ cov_{21} & var_{22} & cov_{23} \\ cov_{31} & cov_{32} & var_{33} \end{bmatrix}.$$

After obtaining the covariance matrix, we can proceed to construct asymptotic confidence intervals for the q-Weibull distribution parameters. These intervals provide an estimate of the parameter values with a (1- α)100% level of confidence. The following are the asymptotic confidence intervals for the parameters q, k, and λ:

$$CI[q:(1-\alpha)100\%] = \left[\hat{q} + z_{\frac{\alpha}{2}}\sqrt{var_{11}}, \hat{q} + z_{1-\frac{\alpha}{2}}\sqrt{var_{11}}\right],$$

$$CI[k:(1-\alpha)100\%] = \left[\hat{k} + z_{\frac{\alpha}{2}}\sqrt{var_{22}}, \hat{k} + z_{1-\frac{\alpha}{2}}\sqrt{var_{22}}\right],$$

$$CI[\lambda:(1-\alpha)100\%] = \left[\hat{\lambda} + z_{\frac{\alpha}{2}}\sqrt{var_{33}}, \hat{\lambda} + z_{1-\frac{\alpha}{2}}\sqrt{var_{33}}\right].$$

In this context, $\hat{q}$, $\hat{k}$, and $\hat{\lambda}$ represent the ML estimators of the parameters. Meanwhile, $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}}$ correspond to the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the standard normal distribution, respectively. While var_{11}, var_{22}, and var_{33} correspond to the diagonal elements of the covariance matrix $Cov(q, k, \lambda)$.

BAYESIAN ANALYSIS OF THE *q*-WEIBULL DISTRIBUTION

In this section, we outline the Bayesian approach to estimate the parameters of the q-Weibull distribution. To begin, we assume that the prior distributions for the parameters (q, k, λ) are independent, meaning we can express the joint prior as the product of individual priors [21], i.e, $\pi(q, k, \lambda) = \pi(q)\pi(k)\pi(\lambda)$.

The main objective is to estimate the Bayesian values of the parameters: q, k, and λ. To proceed, we assign suitable prior distributions to each parameter according to their nature.

- For the shape parameter q, we consider a Beta prior, written as $\pi(q)$ = Beta $(q|q_1,q_2)$, where q_1 and q_2 are known hyper-parameters. This choice is natural since the Beta distribution is defined on the interval [0,1], which coincides with the admissible range of q.
- For the parameter k, we use a Gamma prior, denoted by $\pi(k) = \Gamma(k|k_1,k_2)$, with k_1 and k_2 are known hyperparameters. The Gamma distribution is versatile and can accommodate a wide range of shapes, making it a common choice for modelling k.
- Similarly, for the scale parameter λ, we assign a prior distribution, and the Gamma distribution is a popular choice, represented as $\pi(\lambda) = \Gamma(\lambda|\lambda_1, \lambda_2)$, with known hyper-parameters λ_1 and λ_2. The Gamma distribution is highly flexible and can accommodate a wide range of shapes, making it suitable for modeling positive parameters. It is also commonly used for scale-type parameters such as λ.

Remark: The hyperparameters of the priors are selected to balance computational convenience with limited prior influence. In particular, a Beta (1,1) prior is adopted for q, corresponding to a uniform distribution on [0,1], which reflects the absence of strong prior information. For k and λ, $\Gamma(0.1,0.1)$ priors were chosen to provide sufficient variance for robust parameter exploration. A sensitivity analysis, varying these hyper-parameters (e.g., Beta (0.5, 2) for q and $\Gamma(1,1)$ for k and λ), demonstrated that posterior estimates remained stable across the tested ranges, confirming the reliability of the chosen priors.

Using the Bayes theorem, the joint posterior distribution for (q, k, λ) upon which inference is based, is given by

$$\pi(q,k,\lambda|x) \propto L(q,k,\lambda|x)\pi(q)\pi(k)\pi(\lambda),$$

where $L(q, k, \lambda|x)$ is defined in equation (6).

The conditional posterior distributions are

$$q|x,k,\lambda \propto (2-q)^n \exp\left(\sum_{i=1}^n \frac{1}{1-q}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right]\right) q^{q_1-1}(1-q)^{q_2-1},$$

$$k|x,q,\lambda \propto \frac{k^{n+k_1-1}}{\lambda^{kn}} e^{-kk_2} \exp\left(\sum_{i=1}^n (k-1)\ln(x_i) + \frac{1}{1-q}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right]\right),$$

$$\lambda|x,q,k \propto \lambda^{\lambda_1-nk-1} e^{-\lambda\lambda_2} \exp\left(\sum_{i=1}^n \frac{1}{1-q}\ln\left[1-(1-q)\left(\frac{x_i}{\lambda}\right)^k\right]\right).$$

The complexity of the posterior distribution precludes the derivation of closed-form expressions for its properties. Consequently, our approach is to employ Markov chain Monte Carlo (MCMC) methods to obtain samples from this distribution, facilitating the extraction of parameter inferences.

Yet, the conditional posterior distribution for the parameters q, k, λ does not conform to any readily available closed distribution, and its resampling procedure is not straightforward. In such instances, the conventional Bayesian approach involves the utilization of the Metropolis-Hastings algorithm.

Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm represents the most widely recognized instances of MCMC techniques. This algorithm is employed to generate samples from arbitrary distributions for which we lack a direct method for generating random samples. In a manner analogous to acceptance-rejection sampling, the MH algorithm posits that at each iteration of the process, a potential value can be drawn from a proposal distribution.

Consequently, the candidate value is accepted or rejected during each iteration, contingent upon the calculation of an appropriate acceptance probability. This iterative process is essential for steering the Markov chain toward the target distribution, making it an effective method for sampling from complex models. For a more detailed explanation of the Metropolis–Hastings algorithm, readers are referred to the foundational works [21–23].

To update parameter q using the Metropolis-Hastings algorithm, suppose that (q, k, λ) represents the current state of the Markov chain. Let $q^\star$ be a candidate value generated from a proposal density denoted as $g(q^\star|q)$. The value $q^\star$ is then accepted with a probability denoted as $\Psi(q^\star|q) = \min(1, A_q)$, where A_q is defined as

$$A_q = \frac{L(q^*,k,\lambda)\pi(q^*)}{L(q,k,\lambda)\pi(q)} \frac{g(q|q^*)}{g(q^*|q)}, \quad (7)$$

where $L(.|y)$ represents the likelihood function, as described in equation (4).

In a practical context, the MH algorithm is executed as outlined below

Algorithm 1: Metropolis-Hastings Algorithm

1: **Initialization:** Initialize the iteration count $l = 1$ and set the initial state as: $(q^{(l-1)}, k^{(l)}, \lambda^{(l)})$

2: **While** $l \leq L$

Step 1: Generate a proposal state $q^{\star} \sim g(q^{\star}|q^{(l-1)})$.

Step 2: Calculate acceptance probability $\Psi(q^{\star}|q^{(l-1)}) = \min(1, A_q)$, where A_q is given in equation (7);

Step 3: Generate a random number $u \sim \mathcal{U}(0,1)$

If $u \leq \Psi\ (q^{\star}|q^{(l-1)})$

Accept $q^{\star}$ and set $q^{(l)} = q^{\star}$ **else**

Reject $q^{\star}$ and set $q^{(l)} = q^{(l-1)}$

end While

Two Typical Selections for *g*(.)

To effectively implement the MH algorithm, it is essential to define the candidate-generating density $g(q^{\star}|q)$. Typically, $g(.)$ is selected to ensure it allows for straightforward sampling. In the following section, we elaborate on two commonly utilized choices for $g(.)$.

- *Independent Metropolis-Hastings (IMH):* When $g(q^{\star}|q) = g(q^{\star})$, meaning that the candidate-generating density is independent of the current value, we obtain a special case of the original algorithm. In this simplified scenario, the expression for A_q becomes

$$A_q = \frac{L(q^*, k, \lambda)\pi(q^*)}{L(q, k, \lambda)\pi(q)} \frac{g(q)}{g(q^*)}$$

To address this situation, one can configure $g(q^{\star})$ to be the prior distribution, denoted as $g(q^{\star}) = \pi(q^{\star})$. Consequently, the expression for A_q can be derived by considering the likelihood ratio

$$A_q = \frac{L(q^*,k,\lambda)}{L(q,k,\lambda)}. \quad (8)$$

The implementation of this algorithm is detailed in the following manner.

Algorithm 2: Independent Metropolis-Hastings algorithm

1: **Initialization:** Initialize the iteration count $l = 1$ and set the initial state as: $(q^{(l-1)}, k^{(l)}, \lambda^{(l)})$

2: **While** $l \leq L$

Step 1: Generate a proposal state $q^{\star} \sim Beta(q_1, q_2)$

Step 2: Calculate acceptance probability $\Psi(q^{\star}|q^{(l-1)}) = \min(1, A_q)$, where A_q is given in (8);

Step 3: Generate a random number $u \sim \mathcal{U}(0,1)$

If $u \leq \Psi\ (q^{\star}|q^{(l-1)})$

Accept $q^{\star}$ and set $q^{(l)} = q^{\star}$ **else**

Reject $q^{\star}$ and set $q^{(l)} = q^{(l-1)}$

end While

- *Random Walk Metropolis (RWM):* When the candidate generating-density, denoted as $g(q^{\star}|q)$, exhibits symmetry and the likelihood of transitioning from state q to $q^{\star}$ depends solely on the distance between them, that is, $g(q^{\star}|q) = d(|q\text{-}q^{\star}|)$, with d representing a symmetric density, the expression for A_q can be simplified as follows

$$A_q = \frac{L(q^*,k,\lambda)\pi(q^*)}{L(q,k,\lambda)\pi(q)}, \quad (9)$$

While the choice of prior distributions as the candidate generating-density is theoretically elegant, it can lead to numerous rejections of proposed moves, resulting in slow convergence of the algorithm. This issue is particularly pronounced when prior information is unavailable and the prior distribution has a high variance. By employing the RWM method, we can explore the vicinity of the current chain value to propose a new one, offering a more efficient convergence technique.

To implement this approach, we generate a new value $q^{\star}$ by setting it as $q^{\star} = q + \varepsilon$, where ε is a random perturbation drawn from a Normal distribution with a mean of 0 and a variance of σ_q^2, represented as $\varepsilon \sim \mathcal{N}(0, \sigma_q^2)$. This implies that $q^* \sim \mathcal{N}(q, \sigma_q^2)$. The algorithm is implemented as follows

Algorithm 3: Random Walk Metropolis algorithm

1: **Initialization:** Initialize the iteration count $l = 1$ and set the initial state as: $(q^{(l-1)}, k^{(l)}, \lambda^{(l)})$

2: **While** $l \leq L$

Step 1: Generate $\varepsilon \sim \mathcal{N}(0, \sigma_q^2)$ and set $q^{\star} = q^{(l-1)} + \varepsilon$

Step 2: Calculate acceptance probability $\Psi(q^{\star}|q^{(l-1)}) = \min(1, A_q)$, where A_q is given in (9);

Step 3: Generate a random number $u \sim \mathcal{U}(0,1)$

If $u \leq \Psi\ (q^{\star}|q^{(l-1)})$

Accept $q^{\star}$ and set $q^{(l)} = q^{\star}$ **else**

Reject $q^{\star}$ and set $q^{(l)} = q^{(l-1)}$

end While

The parameter σ_q^2 plays a crucial role in the RWM algorithm, as it determines the size of the random perturbations added to the current state when proposing new moves in the Markov chain. In other words, it controls the step size of the chain as it explores the state space. Selecting the appropriate value for σ_q^2 is of paramount importance because it profoundly influences the efficiency of the MCMC algorithm. This influence operates at both extremes of σ_q^2:

Small σ_q^2: When σ_q^2 is too small, the random perturbations are minor, resulting in a high acceptance rate for proposed states. While this may seem desirable, it can slow down the exploration of the entire state space because the Markov chain takes small, cautious steps.

Large σ_q^2: on the other hand, if σ_q^2 is excessively large, can result in a high rejection rate for proposed states. This can dramatically slow down the convergence of the Markov chain because most steps are rejected, and it takes longer for the chain to explore the space effectively.

Selecting the Right σ_q^2 values is a crucial step in the process. The objective, as proposed in [19], is to find a value where the acceptance ratio (the proportion of proposed states that are accepted) falls within a specific range, typically between 20% and 30%. According to these authors, this range strikes a good balance between playing it safe and efficiently getting to the solution.

SIMULATION STUDY

Maximum Likelihood Estimation (MLE) for q-Weibull Distribution Parameters

We conducted a comprehensive Monte Carlo simulation to evaluate the performance of the ML estimators for the q-Weibull parameters under different settings. The study necessitates specified values for q, k and λ.

For each designated parameter setting, MLEs ($\hat{q}$, $\hat{k}$, $\hat{\lambda}$) are computed from a random sample with the sample size set at 200. This process is replicated 1000 times for each configuration. To gauge the efficacy of the MLEs, metrics such as mean squared error (MSE), Biases^2, and median absolute error (MedAE) are calculated.

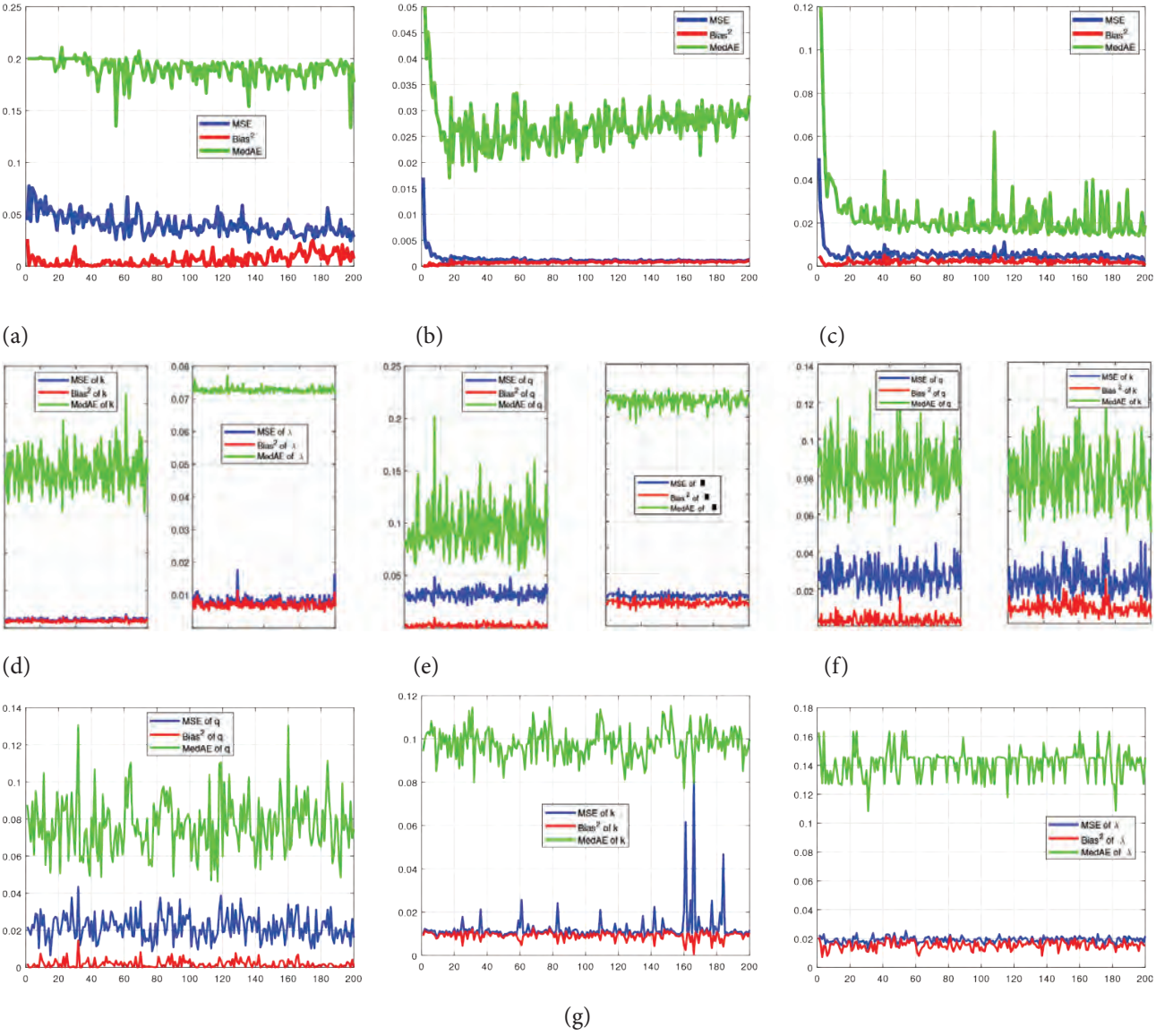


Figure 2. Comparison of MSE, Bias^2, and MedAE for the MLE of the parameters q (a), k (b), λ (c), (k, λ) fixed (d), (q, λ) fixed (e), (q, k) fixed (f), (q, k, λ) non fixed (g), for increasing n.

Figure 2 provides a comprehensive visual representation of the performance of MLE for the parameters q, k, and λ of the q-Weibull distribution under various conditions. The panels illustrate how MSE, Bias^2, and MedAE evolve as the sample size increases and under different parameter constraints.

- Panels (a), (b), and (c): These show the estimation performance when q, k, and λ are independently estimated while the other parameters are fixed. As expected, larger sample sizes lead to more accurate estimates, both MSE and Bias^2 drop noticeably, and the MedAE follows suit, showing that the estimates become not only more precise but also more robust.
- Panels (d), (e), and (f): show what happens when we fix two parameters and estimate just one. Compared to estimating each parameter independently, the error metrics here are lower. This makes sense: with fewer parameters to estimate, the model has more stability.
- Panels (g), (h) and (i): tackle the hardest scenario, estimating all three parameters at once, with no constraints. In this case, the error metrics are higher, which is unsurprising given the increased complexity of estimating multiple parameters simultaneously.

Overall, these results underscore how both sample size and the number of freely estimated parameters influence the accuracy and reliability of MLE for the q-Weibull distribution. They also offer a clear look at the inherent trade-offs in parameter estimation.

Table 3 summarizes the outcomes of a parameter estimation study conducted across three distinct cases. For each case, the table presents the estimated parameter value, along with its standard deviation (Std), mean absolute error (MAE), and the corresponding confidence interval. The table shows that the estimates of the parameters are generally more accurate when more parameters are fixed. For example, the estimates of q are more accurate when k and λ are fixed, and the estimates of k are more accurate when q and λ are fixed. This is because fixing parameters reduces the number of degrees of freedom and makes the estimation problem more tractable. The table also shows that the confidence intervals of the estimates are generally wider when more parameters are fixed. This is because fixing parameters reduces the amount of information available about the parameters, which makes it more difficult to estimate their values accurately.

Table 3 suggests that fixing parameters can be an effective way to improve the accuracy of estimates of distribution parameters. However, it is important to note that fixing parameters can also reduce the width of the confidence intervals of the estimates. Therefore, it is important to choose the method of estimation that is most appropriate for the specific problem at hand.

Metropolis-Hastings (MH) Algorithm for q-Weibull Distribution Parameters

This section will investigate the effectiveness of the MH algorithm in estimating the parameters of the q-Weibull distribution. The MH algorithm will be implemented and compared to MLE in terms of estimation accuracy and computational efficiency. The performance of the MH algorithm will be evaluated under various sample sizes and parameter values.

To demonstrate and contrast the effectiveness of the previously mentioned methods, random samples of sizes n = 25, 50, 100, and 200 were generated from the q-Weibull distribution. This approach accommodates varying data set sizes, encompassing small, medium, and large datasets.

The value of the analogue parameter q was set to 0.25 and 0.5, while the shape parameter k was held constant at 1 and 2. The scale parameter λ was also fixed at 2 and 1. We generated M=1000 distinct artificial datasets and

Table 3. Estimation results for different cases

Cases	Fixed parameters	Estimate	Std	MAE	CI
Case 1	k and λ	$\hat{q}$ = 0.4487	0.1937	0.0513	[0.3676 0.5298]
	q and λ	$\hat{k}$ = 0.9714	0.0137	0.0286	[0.8723 1.0802]
	q and k	$\hat{\lambda}$ =1.9569	0.0444	0.0431	[0.9653 3.0086]
Case 2	q	$\hat{k}$ =0.9534	0.0174	0.0466	[0.8054 1.1539]
		$\hat{\lambda}$ =1.9246	0.0093	0.0754	[0.8024 3.0050]
	k	$\hat{q}$ =0.4420	0.1303	0.0580	[-0.4584 1.5982]
		$\hat{\lambda}$ =1.8260	0.0113	0.1740	[-0.4424 4.0896]
	λ	$\hat{q}$ =0.5726	0.1599	0.0726	[0.4803 0.6649]
		$\hat{k}$ =1.0946	0.1024	0.0946	[1.0116 1.1776]
Case 3	-	$\hat{q}$ =0.5542	0.1345	0.0541	[0.1682 0.8571]
		$\hat{k}$ =0.9006	0.0295	0.0995	[0.7593 1.0562]
		$\hat{\lambda}$ =1.8650	0.0291	0.1350	[1.0727 2.6724]

estimated the parameters using MLE, IMH, and RWM. Hyperparameter values were fixed at $q_1 = q_2 = k_1 = k_2 = \lambda_1 = \lambda_2 = 0.1$ to achieve a prior distribution with a substantial variance.

For the m^{th} generated artificial dataset, we apply IMH and RWM, with fixed iterations $L = 55000$ and a burn-in of $B = 5000$. With these settings, the estimates $\left(\tilde{q}^{(m)}, \tilde{k}^{(m)}, \tilde{\lambda}^{(m)}\right)$ for (q, k, λ) are obtained by averaging the values generated. In other words:

$$\tilde{q}^{(l)} = \frac{1}{L}\sum_{l=1}^{L} q^{(m)}, \quad \tilde{k}^{(m)} = \frac{1}{L}\sum_{l=1}^{L} k^{(l)}, \text{ and } \tilde{\lambda}^{(m)} = \frac{1}{L}\sum_{l=1}^{L} \lambda^{(l)}$$

for $m = 1, \ldots, M$. The results are presented using the average of the M parameter estimates, denoted by $\hat{q}$, $\hat{k}$, and $\hat{\lambda}$.

Tables 3 and 4 display the mean estimates (Est.) and corresponding Standard deviations (Std), Mean Absolute Errors (MAE), and Confidence Interval (CI) values organized by method, considering q values of 0.25 and 0.5, k = 2,1 and λ = 1, 2, respectively. Additionally, the percentage of accepted values (%*acc.*) for the IMH and RWM methods is presented in these tables. Notably, the tables highlight in bold font the smaller standard deviations, mean absolute errors, and widths of confidence intervals for each sample size.

In the analysis of the presented Tables 3 and 4, a noteworthy observation is that the IMH and RWM methods appear to emerge as the most accurate and precise when compared to MLE. This conclusion is based on several indicators across different sample sizes and true parameter values.

The results are clear: IMH and RWM are no slouches. They deliver estimates that are just as accurate as MLE (sometimes even better) with standard deviations that stay in check and MAE that's right in line. Their confidence intervals may be a tad wider, but they still do their job, covering the true parameters every time. This flips the script on the idea that MLE is always superior. Here, IMH and RWM step up, proving they're more than capable of delivering accurate, precise estimates across different sample sizes and parameter settings.

APPLICATION

Within this section, we assess the efficacy of MLE, IMH, and RWM on two datasets that are publicly accessible. The dataset that serves as the focal point of this study, as detailed in Table 9, was originally introduced in [24]. Widely recognized and frequently employed in statistical analyses. This dataset consists of real-world data pertaining to the fatigue failure times of N 304 6061-T6 aluminium coupons. These coupons were subjected to oscillation at 18 cycles per second (cps) under two distinct constant levels of maximum stress per cycle: 31,000 psi for 1 second (1s) and 26,000 psi for 2 seconds (2s). Each stress level produced a comprehensive set of failure data, with $n_1 = 101$ failures observed at 31,000 psi and $n_2 = 102$ failures observed at 26,000 psi, respectively [6]. The outcomes are derived utilizing the Matlab programming language.

Table 3. Average of estimates, Std, MAE and CI by method ($q = 0.25$, $k = 1$, $\lambda = 2$)

Method	Statistic	Sample size											
		25			**50**			**100**			**200**		
		$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$
MLE	Est.	0.2617	0.9000	1.9604	0.25823	0.9234	2.0377	0.3415	0.9000	1.9375	0.26254	0.9001	2.0171
	Std	0.0001	0.0007	0.0002	0.0010	0.0009	0.0001	0.0012	0.0006	0.0001	0.0001	0.0009	0.0002
	MAE	0.3111	0.1000	0.0396	1.7500	0.0999	0.0377	2.0356	0.0915	0.1000	0.0625	1.7500	0.0999
	CI	[0.1254, 1.5346]	[0.5752, 1.2249]	[0.2743, 4.1951]	[0.2134, 0.2918]	[0.7208, 1.0794]	[1.7208, 2.3794]	[-0.2141, 0.8970]	[0.7464, 1.0537]	[0.8511, 3.0239]	[0.2468, 0.2756]	[0.8121, 0.9881]	[1.9854, 2.1761]
IMH	Est.	0.2400	0.8995	1.3835	0.2591	1.0167	1.8199	0.2566	1.0003	1.9832	0.2523	1.0530	2.0711
	Std	0.0348	0.0133	0.3639	0.0317	0.0183	0.6769	0.0013	0.0317	0.1570	0.0020	0.0034	0.0365
	MAE	0.0099	0.1005	0.6165	0.0091	0.0167	0.1801	0.0066	0.0003	0.0169	0.0023	0.0530	0.0711
	CI	[0.2154, 0.2646]	[0.8901, 0.9089]	[1.1262, 1.6408]	[0.2238, 0.2851]	[1.0055, 1.0378]	[1.0385, 2.2246]	[0.2553, 0.2579]	[0.9664, 1.0292]	[1.8791, 2.1637]	[0.2501, 0.2540]	[1.0497, 1.0566]	[2.0311, 2.1028]
	% acc	0.0072	0.0018	0.0018	0.0055	0.0054	0.0036	0.8076	0.0390	0.0885	0.8166	0.6323	0.6875
RWM	Est.	0.2672	0.9529	1.9026	0.2581	1.0119	1.7695	0.2579	1.0335	2.2098	0.2514	1.0544	2.0613
	Std	0.0065	0.0366	0.0272	0.0553	0.0791	0.8563	0.0037	0.0327	0.4083	0.0014	0.0047	0.0658
	MAE	0.0172	0.0471	0.0974	0.0081	0.0119	0.2305	0.0079	0.0335	0.2098	0.0014	0.0544	0.0613
	CI	[0.2627, 0.2718]	[0.9270, 0.9787]	[1.8834, 1.9218]	[0.2094, 0.3183]	[0.9497, 1.1009]	[0.7902, 2.3774]	[0.2547, 0.2620]	[1.0001, 1.0654]	[1.7554, 2.546]	[0.2504, 0.2529]	[1.0514, 1.0598]	[2.0129, 2.1363]
	% acc	0.0072	0.0018	0.0001	0.0036	0.0018	0.0036	0.8060	0.0335	0.1062	0.8147	0.6292	0.6852

Table 4. Average of estimates, Std, MAE and CI by method (q = 0.5, k = 2, λ = 1)

Method	Statistic	Sample size											
		25			50			100			200		
		$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$	$\hat{q}$	$\hat{k}$	$\hat{\lambda}$
MLE	Est.	0.2434	1.8005	1.0233	0.4811	1.8002	0.9621	0.4786	1.8001	0.9792	0.4062	1.800	0.9864
	Std	0.3025	0.0001	0.0273	0.0067	0.0026	0.0001	0.1800	0.0004	0.0169	0.1023	0.0002	0.0062
	MAE	0.2566	0.1995	0.0234	0.0189	0.1998	0.0379	0.0214	0.1999	0.0208	0.0938	0.1999	0.0136
	CI	[-0.2540, 1.4416]	[1.0857, 2.5294]	[0.4514, 1.5906]	[-0.1785, 1.1243]	[1.3658, 2.2435]	[0.6102, 1.3148]	[-2.0702, 3.2515]	[-0.1746, 3.7901]	[-0.9432, 2.9254]	[0.0922, 0.8648]	[1.5457, 2.0544]	[0.7573, 1.2243]
IMH	Est.	0.4955	2.1427	1.0138	0.50128	2.0597	1.0266	0.5088	2.0138	1.0424	0.5028	2.0166	0.9941
	Std	0.0658	0.1562	0.0045	0.0033	0.0063	0.0217	0.0017	0.0062	0.0168	0.0037	0.0037	0.0149
	MAE	0.0471	0.1427	0.0138	0.0013	0.0597	0.0266	0.0088	0.0138	0.0424	0.0028	0.0166	0.0059
	CI	[0.4336, 0.5625]	[1.9635, 2.2652]	[0.9745, 1.0637]	[0.4568, 0.5027]	[2.0532, 2.0792]	[1.0175, 1.0506]	[0.5065, 0.5118]	[2.0154, 2.0235]	[1.0236, 1.0604]	[0.5001, 0.5054]	[2.0140, 2.0192]	[0.9836, 1.0047]
	% acc	0.4341	0.0065	0.0392	0.5469	0.4559	0.6517	0.6373	0.7425	0.804	0.6963	0.8310	0.8601
RWM	Est.	0.5169	2.2353	1.2019	0.5025	2.0669	1.043	0.5141	2.0157	1.0384	0.5070	2.0252	1.0045
	Std	0.0518	0.4306	0.3391	0.0012	0.0146	0.0287	0.0026	0.0069	0.0062	0.0032	0.0005	0.0149
	MAE	0.0169	0.2353	0.2019	0.0025	0.0669	0.0430	0.0141	0.0157	0.0384	0.0070	0.0252	0.0045
	CI	[0.4725, 0.5702]	[1.9171, 2.7254]	[0.9382, 1.5804]	[0.5002, 0.5095]	[2.0645, 2.0852]	(1.02, 1.08)	(0.51, 0.52)	(2.01, 2.02)	(1.03, 1.04)	[0.5048, 0.5093]	[2.0248, 2.0255]	[0.9939, 1.0151]
	% acc	0.4213	0.0060	0.0714	0.5406	0.4594	0.6458	0.6382	0.7413	0.8026	0.6964	0.8327	0.8612

To evaluate the performance of methods, we examine the Root Mean Square Error (RMSE) concerning the empirical distribution function, as expressed by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\hat{F}(x_i) - F(x_i)\right]},$$

where, $\hat{F}(x)$ is derived by substituting the estimates of parameters q, k and λ obtained through each respective method. Meanwhile, $F(x_i)$ represents the empirical distribution function obtained from Kaplan-Meier estimates for each i = 1, ..., n. The method exhibiting the lowest RMSE is considered the most effective among the candidate methods for fitting the model.

Tables 6 and 7 summarize the parameter estimates and RMSE values for various distributions, including Fréchet, Ishita, Birnbaum–Saunders, Weibull, and q-Weibull, applied to the fatigue life data at stress levels 31,000 psi and 26,000 psi, respectively. Across both stress levels, the q-Weibull model came out on top, with RWM consistently delivering the lowest RMSE among all the distributions we tested. This really underscores how robust and reliable the q-Weibull model is for capturing fatigue life data, especially when stacked against the alternatives.

Figures 3 and 4 compare survival functions for fatigue life data at two stress levels (31,000 psi and 26,000 psi) across several distributions and estimation methods. The RWM and IMH curves are nearly indistinguishable, confirming their consistency. For clarity, we've focused on the most informative comparisons. The Kaplan-Meier step

Table 6. Parameter estimates and RMSE by method for the indicated models for the fatigue life data at stress level 31,000 psi.

Parameter	Fréchet	Ishita	Birnbaum–Saunders	Weibull	q-Weibull		
					MLE	IMH	RWM
$\hat{q}$	-	-	-	-	0.86643	0.88341	0.86764
$\hat{k}$	5.0575	422.22	131.819	6.07343	7.08523	6.74273	6.87903
$\hat{\lambda}$	120.7822	-	0.170385	143.167	146.06946	146.07199	148.42818
RMSE	0.054699	0.43237	0.18215	0.04053	0.03287	0.02596	0.02429

Table 7. Parameter estimates and RMSE by method for the indicated models for the fatigue life data at stress level 26,000 psi.

Parameter	Fréchet	Ishita	Birnbaum–Saunders	Weibull	*q*-Weibull		
					MLE	IMH	RWM
$\hat{q}$	-	-	-	-	0.82207	0.87516	0.853304
$\hat{k}$	5.5549	105.55	392.762	7.0075	6.602151	6.852165	6.806042
$\hat{\lambda}$	361.3133	-	0.161448	424.3782	435.30069	432.82873	432.59715
RMSE	0.053135	0.5003	0.017363	0.02782	0.024201	0.024981	0.024061

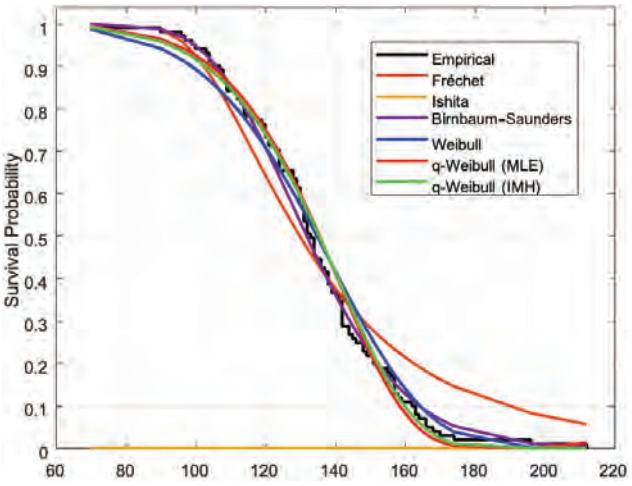


Figure 3. Survival probability estimates for fatigue life data at stress levels 31,000 psi, using MLE, IMH, and the Weibull model. Bayesian methods (IMH) show closer alignment with empirical data, reflected in lower RMSE values.

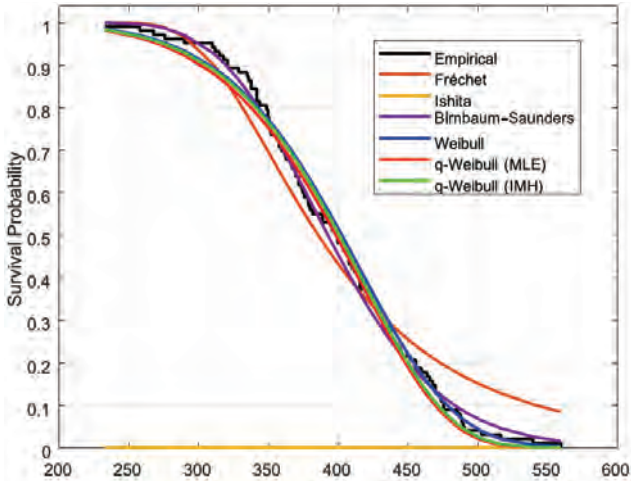


Figure 4. Survival probability estimates for fatigue life data at stress levels 26,000 psi, using MLE, IMH, and the Weibull model. Bayesian methods (IMH) show closer alignment with empirical data, reflected in lower RMSE values.

function provides the empirical benchmark. The *q*-Weibull model, particularly when estimated with Bayesian methods like IMH. It aligns with the Kaplan-Meier curve better than any other distribution, indicating a superior fit.

This showcases the advantage of Bayesian estimation: leveraging prior information and iterative learning to better capture the data. The other distributions (Fréchet, Ishita, Birnbaum–Saunders, and Weibull) perform reasonably well but show noticeable deviations, especially in the tails. These deviations highlight the limitations of traditional frequentist methods and other fixed-form distributions in capturing complex survival patterns. Together, these figures highlight just how well the *q*-Weibull model and Bayesian methods perform, particularly for datasets where survival behaviour is variable or follows a non-linear path. By incorporating prior distributions and employing iterative sampling techniques, Bayesian methods achieve a closer alignment with empirical observations, making them especially effective for real-world reliability analysis. This comparison further underscores just how robust and adaptable the q-Weibull model really is it consistently captures complex survival dynamics across a range of scenarios.

Figures 5 and 6 show the traceplots, ergodic means, and autocorrelation times for the parameters q, k and λ under IMH and RWM, respectively. These results come from the fatigue data at 31,000 psi (the corresponding plots for 26,000 psi are in the Appendix Table 1, Figures 7 and 8).

In stark contrast, IMH presents a scenario characterized by inadequate mixing and unsatisfactory autocorrelation and stability in terms of ergodic mean. These observations, coupled with the consideration of the RMSE value, collectively point towards the superior performance of RWM in comparison to MLE for these datasets. Conversely, IMH falls short in terms of performance, as indicated by its suboptimal mixing and inadequate autocorrelation and stability for ergodic mean.

Table 8 presents a comparative summary of the three estimation Maximum Likelihood Estimation (MLE) Independent Metropolis-Hastings (IMH) and Random Walk Metropolis (RWM), based on the results obtained in our study.

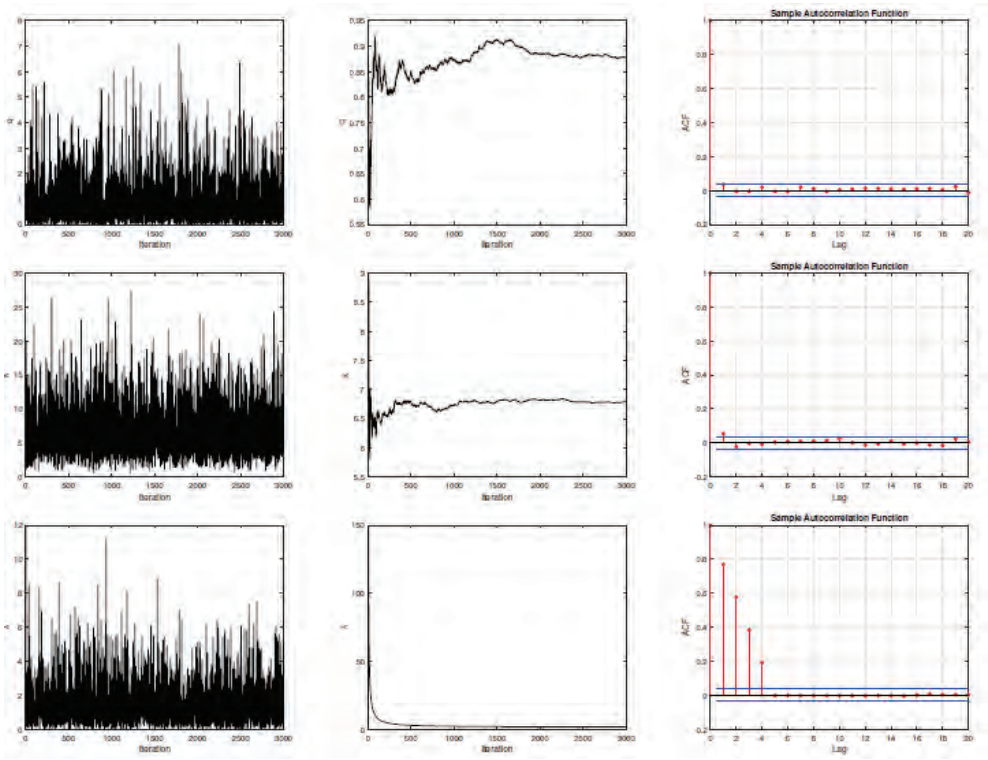


Figure 5. Traceplots, ergodic mean and Autocorrelation by IMH of parameters q, k and λ for fatigue life data at the stress levels 31000 psi.

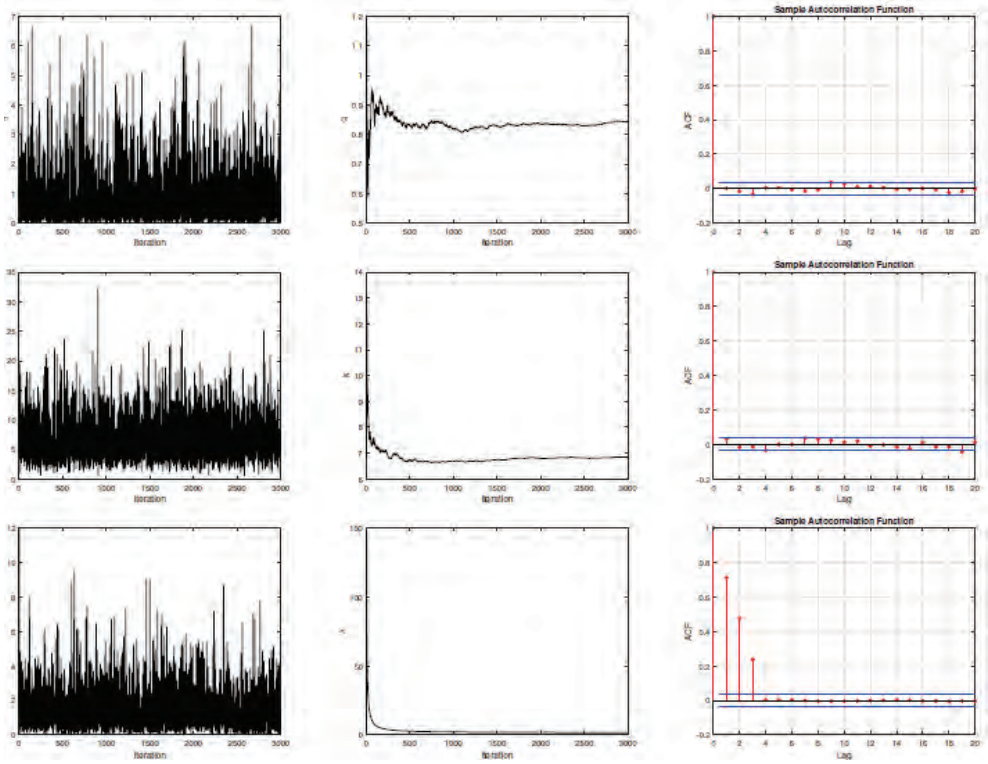


Figure 6. Traceplots, ergodic mean and Autocorrelation by RWM of parameters q, k and λ for fatigue life data at the stress levels 31000 psi.

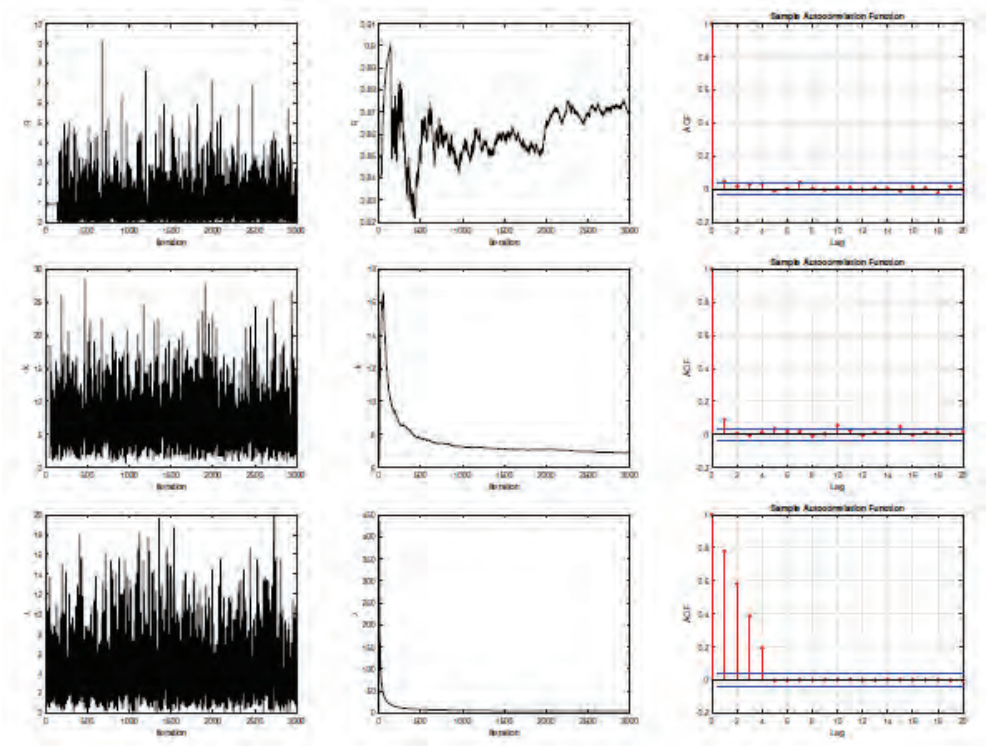


Figure 7. Traceplots, ergodic mean and autocorrelation by IMH of parameters q, k and λ for fatigue life data at the stress levels 26000 psi.

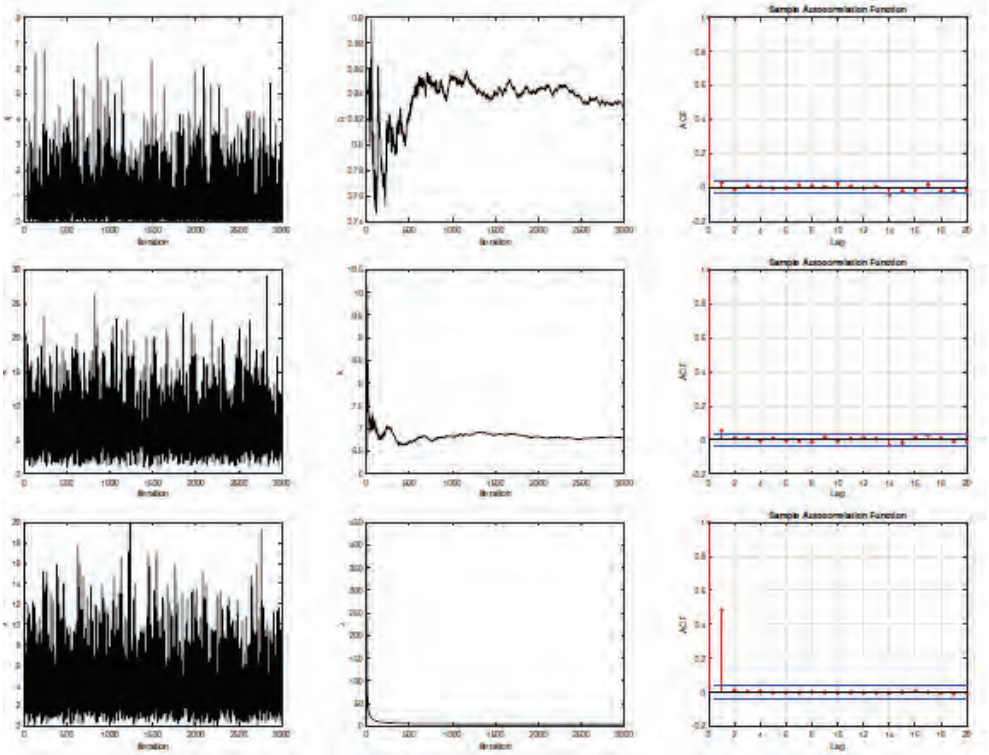


Figure 8. Traceplots, ergodic mean and Autocorrelation by RWM of parameters q, k and λ for fatigue life data at the stress levels 26000 psi.

Table 8. Comparison of MLE, metropolis-hastings (MH), ındependent metropolis-hastings (IMH), and random walk metropolis (RWM) methods based on results

Method	Performance	Advantages	Limitations
MLE	Competitive performance for small datasets; moderate RMSE in real-data results.	Straightforward implementation; widely used and well-understood approach in parameter estimation.	May yield less accurate estimates than Bayesian methods (IMH, RWM), especially in the context of large or complex datasets.
IMH	Good accuracy and variability for larger datasets, but mixing and autocorrelation could be better for some parameters.	Effective when the candidate distribution closely matches the posterior; leads to fewer rejections than MH.	Performs well only if the proposal distribution is carefully chosen; otherwise, the ergodic means can become unstable.
RWM	Best performance overall: lowest RMSE for both simulation and real-data results; robust convergence.	Excellent mixing properties, efficient exploration, stable results across sample sizes and stress levels.	Requires careful tuning of step size (proposal variance); large or small values impact convergence.

CONCLUSION

This study bridges critical gaps in the application of the *q*-Weibull distribution by introducing and evaluating Bayesian computational methods for parameter estimation. We compared three methods: MLE, Independent Metropolis-Hastings, and Random Walk Metropolis. Using Monte Carlo simulations and real-world datasets, we evaluated them on standard deviation, mean absolute errors, and confidence intervals.

The q-Weibull distribution, estimated with Bayesian methods, consistently outperformed alternatives like Fréchet, Ishita, Birnbaum–Saunders, and Weibull. RWM, in particular, stood out, it had the lowest RMSE and matched the empirical survival probabilities better than any other method. This shows just how robust and flexible the *q*-Weibull is, and it makes a strong case for MCMC methods as viable alternatives to traditional approaches.

Bayesian methods, particularly IMH and RWM, are computationally intensive. MCMC algorithms run on iterations, which means they demand significant time and computing power, especially when datasets are large. It's a trade-off, but one we think is justified by the gains in accuracy and robustness, even if it can be a hurdle for some real-world applications. The source code employed in both the simulation study and practical applications was implemented in Matlab software.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Yu J. Prediction of survival time for breast cancer patients. Dalhousie University, Halifax, Nova Scotia 2022.

[2] Barraza-Contreras JM, Piña-Monarrez MR, Molina A. Fatigue-life prediction of mechanical element by using the Weibull distribution. Appl Sci 2020;10:6384. [CrossRef]

[3] Sarkar A, Deep S, Datta D, Vijaywargiya A, Roy R, Phanikanth VS. Weibull and generalized extreme value distributions for wind speed data analysis of some locations in India. KSCE J Civ Eng 2019;23:3476–3492. [CrossRef]

[4] Wiradinata T, Graciella F, Tanamal R, Soekamto YS, Saputri TRD. Post-pandemic analysis of house price prediction in Surabaya: a machine learning approach. J Southwest Jıaotong Univ 2022;57:562–573. [CrossRef]

[5] Sadok I, Zribi M. Image restoration using Weibull particle filters. In: 2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS); 2022 Oct. p. 1–6. IEEE. [CrossRef]

[6] D'Anna G, Giorgio M, Riccio A. Estimating fatigue life of structural components from accelerated data via a Birnbaum-Saunders model with shape and scale stress dependent parameters. Procedia Eng 2016;167:10–17. [CrossRef]
[7] Jose KK, Naik SR. On the q-Weibull distribution and its applications. Commun Stat Theory Methods 2009;38:912–926. [CrossRef]
[8] Picoli Jr S, Mendes RS, Malacarne LC. q-exponential, Weibull, and q-Weibull distributions: an empirical analysis. Physica A Stat Mech its Appl 2003;324:678–688. [CrossRef]
[9] Vuorenmaa T. A q-Weibull autoregressive conditional duration model and threshold dependence. University of Helsinki, RUESG and HECER 2006;117:1–61.
[10] Sadok I. On the q-Rayleigh distribution and its applications. Reliab Theory Appl 2024;19:588–602.
[11] Assis EM, Borges EP, Vieira de Melo SA. Generalized q-Weibull model and the bathtub curve. Int J Qual Reliab Manag 2013;30:720–736. [CrossRef]
[12] Douini H, Sadok I. A study on statistical properties of a new class of q-Fréchet distribution. Math Appl Sci Eng 2020:1–23.
[13] Costa UMS, Freire VN, Malacarne LC, Mendes RS, Picoli Jr S, De Vasconcelos EA, et al. An improved description of the dielectric breakdown in oxides based on a generalized Weibull distribution. Physica A Stat Mech its Appl 2006;361:209–215. [CrossRef]
[14] Al-Essa LA, Saboor A, Tahir MH, Khan S, Jamal F, Elhassanein A. Bivariate q-generalized extreme value distribution: a comparative approach with applications to climate related data. Heliyon 2024;10:8. [CrossRef]
[15] Agrawal S, Vats D, Łatuszyński K, Roberts GO. Optimal scaling of MCMC beyond Metropolis. Adv Appl Prob 2023;55:492–509. [CrossRef]
[16] Zribi M, Sadok I, Marhaba B. Kernel-diffeomorphism Bayesian bootstrap filter to reduce speckle noise on SAR images. Comput Stat 2025;40:3613–3643. [CrossRef]
[17] Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: a review for statisticians. J Am Stat Assoc 2017;112:859–877. [CrossRef]
[18] Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. Am Stat 1995;49:327–335. [CrossRef]
[19] Saraiva EF, Suzuki AK. Bayesian computational methods for estimation of two-parameters Weibull distribution in presence of right-censored data. Chilean J Stat 2017;8:25–43.
[20] Xu M, Droguett EL, Lins ID, das Chagas Moura M. On the q-Weibull distribution for reliability applications: an adaptive hybrid artificial bee colony algorithm for parameter estimation. Reliab Eng Syst Saf 2017;158:93–105. [CrossRef]
[21] Sadok I. Non-informative Bayesian dispersion particle filter. J Innov Appl Math Comput Sci 2023;3:173–189.
[22] Sadok I, Zribi M, Masmoudi A. Non-informative Bayesian estimation in dispersion models. Hacettepe J Math Stat 2023;53:251–268. [CrossRef]
[23] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. 1970;57:97–109. [CrossRef]
[24] Birnbaum ZW, Saunders SC. Estimation for a family of life distributions with applications to fatigue. J Appl Prob 1969;6:328–347. [CrossRef]
[25] Méndez-González LC, Rodríguez-Picón LA, González-Hernández IJ, Pérez-Olguín IJC, Quezada-Carreón AE. The Fréchet-Chen distribution and its applications to reliability analysis of electronic devices. Preprints 2024. doi: 10.20944/preprints202409.1956.v12024. [CrossRef]
[26] Hassan A, Dar SA, Para BA. A new generalization of Ishita distribution: properties and applications. J Appl Prob Stat 2019;13:53–67.
[27] Sadok I, Masmoudi A, Zribi M. Integrating the EM algorithm with particle filter for image restoration with exponential dispersion noise. Commun Stat Theory Methods 2023;52:446–462. [CrossRef]
[28] Sadok I, Masmoudi A. New parametrization of stochastic volatility models. Commun Stat Theory Methods 2022;51:1936–1953. [CrossRef]

Appendix Table 1. Failure data collected

Failure data collected at 31000 psi																
70	90	96	97	99	100	103	104	104	105	107	108	108	108	109	109	112
112	113	114	114	114	116	119	120	120	120	121	121	123	124	124	124	124
124	128	128	129	139	130	130	130	131	131	131	131	131	132	132	132	133
134	134	134	134	134	136	136	137	138	138	138	139	139	141	141	142	142
142	142	142	142	144	144	145	146	148	148	149	151	151	152	155	156	157
157	157	157	158	159	162	163	163	164	166	166	168	170	174	196	212	
Failure data collected at 26000 psi																
233	258	268	276	290	310	312	315	318	321	321	329	335	336	338	338	342
342	342	344	349	350	350	351	351	352	352	356	358	358	360	362	363	366
367	370	370	372	372	374	375	376	379	379	380	382	389	389	395	396	400
400	400	403	404	406	408	408	410	412	414	416	416	416	420	422	423	426
428	432	432	433	433	437	438	439	439	443	445	445	452	456	456	460	464
466	468	470	470	473	474	476	476	486	488	489	490	491	503	517	540	560

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00174

Research Article

Sustainability-financial assessment indicators for renewable energy systems using fuzzy SWARA/WASPAS approach

Razieh KESHAVARZFARD[1,*], Mohsen ABOLHASANI[1]

[1]*Department of Industrial Engineering, North Tehran Branch, Islamic Azad University, Tehran, 1477893855, Iran*

ARTICLE INFO

Article history
Received: 31 October 2024
Revised: 15 January 2025
Accepted: 19 February 2025

Keywords:
Solar PV; Wind; Fuel Cell; Renewable Energy; Sustainable Indicators; MCDM; Fuzzy SWARA; Fuzzy WASPAS

ABSTRACT

Renewable energy has a significant share in the world's energy supply today. The need for energy due to growth and development, along with issues such as greenhouse gas emissions, declining fossil reserves and the growing need for energy, have led countries to turn to use renewable energy. As uncertainty is an inseparable part of decision-making process, we assess and prioritize renewable energy technologies using fuzzy multi-criteria decision-making techniques. In this study, sustainability, technical and financial indices are evaluated then four types of renewable systems are ranked, using hybrid Fuzzy Step-wise Weight Assessment Ratio Analysis/ Fuzzy Weighted Aggregated Sum Product Assessment (F-SWARA/F-WASPAS) method. The results show that among the main criteria, the technology criterion with a weight of 0.402 has gained the first rank, the resources criterion with a weight of 0.253 has gained the second rank and the social criterion with a weight of 0.161 has gained the third rank. Among the sub-criteria, system efficiency, capacity factor and material intensity are ranked first to third, respectively. Among the options, solid oxide fuel cells have won first place.

Cite this article as: Keshavarzfard R, Abolhasani M. Sustainability-financial assessment indicators for renewable energy systems using fuzzy SWARA/WASPAS approach. Sigma J Eng Nat Sci 2026;44(2):1111–1126.

INTRODUCTION

Greenhouse gas emissions from the combustion of fossil fuels have many negative effects on the weather, leading to floods, hurricanes, and other natural disasters. This environmental pressure caused by the continuous production of greenhouse gas emissions, affects future energy production. The main sources of these greenhouse gas emissions are non-renewable energies. They are energies whose consumption causes their reserves to decrease, and meanwhile the formation of this type of resources requires a long time (several hundred thousand years); such as crude oil, gas, coal and all fossil energies [1]. Another option for energy production is renewable energy technologies that use natural energy that can be obtained from the sun, wind, water, ocean (waves, tides, current sea flow and ocean energy thermal system) and geothermal. Renewable energy is any type of energy that can be used without decreasing the reservoirs that supply it. Up to 80% of the energy demand by 2050 could be produced from renewable source [2]. Renewable energies are clean,

*Corresponding author.
*E-mail address: r.keshavarzfard@iau.ac.ir

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

plentiful and reliable, and if they are developed correctly, they can play an important role as sustainable energy sources in achieving the sustainable development goals of countries [3]. It should be known that if the goal is to replace 100% of fossil energy with sustainable energy, it will not be possible in the near future. But if the goal is to save fossil fuel, the existing resourses should be used more appropriately. This is completely practical and possible in the current conditions, both in terms of technology and economic [4]. Among of all types of sustainable energies (wind, solar, fuel cells, etc.) sun is not only a great source of energy, but also the beginning of life and the source of all other energies. According to scientific estimates, about 6000 million years of sun's life have passed and every second about 2.4 million tons of the sun's mass is converted into energy. Considering the sun's weight, which is about 333 thousand times the earth's weight, this luminous sphere can be considered as a huge source of energy for the next 5 billion years. The sun is one of the important sources of energy that should be turned to, because it does not require advanced and expensive technologies and can be used as a useful source and supplier of energy in most parts of the world [5]. In this era, solar energy is used and utilized by different systems and for different purposes, which include: 1. Photobiological systems 2. Photochemical systems 3. Photovoltaic systems 4. Thermal and refrigeration systems [6]. Wind is another renewable source of energy that is obtained from the heat of the sun. Approximately 2.5% of the solar energy reaching the earth is converted into wind energy. Wind on the earth is an important factor for the heat exchange, humidity and the transfer of particulate and non-particulate particles from one point to another, which climatically provides human comfort or disturbs it, both in terms of heat and in terms of behavioral comfort [7]. In America, Germany, Holland and Denmark, a significant part of electricity is provided by using large wind power plants. Wind energy has always grown rapidly in energy sources and has a lot of potential. The wind energy cost is very low and the location of wind energy is a pollution-free solution to the energy crisis [8]. Many researches have been working on approaches to prioritize or rank different types of sustainable energy systems based on specific criteria. For instance, [9] developed a sustainability index approach based on technical, economic and environmental criteria to select three energy storage systems (lithium polymer batteries, fuel cells and lead acid batteries) for intermittent and renewable energy sources (solar and wind). They used a new weighting method to determine the sustainability index of energy storage systems. Fuel cell was chosen as the best energy storage option based on the sustainability index presented in this study. Hacatoglu et al. [10] presented an integrated sustainability index for energy systems using multidimensional sustainability criteria. The stability index was determined using normalization, weighting and aggregation of stability indices. This system was applied to autonomous and hybrid power systems in southern Ontario, Canada. The solar-hydrogen energy system was the best hybrid system compared to the other tested systems according to the sustainability index. Also, the results show that climate change and ozone depletion indicators, reasonable price, commercial viability and land area have the greatest impact on the integrated sustainability index. The results obtained during this study show that the integrated sustainability index has been a useful tool for decision analysis to predict single or combined sustainable energy systems. Papilo et al. [11] developed sustainability indicators for palm oil-based bioenergy. In their study, they used ten indicators for environmental, social and economic factors. The results obtained for the development of bioenergy show that there is a need to balance three aspects and each sustainability index. Dimodke et al. [12] addressed the optimal mapping of hybrid renewable energy systems for locations, using a multi-criteria decision-making algorithm. They carried out an optimal mapping of combined energy systems, based on wind and PV, for households in six locations in the southern geopolitical zone of Nigeria. Optima combined energy systems are capable of providing 7.23 kilowatt hours per day of household energy needs. The combined energy system for each location is optimized based on the calculation of HOMER software and the Technique for Order of Preferences by Similarity to Ideal Solution (TOPSIS) algorithm, which takes into account technical, economic, environmental and socio-cultural criteria. Bauman et al. [13] reviewed Multi Criteria Decision Making (MCDM) methods for evaluating energy storage systems for renewable energy systems. The main objective of this study was to select a suitable and stable EES using the MCDM method for grid-connected renewable power systems. Technical, economic, social and environmental aspects are included in this analysis. Mayer et al. [14] addressed the environmental and economic multi-objective optimization of the combined renewable energy system at the household level by genetic algorithm. This paper aims to fill this gap by presenting a multi-objective design framework for household-scale systems based on technical modeling of several typical components. Solar photovoltaic, wind turbine, solar heat collector, heat pump, heat storage, battery and as a new topic, thermal insulation thickness are considered. They proved that solar photovoltaic is the most competitive technology to reduce the environmental impact of grid-connected systems. However, off-grid systems benefit the most from a balanced mix of different renewable energy sources. Nguyen et al. [15] addressed multi-objective decision making and optimal sizing of a hybrid renewable energy system to meet dynamic energy needs in a wastewater treatment plant. Considering the contemporary economy, system reliability and environmental policies to meet dynamic energy needs in a wastewater treatment plant, this study presents an intelligent management method for optimal measurement and power

management of photovoltaic-wind turbines with hydrogen and battery storage. They presented a multi-objective approach based on fuzzy decision-making method to solve the optimization problem. Osturgard et al. [16] investigated sustainable development using renewable energy systems. This research was conducted with the aim of sustainable development using renewable energy systems with special attention to technology evaluation, pricing and regulation and systems analysis. Also case studies and model development from Austria, Cape Verde, Colombia and Iran are presented with different focal points. Ghenai et al. [17] investigated sustainability indicators for renewable energy systems using multicriteria decision-making model and extended SWARA/ARAS[1] hybrid method. Hashemkhani Zolfani et al. [18] proposed a modified and extended version of the traditional Fuzzy SWARA (F-SWARA) to identify the criteria weights called Improved Fuzzy SWARA (IMF-SWARA), and suggested applying the Fuzzy Multi-Attributive Border Approximation area Comparison (F-MABAC) technique to determine the preference ratings of the alternatives. Alghassab [19] evaluated sustainable renewable energy using Fuzzy Analytic Hierarchy Process (FAHP) and Fuzzy TOPSIS (FTOPSIS). Cutore et al. [20] presented an optimization model to orient energy experts and urban planners in the capacity sizing and flow management of Renewable Energy Communities (RECs) under the Italian regulatory framework. Thirumalaivasan et al. [21] employed a life cycle assessment method to assess the system's sustainability, encompassing energy, economic, environmental, and social dimensions. They explorated how varying meteorological conditions and geographical locations impact on the system's sustainability. Liu et al. [22] evaluated the sustainability and environmental features of renewable energies. They focused on biogas and biohydrogen (Biofuels) production. Mardani et al. [23] analyzed SWARA and WASPAS methods with recent fuzzy developments. Zheng et al. [24] investigated spherical fuzzy soft rough average aggregation operators and applied it to multi-criteria decision making. Mahmood et al. [25] developed Bonferroni mean operators using bipolar complex fuzzy setting and and presented a decision making method. Garg et al. [26] applied an MADM method based on Aczel-Alsina power aggregation operators under bipolar fuzzy information to the quantum computing subject. Mahmood and Rehman [27] investigated the role and impact of AI in the field of healthcare, based on bipolar complex fuzzy Schweizer–Sklar power aggregation operators. Naeem et al. [28] evaluated the sources of renewable energy considering the artificial data in the structure of bipolar complex fuzzy set. Mahmood et al. [29] presented a framework of bipolar complex intuitionistic fuzzy N-Soft sets Using TOPSIS. Adan and Filik [30] proposed a techno-economic assessment of a standalone and on grid hybrid power system. Yolcu [31] investigated the applications of intuitionistic fuzzy hypersoft topology to multi-criteria decision-making. Güven and Yörükeren [32] designed an off-grid renewable energy system, consisting solar, wind, battery and diesel components and proposed a model to optimize the designed energy system.

The above literature study shows that renewable energy systems are prioritized by many researches, in many various ways. Actually, there are different types of renewable energies that can be evaluated using any of MCDM methods. To the best of our knowledge, no research has been done to use SWARA/WASPAS approach in the field of evaluating renewable energy system considering five kinds of indicators. The nearest study to this article is written by Ghenai et al. [17]. They ranked four systems consisting solar energy, wind energy, phosphoric acid fuel cells and solid oxide fuel cells in terms of five main criteria using SWARA/ARAS method. We ranked those four alternatives adding three financial subcriteria as financial indicator. Also, we found the fuzzy hybrid SWARA/WASPAS more accurate and used that approach with all decisions made by Iranian experts in terms of existing conditions in Iran. In this article, the sustainability and technical indicators for choosing renewable systems taken from the article [17] as well as financial criteria are evaluated and prioritized by using the decision-making method with multiple criteria considering uncertainty. For this purpose, we use hybrid Fuzzy Step-wise Weight Assessment Ratio Analysis/Weighted Aggregated Sum Product Assessment (F-SWARA/WASPAS) method. The reminder of this paper is as follows: In section 2 problem description is presented. Section 3 consists of hybrid SWARA/WASPAS methodology. In section 4 results and discussion are discussed and finally in section 5 conclusion and future researches are presented.

PROBLRM DESCRIPTION

In the recent decades, energy planning techniques have changed. Reducing cost has been the main goal of choosing type of energy for using in a regional but recently multiple and even contradictory goals are incorporated. This complex environment highlights the multi-criteria nature of the problem. Therefore, adopting the policy of replacing renewable energies instead of fossil fuels should be done in a multi-criteria environment. Multi-Criteria Decision Making (MCDM) is a suitable approach to deal with complex problems with high uncertainty. Many methods have been developed and proposed to solve these problems. However, there is still the problem of choosing an appropriate method in different decision-making situations. In none of the proposed methods as the best and most appropriate method, all decision situations are considered. In simpler terms, although the use of multi-criteria decision-making

1 Additive Ratio Assessment

methods depends on the purpose of the research and especially on the accuracy and reliability of the data, the correct selection of each of these methods requires a set of evaluation criteria that must be carefully studied. Therefore, in this research, new results will be presented in the evaluation of sustainability-financial-technical indicators for renewable energy systems. In this study, MCDM will be used under uncertainty to evaluate renewable energy systems against sustainability-financial-technical indicators and their sub-indices. In the first stage, the indicators and sub-indices of sustainability will be weighted using the SWARA method, and then the ranking of renewable energy systems will be done using the WASPAS technique.

RENEWABLE POWER SYSTEMS

In this section, the evaluated alternatives are introduced which are four types of renewable energy technologies. We consider four common renewable energies as follow [17]:

1. Solar energy systems (S1)
2. Wind energy systems (S2)
3. Phosphoric acid fuel cells (S3)
4. Solid oxide fuel cells (S4)

SUSTAINABILITY INDICATORS

In this section we use five common sustainability criteria by investigating the literature review mostly from [17]. These indicators include Resource, Environmental, Economic-Financial, Social and Technological. A complete description of fourteen subcriteria can be found in [17]. In this study three financial criteria are added as financial subcriteria of economic-financial indicator. These subcriteria are: Net present value, Internal return rate and Payback period. As a result, mentioned criteria and their seventeen subcriteria are as follow:

A. Resource Indicator
A1. Area Intensity [m^2/kW]
A2. Material Intensity [kg/kW]
A3. Energy Intensity (construction) [MJ/kW]
A4. Energy Intensity (fuel) [MJ/kWh]
B. Environmental Indicator
B1. CO_2 Intensity (construction) [kg/kW]
B2. CO_2 Intensity (fuel) [kg/kWh]
C. Economic-Financial Indicator
C1. Capital Intensity (construction) [US$/kW]
C2. Capital Intensity (fuel) [US$/kWh]
C3. Delivered cost [US$/kWh]
C4. Net present value
C5. Internal return rate
C6. Payback period
D. Social Indicator
D1. Current installed capacity [GW]
D2. Growth rate [% fraction/year]
E. Technology Indicator
E1. Capacity factor [%]
E2. System efficiency [%]
E3. Lifetime [years]

METHODOLOGY- HYBRID FUZZY SWARA/ WASPAS

After determining the criteria and sub-criteria used in this research, at first, these indicators are weighted using the fuzzy SWARA technique, and the most important indicators are introduced. Then, using the fuzzy WASPAS technique, the mentioned renewable energy systems are ranked.

SWARA METHOD

The SWARA method is one of the multi-criteria decision-making methods aiming to calculate the weight of criteria and sub-criteria. The SWARA method was introduced by Keršuliene et al. [33]. In this method, criteria are ranked based on their value. The most important criterion is given the first rank and the least important criterion is given the last rank. Finally, the criteria are prioritized based on average values of relative importance. This technique is based on the opinions of experts and is a completely judgemental method. In this method, experts play an important role in determining the weight of criteria [34].

FUZZY SWARA METHOD

The algorithm of this technique is same as the SWARA method, but it is used in a fuzzy environment. As mentioned earlier, the purpose of the SWARA method is to calculate the weight of the factors, so it is of particular importance. Therefore, by implementing this method in a fuzzy environment, the ambiguities in the experts' responses will be decreased and the results will be more accurate. The steps of the fuzzy SWARA method are given below [35]:

Step 1- Each of the criteria is given a score based on five fuzzy linguistic expressions from the point of view of each decision maker as in Table 1.

Step 2- We sort the research factors according to their importance in descending order based on the average of the previous stage. In fact, the output of this stage is the fuzzy average of the previous stage, which is called S_j and is calculated based on the following equation:

Table 1. Linguistic expressions and fuzzy numbers

Language expressions	Triangular fuzzy numbers
Very low	(0,0,0.25)
Low	(0,0.25,0.5)
Moderate	(0.25,0.5,0.75)
High	(0.5,0.75,1)
Very high	(0.75,1,1)

$$S_j = (\sum_i \frac{a_{ij}}{n}, \sum_i \frac{b_{ij}}{n}, \sum_i \frac{c_{ij}}{n}) \quad (1)$$

The fuzzy number (a_{ij}, b_{ij}, c_{ij}) is the triangular fuzzy number for language expression of ith decision maker for jth criterion.

Step 3- Calculating the coefficient Kj, assuming $\tilde{1}$= (1,1,1). This coefficient is calculated using equation (2):

$$\widetilde{K_j} = \begin{cases} \tilde{1} & j = 1 \\ \tilde{S}_j & j > 1 \end{cases} \quad (2)$$

Step 4- Calculating fuzzy weights (qj). Fuzzy weights are obtained from equation (3):

$$\tilde{q}_j = \begin{cases} \tilde{1} & j = 1 \\ \frac{\tilde{q}_{j-1}}{\tilde{k}_j} & j > 1 \end{cases} \quad (3)$$

Step 5- Calculating relative weights.

$$\widetilde{W_j} = \frac{\widetilde{q_j}}{\sum_{k=1}^{n} \widetilde{q_k}} \quad (4)$$

The output of this step is relative fuzzy weights (w_{jl},w_{jm},w_{ju}). To convert these weights into crisp numbers, equation (5) is used.

$$W_{crisp} = \frac{(w_j^m - w_j^l) + (w_j^u - w_j^l)}{3} + w_j^l \quad (5)$$

WASPAS METHOD

WASPAS method is one of the new multi-criteria decision-making methods which was introduced by [36]. This method is a combination of two models (weighted sum model) and (weighted multiplication model) and has more accuracy compared to independent methods.

FUZZY WASPAS METHOD

Fuzzy WASPAS technique was presented in [37]. The algorithm of this method is almost similar to the WASPAS method, but it is implemented in a fuzzy environment. Also in this research, verbal expressions and fuzzy numbers listed in Table 2 have been used [38].

The steps of the fuzzy WASPAS method are as follows**:** First, we form the current situation matrice based on the designed indicators. Then, the de-scaling of the decision matrice is discussed based on the following two equations:

$$\bar{x}_{ij} = \frac{\tilde{x}_{ij}}{\max_i \tilde{x}_{ij}} \quad for\ beneficial\ criteria \quad (6)$$

$$\bar{x}_{ij} = \frac{\min_i \tilde{x}_{ij}}{\tilde{x}_{ij}} \quad for\ non-beneficial\ criteria \quad (7)$$

In the WASPAS method, a common criterion from optimization, follows two optimal criteria. The first optimization criterion, that is the weighted average success criterion, is similar to the Weighted Sum Method (WSM) method. It is a popular and acceptable multi-criteria decision-making approach that is used to evaluate considered alternatives regarding to a set of decision-making criteria. Based on the WSM method, the total relative importance of the i-th alternative is calculated as below:

$$Q_i = \sum_{j=1}^{n} \bar{\tilde{x}}_{ij} \widetilde{w}_j \quad (8)$$

In this equation, W_j is the j-th criterion weight.

On the other hand, according to the Weighted Product Method (WPM) method, the relative value of all alternatives is calculated using the following equation:

$$P_i = \prod_{j=1}^{n} (\bar{\tilde{x}}_{ij})^{\widetilde{w}_j} \quad (9)$$

In the next step, we deffuzy the value of P_i and Q_i using the following equation:

$$Q_i = \frac{1}{3}(Q_{i\alpha}, Q_{i\beta}, Q_{i\gamma}) \quad (10)$$

Table 2. Verbal expressions and corresponding fuzzy numbers to rank alternatives [38] [created by author]

no	Priorities	Fuzzy equivalent of priorities		
		Lower limit (L)	Moderate (M)	Upper limit (U)
1	Very low	1	1	3
2	Low	1	3	5
3	Moderate	3	5	7
4	High	5	7	9
5	Very high	7	9	11

$$P_i = \frac{1}{3}(P_{i\alpha}, P_{i\beta}, P_{i\gamma}) \quad (11)$$

Then a general criterion for integrating WSM and WPM would be as follows:

$$K_i = \lambda Q_i + (1-\lambda)P_i = \lambda \sum_{j=1}^{n} \bar{x}_{ij} w_j + (1-\lambda) \prod_{j=1}^{n} \left(\bar{x}_{ij}\right)^{w_j} \quad (12)$$

If λ=0, the WASPAS model will become the WPM model; If λ=1, the WASPAS model becomes the WSM model. For deciding the optimal value, λ is calculated from the following equation [36]:

$$\lambda = \frac{\sum_{i=1}^{m} P_i}{\sum_{i=1}^{m} P_i + \sum_{i=1}^{m} Q_i} \quad (13)$$

RESULTS AND DISCUSSION

In this part, the research data is analyzed. The aim of this research is to rank renewable systems based on sustainability, technical and financial indicators. To achieve this goal, first the factors are extracted from the literature review, then using the fuzzy SWARA weighting method all criteria and suncriteria are weighted and finally the alternatives of the research are ranked using the fuzzy WASPAS method. Based on the review of the literature and research background, 17 indicators of sustainability-financial-technical, affecting renewable systems are identified and extracted in five main dimensions. Using five Likert scale (1 = very little importance, 2 = little importance, 3 = moderate importance, 4 = high importance, 5 = very high importance), we asked 20 experts in this area to give appropriate score to each index, considering the geographical, environmental and financial conditions in Iran. Preliminary results from the experts' comments are given in Table 3.

Table 3 shows the number of experts' opinions on research indicators. To fuzzify the numbers, we first convert them into a fuzzy number based on the spectrum in Table 1, then based on the the fuzzy average is obtained from the scores, and then the fuzzy average is converted into a crisp number by equation (5). The results of all fuzzification calculations are given in Table 4. For example, the criterion of row 1 of fuzzy score calculations are as follows:

0 people gave very low points, 1 person gave low points, 10 people gave average points, 0 people gave high points and 9 people gave very high points. Therefore, the fuzzy and non-fuzzy score is as follows:

$$Score_{Fuzzy} = \frac{0*(0,0,0.25)+1*(0,0.25,0.5)+10*(0.25,0.5,0.75)+0*(0.5,0.75,1)+9*(0.75,1,1)}{20} = (0.273,0.505,0.738)$$

$$Score_{Crisp} = \frac{0.463+0.713+0.85}{3} = 0.675$$

Table 3. The results of experts' opinions

Criteria	Sub-criteria code	Sub-criteria	Importance level				
			Very Low	Low	Moderate	High	Very high
Resource	A1	Area intensity	0	1	10	0	9
	A2	Material intensity	0	2	4	7	7
	A3	Energy intensity (construction)	3	1	11	0	5
	A4	Energy intensity (fuel)	4	5	0	2	9
Environmental	B1	CO_2 intensity (construction)	7	1	4	0	8
	B2	CO_2 intensity (fuel)	3	5	4	0	8
Economic-Financial	C1	Capital intensity (construction)	6	3	6	0	5
	C2	Capital intensity (fuel)	2	2	4	3	9
	C3	Delivered cost	7	2	2	3	6
	C4	Net present value	9	5	0	0	6
	C5	Internal return rate	4	1	1	7	7
	C6	Payback period	7	2	0	0	11
Social	D1	Current installed capacity	3	2	1	0	14
	D2	Growth rate	6	3	1	4	6
Technology	E1	Capacity factor	4	0	0	8	8
	E2	System efficiency	0	2	0	4	14
	E3	Lifetime	0	6	6	1	7

Table 4. Fuzzy score results

Criteria	Sub-criteria code	Sub-criteria	Fuzzy average	Non-fuzzy average
Resource	A1	Area intensity	(0.463,0.713,0.85)	0.675
	A2	Material intensity	(0.488,0.738,0.9)	0.708
	A3	Energy intensity(construction)	(0.325,0.538,0.725)	0.529
	A4	Energy intensity(fuel)	(0.388,0.588,0.725)	0.567
Environmental	B1	CO_2 intensity (construction)	(0.35,0.513,0.663)	0.508
	B2	CO_2 intensity (fuel)	(0.35,0.563,0.713)	0.542
Economic-Financial	C1	Capital intensity(construction)	(0.263,0.438,0.625)	0.442
	C2	Capital intensity(fuel)	(0.463,0.688,0.825)	0.658
	C3	Delivered cost	(0.325,0.488,0.663)	0.492
	C4	Net present value	(0.225,0.363,0.538)	0.375
	C5	Internal return rate	(0.45,0.65,0.813)	0.638
	C6	Payback period	(0.413,0.575,0.688)	0.558
Social	D1	Current installed capacity	(0.538,0.75,0.825)	0.704
	D2	Growth rate	(0.338,0.513,0.688)	0.513
Technology	E1	Capacity factor	(0.5,0.7,0.85)	0.683
	E2	System efficiency	(0.625,0.875,0.95)	0.817
	E3	Lifetime	(0.363,0.613,0.775)	0.583

FUZZY SWARA METHOD RESULTS

In this section, the weight and importance of indicators are determined using the SWARA method. The Fuzzy SWARA method is implemented separately on the sub-criteria of each main criterion, which is given below.

FUZZY SWARA RESULTS OF "RESOURCES" SUB-CRITERIA

The first step in this method is to sort the indicators according to their importance in descending order (from high to low), using the average scores given in table 4. This process is carried out as shown in table 5. In this section the fuzzy average is the same as S_j.

In the next step, using equations (2) to (4), we calculate the weight of the criteria based on the opinions of each expert. The results are shown in Table 6 and Fig. 1. The W_j column shows the weight of the criteria. For example, for the "area intensity" criterion, it is calculated as follows:

$$K_{A1} = (1,1,1) + S_{A1} = (1.463,1.713,1.85)$$

$$q_{Q3} = \frac{q_{j-1}}{K_j} = \frac{q_{A2}}{K_{A1}} = \frac{(1,1,1)}{(1.463,1.713,1.85)} = (0.541,0.584,0.684)$$

$$W_{Q3} = \frac{q_j}{\sum q_j} = \frac{q_{A1}}{\sum q_j} = \frac{(0.541,0.584,0.684)}{(2.036,2.191,2.548)} = (0.212,0.267,0.336)$$

Also, to defuzzify W based on equation (5), we have:

$$W_{crisp} = \frac{\left(W_j^m - W_j^l\right) + \left(W_j^u - W_j^l\right)}{3} + W_j^l = \frac{(0.267-0.212)+(0.336-0.212)}{3} + 0.212 = 0.272$$

According to Table 6, the material intensity with a weight of 0.439 has won the first rank. The area intensity with a weight of 0.267 has obtained the second place and the energy intensity (fuel) with a weight of 0.175 has achieved the third place. Figure 1 shows the priority of other sub-criteria.

Table 5. Descending order of resource indices

Sub-criteria code	Sub-criteria	Fuzzy average
A2	Material intensity	(0.488, 0.738, 0.9)
A1	Area intensity	(0.463, 0.713, 0.85)
A4	Energy intensity(fuel)	(0.388, 0.588, 0.725)
A3	Energy intensity(Construction)	(0.325, 0.538, 0.725)

Table 6. The weights of resource sub-criteria

Sub-criteria code	Sub-criteria	S_j	K_j	Q_j	Fuzzy W_j	Non-fuzzy weight	Normal non-fuzzy weight
A2	Material intensity	-	(1,1,1)	(1,1,1)	(0.392, 0.456, 0.491)	0.447	0.439
A1	Area intensity	(0.463, 0.713, 0.85)	(1.463, 1.713, 1.85)	(0.541, 0.584, 0.684)	(0.212, 0.267, 0.336)	0.272	0.267
A4	Energy intensity(fuel)	(0.388,0.588,0.725)	(1.388,1.588,1.725)	(0.313,0.368,0.493)	(0.123,0.168,0.242)	0.178	0.175
A3	Energy intensity (construction)	(0.325,0.538,0.725)	(1.325,1.538,1.725)	(0.182,0.239,0.372)	(0.071,0.109,0.183)	0.121	0.119

FUZZY SWARA RESULTS OF "ENVIRONMENTAL" SUB-CRITERIA

In a similar way, calculations were made for the environmental sub-criteria which are given in Table 7 and Fig. 2.

According to Table 7, among the environmental sub-criteria, CO_2 intensity (fuel) with a weight of 0.599 is ranked first and CO_2 intensity (construction) with a weight of 0.401 is ranked second.

FUZZY SWARA RESULTS OF "ECONOMIC-FINANCIAL" SUB-CRITERIA

In a similar way, calculations were made for the economic sub-criteria which are given in Table 8 and Fig. 3.

According to Table 8, among the economic-financial sub-criteria, the capital intensity (fuel) with a weight of 0.374 has gained the first rank. The Internal return rate with a weight of 0.233 has been second and the Payback period with a weight of 0.152 has got the third place.

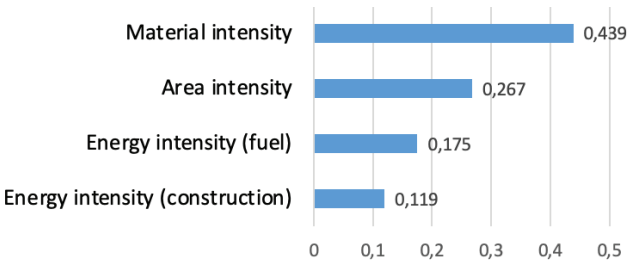


Figure 1. Weights of resource sub-criteria.

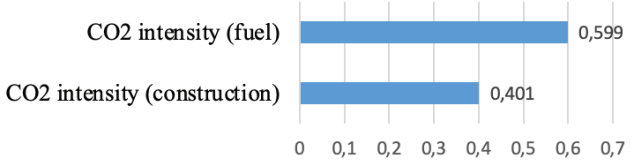


Figure 2. Final weights of environmental sub-criteria.

Table 7. Weights of environmental sub-criteria

Sub-criteria code	Sub-criteria	S_j	K_j	Q_j	Fuzzy W_j	Non-fuzzy weight	Normal non-fuzzy weight
B2	CO_2 intensity (fuel)	-	(1,1,1)	(1,1,1)	(0.574,0.602,0.624)	0.600	0.599
B1	CO_2 intensity (construction)	(0.35,0.513,0.663)	(1.35,1.513,1.663)	(0.602,0.661,0.741)	(0.346,0.398,0.463)	0.402	0.401

Table 8. Economic-financial sub-criteria weights

Sub- criteria code	Sub- criteria	S_j	K_j	Q_j	W_j Fuzzy	Non-fuzzy weight	Normal non-fuzzy weight
C2	Capital intensity (fuel)	-	(1,1,1)	(1,1,1)	(0.325,0.39,0.44)	0.385	0.374
C5	Internal return rate	(0.45,0.65,0.813)	(1.45,1.65,1.813)	(0.552,0.606,0.69)	(0.179,0.237,0.303)	0.240	0.233
C6	Payback period	(0.413,0.575,0 .688)	(1.413,1.575,1.688)	(0.327,0.385,0.488)	(0.106,0.15,0.215)	0.157	0.152
C3	Delivered cost	-	(1.325,1.488,1.663)	(0.197,0.259,0.368)	(0.064,0.101,0.162)	0.109	0.106
C1	Capital intensity (construction)	(0.263,0.438,0.625)	(1.263,1.438,1.625)	(0.121,0.18,0.292)	(0.039,0.07,0.128)	0.079	0.077
C4	Net present value	(0.225,0.363,0.538)	(1.225,1.363,1.538)	(0.079,0.132,0.238)	(0.026,0.052,0.105)	0.061	0.059

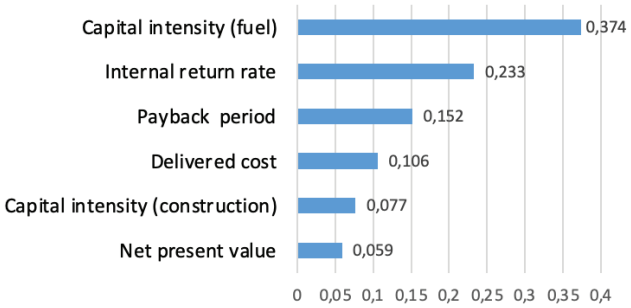


Figure 3. Final weights of economic-financial sub-criteria.

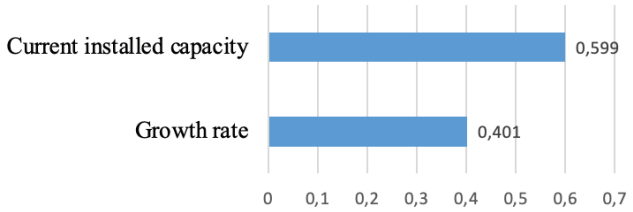


Figure 4. Weights of social sub-criteria.

Table 9. The weights of social sub-criteria

Sub-criteria code	Sub-criteria	S_j	K_j	Q_j	W_j Fuzzy	Non-fuzzy weight	Normal non-fuzzy weight
D1	Current installed capacity	-	(1,1,1)	(1,1,1)	(0.572,0.602,0.628)	0.601	0.599
D2	Growth rate	(0.338,0.513,0.688)	(1.338,1.513,1.688)	(0.593,0.661,0.748)	(0.339,0.398,0.469)	0.402	0.401

FUZZY SWARA RESULTS OF "SOCIAL" SUB-CRITERIA

In a similar way, calculations were made for the social sub-criteria which are given in Table 9 and Fig. 4.

According to Table 9, among the social sub-criteria, the current installed capacity with a weight of 0.599 is ranked first and the growth rate with a weight of 0.499 has been ranked second.

FUZZY SWARA RESULTS OF "TECHNOLOGY" SUB-CRITERIA

In a similar way, calculations were made for technology sub-criteria which are given in Table 10 and Fig. 5.

According to Table 10, among technology sub-criteria, system efficiency with a weight of 0.502 has been ranked first. The capacity factor with a weight of 0.302 has won the second place and the lifetime with a weight of 0.196 has won the third place.

MAIN CRITERIA WEIGHTS

To calculate the weight of the main criteria, first, the Likert scores of the sub-criteria of each criterion are averaged to determine the score of the main criterion, then the criteria are sorted in descending order based on the average scores in Table 4, and then in a similar way, with the SWARA algorithm the weight of the criteria is calculated. The results are shown in Table 11 and Fig. 6. Accordingly, the technology criterion with a weight of 0.402 has gained

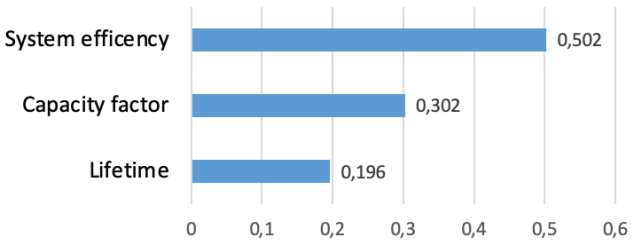


Figure 5. Weights of technology sub-criteria.

Table 10. The weights of technology sub-criteria

Sub-criteria code	Sub-criteria	S_j	K_j	Q_j	W_j Fuzzy	Non-fuzzy weight	Normal non-fuzzy weight
E2	System efficiency	-	(1,1,1)	(1,1,1)	(0.464,0.512,0.542)	0.506	0.502
E1	Capacity factor	(0.5,0.7,0.85)	(1.5,1.7,1.85)	(0.541,0.588,0.667)	(0.251,0.301,0.361)	0.304	0.302
E3	Lifetime	(0.363,0.613,0.775)	(1.363,1.613,1.775)	(0.305,0.365,0.489)	(0.141,0.187,0.265)	0.198	0.196

the first rank, the resource criterion with a weight of 0.253 has earned the second rank, and the social criterion with a weight of 0.161 has earned the third rank.

FINAL WEIGHTS OF SUB-CRITERIA

The final weights of the sub-criteria is obtained by multiplying the the main criteria weight by the relative weight of the sub-criteria calculated in the previous step, which is given in Table 12.

RESULTS OF FUZZY WASPAS METHOD

In this section, 4 types of energy are used to rank research alternatives. These 4 types of energy are: Solar energy systems (S1), Wind energy systems (S2), Phosphoric acid fuel cells (S3) and Solid oxide fuel cells (S4).

First step: establishing the decision matrice

In this step, we form the decision matrice of opinions. The decision matrice of the WASPAS method is a matrice consisting of criteria (sub-criteria) and research alternatives. This decision matrice is completed by 20 experts and then it is integrated by arithmetic mean method. The fuzzy feedback decision matrice is given in Table 13. In this matrice, there are 17 research sub-criteria in the column and 4 alternatives in the row.

Second step: normalization of the decision matrice

In the second step, we normalize the decision matrice based on equations (6) and (7). For example, for cell A11, which is the intersection of criterion A1 and alternative S1, normalization is done as follows. First, the fuzzy numbers of the criterion column S1 should be divided by the maximum of the upper limits of the fuzzy numbers of the criterion column A1. The maximum value is equal to 7.9. For example, for the first level, the calculations are as follows:

$$A_{11}^{Normal} = \frac{(4.1,5.9,7.9)}{7.9} = (0.519,0.747,1)$$

Calculations are done in a similar way for other cells, which are given in Table 14.

Third step: determining the weighted sum and weighted multiplication values

In the third step, using equations (8) and (9), we calculate the values of WSM (weighted sum) (Q) and WPS model (weighted product) (P). And then, we deffuzify these values using equations (10) and (11), which is given in table 15.

Fourth step: determining the final score and ranking the options

In the fourth step, the score of each attribute is calculated using equation (10), but before that, the value of λ is calculated based on equation (2), which is given in table 16 and Fig. 17.

$$\lambda = \frac{\sum_{i=1}^{m} P_i}{\sum_{i=1}^{m} P_i + \sum_{i=1}^{m} Q_i} = \frac{2.61}{2.61 + 2.634} = 0.498$$

Table 11. The weights of the main criteria

Code	Main criteria	S_j	K_j	Q_j	W_j Fuzzy	Normal non-fuzzy weight	Non-fuzzy weight
E	Technology	-	(1,1,1)	(1,1,1)	(0.354,0.42,0.461)	0.402	0.411
A	Resource	(0.416,0.644,0.8)	(1.416,1.644,1.8)	(0.556,0.608,0.706)	(0.196,0.255,0.326)	0.253	0.259
D	Social	(0.438,0.631,0.756)	(1.438,1.631,1.756)	(0.316,0.373,0.491)	(0.112,0.157,0.226)	0.161	0.165
C	Economic-Financial	(0.356,0.533,0.692)	(1.356,1.533,1.692)	(0.187,0.243,0.362)	(0.066,0.102,0.167)	0.109	0.112
B	Environmental	(0.35,0.538,0.688)	(1.35,1.538,1.688)	(0.111,0.158,0.268)	(0.039,0.066,0.124)	0.075	0.076

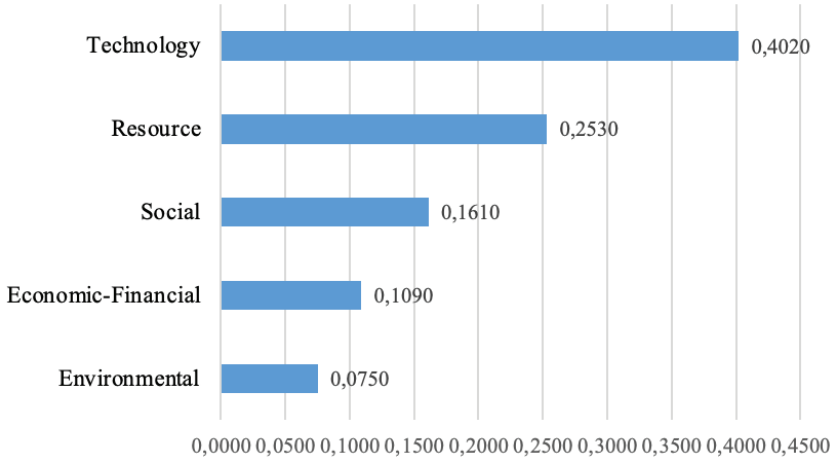


Figure 6. The main criteria weights.

Table 12. Weights and final ranking of sub-criteria

Criteria	Criteria weight	Sub-criteria	Relative weight	Final weight	Final rank
Resource	0.253	Area intensity	0.267	0.0676	6
		Material intensity	0.439	0.1112	3
		Energy intensity (construction)	0.119	0.0301	11
		Energy intensity (fuel)	0.175	0.0442	9
Environmental	0.075	CO_2 intensity (construction)	0.401	0.299	12
		CO_2 intensity (fuel)	0.599	0.0447	8
Economic-Financial	0.109	Capital intensity (construction)	0.077	0.0084	16
		Capital intensity (fuel)	0.374	0.0408	10
		Delivered cost	0.106	0.0115	15
		Net present value	0.059	0.0064	17
		Internal return rate	0.233	0.0254	13
		Payback period	0.152	0.0166	14
Social	0.161	Current installed capacity	0.599	0.096	4
		Growth rate	0.401	0.0646	7
Technology	0.402	Capacity factor	0.302	0.1214	2
		System efficiency	0.502	0.2017	1
		Lifetime	0.196	0.0788	5

Table 13. Fuzzy WASPAS decision matrice

	A1	A2	A3	...	E1	E2	E3
S1	(4.1,5.9,7.9)	(2.6,4.3,6.3)	(3,4.4,6.4)	...	(3.2,5.1,7.1)	(4.2,6,8)	(3.2,4.7,6.7)
S2	(3.9,5.6,7.6)	(3.3,5,7)	(3.1,4.6,6.6)	...	(3.2,4.9,6.9)	(2.4,3.8,5.8)	(3.4,4.8,6.8)
S3	(3.1,4.5,6.5)	(3.5,5.1,7.1)	(4,5.7,7.7)	...	(3,4.5,6.5)	(2,3.7,5.7)	(3.5,5.2,7.2)
S4	(4,5.6,7.6)	(3.8,5.5,7.5)	(4.7,6.4,8.4)	...	(2.7,4.1,6.1)	(3.2,4.8,6.8)	(3.1,4.4,6.4)

Table 14. Fuzzy WASPAS normal matrice

	A1	A2	A3	...	E1	E2	E3
S1	(0.519,0.747,1)	(0.347,0.573,0.84)	(0.357,0.524,0.762)	...	(0.451,0.718,1)	(0.525,0.75,1)	(0.444,0.653,0.931)
S2	(0.494,0.709,0.962)	(0.44,0.667,0.933)	(0.369,0.548,0.786)	...	(0.451,0.69,0.972)	(0.3,0.475,0.725)	(0.472,0.667,0.944)
S3	(0.392,0.57,0.823)	(0.467,0.68,0.947)	(0.476,0.679,0.917)	...	(0.423,0.634,0.915)	(0.25,0.463,0.713)	(0.486,0.722,1)
S4	(0.506,0.709,0.962)	(0.507,0.733,1)	(0.56,0.762,1)	...	(0.38,0.577,0.859)	(0.4,0.6,0.85)	(0.431,0.611,0.889)

Table 15. WSM and WPM values

	Q (fuzzy)	P (fuzzy)	Q (crisp)	P (crisp)
A1	(0.443,0.662,0.929)	(0.436,0.655,0.924)	0.678	0.672
A2	(0.416,0.617,0.883)	(0.409,0.61,0.878)	0.639	0.633
A3	(0.403,0.613,0.88)	(0.392,0.606,0.874)	0.632	0.624
A4	(0.457,0.666,0.933)	(0.453,0.663,0.93)	0.685	0.682

Table 16. Scores of attributes and their ranking

Code	Alternatives	K	k Normal	Rank
A1	Phosphoric acid fuel cells	0.675	0.257	2
A2	Wind energy systems	0.636	0.242	3
A3	Solar energy systems	0.628	0.239	4
A4	Solid oxide fuel cells	0.684	0.261	1

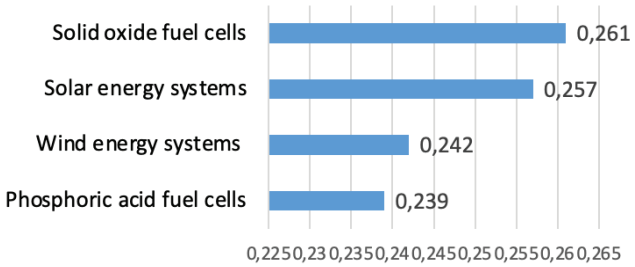


Figure 7. Weights and final ranking of alternatives.

As a result, the solid oxide fuel cells ranked as the first attribute in the sustainability evaluation among the selected green power generation systems. This is followed by solar energy systems, the land-based wind energy systems and phosphoric acid fuel, as follows:

1. Solid oxide fuel cells.
2. Solar energy systems [poly-silicon]
3. Wind energy systems [Land-based].
4. Phosphoric acid fuel cells.

In this study, the hybrid Fuzzy SWARA/WASPAS method is done for sustainability-financial evaluation of renewable energy systems in order to help sustainable energy development. Solid oxides fuel cells (SOFCs) running with natural gas or propane as fuel are the most efficient fuel cells for electricity generation with low emissions [17]. As natural gas is one of natural resources in Iran, it completely makes sense that our experts and decision makers ranked it over other indicators. So, it has gained the first rank, whereas in [17] it ranked as the second attribute, using SWARA/ARAS methd without considering three financial subcriteria. Also, solar energy systems have no greenhouse gas emissions during operation and construction, so their contribution to global warming is negligible. Solar energy that is always available in some regions in Iran has gained the second rank in contrast with the last place in [17].

CONCLUSION

In this study, we have investigated four types of renewable energy systems considering five main sustainability-financial-technical indicators and their seventeen sub-criteria. The evaluated technologies consisting land-based wind energy systems, solid oxide fuel cells, phosphoric acid fuel cells, and poly-silicon solar energy systems are considered as alternatives. The five sustainability indicators were resource, environmental, economic-financial, social and technological. Also, their seventeen subcategories were: area, material, energy-construction, energy-fuel, CO_2 construction, CO_2-fuel, capital-construction, capital-fuel, delivered cost of energy, current installed capacity, growth rate, capacity factor, net present value, Internal return rate, payback period, system efficiency, and lifetime. To aassess the importance of these criteria and their subcriteria considering vagueness of the data we used the Fuzzy Step-wise Weight Assessment Ratio Analysis (F-SWARA) method consulting twenty energy experts according to the current state of environment, economy and society in Iran. Then after finding all weights, we used Fuzzy Weighted Aggregated Sum Product Assessment (F-WASPAS) to prioritize the alternatives. The final results of the study showed that the most to the least preferable renewable energy system are: system efficiency, capacity factor, material intensity, current installed capacity, lifetime, area intensity, growth rate, CO_2 intensity-fuel, energy intensity-fuel, capital intensity-fuel, energy intensity-construction, CO_2 intensity-construction, internal efficincy rate, payback period, delivered cost of energy, capital intensity-construction and net present value. The ranking of the different renewable energy systems showed that solid oxide fuel cells are ranked first with respect to sustainability indicators and their sub-indicators, followed by solar energy systems, then wind energy systems, and lastly phosphoric acid fuel cells. The limitations of this study can be manifested through the small number of considered nrenewable energy systems, as there are many other technologies to investigate and prioritize. Also, in real world situations various criteria and subcriteria are involved in making decisions, depending on the region. It is suggested that in the future studies, the uncertainty of the data should be entered into the model in gray form and its effect on the criteria and sub-criteria should be investigated. Also, other multi-indicator decision making models can be used to compare the results with the proposed model. Finally, more criteria and sub-criteria can be used in future researches.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. Energy Strategy Rev 2019;24:38–50. [CrossRef]

[2] Ghenai C, Janajreh I. Comparison of resource intensities and operational parameters of renewable, fossil fuel, and nuclear power systems. Int J Thermal Environ Eng 2013;5:95–104.

[3] Erdiwansyah, Mamat R, Sani MSM, Sudhakar K. Renewable energy in Southeast Asia: Policies and recommendations. Sci Total Environ 2019;670:1095–1102. [CrossRef]

[4] Benedek J, Sebestyén TT, Bartók B. Evaluation of renewable energy sources in peripheral areas and renewable energy-based rural development. Renew Sustain Energy Rev 2018;90:516–535. [CrossRef]

[5] Gong J, Li C, Wasielewski MR. Advances in solar energy conversion. Chem Soc Rev 2019;48:1862–1864. [CrossRef]

[6] Ager JW, Lapkin AA. Chemical storage of renewable energy. Science (1979) 2018;360:707–708. [CrossRef]

[7] Marugán AP, Márquez FPG, Perez JMP, Ruiz-Hernández D. A survey of artificial neural network in wind energy systems. Appl Energy 2018;228:1822–1836. [CrossRef]

[8] Shoaib M, Siddiqui I, Rehman S, Khan S, Alhems LM. Assessment of wind energy potential using wind energy conversion system. J Clean Prod 2019;216:346–360. [CrossRef]

[9] Raza SS, Janajreh I, Ghenai C. Sustainability index approach as a selection criteria for energy storage system of an intermittent renewable energy source. Appl Energy 2014;136:909–920. [CrossRef]

[10] Hacatoglu K, Dincer I, Rosen MA. Sustainability assessment of a hybrid energy system with hydrogen-based storage. Int J Hydrogen Energy 2015;40:1559–1568. [CrossRef]

[11] Papilo P, Marimin, Hambali E, Sitanggang IS. Sustainability index assessment of palm oil-based bioenergy in Indonesia. J Clean Prod 2018;196:808–820. [CrossRef]

[12] Diemuodeke EO, Addo A, Oko COC, Mulugetta Y, Ojapah MM. Optimal mapping of hybrid renewable energy systems for locations using multi-criteria decision-making algorithm. Renew Energy 2019;134:461–477. [CrossRef]

[13] Baumann M, Weil M, Peters JF, Chibeles-Martins N, Moniz AB. A review of multi-criteria decision making approaches for evaluating energy storage systems for grid applications. Renew Sustain Energy Rev 2019;107:516–534. [CrossRef]

[14] Mayer MJ, Szilágyi A, Gróf G. Environmental and economic multi-objective optimization of a household level hybrid renewable energy system by genetic algorithm. Appl Energy 2020;269:115058. [CrossRef]

[15] Nguyen HT, Safder U, Nhu Nguyen XQ, Yoo CK. Multi-objective decision-making and optimal sizing of a hybrid renewable energy system to meet the dynamic energy demands of a wastewater treatment plant. Energy 2020;191:116570. [CrossRef]

[16] Østergaard PA, Johannsen RM, Duic N. Sustainable Development using renewable energy systems. Int J Sustain Energy Plan Manag 2020;29:1–6. [CrossRef]

[17] Ghenai C, Albawab M, Bettayeb M. Sustainability indicators for renewable energy systems using multi-criteria decision-making model and extended SWARA/ARAS hybrid method. Renew Energy 2020;146:580–597. [CrossRef]

[18] Hashemkhani Zolfani S, Görçün ÖF, Küçükönder H. Evaluating logistics villages in Turkey using hybrid improved fuzzy swara (Imf swara) and fuzzy mabac techniques. Technol Econ Dev Economy 2021;27:1582–612. [CrossRef]

[19] Alghassab M. Quantitative assessment of sustainable renewable energy through soft computing: Fuzzy AHP-TOPSIS method. Energy Rep 2022;8:12139–12152. [CrossRef]

[20] Cutore E, Volpe R, Sgroi R, Fichera A. Energy management and sustainability assessment of renewable energy communities: The Italian context. Energy Convers Manag 2023;278:116713. [CrossRef]

[21] Thirumalaivasan N, Nangan S, Kanagaraj K, Rajendran S. Assessment of sustainability and environmental impacts of renewable energies: Focusing on biogas and biohydrogen (Biofuels) production. Process Safety and Environmental Protection 2024;189:467–485. [CrossRef]

[22] Liu L, Zhai R, Xu Y, Hu Y, Liu S, Yang L. Comprehensive sustainability assessment and multi-objective optimization of a novel renewable energy driven multi-energy supply system. Appl Therm Eng 2024;236:121461. [CrossRef]

[23] Mardani A, Nilashi M, Zakuan N, Loganathan N, Soheilirad S, Saman MZM, et al. A systematic review and meta-Analysis of SWARA and WASPAS methods: Theory and applications with recent fuzzy developments. Appl Soft Comput 2017;57:265–292. [CrossRef]

[24] Zheng L, Mahmood T, Ahmmad J, Rehman UU, Zeng S. Spherical Fuzzy Soft Rough Average Aggregation Operators and Their Applications to Multi-Criteria Decision Making. IEEE Access 2022;10:27832–27852. [CrossRef]

[25] Mahmood T, Rehman UU, Ali Z, Aslam M. Bonferroni mean operators based on bipolar complex fuzzy setting and their applications in multi-attribute decision making. AIMS Mathematics 2022;7:17166–17197. [CrossRef]

[26] Garg H, Mahmood T, ur Rehman U, Nguyen GN. Multi-attribute decision-making approach based on Aczel-Alsina power aggregation operators under

bipolar fuzzy information & its application to quantum computing. Alexandria Engineering Journal 2023;82:248–259. [CrossRef]

[27] Mahmood T, Rehman U. Digital technology implementation and impact of artificial intelligence based on bipolar complex fuzzy Schweizer–Sklar power aggregation operators. Appl Soft Comput 2023;143:110375. [CrossRef]

[28] Naeem M, Mahmood T, Rehman U, Mehmood F. Classification of renewable energy and its sources with decision-making approach based on bipolar complex fuzzy frank power aggregation operators. Energy Strategy Rev 2023;49:101162. [CrossRef]

[29] Mahmood T, Rehman UU, Shahab S, Ali Z, Anjum M. Decision-Making by Using TOPSIS Techniques in the Framework of Bipolar Complex Intuitionistic Fuzzy N-Soft Sets. IEEE Access 2023;11:105677–105697. [CrossRef]

[30] Adan HK, Başaran Filik Ü. Technical and economic evaluation of a standalone and on grid hybrid renewable energy: A case study at Eskişehir Technical University. Sigma J Eng Nat Sci 2021;39:184–194. [CrossRef]

[31] Yolcu A. Intuitionistic fuzzy hypersoft topology and ıts applications to multi-criteria decision-making. Sigma J Eng Nat Sci 2023;41:106–118. [CrossRef]

[32] Güven AF, Yörükeren N. A comparative study on hybrid GA-PSO performance for stand-alone hybrid energy systems optimization. Sigma J Eng Nat Sci 2024;42:1410–1438. [CrossRef]

[33] Keršulienė V, Kazimieras Zavadskas E, Turskis Z. Selection of rational dispute resolution method by applying new step-wise weight assessment ratio analysis (Swara). J Business Econ Manag 2010;11:243-225. [CrossRef]

[34] Rani P, Mishra AR, Krishankumar R, Mardani A, Cavallaro F, Ravichandran KS, et al. Hesitant fuzzy SWARA-complex proportional assessment approach for sustainable supplier selection (HF-SWARA-COPRAS). Symmetry (Basel) 2020;12. [CrossRef]

[35] Mavi RK, Goh M, Zarbakhshnia N. Sustainable third-party reverse logistic provider selection with fuzzy SWARA and fuzzy MOORA in plastic industry. Int J Adv Manufact Technol 2017;91:2401–2418. [CrossRef]

[36] Zavadskas EK, Turskis Z, Antucheviciene J, Zakarevicius A. Optimization of Weighted Aggregated Sum Product Assessment. Elektronika Ir Elektrotechnika 2012;122:3–6. [CrossRef]

[37] Turskis Z, Zavadskas EK, Antucheviciene J, Kosareva N, Turskis Z, Zavadskas K, et al. A hybrid model based on fuzzy AHP and fuzzy WASPAS for construction site selection. vol. 10. 2015. [CrossRef]

[38] Pamucar D, Deveci M, Canıtez F, Lukovac V. Selecting an airport ground access mode using novel fuzzy LBWA-WASPAS-H decision making model. Eng Appl Artif Intell 2020;93:103703. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1127–1137, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2031

Research Article

Dynamic mechanical, morphological, and thermal characteristics of polypropylene/red mud polymer composites

Beril EKER[1,*], Özlem YAĞCI[2], Elif ULUTAŞ[3], Munir TAŞDEMIR[3], Volkan Murat YILMAZ[4]

[1]Science and Technology Application and Research Center, Yıldız Technical University, Istanbul, 34220, Türkiye
[2]Department of Physics, Yildiz Technical University, Istanbul, 34220, Türkiye
[3]Department of Metallurgical and Materials Engineering, Faculty of Technology, Marmara University, Istanbul, 34722, Türkiye
[4]Central Research Laboratory Research & Application Center, Bartın University, Bartın, 74100, Türkiye

ARTICLE INFO

Article history
Received: 03 November 2025
Revised: 12 January 2026
Accepted: 14 February 2026

Keywords:
Composites; Polypropylene; Red Mud; Storage Modulus; Thermal Properties

ABSTRACT

This study aims to examine how different amounts of red mud (RM) (0, 10, 20, 30, and 40 wt.%) affect the dynamic mechanical and thermal characteristics of polypropylene (PP)/red mud (RM) polymer composites. Fourier Transform Infrared Spectroscopy (FTIR) was used to confirm the presence of PP and RM in the samples. The distribution of RM in the PP matrix was analysed using Scanning Electron Microscopy (SEM). X-ray Diffraction (XRD) results showed that crystallinity of polypropylene increased from 80.04% for pure PP to 94.94% for the composite containing 40 wt% red mud. Thermogravimetric Analysis (TGA) was employed to analyse the thermal stability of the polypropylene (PP)/red mud (RM) polymer composites. Specifically, it was used to assess the thermal stability of the composites, with results indicating excellent stability for the 40 wt.% RM composites. Differential Scanning Calorimetry (DSC) was used to compare the melting temperatures (Tm) of the composites. There is a slight increase in Tm, from 169.08 °C for pure polypropylene to 171.45 °C for the composite with 40 wt% red mud. Furthermore, Dynamic Mechanical Analysis (DMA) results revealed a 58% reduction from the storage modulus of the pure polypropylene (6.28×10^8 Pa) to the storage modulus of the composite containing 40 wt% red mud (2.60×10^8 Pa).
The research highlights that red mud (RM) waste is a low-cost, sustainable filler material in a polypropylene matrix. It enhances thermal stability and structural performance for various engineering applications.

Cite this article as: Eker B, Yağcı Ö, Ulutaş E, Taşdemir M, Yılmaz VM. Dynamic mechanical, morphological, and thermal characteristics of polypropylene/red mud polymer composites. Sigma J Eng Nat Sci 2026;44(2):1127–1137.

***Corresponding author.**
*E-mail address: beker@yildiz.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

The demand for eco-friendly polymers and polymer composites derived from renewable resources is increasing. Growing environmental concerns and the depletion of fossil resources drive this demand. The use of waste materials has gained increasing attention due to their advantages from both environmental and economic perspectives. The chemical composition of red mud is 37.67% Fe_2O_3, 5.55% TiO_2, 20.42% Al_2O_3. Red mud has reddish-brown colouration and a very fine grain size. The characteristic fine grains of red mud favour its use as an admixture in mortar and concrete. The mineralogical composition of red mud varies with the bauxite source and production method.

Red mud is a waste byproduct produced during the extraction of alumina from bauxite through the Bayer process. Acidic, electrothermic, and basic methods are used for alumina production. The demand for alumina has increased with global Al consumption. For one tonne of alumina or 0.5 tonnes of aluminium metal produced, approximately one tonne of red mud is generated. Red mud is the most significant waste issue in this industry due to the increasing production of aluminium. Between 120 and 150 million tons of red mud are produced annually [1]. Disposing of red mud is challenging due to its toxicity and high alkalinity. Red mud is a byproduct of aluminium production and contains hazardous substances, including heavy metals and radioactive materials. There are various disposal methods as follows: 1) Ocean dumping: the mud is released into coastal waters, 2) lagoons: storing the mud in large ponds, 3) dry stacking: the mud is stored in a dry form, 4) dry cake disposal: solidifying the mud before disposal [2]. As aluminium production increases, bauxite consumption and red mud generation also increase. The accumulation of red mud causes several environmental problems. Under hot, dry conditions, dried red mud particles can cause air pollution through wind dispersion. Moreover, red mud stored in dams or disposal sites may damage biological and ecological systems. Red mud is hazardous industrial waste composed of iron compounds and heavy metals. If disposal dams are not constructed and managed in compliance with applicable legislation and engineering standards, red mud will irreversibly harm living organisms and the environment. Therefore, the industrial utilization of red mud is a critical strategy to mitigate its environmental harm, and research into its safe and sustainable reuse is ongoing [3,4].

Red mud has been assessed in various industrial sectors, including building materials production [5,6], catalysis [7], as well as in the development of materials like metal and/or polymer matrix composites, geopolymers [8,9], iron, steel manufacturing [10], and ceramics [11].

With the utilization of red mud, environmental pollution will be prevented, the storage problem will be eliminated, the cost of aluminium production will be reduced, and an economic contribution will be provided to the metallurgy, construction and chemical industries. Red mud has various potential uses, such as an adsorbent, construction material, and pigment in the ceramics sector. It can additionally function as an absorbent for exhaust gases. Another possible use is the recovery of precious elements from red mud, including iron, aluminium, titanium, and rare-earth metals such as scandium, cerium, and lanthanum.

The first article on red mud was published in 1976, but research on this topic gained much attention only in 2010. Since then, the number of articles on red mud has increased rapidly. This rise is primarily due to environmental concerns about hazardous red mud and interest in adopting circular-economy practices to manage it. This upward trend is expected to persist because current research doesn't provide enough information to achieve full industrial utilisation of red mud. However, with further research, the full potential of red mud can be unlocked, inspiring a new era of sustainable waste management and circular economy practices [12].

Red mud can be added to polymer materials as a reinforcement and filler. In recent years, red mud has been incorporated into polymer materials, thereby expanding the use of newly developed blends. Red mud is a plentiful and low-cost reinforcing material for thermoplastics used in construction, electrical applications, decorative household applications, and the packaging industry [13].

Zhang et al.'s study reported that as the red mud content in PP-based composites increased, bending strength, tensile strength, heat deflection temperature, and Vicat softening point all improved, whereas shock resistance decreased. In their research, the PP composites reinforced with 15 wt% red mud exhibited the highest tensile strength [14].

Akıncı et al. investigated the impact of red mud content (10, 20, 30, 40, and 50 wt.%) on isotactic polypropylene/red mud composites in terms of tension, three-point bending, and hardness test results. Their findings indicated that as the red mud content increased, the composites' hardness improved, whereas bending strength, tensile strength, and elongation at failure decreased [15]. Most existing studies focus on either mechanical or thermal properties individually, often within a limited range of filler loadings [16,17]. Recent studies have also examined the tribological and microstructural performance of filler-matrix interaction and dispersion quality in terms of overall composite performance [18,19].

The originality of this study lies in the use of untreated red mud, an industrial waste by-product, as a filler in polypropylene matrix without surface modification or compatibilizers. Most previous studies focus on chemically treated red mud or systems designed for better dispersion. This work provides a direct evaluation of the effects of raw red mud on the morphological, mechanical, and thermomechanical properties of PP.

The main goal is to explore the potential use of waste RM as a low-cost and sustainable filler in PP matrix. The existing literature indicates that PP composites incorporating RM have primarily been examined for their limited mechanical or thermal properties. In contrast, this study systematically

investigates the combined effects of varying RM additions on the dynamic mechanical behaviour, crystalline structure, and thermal stability of the composites. In this context, the study aims not only to evaluate the effect of RM on the overall performance of PP composites but also to reveal the structure–property relationships through detailed analyses. This approach, based on recycling industrial waste, contributes to both reducing material costs and mitigating environmental impacts associated with RM storage. The results obtained are expected to enable RM to find broader applications in the development of high-performance polymer composites. Figure 1 shows graphical presentation of the study.

EXPERIMENTAL

Materials

The red mud was obtained from Güray Seramik Company in Avanos, Nevşehir, Turkey. The PP (Moplen EP 3307) was sourced from Lyondell Basell. It has a density of 0.900 g/cm^3, an MFI of 15 g/10 min (230°C, 2.16 kg), and a heat deflection temperature (0.45 MPa, unannealed) of 95.0°C.

Production of The Composites

The red mud was pre-dried at 85°C for 24 hours to break down associated hydrates. Then, Siemens Simatic C7-621 is used to produce powders. The particle size of the red mud ranged from 38 to 250 micrometres. Solid compositions were mechanically premixed with an LB-5601 liquid-solids batch blender for 15 minutes. The composites were produced using a Microsan extruder at temperatures of 170-220°C, a pressure of 25-30 bar, and a rotational speed of 23 rpm. Following extrusion, the polymer composites were dried in a vacuum oven at 80°C for 24 hours. Finally, test samples were produced using an injection moulding machine at temperatures of 170–220°C, a pressure of 90–120 bar, and a screw speed of 23 rpm.

Table 1 presents the percentages of polypropylene and red mud for each group.

Table 1. The makeup of the various polymer composite formulations.

Groups	PP (wt %)	Red Mud (RM) (wt %)
1	100	-
2	90	10
3	80	20
4	70	30
5	60	40

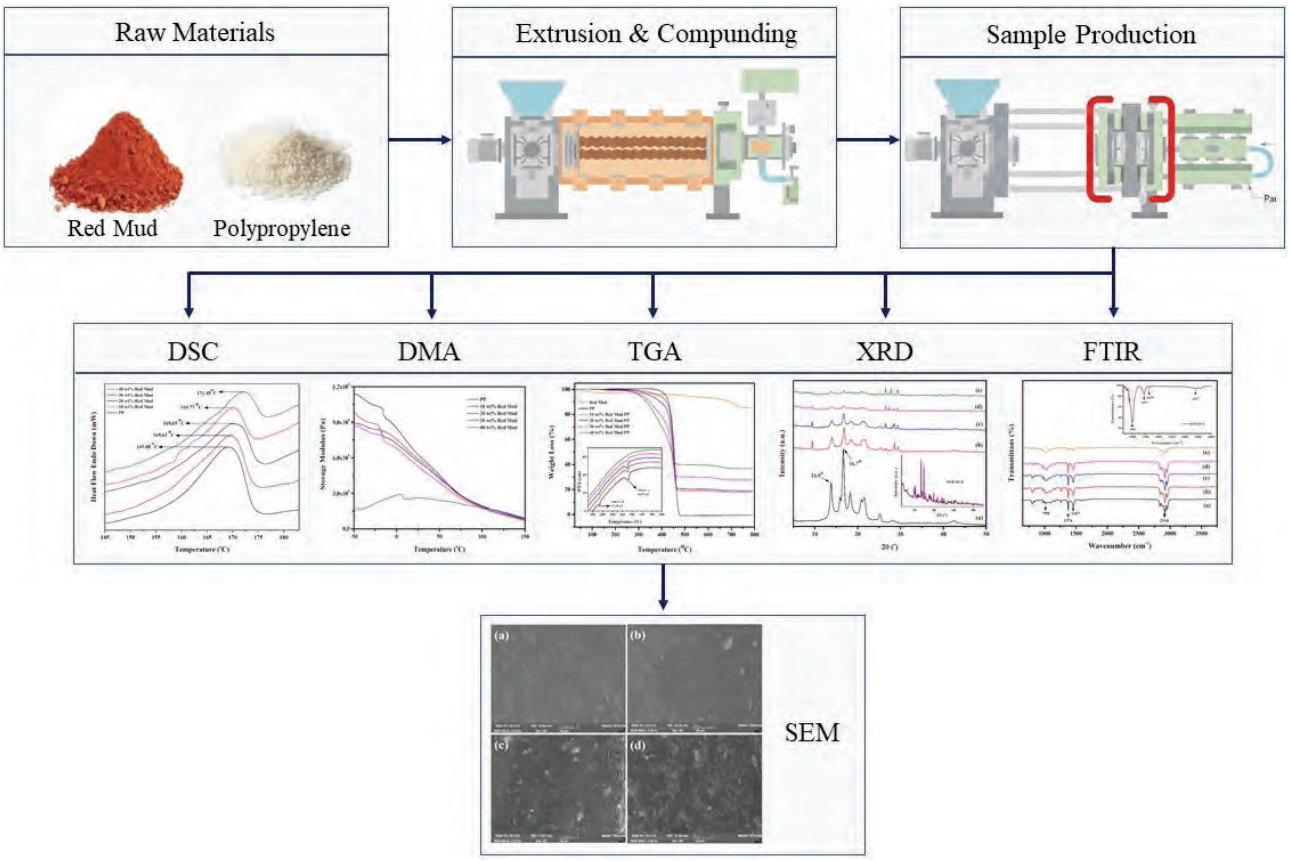


Figure 1. Graphical presentation of the study.

CHARACTERIZATION METHODS

Fourier Transform Infrared Analysis

The FTIR Spectrum 100 Perkin Elmer was utilized to analyze the samples in transmittance mode across a wavelength range of 600–4000 cm^{-1}.

Surface Morphology

The examination was conducted using an EVO LS Zeiss operating at 3 kV.

X-Ray Diffraction Examination

The measurements were carried out using a Panalytical X'PERT PRO X-ray diffractometer equipped with monochromatic CuKα radiation, functioning at 45 kV and 40 mA under ambient conditions.

Dynamic Mechanical Analysis

The DMA 8000 Perkin Elmer was employed for the analysis. The deformation mode used was single cantilever bending with a 50 µm amplitude, a frequency of 1 Hz, a heating rate of 5°C/min, and a temperature range of -50°C to 150°C.

Thermogravimetric Analysis

The EXSTAR TG/DTA 6300 Seiko device was employed for the assessments. Samples weighing 5-10 mg were placed in a platinum crucible under a nitrogen environment and heated from ambient temperature to 800°C at a rate of 10°C/min.

Differential Scanning Calorimetry Examination

The DSC evaluations of the samples (8±1 mg, n=5) were performed using a Perkin-Elmer Diamond DSC 8230 in a nitrogen atmosphere at a flow rate of 50 mL/min. The gathered data was then processed using Pyris Software. The specimens were heated at 10 °C/min over the 20–200 °C temperature range.

RESULTS AND DISCUSSION

Fourier Transform Infrared Analysis

FTIR spectroscopy uses infrared light to identify chemical bonds in molecules for organic and inorganic materials [20,21]. Thus, it enables the content of the tested sample to be determined. FTIR spectra of PP/RM composites blended in varying proportions (0-40 wt% RM) were shown in Figure 2.

The main transmittance bands observed in the FTIR spectrum in this study were consistent with those reported in the literature [22]. 2918 cm^{-1} (CH_2), 2951 cm^{-1} (CH_3), 2839 cm^{-1} (CH_2), 1457 cm^{-1} (CH_2), 1376 cm^{-1} (CH_3), 1359 cm^{-1} (CH_2- CH), 1330 cm^{-1} (CH_2- CH), 998 cm^{-1} (CH_3), 973 cm^{-1} (CH_3, C–C), 841cm^{-1} (C–H), 808 cm^{-1} (C–C), 1304 cm^{-1}, 1223 cm^{-1}, 1261 cm^{-1}, 1168 cm^{-1}, and 1154 cm^{-1} (CH_3), (CH_2), (CH), respectively. The intensity of the characteristic PP bands decreased as RM was mixed with PP in increasing proportions. In the FTIR spectrum of RM, the absorption bands were seen at 3437, 2920, 2853, 1745, 1639, 1578, 1560, 1542, 1452, 1402, 995, 802 and 654 cm^{-1}

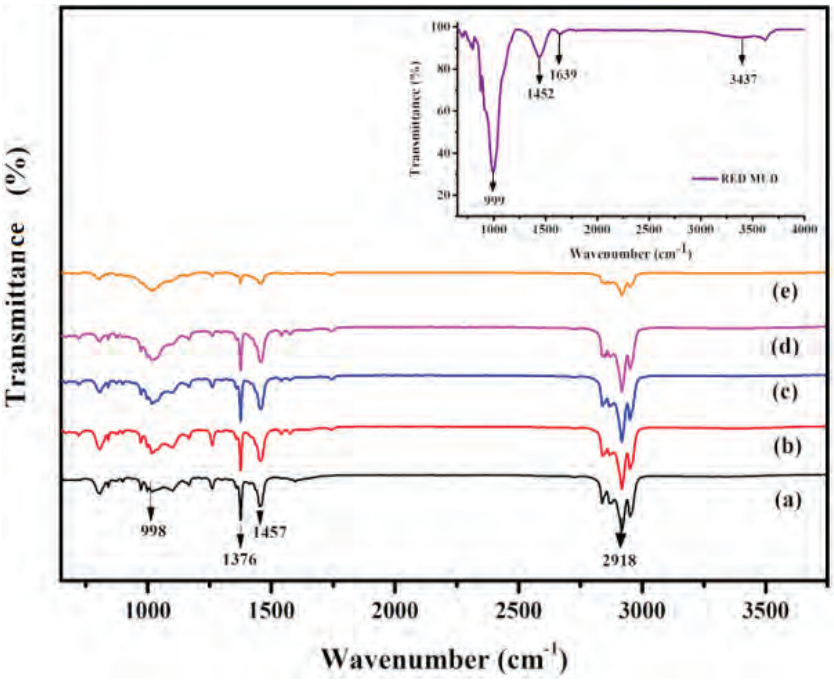


Figure 2. The FTIR spectra of (a) pure PP, (b) 10 wt%, (c) 20 wt%, (d) 30 wt%, (e) 40 wt% red mud powder mixed PP composites.

[23]. The band at 3437 cm^{-1} corresponds to an envelope of hydrogen-bonded surface OH groups. The bands at 2920 and 2853 cm^{-1} were assigned to the CH_2 symmetric or asymmetric stretch of the organic hydrocarbon compound in the red mud. The absorption band at 1745 cm^{-1} corresponds to the C=O double bond. The absorption band at 1639 cm^{-1} is attributed to the bending vibration of H-O-H from water in the structure of the aluminosilicate material. It is estimated that the band at 1452 cm^{-1} may be due to the carbonate component originating from calcite. The band at 999 cm^{-1} can be attributed to the stretching of the Si–O–Si bond [24-26]. FTIR measurements confirmed that RM was successfully incorporated into the PP matrix. The interaction between PP and RM is mainly physical.

Surface Morphology

Figure 3 shows the surfaces of PP/RM composites with different weight ratios (0, 10, 30, and 40 wt%) of RM. The images reveal that RM particle clusters are mixed into the PP, and as the amount of RM in the composite increases, there is an increase in the agglomeration of the powder (Figure 3b-d). The rise in RM concentration leads to the formation of agglomerates, thereby reducing the storage modulus. This result was confirmed by the SEM images, which supported the DMA result.

X-Ray Diffraction Examination

The XRD pattern of the RM, pure and 10 wt.%,20 wt.%, 30wt.% and 40 wt% red mud mixed PP matrix is given in Figure 4. It can be seen from the XRD pattern of RM, its composition is composed of many different components in accordance with the literature. From Figure 4, the main components of RM are Quartz (SiO_2), Dicalcium silicate (Ca_2SiO_4), Hematite (Fe_2O_3), Calcite ($CaCO_3$), Rulite (TiO_2), Sodium aluminium silicate ($Na(AlSiO_4)$), Gibbsite ($Al(OH)_3$) and Goethite (FeO(OH)). Phase identification was performed by qualitative peak matching using the ICDD database within X'Pert HighScore software, and Rietveld refinement was not applied, as the XRD analysis was intended to provide qualitative phase identification rather than quantitative phase analysis for the multiphase red mud–filled polypropylene composites. [27]. In the XRD

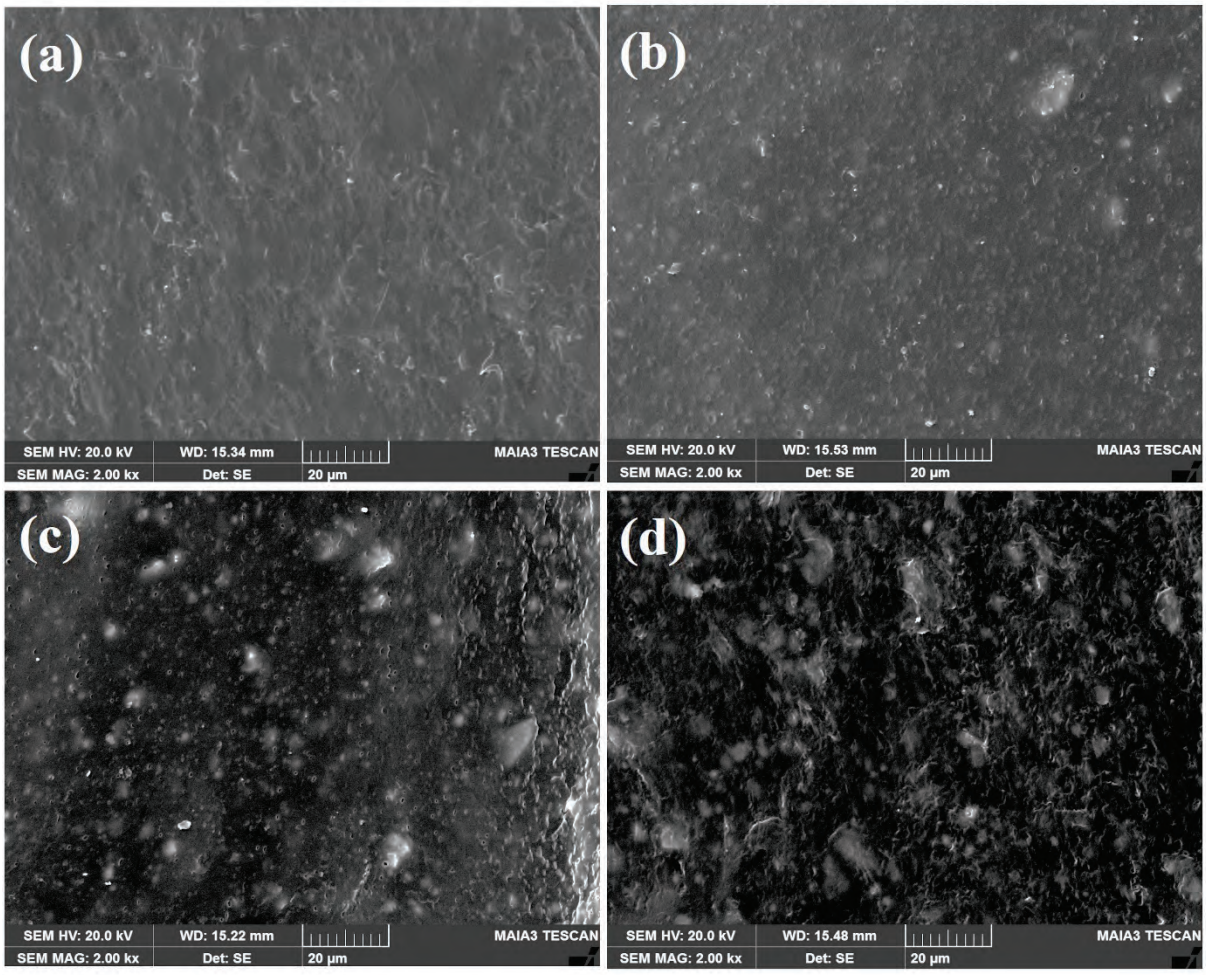


Figure 3. SEM images of (a) pure PP, (b) 10 wt%, (c) 30 wt%, (d) 40 wt% red mud powder mixed PP composites.

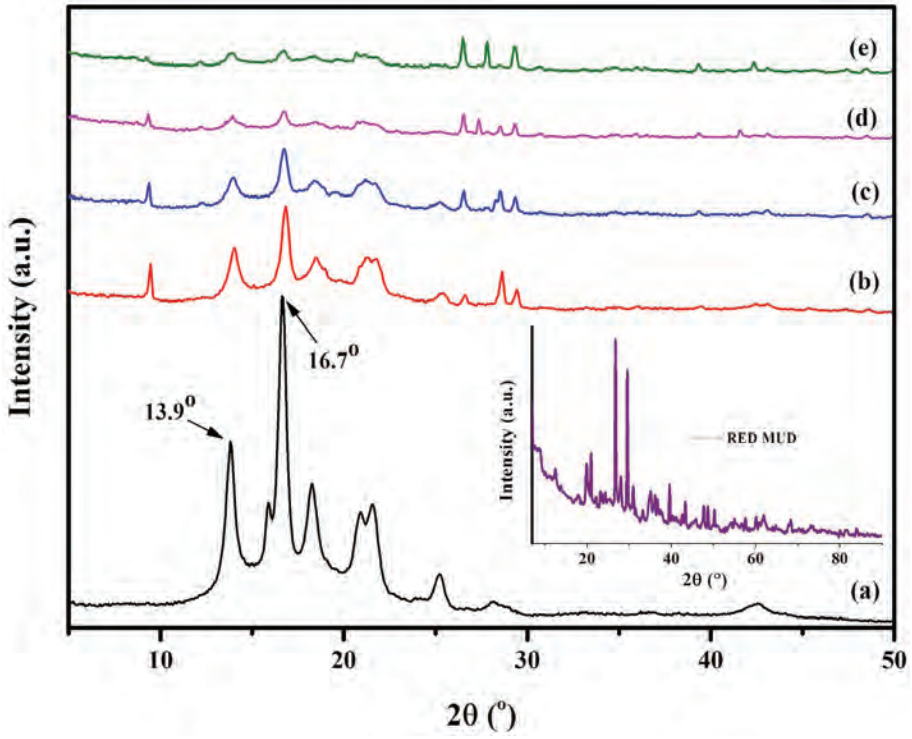


Figure 4. XRD patterns of (a) pure PP, (b) 10 wt%, (c) 20 wt%, (d) 30 wt%, (e) 40 wt% red mud powder mixed PP composites.

pattern of the PP matrix given in Figure 4a, two different crystal planes given in the literature were observed in 2θ of 13.9° and 16.7° [28].

Upon analysis of the pattern, it was observed that the addition of red mud (RM) led to an increase in the crystallization degree of polypropylene (PP). The intensity of the PP peaks decreased as the ratio of RM mixed with PP increased. New peaks appeared in the spectrum due to the mineral substances in the red mud. It is known that the crystalline nature of the thermoplastic polymer affects its thermal properties, and thus, modifying the crystallization process can improve its thermal properties [29-31]. After XRD measurement, the percentage of crystallinity of red mud-reinforced polypropylene composites was determined using the following equation [32]:

$$Percent\ Crystallinity = \frac{I_{crystalline}}{I_{crystalline} + I_{amorphous}} \times 100 \quad (1)$$

From the results obtained using Equation 1, the percentage of crystallinity increased with increasing red mud content. The percent crystallinity were 90,29% for RM, 80,04% for PP, 80,18% for 10.0 wt% PP/RM composite, 83,75% for 20.0 wt% PP/RM composite, 87,34% for 30.0 wt% PP/RM composite and 94,94% for 40.0 wt% PP/RM composite (Fig. 5).

The material's crystallinity increased proportionally with the increasing concentration of highly crystalline RM in the matrix. This relationship is reinforced by the detection of an increasing number of new peaks in the XRD spectrum as the RM ratio increases, indicating a higher degree of crystallinity.

Dynamic Mechanical Analysis

Figure 6 compares storage modulus against temperature for pure polypropylene (PP), and PP compounded with 10 wt.%, 20 wt.%, 30 wt.%, and 40 wt.% red mud (RM). The storage modulus (E') of the RM-doped composites was lower than that of pure PP at lower temperatures. At 25°C, E' of pure PP was measured at 6.28 x 10^8 Pa. However, the composite with 40 wt.% RM showed a storage modulus of only 2.60 x 10^8 Pa at the same temperature, representing a 58% decrease. This reduction is likely due to the increased agglomeration of red mud particles within the polypropylene matrix. As the filler loading increases, particle-particle interactions become more dominant than particle-matrix interactions [33,34]. The agglomeration reduces the efficiency of stress transfer between the matrix and the filler. Also, the lack of a compatibiliser or surface treatment weakens interfacial adhesion. This heterogeneous dispersion reduces stiffness.

Thermogravimetric Analysis

PP is a polymer with low thermal stability. This feature limits its application areas. PP is mixed with various additives to improve its thermal properties. This study aimed to

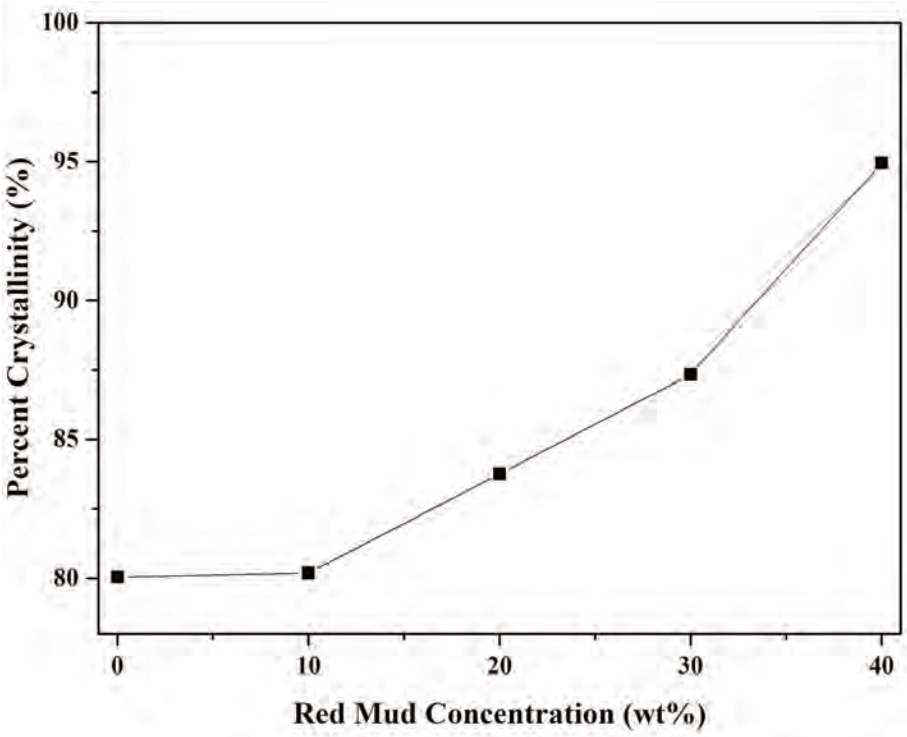


Figure 5. Percent crystallinity depending on red mud content (0, 10, 20, 10, and 30 wt%).

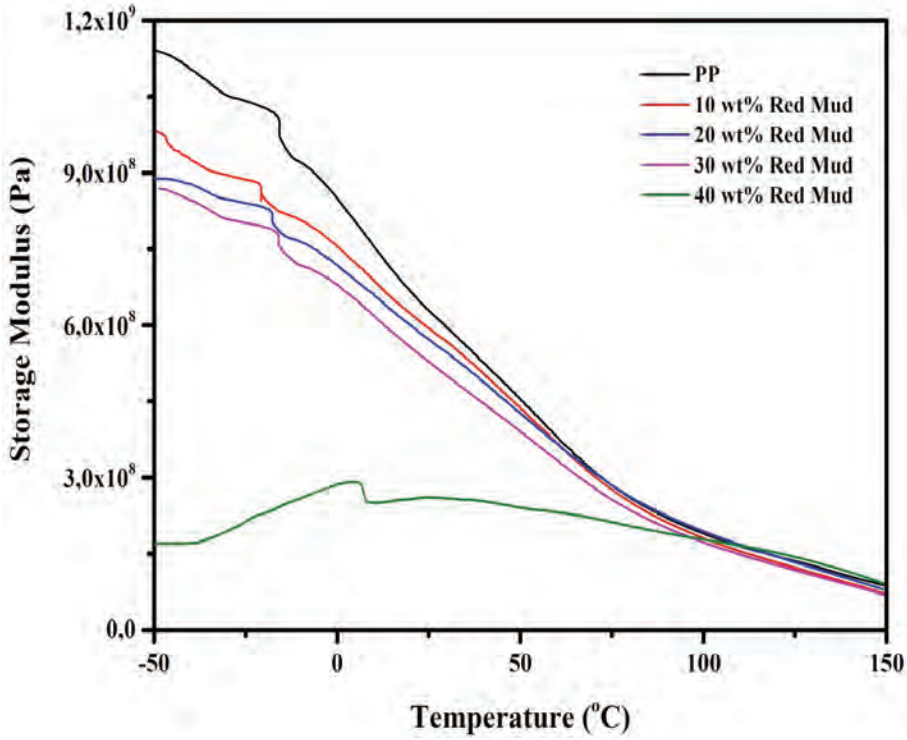


Figure 6. Storage modulus of PP/RM composites containing varying amounts (0-40 wt%) of red mud powder.

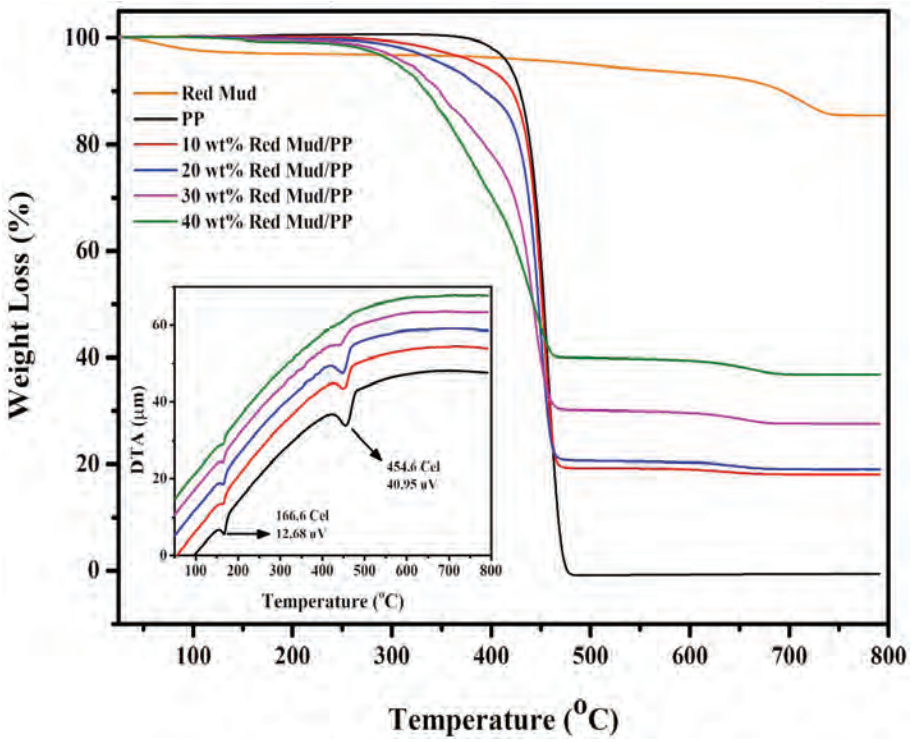


Figure 7. TGA thermograms of RM and PP/RM polymer composites with different percentages (0-40 wt%) of red mud powder.

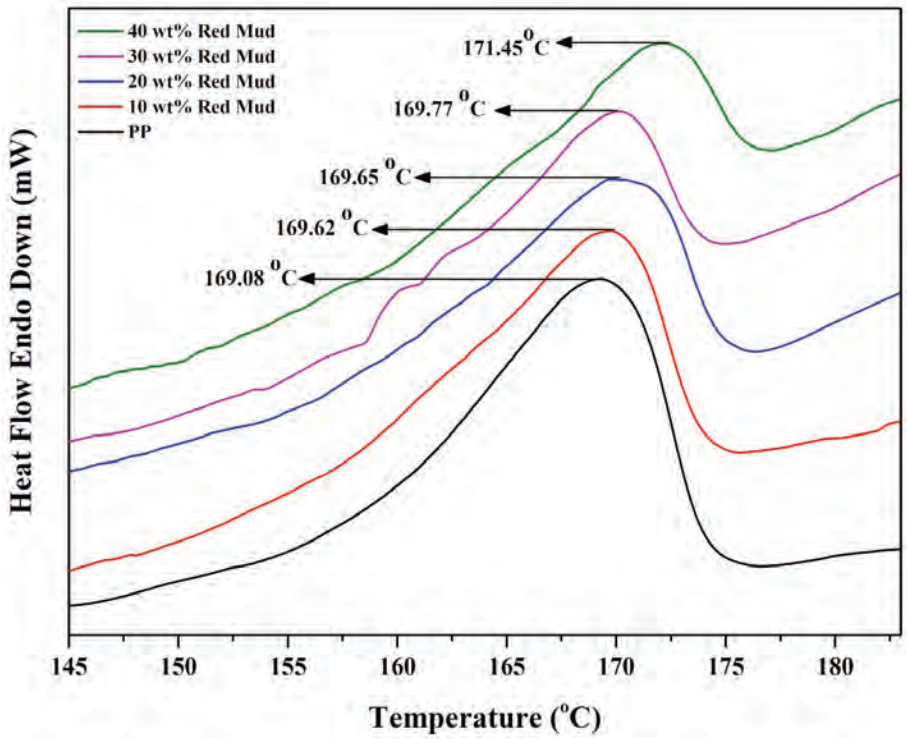


Figure 8. DSC measurement results of PP/RM polymer composites with different percentages (0-40 wt%) of red mud powder.

enhance PP's thermal stability by incorporating red mud, a low-cost waste material. Thermal behaviour of the RM, pure and 10 wt.%,20 wt.%, 30wt.% and 40 wt% red mud mixed PP matrix is shown in Figure 7. For RM, the 2.98% mass loss observed up to 200 °C is attributable to the loss of physically adsorbed water. The 3.62% mass reduction within the temperature interval of 200-600 °C is a result of the dehydroxylation of the FeO(OH) and AlO(OH) phases in the RM. The 7.98% mass reduction observed at temperatures up to 800 °C can be attributed to the degradation of the sodalite group in RM [35]. According to the TGA curve of PP, it's clear that all the mass loss occurs suddenly within a specific temperature range, between 374 °C and 476 °C. Incorporating red mud into PP resulted in mass degradation in two steps. Between 25 °C and 800 °C, the mass degradation of the composite decreased with increasing red mud ratio. With increasing temperatures spanning 200-500 °C, weight losses were 80.8% for 90wt/10wt composite, 79.32% for 80wt/20wt composite, 69.73% for 70wt/30 wt% composite and 59.96% for 60wt/40wt composite. Within 500-700 °C, weight losses were 1.14% for 90wt/10wt composite, 1.6% for 80wt/20wt composite, 2.68% for 70wt/30wt composite and 3.22% for 60wt/40wt composite. When measured between room temperature and 800 °C, the 60 wt/40 wt composite exhibited excellent thermal stability, as it showed the lowest mass loss. The TGA results unambiguously indicate that incorporating red mud significantly enhances the thermal stability of polypropylene [36,37].

Differential Scanning Calorimetry Examination

Figure 8 shows the DSC melting curves for pure PP and PP/RM composites with different mixing ratios. The melting temperatures (T_m) of all samples (pure PP, 10 wt.%, 20 wt.%, 30wt.%, and 40 wt% red mud mixed PP matrix) were 169.08, 169.62, 169.65, 169.77 and 171.45 °C, respectively. The Tm of the PP matrix shifted to higher temperatures after mixing with RM, indicating that PP and RM mix well and interact.

CONCLUSION

This study lies in the systematic evaluation of untreated red mud, an industrial waste by-product, as a filler in polypropylene composites without the use of surface modification or compatibilizers. Unlike most previous studies that focus on chemically treated red mud or dispersion-optimized systems, this work provides a direct assessment of the intrinsic effect of raw red mud on the morphological, mechanical, and thermomechanical properties of PP composites.The different crystal structures and mechanical properties of RM influenced the PP-based composites' thermal and dynamic mechanical properties. SEM images revealed agglomeration with increased RM content. At 25 °C, the storage modulus of pure polypropylene is 6.28×10^8 Pa, whereas it decreases to 2.60×10^8 Pa for the 60 wt/40 wt composite. There is a reduction of about 58%. DMA results suggested that the decreased storage modulus of the composites was due to this agglomeration. FTIR analysis confirmed that the composites comprised PP and red mud. The intensity of the PP peaks decreased with increasing RM ratios, whereas new peaks appeared in the spectrum due to red mud. The percentage of crystallinity increased with RM mixed into the matrix, as evidenced by new peaks in the XRD pattern. According to X-ray tests, crystallinity increases from 80.04% for pure polypropylene to 94.94% for the 60 wt/40 wt composite. The melting temperature rises slightly from 169.08 °C to 171.45 °C as red mud content increases. The shift of Tm to higher temperatures in the DSC result indicated a good mixing of RM with PP. Thermogravimetric analysis indicates the composites are more thermally stable, with less mass loss. The 60 wt/40 wt composite exhibits the lowest mass loss up to 800 °C. TGA results demonstrated that low-cost red mud improved the thermal properties of PP, although it exhibited low thermal stability. These results suggest that untreated red mud can be an effective, low-cost filler for polypropylene composites for engineering applications.

From an economic standpoint, utilizing waste materials can help to create new industries and business opportunities while also reducing costs associated with waste disposal. Overall, the utilization of waste materials has the potential to be a win-win for both the environment and the economy. These results suggest that untreated red mud can be an effective, low-cost filler for polypropylene composites for engineering applications.

ACKNOWLEDGMENTS

Thanks to Yıldız Technical University, Science and Technology Application and Research Center, Bartın University Central Research Laboratory and Application Center for enabling the use of the lab equipment and facilities. The authors are also grateful to Marmara University for providing laboratory facilities.

AUTHOR CONTRIBUTIONS

All authors contributed equally to both the research implementation and the manuscript preparation. All authors have accepted responsibility for the entire content of this submitted manuscript and have approved the submission.

DATA AVAILABILITY STATEMENT

All data are included in this manuscript.

CONFLICT OF INTEREST

There is no any conflict of interest.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Swain B, Akcil A, Lee JC. Red mud valorization an industrial waste circular economy challenge; review over processes and their chemistry. Crit Rev Environ Sci Technol 2022;52:520–570. [CrossRef]

[2] Akcil A, Akhmadiyeva N, Abdulvaliyev R, Meshram A, Meshram P. Overview on extraction and separation of rare earth elements from red mud: Focus on scandium. Miner Process Extr Metall Rev 2018;39:145–151. [CrossRef]

[3] Kucukdogan N, Aydın L, Sutcu M. Theoretical and empirical thermal conductivity models of red mud filled polymer composites. Thermochim Acta 2018;665:76-84. [CrossRef]

[4] Yılmaz VM. Altın ve bakır sektöründe atık yönetimi. In: Çevresel Açıdan Sektörel Atık Yönetimi ve Uygulamaları III. Ankara: Nobel Akademik Yayıncılık; 2024. p. 121–144.

[5] Yan P, Chen B, Haque MA, Liu T. Influence of red mud on the engineering and microstructural properties of sustainable ultra-high performance concrete. Constr Build Mater 2023;396:132404. [CrossRef]

[6] Wu P, Liu X, Zhang Z, Wei C. Properties of red mud-filled and modified resin composites. Constr Build Mater 2023;409:133984. [CrossRef]

[7] Liu Z, Yang Y, Xie M, Cheng M, Yang R, Huang Z, et al. TG-FTIR-Py-GCMS analysis and catalytic pyrolysis mechanism of textile waste by red mud catalyst for liquid fuel production. Sci Total Environ 2024;952:175874. [CrossRef]

[8] Uysal M, Kuranlı ÖF, Aygörmez Y, Canpolat O, Coşgun T. The effect of various fibers on the red mud additive sustainable geopolymer composites. Constr Build Mater 2023;363:129864. [CrossRef]

[9] Yin H, Liu J, Zhou X, Qi H, Liu S, Pang S. Flexural properties of fiber-reinforced alkali slag-red mud geopolymer. Constr Build Mater 2023;370:130708. [CrossRef]

[10] Mian MAA, Taotao S, Min D, Wentao X, Yuchen Y, Imran A, et al. Technologies for recovery of iron from red mud: Processes, challenges, and opportunities. Sustain Mater Technol 2024;41:e01053. [CrossRef]

[11] Jiahai B, Li C, Du Q, Dong C. Fabrication and properties of self-foamed glass ceramics from red mud and ceramic tile polishing waste. J Sustain Metall 2024;10:1559–1571. [CrossRef]

[12] Niu A, Lin C. Trends in research on characterization, treatment and valorization of hazardous red mud: A systematic review. J Environ Manage 2024;351:119660. [CrossRef]

[13] Bhat AH, Abdul HPS, KA. Thermoplastic polymer based modified red mud composites materials. In: Attef B, (editor). Advances in Composite Materials - Ecodesign and Analysis. London: IntechOpen; 2011. [CrossRef]

[14] Zhang Y, Zhang A, Zhen Z, Lv F, Chu PK, Ji J. Red mud/polypropylene composite with mechanical and thermal properties. J Compos Mater 2011;45:2811–2816. [CrossRef]

[15] Akıncı A, Akbulut H, Yılmaz F. Mechanical properties of cost-effective polypropylene composites filled with red-mud particles. Polym Polym Compos 2008;16:439–446. [CrossRef]

[16] Ravi Kumar M, Kumar V, Prasad CD, Sridevi G, CR A, Kumar A, et al. Estimating and parametrically improving the microstructure, hardness, and wear resistance of SiC-CeO2 reinforcements on hot rolled Al7075 hybrid composites. Results Eng 2025;26:104634. [CrossRef]

[17] Ravikumar M, Naik R, Vinod BR, Chethana KY, Rammohan YS. Study on nanosized Al2O3 and Al2O3-SiC on mechanical, wear and fracture surface of Al7075 composites for soil anchoring applications. Mater Phys Mech 2023;51:24–41.

[18] Ravikumar M, Chethana KY, Vinod BR, Suresh R. Nano-sized SiC-Gr particulates reinforced AI7075 hybrid composite: Experimental studies and analysis of quenching agents. J Tribol 2023;38:82–99.

[19] Ravi Kumar M, Suresh R. Study on mechanical and machinability characteristics of n-AI2O3/SiC-reinforced AI7075 composite by design of experiment technique. Multiscale Multidiscip Model Exp Des 2023;6:747–760. [CrossRef]

[20] Gümüş BE, Yağci Ö, Taşdemir M. High-density polyethylene/artichoke leaf powder polymer composites: Dynamic mechanical, morphological and thermal properties. Iran Polym J 2022;1:1–11. [CrossRef]

[21] Benli M, Gümüş BE, Kahraman Y, Yağcı Ö, Erdoğan D, Huck O, et al. Thermal, structural and morphological characterization of dental polymers for clinical applications. J Prosthodont Res 2021;65:176–185. [CrossRef]

[22] Gumus BE, Yagci O, Erdogan DC, Tademir M. Dynamical mechanical properties of polypropylene composites filled with olive pit particles. J Test Eval 2019;47:2551–2561. [CrossRef]

[23] Singh S, Aswath MU, Das Biswas R, Ranganath RV, Choudhary HK, Kumar R, et al. Role of iron in the enhanced reactivity of pulverized red mud: Analysis by Mössbauer spectroscopy and FTIR spectroscopy. Case Stud Constr Mater 2019;11:e00266. doi:10.1016/j.cscm.2019.e00266. [CrossRef]

[24] Gotić M, Musić S. Mössbauer FT-IR and FE SEM investigation of iron oxides precipitated from FeSO4 solutions. J Mol Struct 2007;834–836:445–453. [CrossRef]

[25] Christou C, Agapiou A, Kokkinofta R. Use of FTIR spectroscopy and chemometrics for the classification of carobs origin. J Adv Res 2018;10:1–8. [CrossRef]

[26] Al Bakri Abdullah MM, Hussin K, Bnhussain M, Ismail KN, Yahya Z, Razak RA. Fly ash-based geopolymer lightweight concrete using foaming agent. Int J Mol Sci 2012;13:7186–7198. [CrossRef]

[27] Nath H, Sahoo A. A study on the characterization of red mud. Int J Appl Bio-Engineering 2014;8:1–4. [CrossRef]

[28] Alonso M, Velasco JI, De Saja JA. Constrained crystallization and activity of filler in surface modified talc polypropylene composites. Eur Polym J 1997;33:255–262. [CrossRef]

[29] Labour T, Gauthier C, Séguéla R, Vigier G, Bomal Y, Orange G. Influence of the β crystalline phase on the mechanical properties of unfilled and CaCO3-filled polypropylene. I. Structural and mechanical characterisation. Polymer 2001;42:7127–7135. [CrossRef]

[30] Velasco JI, Morhain C, Martinez AB, Rodriguez-Peréz MA, De Saja JA. The effect of filler type, morphology and coating on the anisotropy and microstructure heterogeneity of injection-moulded discs of polypropylene filled with aluminium and magnesium hydroxides. Part 1. A wide-angle X-ray diffraction study. Polymer 2002;43:6805–6811. [CrossRef]

[31] Akinci A, Akbulut H, Yilmaz F. The effect of the red mud on polymer crystallization and the interaction between the polymer-filler. Polym Plast Technol Eng 2007;46:31–36. [CrossRef]

[32] Yağci Ö, Eker Gümüş B, Taşdemir M. Thermal, structural and dynamical mechanical properties of hollow glass sphere-reinforced polypropylene composites. Polym Bull 2021;78:3089–3101. [CrossRef]

[33] Mothé CG, Monteiro DFJ, Mothé MG. Dynamic mechanical and thermal behavior analysis of composites based on polypropylene recycled with vegetal leaves. Mater Sci Appl 2016;7:349–357. [CrossRef]

[34] Zhang J, Sun C, Li P, Ming L, Jiang H, Yao Z. Experimental study on rheological properties and moisture susceptibility of asphalt mastic containing red mud waste as a filler substitute. Constr Build Mater 2019;211:159–166. [CrossRef]

[35] Kazak O, Eker YR, Akin I, Bingöl H, Tor A. Green preparation of a novel red mud@carbon composite and its application for adsorption of 2,4-dichlorophenoxyacetic acid from aqueous solution. Environ Sci Pollut Res 2017;24:23057–23068. [CrossRef]

[36] Reis JML. Fracture and flexural assessment of red mud in epoxy polymer mortars. Mater Struct 2015;48:3929–3936. [CrossRef]

[37] Kazak O, Tor A, Akin I, Arslan G. Preparation and characterization of novel polysulfone-red mud composite capsules for the removal of fluoride from aqueous solutions. RSC Adv 2016;6:86673–86681. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.1958

Research Article

Assessment of future renewable energy use and CO_2 emissions of Türkiye and other countries in terms of climate targets of international agreements: An estimate with anfis

Abdullah Zübeyr ŞEKERCİ[1,*], Berk AYVAZ[2]

[1]*Department of Industrial Engineering, Istanbul Technical University, Istanbul, 34469, Türkiye*
2Department of Industrial Engineering, Istanbul Ticaret University, Istanbul, 34134, Türkiye

ARTICLE INFO

Article history
Received: 24 July 2025
Revised: 18 October 2025
Accepted: 17 November 2025

Keywords:
ANFIS; Climate Protokols; Emission; Forecasting; RES; Türkiye

ABSTRACT

Global climate change is largely driven by increasing fossil fuel use and related emissions. The largest contributors include the United States, China, the European Union, and India, which together account for a significant share of global emissions. International agreements such as the Kyoto Protocol (2009) and the Paris Climate Agreement (2016) aim to reduce emissions and promote renewable energy use; however, they impose similar targets on countries with different energy structures and development levels. this study examines changes in renewable energy shares and co_2 emissions (ce) for selected countries and regions. using three input variables and one output variable commonly adopted in the literature, projections for the 2025–2030 period are obtained with the adaptive network-based fuzzy inference system (ANFIS). The results indicate that Türkiye increases its renewable electricity share from 35% in 2021 to 50% in 2030, while reducing emissions from 12% to 6%. Similar but less pronounced trends are observed in the European Union and the United States. In contrast, India shows an increase in emissions with limited improvement in renewable energy use, whereas China exhibits moderate growth in both indicators. Overall, the findings suggest that while some regions align with international targets, others remain misaligned. Accordingly, defining country-specific targets based on emission levels and renewable energy capacity may lead to more effective and realistic policy outcomes.

Cite this article as: Şekerci AZ, Ayvaz B. Assessment of future renewable energy use and CO_2 emissions of Türkiye and other countries in terms of climate targets of international agreements: An estimate with anfis. Sigma J Eng Nat Sci 2026;44(2):1138–1152.

INTRODUCTION

With rapid industrial and technological growth in the early 21st century, global emissions and temperatures have increased significantly. Fossil sources (FS) remain the primary source of these emissions. For this reason, renewable energy sources are increasingly promoted as alternatives due to their lower environmental impact.

***Corresponding author.**
*E-mail address: abdullahzsekerci@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Figure 1 presents the distribution of electricity generation by source across different countries and regions [1]. When nuclear energy is included, the global average renewable share in electricity is approximately 37%, while the direct renewable share is around 31%. In comparison, Türkiye exceeds both averages, with a direct renewable share of about 42%. Over the past decade, this level has been supported by significant investments in hydroelectric (HPP), solar (SPP), and wind (WPP) power plants, which have contributed to the increase in renewable electricity generation [2].

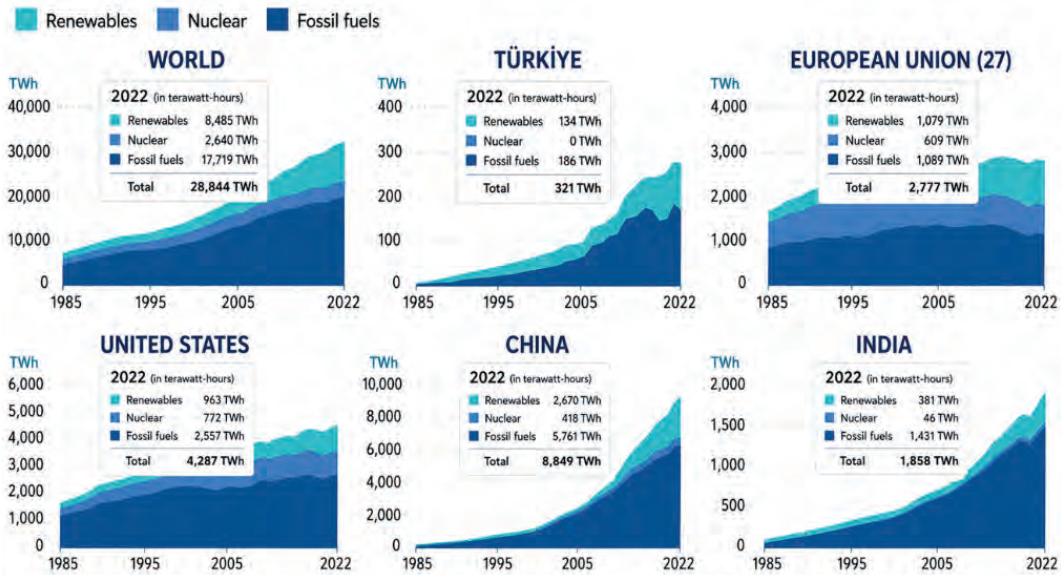
The European Union also remains above the global average, with a combined renewable and nuclear share of approximately 61% and a direct renewable share of 39%. This reflects long-term energy policies that have reduced fossil fuel use while maintaining nuclear energy at a relatively stable level [3]. In the United States, the combined share is about 40%, which is above the global average; however, the direct renewable share remains lower, at around 22%. Although fossil fuel use has slightly declined over the past 15 years, nuclear energy levels have remained stable, and domestic fossil resources such as natural gas and oil continue to limit the expansion of renewable energy. As a result, the expected reduction in CE has not been fully achieved [4].

China also remains below the global average, with a combined share of approximately 35% and a direct renewable share of around 30%. Despite considerable progress in renewable energy development, the strong dependence on coal continues to restrict this transition, leading to increasing emission levels [5]. A similar situation is observed in India, which presents the weakest performance among the countries considered. India has a combined share of approximately 23% and a direct renewable share of about 21%, both below global averages. Although the country aims to generate 30% of its electricity from solar energy by 2040 under the "Hybrid Energy" policy, fossil fuel dependence continues to constrain the growth of renewable energy, similar to the situation in China and the United States [6,7].

International organizations such as the International Energy Agency (IEA) and global agreements, including the Kyoto Protocol (2009) [8] and the Paris Climate Agreement [9], call for increasing renewable energy use and reducing emissions to mitigate temperature rise. However, these frameworks impose similar obligations on countries despite significant differences in fossil fuel use and renewable energy growth rates [10]. Therefore, aligning targets with country-specific energy capacities may lead to more realistic outcomes.

As renewable energy replaces fossil fuels in electricity generation, CO_2 emission levels are expected to decline. To examine this relationship, emission trends for different countries and regions are presented in Figure 2 [7].

In this study, the effects of the above-mentioned protocols on countries and regions and the deficiencies of the protocol texts according to this effect are discussed. For this review, the RES usage rate in electricity generation and the CE increase status of countries and regions are estimated. First of all, inputs and outputs suitable for this study is selected from the literature. Then, data on these parameters are collected for Türkiye, the EU Region, the USA, China and India between 1971-2022. Since renewable electricity production (REP) vary across the specified countries and regions, comparisons are made using REP shares (%) rather than absolute REP amounts (GWh/



*TWh: Terawatt-hours

Figure 1. Electricity generation capacities by sources / countries and regions.

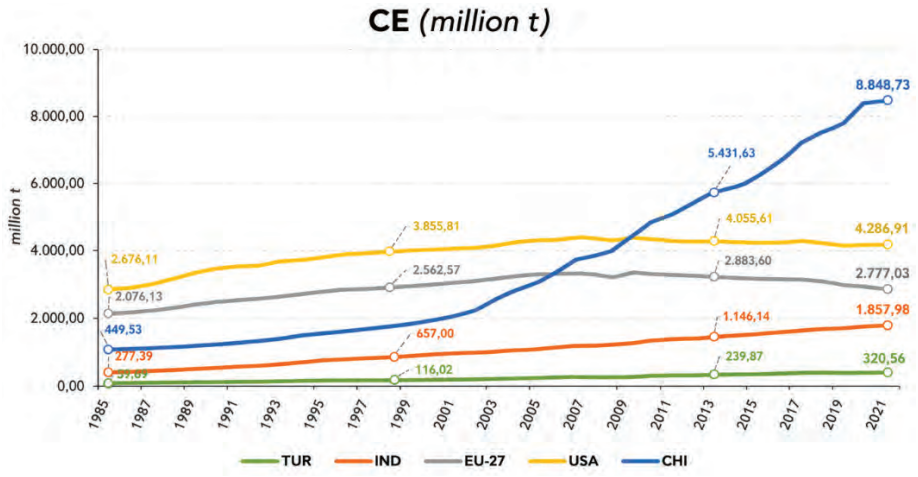


Figure 2. CE by country and region (Mt) / (1971-2022).

TWh). To compute these shares, the total electricity production (TEP) of each country/region must be determined. For forecasting future REP amounts (GWh/TWh), the input variables are Gross Domestic Product (GDP), Population (POP), and Carbon Emissions (CE), and the output variable is the REP amount. Using these three inputs and one output, the values for 2025–2030 are estimated with Adaptive Network-Based Fuzzy Inference System (ANFIS) via MATLAB's Fuzzy Logic Toolbox. For TEP, a simpler ANFIS-based autoregressive approach is used: instead of three inputs, past TEP values, TEP(t), are employed to predict TEP(t+1) using the same fuzzy inference framework implemented in MATLAB's Fuzzy Logic Toolbox. Similar to the REP process, future CE levels were also estimated with ANFIS. The inputs here are GDP, POP, and CE (t). Data between 1971-2022 were processed in MATLAB. CE (t+1) was considered as the output. Similar to the first process, the emissions for 2025-30 were estimated with ANFIS. The important thing in the estimation process is the rate at which the inputs affect the output separately. This effect is understood in the rule base produced at the end of the estimation. If the input data conflict with the output data, then the rule base structure naturally exhibits irregular geometric patterns. On the other hand, a flatter ground will be obtained if there is compatibility. Many studies have been examined while selecting the inputs considered from the literature. The aim is to examine the compliance of countries and regions with the conditions specified in [8,9] and the main target of 1.5°C surface temperature in [9]. As a result, it turns out that some of the countries and regions examined cannot achieve the desired CE reduction or REP increase rate for this target. Energy usage rates vary greatly from country to country. Therefore, it is important that these protocols specify conditions proportional to the usage capacities of the country.

The following section includes the literature review necessary to examine forecasting methods and determine inputs and outputs.

LITERATUR REVIEW

In this section, studies on predicting the energy source level, CE level, and electricity consumption level that will be replaced in the future are examined. Similar estimation methods and input-output parameters are used in these subjects, guiding this study. In addition, one of the main elements sought in the studies examined is that the method is ANFIS or Artificial Neural Network (ANN). Rodrigues et al. [11], in their study published in 2009, used 37 years between 1970 and 2006 as training and test data and made predictions for 2007 with ANFIS. In their studies, POP, GDP, and CE were used as inputs and CE estimates were made for Brazil. Appiah et al. [12] used ANN in their study published in 2018 and made a 2013 CE estimate for China. In the study, the 40 years between 1973 and 2013 for POP, GDP, IE, CE, and EC data were used, and 2014 was estimated. Şekerci and Kara [10], in their study published in 2024, POP, GDP, EC inputs, and CE output were used within ANFIS. Fifty-two years between 1971 and 2022 were used, and 2030 was estimated. In their study published in 2019, Güleryüz [13] made a 2020 industrial EC estimate for Türkiye using data collected between 1990-2019 and GDP, CE, EC and general energy expenditures in industry as inputs. Nassef et al. [14] used the CE data of Saudi Arabia between 1954-2020 as input in their study published in 2023. They estimated the CE data between 2020-30 with ANFIS, FFNN, and LSTM methods. As a result, they found that the CE level decreased by more than 30%. Arevalo et al. [15] in their study published in 2024, estimated the 2050 EC values using Ecuador's POP, TEM, and EC values from previous years. Then, with the EnergyPlan method, they made a plan that meets the electricity completely from RES in 2050. Table 1 and Table 2 show the literature review.

Table 1. Literatur review

Writer(s)	Years	Application*	Method**	Forecasting Year
Rodrigues vd. [11]	2009	Brazil CEE	ANFIS	2007
Özdemir [16]	2011	Türkiye CEE	ANN	2025
Yılmaz and Yılmaz [17]	2013	Türkiye CEE	GP	2020
Baareh [18]	2013	CEE	ANN	2011
Appiah vd. [12]	2018	China CEE	ANN	2014
Güleryüz [13]	2019	Türkiye EC-F	ANFIS & PSO & MLR	2020
Khan and Khan [19]	2019	Pakistan CEE	ANFIS & ANN	2013
Jusoh et al. [20]	2021	Indonesia EC-F	ANFIS	2028
Jena et al. [21]	2021	17 Countries CEE	ANN	2019
Lazım and Pauzi [22]	2021	CEE	ANFIS	2014
Nugraha et al. [23]	2021	Indonesia RES-F	ANFIS	2027
Şekerci [24]	2022	Türkiye CEE	ANFIS	2022
Mutascu [25]	2022	USA CEE	ANN	2022
Ginting et al. [26]	2023	Indonesia RES-F	ANFIS	2023
Nassef et al. [14]	2023	Saudi Arabia CEE	ANFIS & FFNN & LSTM	2030
Behgouy and Uğurenver [27]	2024	USA EC-F	LSTM & NAR	2023
Şekerci and Kara [10]	2024	Türkiye CEE	ANFIS	2030
Arevalo et al. [15]	2024	Ecuador EC-F & RES-F	ANFIS & EnergyPlan	2050

*CEE: CO_2 Emission Estimation, RES-F: RES Level Forecasting, EC-F: Electricity Consumption Forecasting.

**PSO: Partical Swarm Optimization, MLR: Multiple Linear Regression, FFNN: Feed-Forward Neural Network, LSTM: Long Short-Term Memory, GP: Grey Prediction, NAR: Nonlinear Autoregressive.

Table 2. Input and output use in literature

Writer(s)	Inputs*						Output**	
	POP	GDP	IE	TEM	CE	EC	CE	EC
Rodrigues vd. [11]	√	√			√		√	
Özdemir [16]		√	√			√	√	
Yılmaz and Yılmaz					√		√	
Baareh [18]					√	√	√	
Appiah vd. [12]	√	√	√			√	√	
Güleryüz [13]		√	√		√		√	
Khan and Khan [19]				√		√	√	
Jusoh et al. [20]	√	√						√
Jena et al. [21]	√	√	√				√	
Lazım and Pauzi [22]	√	√				√	√	
Nugraha et al. [23]	√	√				√		√
Şekerci [24]	√	√		√		√	√	
Mutascu [25]		√	√			√	√	
Ginting et al. [26]	√	√				√		√
Nassef et al. [14]					√		√	
Behgouy and Uğurenver [27]						√		√
Şekerci and Kara [10]	√	√				√	√	
Arevalo et al. [15]	√			√				√
TOPLAM	***10***	***12***	***5***	***3***	***5***	***11***	***13***	***5***

*POP: Population, GDP: Gross Domestic Product, IE: Import-export, TEM: Temperature, **EC: Energy Consumption

As presented from the literature review above, the most frequently used inputs are POP, and GDP. It can be stated that among these outputs, there is a strong positive correlation between CE, the mentioned output, and the two input variables. In countries with a high level of industrial capacity, as the POP and GDP indicators increase, development accelerates, which in turn promotes industrial growth and raises existing emission levels. In addition to this strong correlation, CE(t+1) is selected as the output variable to examine the emission obligations of the specified countries and regions. The third input variable used to predict future CE values is the historical emission data, CE(t).

In this study, the output variable REP is considered instead of the EC output specified in Table 2. The reason for this is to examine the clean energy transition obligations of the selected countries and regions. EC, on the other hand, refers to total electricity generation, including fossil-based sources. The input variables used for this output are POP, GDP, and CE.

Thus, both this inputs and outputs have gained a real basis from the literature. The next section describes the methodology of the study and its application.

METHODOLOGY

In this section, the methodology is examined in two parts. First, the study methodology is explained using a flow chart. Second, the application methodology is explained using mathematical equations.

Methodology of the Study

The study's methodological flow clearly demonstrates how the protocols operate; accordingly, Figure 3 illustrates this framework.

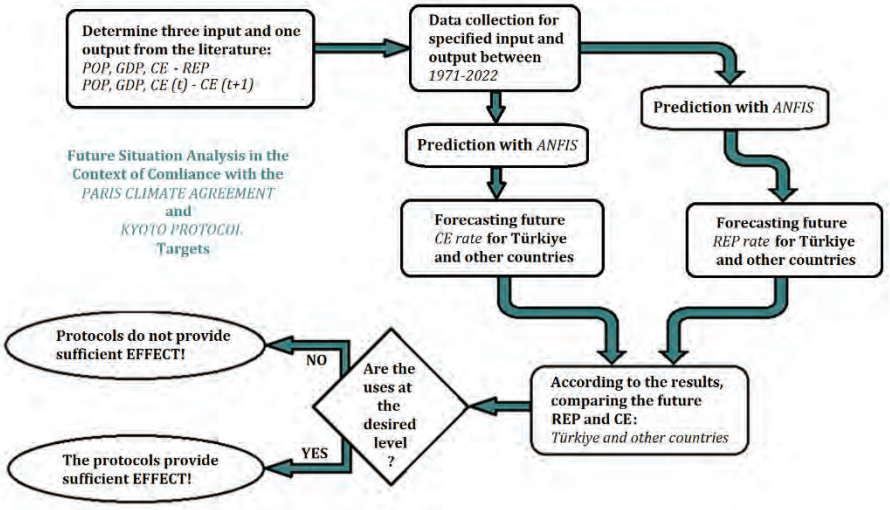


Figure 3. Flow chart of the study.

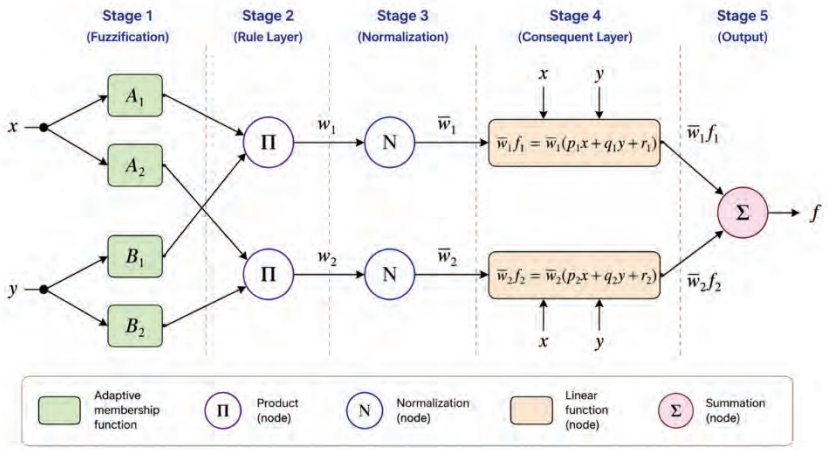


Figure 4. Fuzzy-ANN model of ANFIS method.

The methodological flow of the study is given above. Figure 3 shows how forecast values differ from current values and how countries compare. According to this comparison, the impact of the protocols is examined.

The application methodology used in the study is discussed in the next section.

Application Methodology

The application of this study is the ANFIS method. Therefore, the mathematical methodology of ANFIS is explained in this section [11, 13].

The mathematical model of ANFIS methodology is explained below. [23]:

Layer 1: Degrees of membership of specified inputs are explained [26]:

$$o_i^1 = \mu_{A_i}(x), i = 1, 2, \ldots \quad (1)$$

Membership degrees can be triangular, trapezoidal or Gaussian. Equation-2 shows the graphical bell curve of membership functions (b: number of rules, a: number of inputs, n: number of membership functions):

$$\mu_{A_i}(x) = \frac{1}{1+[(\frac{x-c_i}{a_i})^2]^{b_i}} \quad (2)$$

$$b = n^a \quad (3)$$

Layer 2: The inputs specified in the first layer are blurred and form rules with each other:

$$o_i^2 = \omega_i = \mu_{A_i}(x) * \mu_{B_i}(x), i = 1, 2, \ldots \quad (4)$$

Layer 3: "ω_i" values are normalized.

$$o_i^3 = \overline{\omega}_i = \frac{\omega_i}{\sum_{i=1}^{n} \omega_i}, i = 1, 2, \ldots \quad (5)$$

Layer 4: Rules normalized in the previous layer are subjected to output functions [20].

$$o_i^4 = \overline{\omega}_i f_i = \overline{\omega}_i (p_i x + q_i y + r_i) \quad (6)$$

$\overline{\omega}_i$: Layer 3 output, normalized values

p_i, q_i, r_i : Result parameters - calculated by least squares method

Layer 5: The total contribution of the outputs is calculated [13]:

$$o_i^5 = \sum \overline{\omega}_i f_i \quad (7)$$

The mathematical method of ANFIS is explained with the above equations. The methods used in the forward-feedback feed of the ANFIS structure are shown in Table 3 [10].

METHODOLOGY

This study examines how the requirements of the protocols [8] and [9] affect different countries and regions. These frameworks impose similar obligations on all member states, despite differences in CE levels and renewable electricity production (REP). For this reason, evaluating countries based on proportional indicators, such as the shares of REP and CE, may provide a more meaningful comparison than using absolute values.

To support this approach, changes in REP and CE rates are analyzed for both current and future periods. The application is based on separate estimations of REP and CE using the ANFIS method. The results are then evaluated, and their implications for countries and regions, as well as the effectiveness of the protocols, are discussed in the conclusion section.

Renewable Electricity Production Estimation

As seen in the methodology in Figure 3, three inputs and one output selected from the literature were used in the ANFIS structure. Data from Türkiye, the USA, China, India, and the EU Region were used for inputs and outputs. Applications have been made with ANFIS for all the countries and regions mentioned. However, in this study, only data and application processes for Türkiye are included as a sample study. Within this framework, information about the data is given under subheadings.

It should be noted that, alongside REP, total electricity generation for 2025–2030 was also estimated. Here, a more straightforward method was chosen, electricity generation historical data in ANFIS was considered, and estimation was made with a single input and a single output. Then, REP was estimated with three inputs and one output. Thus, the REP could be determined thanks to the ratio of the two outputs to each other.

Renewable Electricity Production Estimation Data

ANFIS application inputs are considered POP, GDP, and CE, and output is REP. Figure 5 shows POP data from 1971 to 2022 [28].

Table 3. ANFIS methods in feed forward and feedback

	Forward Path	Reverse Path
Antecedent Parameters	Constant	Gradient Descent
Consequent Parameters	Least Squares Estimator	Fixed
Output	Node Outputs	Error Outputs

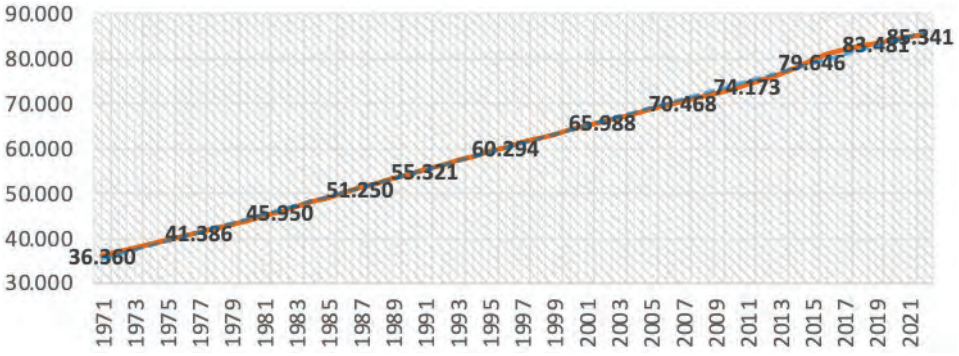


Figure 5. Türkiye POP (1971-2022) (x1000 person).

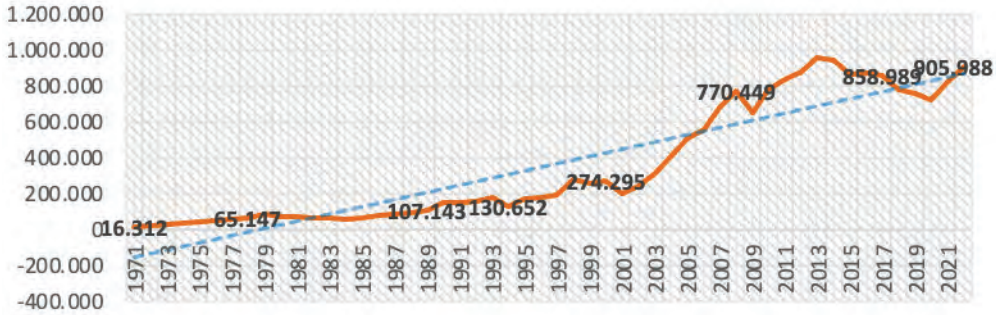


Figure 6. Türkiye / GDP (million $) / (1971-2022).

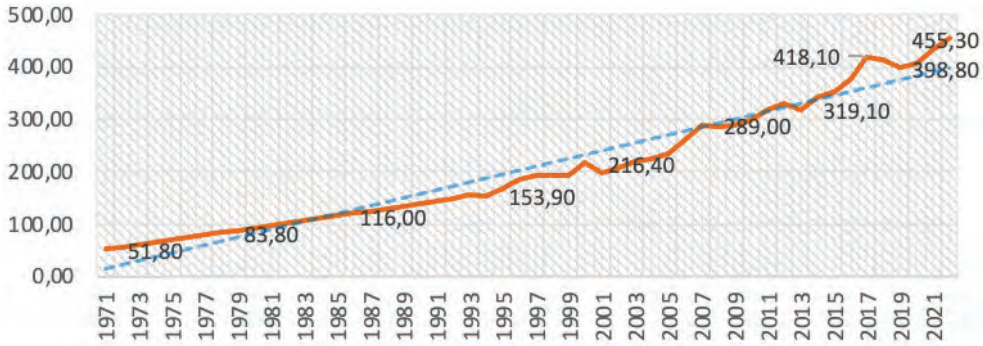


Figure 7. Türkiye / CE (Mt) / (1971-2022).

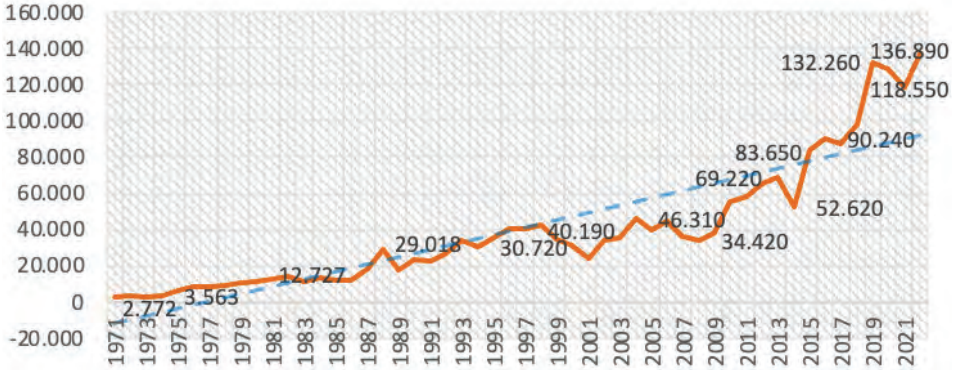


Figure 8. Türkiye / REP (TWh) / (1971-2022).

Figure 6 shows GDP data from 1971-2022 [28] [created by author]

Figure 7 shows CE data from 1971-2022 [28] [created by author]

Figure 8 shows REP between 1971-2022 [1] [created by author]

Using the above input and output data, prediction is made in the next section.

Renewable Electricity Production Estimation Application

Following the approach in previous studies [13], 70% of the dataset was used for training, while the remaining 30% served as test data. Sugeno fuzzy logic base was used in the Fuzzy Logic Toolbox package program of the MATLAB program. In order to obtain a more efficient result from the data, the data was loaded into the interface by mixing it, not according to the increase or decrease rate. As shown in Figure 9, the training and test datasets align well with the output data, indicating consistent performance.

Afterward, the data was epoched 300 times to determine the lowest error in different membership types. Table 4 shows the membership type with the lowest error. The membership type with the lowest error at the end of 300 epochs was determined as "gbellmf" and its membership subset was "4-4-4". Table 4 shows the membership type with the lowest errors.

The error mentioned in Table 4 is the Root Mean Squared Error (RMSE). The reason is that the above graphic data (except POP) deviates too much from the mean after a certain year and contains large values. Since using standard MSE in this example would artificially inflate the error value [29], ANFIS employs RMSE, which is the square root of MSE. More meaningful error values are reached with the square root operation, and a selection is made among these values according to the degree of closeness to "0". Accordingly, both training and test data were subjected to RMSE examination at 300 cycles. Figure 10 shows this process in ANFIS.

Iterations continue until the RMSE value stabilizes; once it does, the compatibility between the datasets is carefully evaluated. The scale used to examine this compatibility is R^2 and correlation. According to the R^2 scale (ranging from 0 to 1), a higher value reflects greater accuracy [29]; therefore, Figure 11 presents each dataset separately. It should

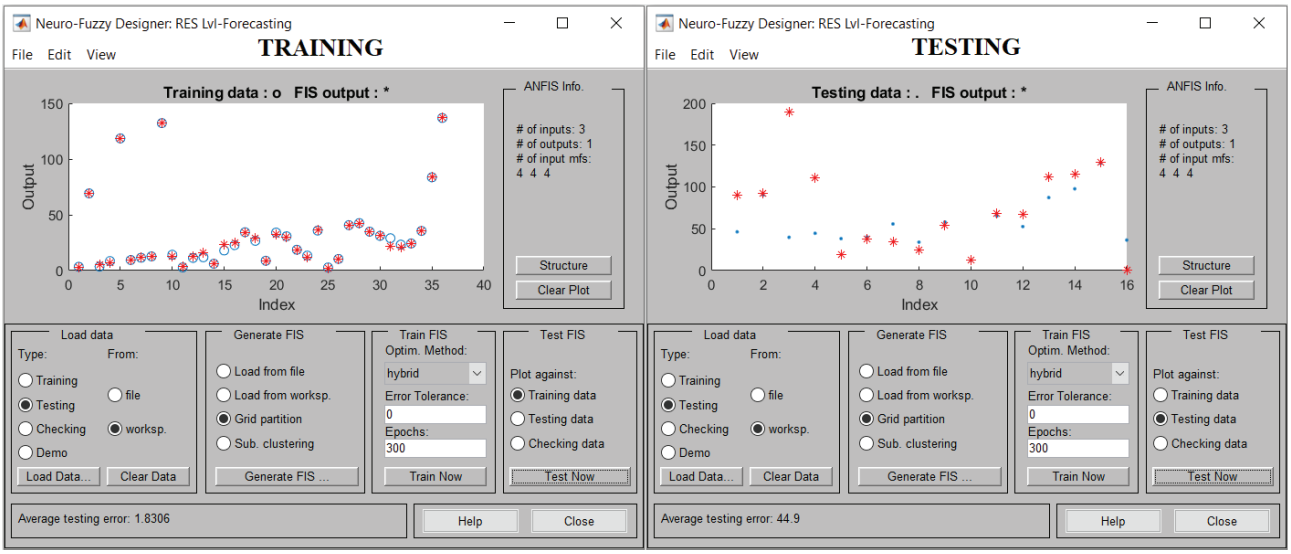


Figure 9. ANFIS / Compatibility of training and testing data with output data.

Table 4. ANFIS / Membership type for lowest error

Membership name	Membership type	Membership subset	RMSE value	RMSE rank
Triangle MF	trimf	3-3-3	2,1838	
Triangle MF	trimf	4-4-4	1,8854	
Trapezoid MF	trapmf	3-3-3	4,0886	
Trapezoid MF	trapmf	4-4-4	3,0863	
Gaussian Bell MF	gbellmf	3-3-3	2,1627	
Gaussian Bell MF	gbellmf	4-4-4	1,8307	min.

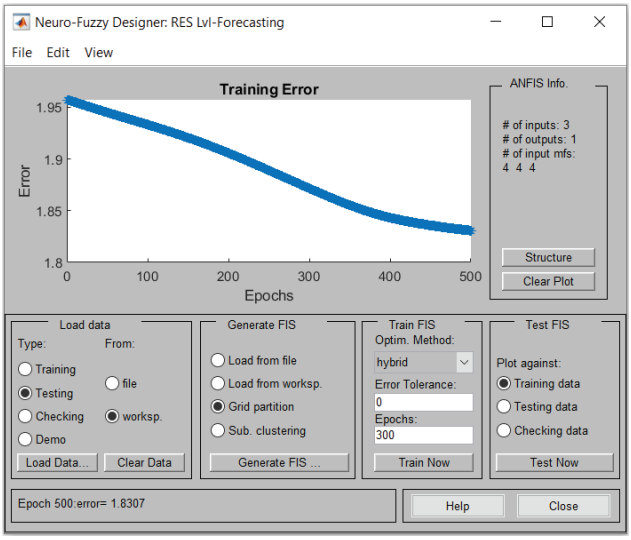
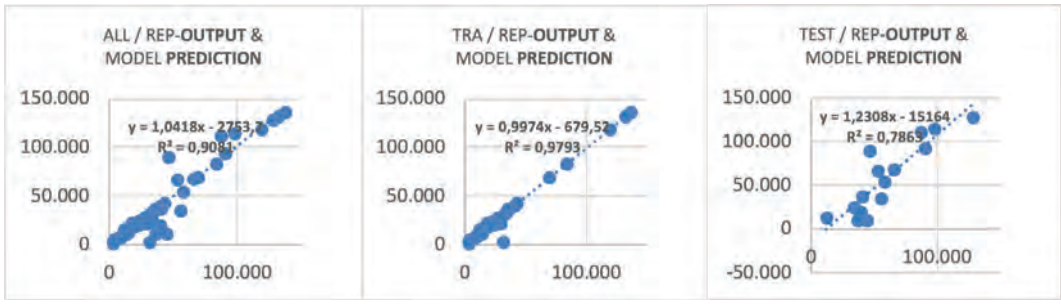


Figure 10. ANFIS / Determining membership type.

be noted that as the correlation coefficient approaches 1, it shows a high relationship between the data.

Figure 11 shows a high correlation for training, testing and all data. Therewithal, the main scale, R^2, is very close to 1 for all data and training data, and this ANFIS declares that the estimation has a high accuracy rate of 90%. Although the test data does not show a very high accuracy rate of 78%, this rate does not pose a significant issue since the main purpose of the test data set is to perform testing in order to measure performance. This rule base and surface is shown in Figure 12.

After these results, there is a surface with an interference pattern formed by the inputs with each other. A steep surface indicates lower accuracy, whereas a flat surface demonstrates higher predictive precision [10]. This "rule surface" is shown on the left in Figure 12. This rule surface is designed on the graph of the prediction values obtained due to the rule formed by the inputs with each other in



	ALL	TRAINING	TEST
Correlation	0,9529	0,9896	0,8868
R^2	0,9081	0,9793	0,7863

Figure 11. Correlation of training-test inputs and output data

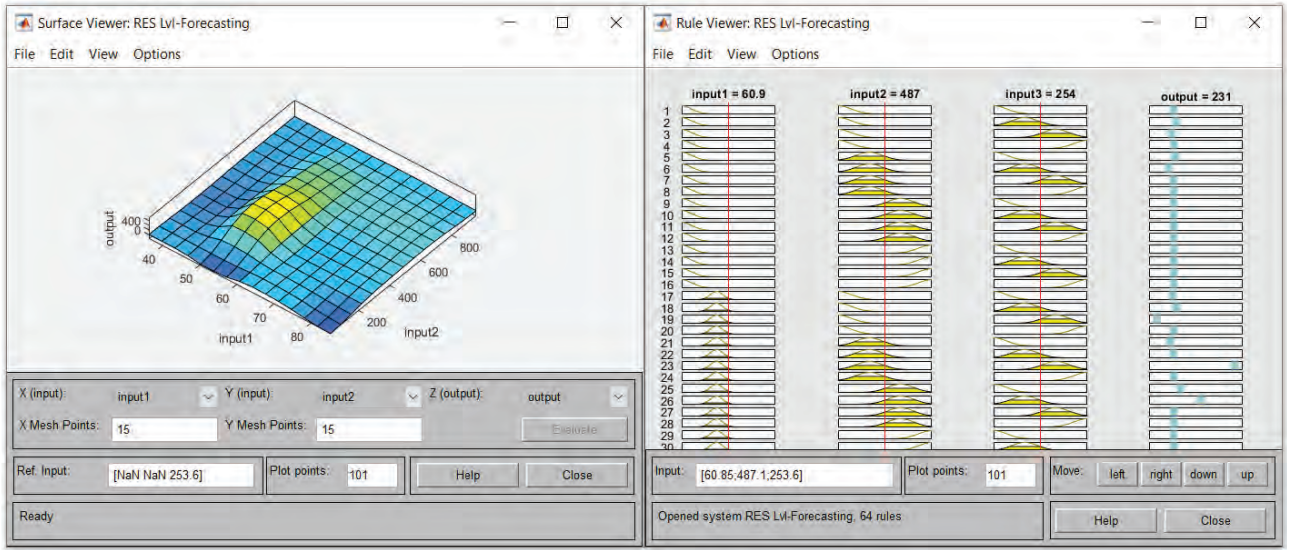


Figure 12. ANFIS / Rule base of the model.

ANFIS. The right representation shows the current inputs and the prediction outputs corresponding to the rules generated by these inputs. This representation is called the "rule base". This rule base is used when making future period forecasts. The estimation results obtained according to this regular structure are given in the "Findings and Discussion" section. However, CE estimations is explained before this section.

CO_2 Emissions Estimation

As shown above, CE estimation was also performed with ANFIS in a similar way to REP estimation. The same three inputs were used. The countries and regions considered for the estimation were also the same. CE data for 1971-2022 and GDP and POP data for this period, shown in Figure 2, were used as inputs. CE (t+1) data were also used as outputs. Again, membership functions with minimum errors were used with ANFIS separately for each country and region.

Estimation results and interpretations are included in the "Findings and Discussion" section.

THE RESEARCH FINDINGS AND DISCUSSION

This section is explained in two parts: "Findings" and "Discussion".

Findings

This study investigates the future REP and CE utilization rates to examine the impact of the protocols on countries and regions. In this regard, these rates are estimated separately for each country and region with ANFIS in the "Application" section, based on the 2025-30 period. The estimation results are shown in Table 5.

As seen in Table 5, the 2023-25-27-30 estimates are given for REP and CE obtained using values between 1971-2022. The left side of the table shows the REP ratio in the electricity produced on a TWh basis. There are three rows for each country and region. The first row shows the REP for the relevant year, and the second row shows the total electricity production for the same year. The third row shows the ratio of the former to the latter, i.e., the proportional REP. According to the future estimate here, Türkiye ranks first with an increase of approximately 15 percent in REP in nine years (2021-30). The closest follower is the EU, which has approximately 12 percent, and China ranks third, with an increase of approximately 11 percent. While the USA did not quite meet the climate protocol expectations with an increase of approximately 2 percent, India could not increase its REP rate in nine years according to the forecast scenario.

On the right side of the table are the CE forecast results for 2021-30. There are two rows for each country and region. The first row shows the CE increase amount by year in million tons, while the second row shows the CE increase rate between years. Accordingly, the EU Region draws the most positive curve with a decrease of approximately 10%. The USA, with a decrease of approximately 8%, and Türkiye, with a decrease of approximately 7%, are the followers. On the other hand, China and India increase their emissions by approximately 4% and 7%, respectively, presenting a negative picture.

In order to understand the increases and decreases in this table more clearly, along with their slopes, a graphical version is presented in Figure 13.

Table 5. Future REP and CE rate forecasting for countries and regions

Regions/Years	REP-rate forecasting					Regions/ Years	CE-rate forecasting				
	21	23	25	27	30		21	23	25	27	30
TÜR-(REP-TWh)	116,5	141,8	172,8	186,9	208,1	TÜR-(Mt)	452,7	446,1	473,8	494,5	520,7
TÜR-(TEP-TWh)	327,9	325,0	351,4	367,9	412,2	*TÜR-(%)*	*12,44*	*-1,46*	*6,20*	*4,37*	*5,30*
TÜR-(REP-%)	*35,53*	*43,63*	*49,18*	*50,81*	*50,48*						
IND-(REP-TWh)	333,0	381,3	427,0	467,9	502,3	IND-(Mt)	2.674,2	2.912,0	3.132,9	3.348,7	3.682,2
IND-(TEP-TWh)	1.714,8	1.896,3	2.127,2	2.302,0	2.590,1	*IND-(%)*	*2,35*	*8,89*	*7,59*	*6,89*	*9,96*
IND-(REP-%)	*19,42*	*20,11*	*20,07*	*20,33*	*19,39*						
EU-27-(REP-TWh)	1.075,3	1.152,3	1.235,0	1.322,8	1.392,9	EU-27-(Mt)	*2.805,8*	*2.700,8*	*2.575,2*	*2.514,1*	*2.410,7*
EU-27-(TEP-TWh)	2.875,0	2.790,5	2.814,9	2.823,1	2.802,7	*EU-27-(%)*	*6,68*	*-3,74*	*-4,65*	*-2,37*	*-4,11*
EU-27-(REP-%)	*37,40*	*41,29*	*43,87*	*46,86*	*49,70*						
USA-(REP-TWh)	862,6	897,4	936,8	1.002,9	1.024,4	USA-(Mt)	5.032,2	5.082,4	4.985,0	4.846,8	4.712,6
USA-(TEP-TWh)	4.153,6	4.284,5	4.346,7	4.503,0	4.555,9	*USA-(%)*	*6,74*	*1,00*	*-1,92*	*-2,77*	*-2,77*
USA-(REP-%)	*20,77*	*20,95*	*21,55*	*22,27*	*22,49*						
CHI-(REP-TWh)	2.449,0	2.876,4	3.384,3	4.002,2	4.792,3	CHI-(Mt)	11.336,2	11.805,0	12.303,4	12.972,9	13.932,2
CHI-(TEP-TWh)	8.534,3	9.285,1	9.911,1	10.871,1	11.987,6	*CHI-(%)*	*3,87*	*4,13*	*4,22*	*5,44*	*7,39*
CHI-(REP-%)	*28,70*	*30,98*	*34,15*	*36,82*	*39,98*						

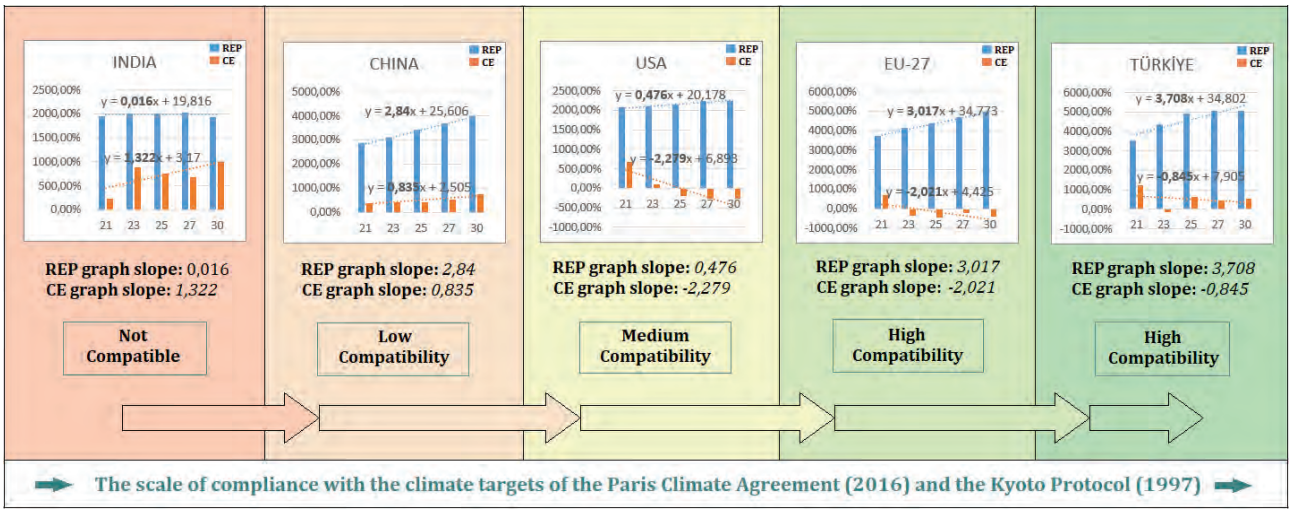


Figure 13. Future projections for REP and CE (2023-2030).

The findings regarding the comparison of REP and CE growth rates of Türkiye and other regions according to their future situations are shown above. It is possible to make some inferences from these findings. These inferences and comments are included in the "Discussion" section.

DISCUSSION

In this study, the conditions that member countries must comply with in order to achieve the climate conditions envisaged by the [8] and [9] protocols are questioned. Countries with different REP and CE levels are subject to the same conditions. However, countries with more emissions should have a higher liability. The first is the status of REP over the years. The second is the status of total CE increase.

First of all, if the current situation is evaluated, as seen in Figure 1 and Figure 2, the most obvious graph belongs to China. In the 1970s, it followed a graph closer to India and Türkiye, and later, by using intensive coal resources in every field from electricity to fuel, it has recently had an emission twice as much as the USA and EU Region. As mentioned above, although the use of RES in electricity has increased in China and even RES policies have gained importance, this increase cannot prevent the increase in CE due to FS. In India, although policies towards RES have gained importance and the RES graph has a good increase slope, the increase in CE due to FS is higher. Similar to China, it cannot prevent this. For these two countries, a high population also plays an active role in the use of resources. The primary objective is to secure the public electricity supply; hence, climate protocols and RES initiatives receive secondary attention. For example, approximately 70% of electricity production in India is provided by coal. When this main source is not sufficient and a deficit occurs, imported coal is used instead of RES [30]. The same applies to China, where the rate is around 61% and increasing [31]. In Türkiye, the CE graph increases with a slight slope. This level can be considered as a typical situation for a developing country. In addition, the RES increase slope is observed to be higher than the CE increase slope. There is an increase slope for coal in electricity, but thanks to the increase in HPP, WPP and SPP installations accompanied by an energy strategy [32], Türkiye suppresses CE in electricity production [31]. Türkiye appears well positioned to meet the requirements of the climate protocols; thus, it aligns with global sustainability goals. The CE level in the EU Region, as seen in Figure 2, is decreasing. This has been triggered by the EU's general need to turn to internal resources due to its decarbonization and energy security policies [33]. In addition, it should be noted that within the EU, it has been decided to reduce even NES, which has been the primary source of electricity for about thirty years, within the scope of these policies [31]. In this way, the FS rate in electricity production has been reduced to 39%. The RES rate is still increasing. This is a positive picture for both the NetZero [34] target and the protocols. Like the EU, the USA actively promotes RES through various incentives; notably, the Inflation Reduction Act plays a key role. According to the IEA estimate, the USA will increase its installed RES capacity from approximately 150 GW to 340 GW between 2023-28. It is predicted that the majority of this installed capacity will be SPPs and HPPs [33]. However, the picture seen until 2022 does not fully reflect this. As stated, there is an increase in all types of RES, and at the same time, the FS ratio in electricity production decreased from approximately 70% in the 2000s to approximately 59% around 2022. However, the subtle nuance here is that there is no decrease in FS resources in this rate decrease. The steady rise in RES share is the primary factor behind the decline in FS use, showing a clear substitution trend. Diminishing coal resources have gradually been replaced by natural gas, as indicated in

Table 6. ANFIS Estimation Performance by Country/Region for REP and CE Processes

	REP Estimation Processes			CE Estimation Processes		
	MF type	**MF subset**	**min. RMSE value**	**MF type**	**MF subset**	**min. RMSE value**
Türkiye	gbellmf	3-3-3	2,1838	gbellmf	4-4-4	1,6879
EU-27	trimf	4-4-4	2,0271	gbellmf	4-4-4	2,3627
USA	gbellmf	4-4-4	2,1634	gbellmf	3-3-3	2,8651
China	gbellmf	3-3-3	6,6402	trimf	3-3-3	7,0215
India	trimf	3-3-3	3,3296	trimf	4-4-4	4,4236

reference [31]. The results indicate that the emission reduction in the USA is not sufficient, but it reaches a moderate level in terms of protocol obligations.

After evaluating the current situations of the countries in question, their future projections are assessed. However, before that, information is provided on the ANFIS processes, training-testing phases, and the membership functions selected for minimum error for the countries and regions not specified in the application section. Table 6 presents information regarding the REP and CE level prediction processes for the countries and regions considered in this study.

The data in Table 6 represent the scaled training–testing datasets of the countries and regions (the numerical values were reduced to comparable magnitudes to avoid outlier errors). These results were obtained from ANFIS after 300 epochs. The models with the minimum MF values serve as the rule bases for both the REP and CE estimation processes, thus providing the foundation for forecasting future values of the countries and regions.

As can be seen, China's minimum error value remains the highest among all. This can be attributed to the steeply increasing trend observed in Figures 1 and 2. Similarly, although the EU-27 REP values are higher than those of Türkiye, the smaller variation and higher consistency among its data lead to a lower minimum error value in Table 6. The same situation applies to the USA, whereas the India case resembles China. The same interpretation pattern can also be applied to the CE process results.

The estimation results obtained based on the rule bases formed by the models with the minimum errors specified in Table 6 are presented in detail in Table 5 and Figure 13. As can be seen in, three graphs are on a positive trend towards 2030. These graphs belong to Türkiye, the EU region, and the USA. Türkiye presents the most optimistic picture with an increase slope of 3,7 in REP. It can be stated at this point that there is an important driving force for Türkiye in this table. The force in question is the energy expenses, which constitute the most oversized item in its current deficit and are mostly imported. To overcome this situation, Türkiye has turned to domestic use, and while continuing its natural gas exploration activities [35], it has also significantly increased its REP. REP reached 42% by the end of 2023 [36]. As a result of the orientation towards emission-free energy, the increase rate in CE level is decreasing with a slope of -0,84. This situation shows its compliance with the protocols.

Apart from Türkiye, two other regions are trying to comply with the protocols. The EU Region is increasing its REP with a slope of 3,0 towards 2030. This increase rate is quite good for the EU Region, which realizes 15% of the world's electricity REP [1]. In addition, the CE increase rate is decreasing with a slope of -2. According to the graphs in Figures 1 and 2, RES strategies have gained significant importance in European countries in the last 15 years [37] and many measures are being taken to reduce emissions. For example, the EU has expanded the "Emissions Trading" agreement to include buildings and roads. It has also established a carbon border adjustment system [33]. Although more than this positive situation is needed for the 2050-net zero [34] target, it will provide a future close to this target. The protocols appear to be effective in the EU region. Similarly, the USA is also offers practices such as tax breaks and incentives for the use of RES technologies [33]. Nevertheless, the increase slope of REP towards 2030 has not exceeded 0,45. Although the increase in REP was not at the desired level, a CE reduction slope of -2.82 was achieved by reducing the use of FS. This is the most positive emission graph in all countries and regions examined. In this context, it can also be said that the effect of the protocols is evident in the USA. On the other hand, China cannot display a net positive outlook since it shows a rapid trend towards both FS and RES towards 2030. The reason for not giving up fossil fuels while moving towards RES is its intensive coal resources. Here, it has both high FS resources and a good REP increase slope [33,38]. Comparatively, China is projected to have approximately five times more REP than the USA in 2030. The slope increase of REP in China is 2.84 which is six times higher than the USA and an increase almost on a par with the EU Region. On the other hand, the CE increase slope is also increasing and is 0,84. Therefore, although REP is growing in China, it can be said that the protocols are not effective due to the positive emission rate.

India, one of the countries examined, is decreasing the REP growth rate and increasing the CE growth rate towards

2030, according to the forecast projection. However, India's energy policies include the target of meeting half of its power generation capacity from RES by 2030. In addition, there is a net-zero target by 2070 [34]. The estimates show that these policies are not fully implemented, and the effect of the protocols is insufficient.

When this study is compared with earlier works in the literature, such as Yılmaz and Yılmaz [17], who employed a fuzzy logic model for a single-country analysis of Türkiye, this study expands the scope to a multi-regional comparative framework that integrates both renewable energy forecasting and international policy evaluation. Similar studies in the literature (such as Kumar et al. [39] for India, Li et al. [40] for China, and Rahman et al. [41] for the EU region) primarily focused on energy–emission trends using separate machine learning or statistical models, yet their forward-looking projections under the same global protocols produced findings consistent with this study. As stated above, these studies also show that Türkiye and the EU-27 appear more compliant with international protocols, while China and India show weaker alignment. In addition, the RMSE values in those studies are higher for China and India, whereas they are lower—and thus more consistent—for Türkiye and the EU-27 countries. However, there are also studies that produce findings not fully consistent with the results. For instance, Şahin [38] obtained similar findings to this study regarding the increase in renewable energy production (REP) in future projections; however, as a consequence of this increase, contrary to the results indicate that the rate of emission growth would slow down.

The present study fills this gap by linking quantitative forecasting (ANFIS) with policy-based interpretation, providing a cross-country perspective on protocol effectiveness.

The 2030 forecast of the countries and regions examined along with the impacts of the protocols and the resulting positive and negative trends, were presented and discussed. The overall results are evaluated in the next section.

CONCLUCION

Today, in the pursuit of sustainability, there is an increasing shift toward next-generation energy sources that aim to achieve both resource abundance and emission reduction. Energy diversification and emission mitigation also vary according to the specific goals of each country and region. This study aims to estimate the future renewable energy use and emission levels of countries and regions in order to examine the impact of international climate protocols on them. Among these agreements, the United States of America, China, and the Europa Union-27 signed the Kyoto Protocol in 1998, while India ratified it in 2002 and Türkiye became a party in 2009. Similarly, all five countries/regions signed the Paris Climate Agreement in 2016. At this point, Türkiye is the main country considered for comparison with other countries and regions. The criteria are the renewable electricity production and the CO_2 emission level. The 2025-27-30 estimate is made with ANFIS, and their future situations are examined. In this context, despite being a developing economy, there is a positive picture in which Türkiye is reducing its fossil sources use and turning to renewable energy sources. The region with the most optimistic picture after Türkiye is the Europe Union, which both reduces its emission slope and increases its renewable electricity production. The United States of America follows these two; although the decrease in the emission slope is good, the renewable electricity production increase remains low. The country that follows is China. Although the renewable energy soruces increase is higher than all other regions, since the fossil sources use is also high at a similar rate, the emission rate remains far from the warning of the protocols. India cannot increase its renewable electricity production, and its emission rate increases considerably with fossil sources use. The findings are consistent with earlier literature in identifying the general link between renewable energy expansion and emission reduction, but the multi-country comparison here reveals a broader perspective: countries with high industrial growth diverge from protocol expectations, while Türkiye and the EU-27 show more balanced adaptation.

According to these results, it should be noted that it would be more effective to transform the standard and common language texts of the protocols into country and region-based texts. High-emitting countries should be subject to more conditions. The renewable energy soruces orientation should be encouraged much more in the text prepared for these countries. In this way, texts should be created separately for countries and regions and specific to their own situations.

At this point, the core idea proposed by this study is that the standards established by international protocols should be able to define adaptable provisions in response to changing conditions. Although the methods and datasets may vary, the fundamental rationale of this study can serve as a foundation for future works.

ACKNOWLEDGEMENTS

This work was supported by the İlim Yayma Foundation.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Our World in Data. Electricity generation from renewables. Available at: https://ourworldindata.org/grapher/electricity-renewables. Accessed on Apr 21 2026.

[2] Onder HG. Renewable energy consumption policy in Turkey: An energy extended input-output analysis. Renew Energy 2021;175:783–796. [CrossRef]

[3] Carfora A, Pansini RV, Scandurra G. Energy dependence, renewable energy generation and import demand: Are EU countries resilient?. Renew Energy 2022;195:1262–1274. [CrossRef]

[4] Shang Y, Han D, Gozgor G, Mahalik MK, Sahoo BK. The impact of climate policy uncertainty on renewable and non-renewable energy demand in the United States. Renew Energy 2022;197:654–667. [CrossRef]

[5] Liu X, Zhao T, Chang CT, Fu CJ. China's renewable energy strategy and industrial adjustment policy. Renew Energy 2021;170:1382–1395. [CrossRef]

[6] Das A, Jani HK, Nagababu N, Kacchwaha SS. A comprehensive review of wind–solar hybrid energy policies in India: Barriers and recommendations. Renew Energy 2020;35:108–121. [CrossRef]

[7] Our World in Data. Annual CO_2 emissions. Available at: https://ourworldindata.org/grapher/annual-CE-emissions-per-country. Accessed on Apr 21 2026.

[8] Republic of Türkiye Ministry of Foreign Affairs. Kyoto Protocol. Available at: https://www.mfa.gov.tr/kyoto-protokolu.tr.mfa. Accessed on Apr 21 2026.

[9] Republic of Türkiye Ministry of Foreign Affairs. Paris Agreement. Available at: https://www.mfa.gov.tr/paris-anlasmasi.tr.mfa. Accessed on Apr 21 2026.

[10] Sekerci AZ, Kara SS. Türkiye CO_2 emission estimation and OECD comparison: An application with ANFIS. Springer 2026.

[11] Rodrigues JAP, Neto LB, Coelho PHGJ, Mello CCBS. Estimating greenhouse gas emissions using computational intelligence. Scitepress 2009;248–250.

[12] Appiah K, Du JG, Appiah R, Quacoe D. Prediction of potential carbon dioxide emissions of selected emerging economies using artificial neural network. J Environ Sci Eng 2018;7:321–335. [CrossRef]

[13] Guleryuz D. Determination of industrial energy demand in Turkey using MLR, ANFIS and PSO-ANFIS. J Artif Intell Syst 2021;3:16–34. [CrossRef]

[14] Nassef AM, Olabi AG, Rezk H, Abdelkareem MA. Application of artificial intelligence to predict CO_2 emissions: Critical step towards sustainable environment. Sustain 2023;15:7648. [CrossRef]

[15] Arevalo P, Cano A, Jurado F. Large-scale integration of renewable energies by 2050 through demand prediction with ANFIS: Ecuador case study. Energy 2024;286:129446. [CrossRef]

[16] Ozdemir V. Determination of Turkey's carbonization index based on basic energy indicators by artificial neural networks. J Fac Eng Archit Gazi Univ 2011;26:9–15.

[17] Yilmaz H, Yilmaz M. Forecasting CO_2 emissions for Turkey by using the grey prediction method. Sigma J Eng Nat Sci 2013;31:141–148.

[18] Baareh AK. Solving the carbon dioxide emission estimation problem: An artificial neural network model. J Softw Eng Appl 2013;6:338–342. [CrossRef]

[19] Khan MZ, Khan MF. Application of ANFIS, ANN and fuzzy time series models to CO_2 emission from the energy sector and global temperature increase. Int J Clim Chang Strateg Manag 2019;11:622–642. [CrossRef]

[20] Jusoh M, Panjaitan CE, Sentosa V. Forecasting of electrical energy consumption in Aceh 2028 using the adaptive neuro fuzzy inference system. J Electron Telecommun Electr Phys Sci 2021;1:13–19.

[21] Jena PR, Managi S. Forecasting the CO_2 emissions at the global level: A multilayer artificial neural network modelling. Energies 2021;14:6336. [CrossRef]

[22] Lazim A, Pauzi HM. Adaptive-neuro fuzzy inference system for CO_2 emissions data. Springer 2021;155–161.

[23] Nugraha YT, Ghabriel K, Arshad SNM. Implementation of ANFIS in forecasting the development of renewable energy in Indonesia in 2027. J Electron Telecommun Electr Phys Sci 2021;1:1–6.

[24] Sekerci AZ. Investigation of the effect of increased CO_2 gas emissions on temperature levels in Turkey: An application with ANFIS and linear regression. Proceedings of the 2nd Rumeli Energy and Design for a Sustainable Environment Symposium. Istanbul, Türkiye; 2022. p.69–181.

[25] Mutascu M. CO_2 emissions in the USA: New insights based on ANN approach. Environ Sci Pollut Res 2022;29:68332–68356. [CrossRef]

[26] Ginting RT, Angin DP, Nugraha YT, Gultom TT, Nainggolan WP, Sitanggang D. Short-term forecast for the growth of Indonesia's new renewable energy using ANFIS. J Inf Syst Comput Sci Prima 2023;6:57–60. [CrossRef]

[27] Beghouy P, Ugurenver A. Electricity energy consumption forecasting using LSTM and NAR neural networks, Sigma J Eng Nat Sci 2024;42:237–245.
[28] The World Bank. Türkiye. Available at: https://data.worldbank.org/country/turkiye. Accessed on Apr 21 2026.
[29] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7:e623. [CrossRef]
[30] PRSIndia. Overview of issues in the power sector in India. Available at: https://prsindia.org/policy/vital-stats/overview-issues-power-sector-india. Accessed on 21 Apr 2026.
[31] Our World in Data. Electricity production by source. Available at: https://ourworldindata.org/grapher/electricity-production-by-source. Accessed on Apr 21 2026.
[32] Republic of Türkiye Ministry of Foreign Affairs. Türkiye's International Energy Strategy. Available at: https://www.mfa.gov.tr/turkeys-energy-strategy.en.mfa. Accessed on Apr 21 2026.
[33] International Energy Agency. World Energy Outlook 2023. Available at: https://www.iea.org/reports/world-energy-outlook-2023. Accessed on Apr 21 2026.
[34] International Energy Agency. Net Zero by 2050. Available at: https://www.iea.org/reports/net-zero-by-2050. Accessed on Apr 21 2026.
[35] Turkish Petroleum Corporation. Oil and Natural Gas Sector Report. Available at: https://www.tpao.gov.tr/file/2309/2022-sektor-raporu-guncel-15766516d52de0444.pdf. Accessed on Apr 21 2026.
[36] SHURA Energy Transformation Center. Türkiye Energy Transformation Outlook. Available at: https://shura.org.tr/wp-content/uploads/2024/04/SHURA-2024-04-Rapor-outlook-2023-rev.pdf. Accessed on Apr 21 2026.
[37] Yasuda Y. Quantifying the reduction in coal and increase in renewables in OECD countries. Renew Sustain Energy Rev 2024;198:114424. [CrossRef]
[38] Sahin U. Forecasting share of renewables in primary energy consumption and CO_2 emissions using grey model. Exp Syst Appl 2022;209:118429. [CrossRef]
[39] Kumar A, Sah B, Singh AR, Deng Y, He X, Kumar P, et al. A review of multi-criteria decision making towards sustainable renewable energy development. Renew Sustain Energy Rev 2019;69:596–609. [CrossRef]
[40] Li J, Fang K, Song J, Peng L. Forecasting carbon emissions of China using a hybrid machine learning model based on improved grey relational analysis and ANFIS. J Clean Prod 2021;279:123841.
[41] Rahman MM. Hasanuzzaman M. Rahim NA, Hosenuzzaman M. Energy, environmental and carbon emission analysis of the European Union. Renew Sustain Energy Rev 2020;133:110099.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2032

Research Article

Optimized mix framework for accelerated carbonation concrete by ground granulated blast furnace slag and silica fume

Ramatai B. SOMWANSHI[1,*], Abhijeet R. NARDEY[1], Padmanabh Arun GADGE[2]

[1]Department of Civil Engineering, G H Raisoni University, Madhya Pradesh, 480337, India
[2]Department of Mechanical Engineering, G H Raisoni University, Madhya Pradesh, 480337, India

ARTICLE INFO

Article history
Received: 27 December 2024
Revised: 20 March 2025
Accepted: 21 March 2025

Keywords:
Acid Attack; Carbonation Concrete; Chloride Permeability; Compression Strength; Taguchi Optimization

ABSTRACT

The proportion of carbon dioxide in the atmosphere has risen enormously, particularly in the last decade. The carbon sequestration technique reduces the carbon dioxide level in the atmosphere. One such method in concrete technology is carbonation curing of concrete during its early strength developing age. During accelerated carbonation, the un-hydrated particles in the concrete combine with carbon dioxide and form carbonate, leading to the accelerated solidification of concrete with improved durability properties. However, if the mix is not selected properly, then it takes a longer curing time than other traditional mixes. Hence, selecting the optimal mix based on the desired outcome and less curing time has tended to gain the finest mechanical properties. In this investigation, a novel approach, termite-based Taguchi optimization, is implemented to predict the mix proportion of concrete from the material parameters of the concrete. Before, the termite model was not implemented for this specific application, nor was Taguchi optimization. The fitness function in the Termite algorithm helps with carbonation curing of the concrete, and the prediction function is executed to determine the endurance parameters of the concrete, which is considered a novelty of the study. The endurance of the carbonation concrete is determined in terms of compression strength, carbonation depth, chloride permeability, water sorptivity, and acid attack. Moreover, the important results are the mix containing 15% silica fume and 60% Ground Granulated Blast Furnace Slag has lower results in the permeability tests of concrete that include Carbonation depth of 4mm, water sorptivity of 0.0125 mm/sec$^{0.5}$ and 619.25 coulombs in the chloride permeability test at the 28th day. In strength parameters, the mix containing 15% silica fume and 40% Ground Granulated Blast Furnace Slag possesses higher compression strength of 47 MPa and strength reduction in the acid attack test of 6.26% at the 28-day test. Since, the mix proportion with 15% silica fume and 40% Ground Granulated Blast Furnace Slag possess higher strength and optimal permeability values. This mix is considered an optimum mix for this research.

Cite this article as: Somwanshi RB, Nardey AR, Gadge PA. Optimized mix framework for accelerated carbonation concrete by ground granulated blast furnace slag and silica fume. Sigma J Eng Nat Sci 2026;44(2):1153–1171.

*Corresponding author.
*E-mail address: ramataibabaraosomwanshirbs@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

The industrialization and urbanization that have taken place in recent decades have been greatly influenced by the building industry [1]. The construction sector generates 5–10% of all jobs worldwide and contributes 5–15% of GDP to each country [2]. The building industry accounts for about 30% of total resource depletion and 40% of overall energy use [3]. Moreover, these construction sectors generate over 30% of the world's waste and 40% of its carbon dioxide emissions [4]. The production of this cement was considered the main reason for carbon dioxide emissions across the globe, and the percentage of emissions was between 7 and 9 [5]. Currently, more than 5% of carbon dioxide emissions globally come from manufacturing facilities of cement production [6]. By using different binding elements in place of OPC, cementation materials can help lower carbon dioxide emissions [7]. Multiple industry wastes can be used in multi-component binder materials for various applications [8]. Much research has been done on concrete construction employing additives to lower the cost and scarcity of common materials [9]. In the building industry, cement is an essential part of concrete, the manufactured resource used in construction most frequently [10]. Approximately 4 billion tonnes of cement are produced annually worldwide, which translates to nearly 29 billion tonnes of concrete in 2015 [11]. The demand for cement is expected to expand steadily between 1990 and 2030. Considerable expansion in production has resulted from the rising demand for cement in modern infrastructure and structures, especially in emerging nations like China, Russia, and Japan [12]. Cement production has developed a capacity for generating 59.5 million tonnes of annual cement due to the anticipation of future demand for expanding infrastructure [13]. Cement prices have increased by almost 150% in the last ten years. As a result, it is crucial to use additional materials whenever possible in place of cement [14]. In the last thirty years, the building industry, especially in the United Kingdom, has taken some actions to minimize the release of toxic gases associated with cement production [15].

Alternatively, consider calcining coal using organic gas instead of petrol, absorbing carbon dioxide with chemicals, creating a more effective clinker grinding method, and implementing sustainable cement production [16]. However, cementitious materials might be a workable plan to cut greenhouse gas emissions dramatically [17]. Waste materials from manufacturing processes, like fly ash a byproduct from power plants, metakaolin, ground granulated blast furnace slag (GGBS), and fumes of silica, are utilized in place of OPC and can significantly reduce greenhouse gas emissions [18]. The particles of pozzolanic can be created through the buildup of chemical materials, such as silica, calcium, alumina, magnesia, and iron, into concentrations of more than seventy percent, according to ASTM [19]. GGBS contains over 70% of several elements, including iron, silica, calcium, alumina, and magnesia [20]. Therefore, fly ash, metakaolin, and GGBS silica fume can utilize the pozzolanic material, which might be used in concrete to replace OPC. Because of their binding qualities, these wastes silica fume, metakaolin GGBS, and fly ash, have drawn attention from researchers as potential additions to cement blends that could reduce the carbon dioxide quantity left on the bottom layers of the atmosphere [21]. Binding material can be used with other byproducts from industries, including rubber, fly ash, slag, foundry sand, waste marble, etc. The majority of these pozzolanic materials originate from manufacturing waste [22].

Industrial wastes, such as slag from the blast furnace and its sub-product GGBS, are most widely used as auxiliary materials in cement-based products for the production of concrete and cement [23]. Apart from its pozzolanic effect, its self-hydration characteristic sets it apart from other compounds [24]. Its color alters to the presence of 30–40% with the oxides of calcium. The certainty that GGBS will be an eco-friendly alternative for the binder in the concrete accounts for its significance [25]. GGBS needs to be disposed of properly because it is a byproduct. By mixing it with fine and coarse aggregates, OPC, and other conventional concrete components, this waste material can reduce its depletion from the concrete mix [26].

Cement-based materials can perform much better and store CO_2 permanently if their carbonation is accelerated. Research indicates that cement-based products cured by CO_2 have greater mechanical performance and durability, while recycled concrete aggregates treated by CO_2 have better qualities than untreated RCAs. Cement has considerable potential to store CO_2, as it can absorb up to 50% of its mass. Cement-based products have reduced CO_2 emissions by over 40%, sequestering 1.4 Gt of CO_2. Volume stability of carbonation-exposed cement-based composites is a concern due to cracking risk. Dimensional stability is less studied than mechanical properties, carbonation depth, and CO_2 uptake capacity. Advancements in microscope and spectroscopic technologies have led to the increased investigation into the carbonation process of cement hydration products and their microstructural impacts on concrete matric. Understanding these variations is essential for intelligent CO_2 sequestration and utilization technologies. Some alternative methods of lessening CO_2 emissions and streamlining concrete-making processes are capturing and utilizing flue gases produced in making cement, stimulating energy-efficient technology in cement manufacture, recycling and utilizing industrial residues, and investigating cement-free building practices. In particular, inorganic CO_2 sequestration has emerged as a relatively innovative mechanism for CO_2 reduction, promoting significant progress and environmental outcomes over the past two decades. Mineral CO_2 sequestration in concrete production can defined as a carbon capture and storage method that essentially entails CO_2 reacting with calcium-containing material to produce calcium carbonates. When atmospheric

carbon dioxide reacts with the alkaline components of concrete, mainly portlandite, and the hydrated calcium silicate (CsSsH) gel, calcium carbonate ($CaCO_3$) is produced. This process is known as carbonation in concrete. The pH level in the pore water of the concrete usually drops due to this chemical reaction, going from extremely alkaline levels over 12 down below 8. The passive layer may be compromised if this reaction spreads to the steel reinforcement, leaving the steel vulnerable to possible corrosion.

Moreover, multiple studies have demonstrated that GGBS replacement in the binder of concrete will not decrease the strength of concrete [27]. Some of the research gaps found in the past literature are:

The carbon emission from the cement sector in India is considered a major source of a sustainable substitute to Portland cement. Materials activated by alkali solutions, including fly ash and GGBS are being investigated by Pandit et al. [28]. These materials have a good chemical composition, are easily accessible, and can be used to repurpose waste effectively. However, occasionally, it might impact the carbonation's passive layer and reduce its passivity.

The carbonation performance of Geopolymer Concrete (GPC), a sustainable green material for civil engineering that is comparable to Portland Cement (OPC), is introduced by Zhao et al. [29]. The merits, mechanism of the carbonation technique, and the evaluation techniques of GPC are presented, along with the correlation of the carbonation characteristics of the GPC with OPC, as well as an analysis of the factors influencing GPC's performance. Additional numerical models are recommended to increase the robustness and lifespan of GPC carbonation. On the other hand, it has an impact on the carbon dioxide content and the cement-to-water ratio

Sun et al. [30] examine replacing regular Portland cement (OPC) with a Slag of carbide (SC) as a catalyst in the excess-sulfate-based phospho-calcium sulfate dihydrate slag cement (EPSC). According to the results, increased SC doping causes low latest strength and an unstable hydration rate, while lower SC dosage lengthens the induction period and improves the final strength. In EPSC applications, a pH range of 11.00–11.75 provides the best hydration conditions for SC, guaranteeing both stability and effectiveness of hydration. On the other hand, air pressure and relative humidity following carbonation may result in certain uncommon situations.

The energy usage and emission of carbon dioxide in cement production led to environmental pollution. Two supplemental materials with cement characteristics (SMC) are added to the concrete to lower energy consumption fly ash (FA) and silica fume (SF). Using FA and SF as SMCs, Padavala et al. [31] focus on creating sustainable concrete in this work. Cement was used to replace FA and SF in two different mixes that were created. Other areas of study included sustainability index, mechanical characteristics, microstructure analysis, and cost-benefit analysis. Nevertheless, there are cases where the concrete cover reinforcement raises the possibility of corrosion.

Thorne et al. [32] investigate how low-carbon concrete with additional cementitious materials (SCMs) affects the environment. According to the results, adding fly ash admixture and limestone lowered carbon emissions but did not affect mechanical qualities. The study offers guidance for creating concrete materials that take into account the availability of nearly carbon-free concrete ingredients on a worldwide scale. Though concrete naturally has a high compressive strength, it is more expensive.

The research gaps that need to be addressed are poor compressive strength, poor bulk density, and other poor mechanical properties exposed in Table 1. This present study needed to use different types of municipal solid wastes and agricultural ashes to replace regular Portland cement, either entirely or partially, which has been relatively minimal and primarily limited to GGBS and pulverized fly ash PFA. Due to the negative environmental effects

Table 1. Summary of literature review

Authors	Methods	findings	Research gaps
Pandit et al. [28]	GGBS with alkali treatment	The compressive strength of GPC is decreased by 5%, and the slag blend was expanded by 0.0963%	Passivity was decreased due to the impact of carbonation.
Zhao et al. [29]	GPC-anti carbonization	Carbonization PH11.23for 28 days with sodium concentration 3%	The carbon emission was too high
Sun et al. [30]	sulfate phosphogypsum slag cement	The hydration environment measured 11 to 11.75PH value, the attained suitable hydration environment 1.3% carbide slag	Durability strength is too low.
Padavala et al. [31]	supplementary cementitious materials with SF	The compressive strength 50Mpa recorded for 7.5% SF and 30% fly ash	Flexural strength is too low
Thorne et al. [32]	supplementary cementitious materials	Higher durability strength by 20.5% and reduced $C0_2$ 21.1%.	Chloride permeability was high.

of burning coal and producing steel, it will soon be more difficult to find GGBS and PFA. GGBS causes concrete to become more carbonated, but at a faster rate as its content increases. This phenomenon is comparable to the fly ash effect. Compared to Portland cement (PC) concrete, this effect is stronger for concrete designed on an equal water/cement (w/c) ratio. The fast phase-out of fossil fuels in coal-fired power plants and steel manufacture and the serious environmental effects accompanying it are making it harder to find GGBS and FA these days. This study's importance is to replace OPC entirely or partially (using alkali activation technology) with other SCMs, such as ashes from volcanoes, woods, or agricultural ashes, as well as municipal solid wastes like glass trash and burned ashes.

The originality and novelty of this study is implementing the hybrid optimization model based on termite and taguchi for finding the optimal mix ratios that give desired mechanical properties in less curing time. The steps to proceed with the novelty solution are described below,

- Initially, the material parameters are defined in the orthogonal array of the system.
- The material parameters are assigned to the novel Termite-based Taguchi Optimization (TbTO) based on the major, minor, and intermediate conditions.
- The novel approach is intended to use finite optimization characteristics and prediction features to enhance quality without altering the quantity.
- Finally, the prediction function generates the key metrics in the optimized concrete, which are analyzed as compression strength, carbonation depth, permeability, and acid attack.
- To validate the study, the key metrics of the optimized concrete are compared with recent concrete models.

The later part of this research is discussed as follows: The recent works on concrete with their problem statements are mentioned in the first section. A novel proposed methodology is mentioned in the 2nd section. The outcomes and their discussions of the suggested model and comparisons are represented in the 3rd section. The 4th section concludes the study with subsequent works.

MATERIALS AND METHODS

The present work aims to implement a novel termite-based Taguchi optimization (TbTO) to optimize the concrete mix. The main aim of adopting this technique is to improve the concrete quality without altering the quantity. Initially, the material parameters of the concrete are assigned to the perpendicular array of the system. After that, the material parameters based on major, minor, and intermediate terms are assigned in the novel approach to determine the optimized mix proportion in the concrete. Finally, the key metrics in the carbonation concrete are determined and correlated with recent concrete specimens, such as Fly Ash Concrete (FAC), Ordinary Portland Cement Concrete (OPCC) [33], and Residual Masonry concrete (RMC) [34]. The carbonation depth of the carbonation concrete is correlated with Carbonation concrete (CC) [35],

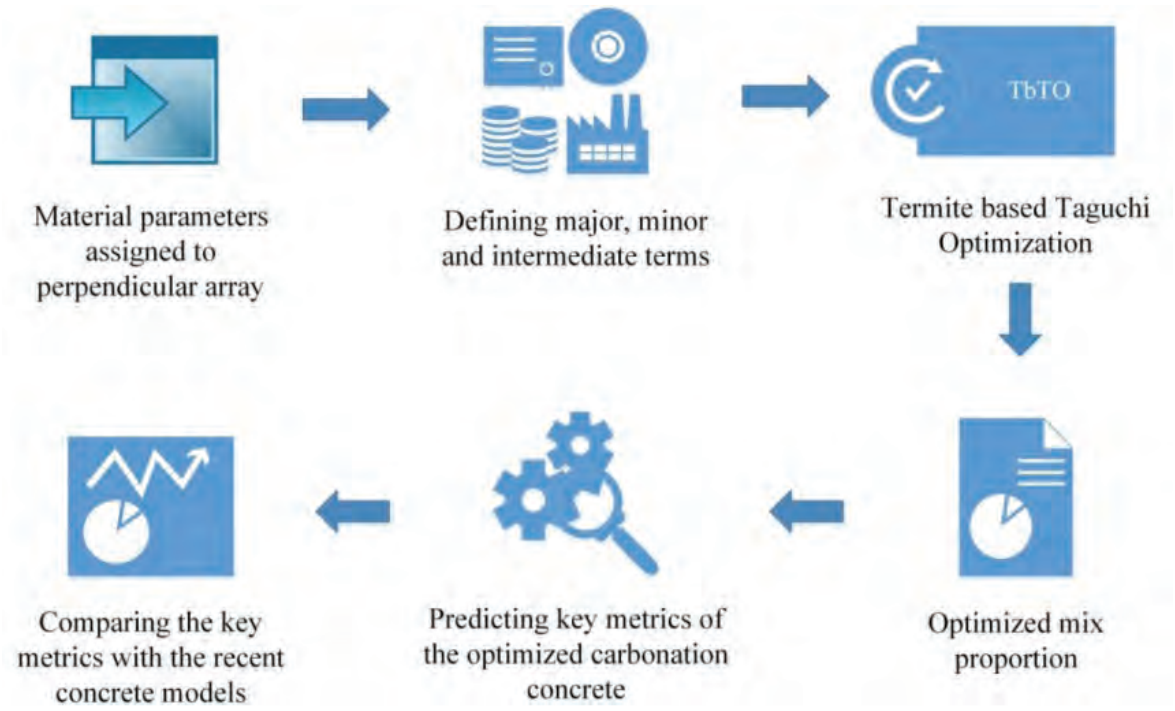


Figure 1. Proposed methodology.

Self Compacting Concrete by Masonry Residue (SCC-MR) [36], and Geopolymer Concrete (GPC) [37]. The chloride permeability of the carbonation concrete is correlated with Geopolymer concrete (GPC), Ordinary Portland Cement Concrete (OPCC) [37], and Fine Rice Husk Ash Concrete (FRHAC) [38]. The water Sorptivity of the carbonation concrete is correlated with Geopolymer Concrete (GPC) [34], Ordinary Portland Cement Concrete (OPCC), Fly Ash Concrete (FAC) and Silica Fume Concrete (SFC) [39]. The acid attack of the carbonation concrete is correlated with Geopolymer concrete (GPC) [40], Fly Ash Concrete (FAC), and Hybrid Fibre Concrete (HFC) [41]. All models were tested with the same proposed mix concentration from the same MATLAB software, and the results were compared. The reason for selecting these components is, which is the latest optimal technology executed by the different researchers.

The proposed methodology is illustrated in Figure 1. Initially, the material parameters of the concrete are assigned to the perpendicular array. Afterward, based on the major, minor, and intermediate input terms, these values are assigned to the novel TbTO. The output gained will be of optimized mix proportion. The key metrics of the concrete in terms of compression strength, carbonation depth, chloride permeability, sorptivity, and acid attack are determined. These key metrics are correlated with recent concrete models for validating the study.

Materials

Grade 40 ordinary Portland cement is implemented as the binder. 30% is the standard consistency of the cement used. 123 minutes is the initial setting time of the cement used, and 270 minutes is the final setting time of cement. 3.1 is the specific gravity of the cement, and 44.5 MPa is the compression strength of the cement used. Natural sand is considered fine aggregate, and its size is less than 4.75mm. 2.64 is the specific gravity of the cement. 0.87% is the water absorption capacity of the fine aggregate. 3.02 is the fineness modulus of the sand used. 10mm is the maximum size of the coarse aggregate, and 3.06 is the specific gravity of the coarse aggregate. 2.04% is the water-absorbing capacity of the coarse aggregate. 2.64 is the modulus of fineness in the coarse aggregate.

GGBS is a byproduct obtained from the steel and iron industry. GGBS reacts with water and cement during hydration and forms additional C-S-H gel and $Ca(OH)_2$. This additional chemical compounds increase the strength of concrete. For these advantages, GGBS is considered for this research. 3.1 is the specific gravity of the GGBS used in this research, and the fineness modulus of the GGBS is 12000 cm^2/g, 680 kg/m^3 is the bulk density of GGBS, and 27 sec is the marsh cone flow value of GGBS. The GGBS particle distribution of d_{10}, $d_{50,}$ and d_{90} are 1.5μm, 4.2 μm and 9μm. Silica fume also enhances the strength of concrete by forming $Ca(OH)_2$ and C-S-H gel. Silica fume is a byproduct obtained from the quartz manufacturing process. The residue of silica fume in the 90μm sieve is 4%.

Mix Proportions

The concrete grade executed in the work was M30 as per IS10262-2019. 0%, 20%, 40%, and 60% by weight of cement, the GGBS is replaced partially. The compression strength increased at the silica fume of 15% and decreased at the 20% replacement level. At this point, free Ca(OH)2 and silica utilization are at an optimum level [42]. Hence, 15% of silica fume is considered the optimum dosage level. Already, 15% SF has been researched, but it has reported dry density and poor compression strength [43]. To address these issues 15% of SF is considered for this study. The mix parameters and their values are represented in Table 2.

The sustainable concrete materials are then replaced in place of cement, and the mix proportions of the M30 concrete grade in terms of cement, water, silica fume, GGBS, fine aggregate, and coarse aggregate are represented in Table 3.

Table 2. Mix parameters and their values as per IS 10262 2019

Mix parameters	Values
Target strength	38.25 MPa
Water by cement ratio	0.43
Water content	169.77 kg/m^3
Cement content	394.813 kg/m^3
Air voids	0.015
Aggregate percentage in consideration with overall concrete	68.633%
Coarse aggregate percentage in consideration with overall aggregate	62%
Fine aggregate percentage in consideration with overall aggregate	38%
Coarse aggregate content	1302.105 kg/m^3
Fine aggregate content	688.526 kg/m^3
Mix proportion	1:1.74:3.3

Table 3. Mix proportions of M30 concrete mixes

Mix proportions	Water (liter/m³)	Cement (kg/m³)	GGBS (kg/m³)	Silica fume (kg/m³)	Fine aggregate (kg/m³)	Coarse aggregate (kg/m³)
0% SF, 0% GGBS	169.77	394.813	0	0	688.526	1302.105
15% SF, 20% GGBS	169.77	256.633	78.96	59.22	688.526	1302.105
15% SF, 40% GGBS	169.77	177.633	157.93	59.22	688.526	1302.105
15% SF, 60% GGBS	169.77	98.703	236.89	59.22	688.526	1302.105

Mixing Preparation and Curing

Based on the mix proportion, the samples are weight-batched, and the materials are placed in the tray. Initially, the coarse aggregate is placed, followed by the fine aggregate, then the fines are placed above the aggregates and the materials undergo dry mixing for 3 minutes. After that, water is added to the dry mix, and the wet mix is carried out. Once the concrete attains a uniform color, the fresh concrete is subjected to a slump cone test. On reaching the true slump, the concrete is placed in the molds with consideration to compaction in three layers. The molds are demoulded after 24 hours, and the specimens are placed in water curing for the particular days of curing. The chemical composition is defined in Table 4, physical composition is exposed in Table 5, and particle size distribution is in Table 6.

Termite based Taguchi Optimization

The Termite-based Taguchi optimization is designed with the functions of termite optimization and taguchi method. This dual model optimization is intended to improve the material quality without altering quantity. Hence, it is a novel approach to concrete technology. First, the material and strength parameters of the carbonation concrete from the experimental data are defined in the perpendicular array of the proposed approach, as expressed in Eqn. (1).

Table 4. Chemical composition

Material	SiO_2	Al_2O_3	Fe2O3	CaO	MgO	SO_3	Na_2O	K_2O
Cement	20.67	4.86	2.87	60.69	1.6	2.34	0.27	0.54
GGBS	38.94	11.16	1.94	30.75	7.65	-	1.4	0.26
Silica Fume	90.87	0.41	1.12	0.35	0.57	0.13	0.42	1.21

Table 5. Physical composition

Material	Specific gravity	Bulk density	Particle size	Surface area (BET)
Cement	3.12	1440 kg/m³	<90μm	232 m²/kg
GGBS	2.74	1260 kg/m³	<10μm	600 m²/kg
Silica fume	2.24	720 kg/m³	<1μm	12500 m²/kg

Table 6. Particle size distribution & physical properties of aggregates

Property	Fine aggregate	Coarse aggregate
Fineness modulus	3.24	8.4
Specific gravity	2.62	2.67
Absorption of water	1.2%	0.8%
Shape of particles	Rounded	Angular
Crushing value	-	19.4
Impact value	-	12.14

$$P_a = \begin{cases} a_1, b_1, c_1, \ldots\ldots \\ a_2, b_2, c_2, \ldots\ldots \\ a_3, b_3, c_3, \ldots\ldots \\ a_m, b_m, c_m, \ldots\ldots \end{cases} \quad (1)$$

Where a_1, a_2, a_3, ... defines the parameter of concrete in the same category and a_1, b_1, c_1, ... defines the different parameters perpendicularly. Afterward, every parameter in the perpendicular array is assigned to the indication to dissonance ratio (I / D) for determining the minor, major, and intermediate terms shown in Eqn. (2), Eqn. (3), and Eqn. (4).

$$I / D_{\min or} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} P^2\right) \quad (2)$$

$$I / D_{major} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} \frac{1}{P^2}\right) \quad (3)$$

$$I / D_{\text{int}ermediate} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} (P-R)^2\right) \quad (4)$$

Where P defines the enhanced response in the system for the particular parameter and R represents the average response of the specified parameter. The I / D for each perpendicular array parameter is taken as an input in the optimization by the fitness function based on the quality characteristics. Major quality is considered for strength parameters, and minor quality is considered for voids in the concrete. Based on such conditions, an optimal mix proportion of concrete is obtained in the proposed approach. Once the mix proportion is obtained, the concrete mix proportion is trained in the carbon curing process for a definite period by the fitness function. The optimization's prediction function is initiated to determine the durability parameters as represented in Eqn. (5).

$$D_p = \begin{cases} [C_s]_{1,3,28,56} \\ [C_p, W_s]_{3,28} \\ [C_d, A_a]_{28,56} \end{cases} \quad (5)$$

Where C_s are the compression strength and carbonation depth of the concrete at 1, 3, 28, and 56 days, C_p, W_s are the chloride permeability and water sorptivity at 3 and 28 days. A_a, C_d is the acid attack residual strength at 28 and 56 days.

Algorithm 1: TbTO

Start
{
initialization()
{
int P_a;
//assigning the material parameters

$$P_a = \begin{cases} a_1, b_1, c_1, \ldots\ldots \\ a_2, b_2, c_2, \ldots\ldots \\ a_3, b_3, c_3, \ldots\ldots \\ a_m, b_m, c_m, \ldots\ldots \end{cases}$$

}
quality parameters()
{
int $I / D_{\min or}$, I / D_{major}, $I / D_{\text{int}ermediate}$;
//fixing the enhanced response for material parameter

$$I / D_{\min or} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} P^2\right)$$

$$I / D_{major} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} \frac{1}{P^2}\right)$$

$$I / D_{\text{int}ermediate} = -10.1 \log_{10}\left(\frac{1}{m}\sum_{y=1}^{m} (P-R)^2\right)$$

//determining I/D
}
testing()
{
int D_p;
//initializing durability parameters

$$D_p = \begin{cases} [C_s]_{1,3,28,56} \\ [C_p, W_s]_{3,28} \\ [C_d, A_a]_{28,56} \end{cases}$$

//finalizing parameters
}
}
End

Initially, the material parameters of the carbonation concrete are assigned to the perpendicular array. Afterward, based on the quality of parameters, they are defined according to the proposed approach in major, minor, and intermediate responses. From the proposed approach the optimized concrete mix proportion is obtained. Subsequently, the carbon curing is carried out in the samples. The key metrics of the concrete in terms of compression strength, carbonation depth, chloride permeability, sorptivity, and acid attack are determined and correlated with the recent concrete models for validating the study. Precisely, by implementing the above steps, accurate outcomes can be obtained.

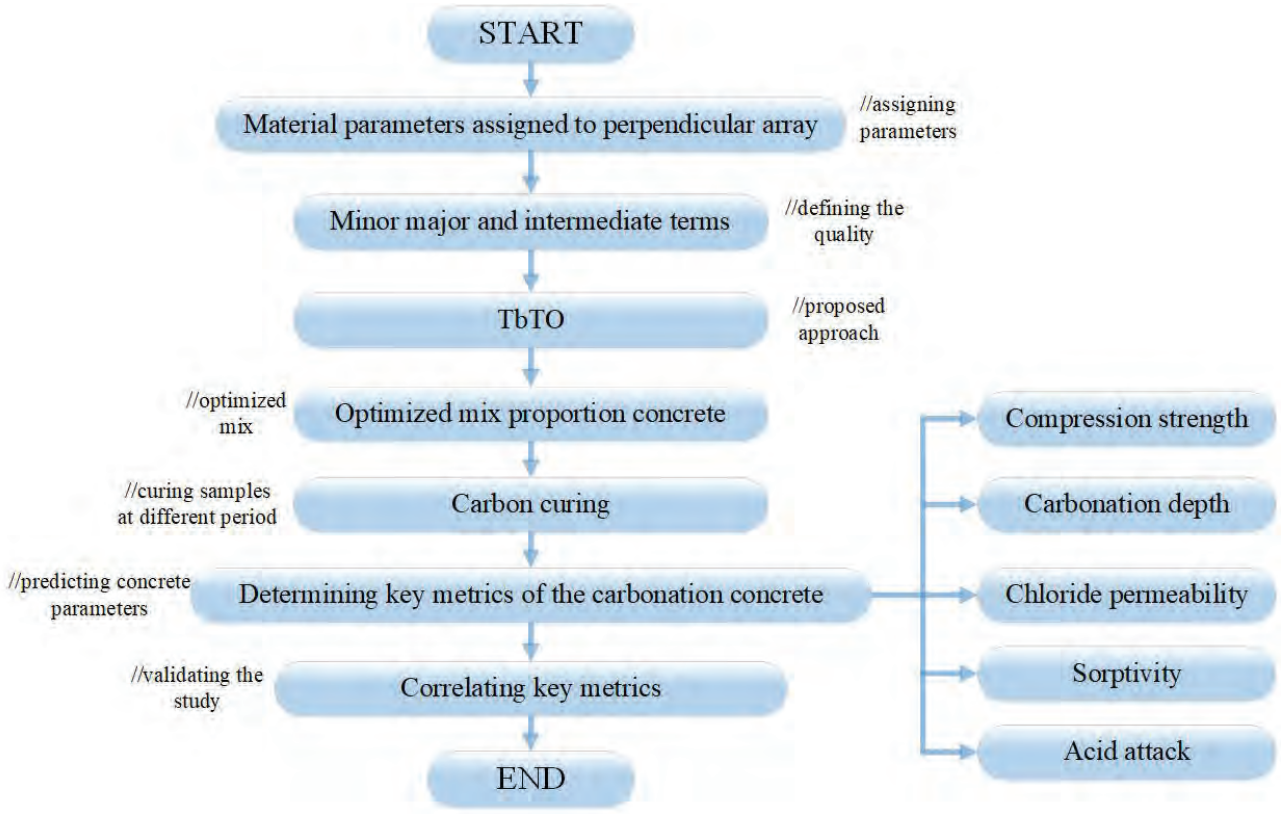


Figure 2. Flow process of the proposed approach.

RESULTS AND DISCUSSION

The proposed technique is executed in the Windows 10 Operating System. The process is carried out in MATLAB R2023a Windows software. The Matrix Laboratory is the full form of MATLAB, computer software for data analysis, algorithm development, and models and applications. It has multi-layers of programming that have operational and functional direction points. It is generally utilized by mathematicians and research scholars to develop complex models in programming. This software is also employed to predict numerical data and perform graphical visualizations. Numerical computations, intelligent data learning and prediction, and data processing in academic studies are the applications of this software. The TbTO approach is developed in MATLAB to determine the optimal mix proportion from the material properties. Thereafter, the optimal mix proportion is trained in the carbon curing process, and the durability parameters are predicted and correlated with the recent concrete studies.

Case Study

To validate the proposed approach, the tests on the carbonation concrete are carried out, and the outcomes are illustrated in order. The tests include compression strength, carbonation depth, chloride permeability, water sorptivity, and acid attack. The specimen size and curing period will also vary for each test. The curing is carried out only by the carbon curing process. The compression strength of the sample is 10×10×10 cm, and the testing is carried out for 1,3,28 and 56 days. The carbonation depth is calculated on 28 and 56 days. The water sorptivity is determined in the concrete specimen with a length of 50 mm and a diameter of 100mm. The chloride permeability in the concrete specimen has dimensions of 50 mm length and 100 mm diameter. The chloride permeability and the water sorptivity tests are carried out in 3 days and 28 days of carbon curing. The acid attack was carried out in a sample size of 10×10×10 cm, and the testing was carried out on 28^{th} and 56^{th} day. The procedure for testing and their results are represented in sequence and order.

The precision in the proposed approach was defined by correlating the outcomes with the recent concrete studies, which included compression strength, carbonation depth, water sorptivity, chloride permeability, and acid attack. For comparison, the carbonation depth, acid attack, and compression strength of the 28^{th} and 56^{th} day results are taken, and for chloride permeability and water sorptivity, 28^{th} day results are considered. The carbonation concrete with 15% silica fume and 40% GGBS has shown better compression strength and acid attack results. The mix with a higher amount of admixture has lower permeability, that is the mix with 15% silica fume and 60% GGBS. This mix proportion

possesses lower carbonation depth, water sorptivity, and chloride permeability.

The mix containing 15% silica fume and 60% GGBS is considered for its superior strength behavior and optimum permeability. The carbonation concrete by the proposed approach is carried out using MATLAB R2021 software with precise specimen sizes and carbon curing conditions to obtain accurate outcomes. Thereafter the outcomes are compared with recent concrete investigations for validating the research. The concrete references used in the comparisons correlate with their material composition, as shown in Table 7.

Compression Strength

The compression strength of concrete on the 28th day defines the grade of concrete. It is expressed in load per area. The ability of the specified dimension cube based on the size of coarse aggregate to resist the compression load without yielding is termed the compression strength of the cube. The compression strength tests are determined in cubes of size 100 × 100 × 100 mm, and the tests are carried out in 1, 3, 28, and 56 days. The load is considered to be applied on the cube's right angle to the concrete placing side as per BIS 516-1959. The rate of loading is 2.4 MPa per second. The compression strength of the carbonation concrete at 1,3,28 and 56 days of curing are represented in Figure 3.

The compression value of the carbonation concrete at 1 day of testing was 5 MPa for the conventional mix. The silica fume is fixed for the mixes at 15%. The mix containing 20% GGBS has a compression value of 5.2 MPa, 40% GGBS has a compression strength of 5.7 MPa, and 60% has a compressive strength of 5.4 MPa. On 3rd day of testing, the conventional mix had a compression value of 10 MPa, the mix having 20% GGBS had a compression value of 10.3 MPa, the mix having 40% GGBS had a compressive value of 11.11 MPa, and the mix containing 60% GGBS has a compressive strength of 11 MPa.

On the 28th day of testing the conventional mix had a compression strength value of 40 MPa, the mix having 20%

Table 7. Comparison of different types of concrete with their material composition

Concrete	Cement (Kg/m^3)	Water (l/m^3)	Fine aggregate (Kg/m^3)	Coarse aggregate (Kg/m^3)	Sustainable material	Replacement material	Replacement material (Kg/m^3)
Fly Ash Concrete (FAC) [33]	286	165	1007	883	Fly ash	cement	72
Ordinary Portland Cement Concrete (OPCC) [33]	358	165	1007	883	-	-	-
Residual Masonry concrete (RMC) [34]	375	202.5	895.11	686.21	Masonry residue	cement	106.4
Self-Compacting Concrete by Masonry Residue (SCC-MR) [34]	375	202.5	895.11	686.21	Masonry residue	cement	106.4
Carbonation concrete (CC) [35]	330	198	719	813	Carbonation curing	-	-
Geopolymer Concrete (GPC) [37]	-	263 (Alkali activate slag)	685	1058	GGBS and MBS	cement	364.5 and 40.5
Fine Rice Husk Ash Concrete (FRHAC) [38]	40	180 (Alkali activate slag)	576	1070	Fine rice husk ash	cement	360
Ordinary Portland Cement Concrete (OPCC) [39]	550	192.5	989.5	662.5	-	-	-
Fly Ash Concrete (FAC) [39]	385	192.5	922.5	615	Fly ash	cement	165
Silica Fume Concrete (SFC) [39]	495	192.5	975	652.5	Silica fume	cement	55
Geopolymer concrete (GPC) [40]	-	141.75	683	1269	Fly ash and GGBS	cement	303.75 and 101.25

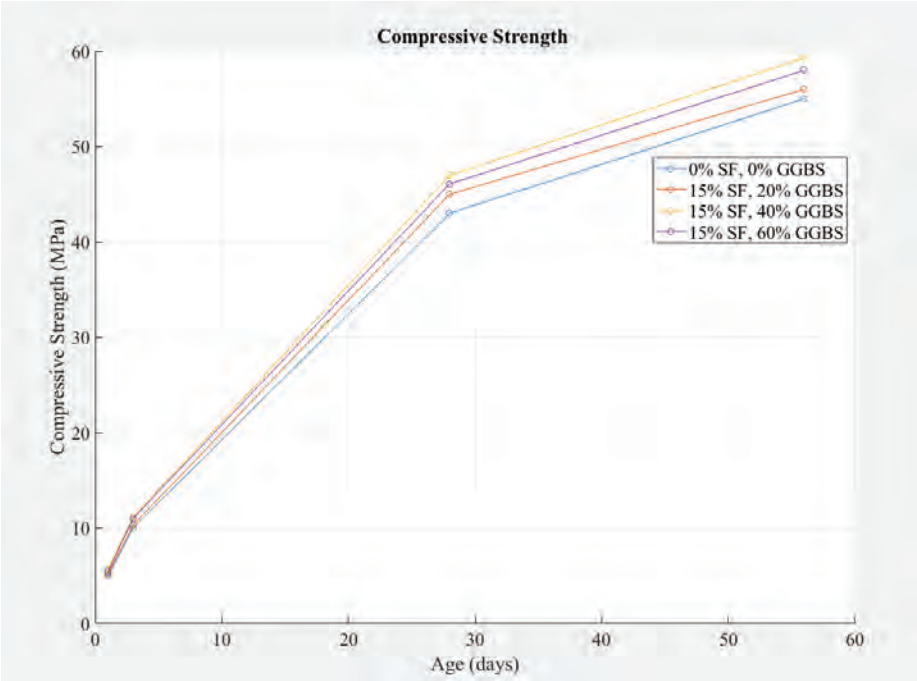


Figure 3. Compression strength of carbonation concrete at 1,3,28 and 56 days of testing.

Table 8. Compression strength of carbonation concrete at 1, 3, 28, and 56 days of testing

Proportions	Compressive strength (MPa)				Error bar
	1 day	3 days	28 days	56 days	±4
0% SF 0% GGBS	5	10	43	55	±6
15% SF 20% GGBS	5.2	10.3	45	56	±3
15% SF 40% GGBS	5.7	11.11	47	59.31	±1
15% SF 60% GGBS	5.4	11	46.09	58	±2

GGBS had a compression strength value of 45 MPa, the mix containing 40% GGBS had a compression strength value of 47 MPa, and the mix containing 60% GGBS has compression strength of 46.09 MPa. On the 56th day in the compression strength testing, the conventional concrete had a compressive strength value of 55 MPa, the mix containing 20% GGBS had a compression strength value of 56 MPa, the mix having 40% GGBS had a compressive strength value of 59.31 MPa and the mix having 60% GGBS has compressive strength value of 58 MPa. The compression strength of the carbonation concrete at 1, 3, 28, and 56 days, along with various mix ratios, are represented in Table 8. Compared with all these tests, the mix containing 15% silica fume (SF) and 40% GGBS has higher compression strength than all the other mix proportions.

The strength of compression in the concrete was represented in MPa. The compression strength in the carbonation concrete on the 28th day and 56th day are compared with FAC, OPCC [33], and RMC [34]. On the 28th day, the strength of compression in the FAC is 32 MPa, the OPCC has the strength of compression of 37 MPa, and the RMC has the strength of compression of 38 MPa.

The proposed approach carbonation concrete has compression strength of 47 MPa at the 28th-day test. The graphical representation of the compression strength in the carbonation concrete on the 28th-day and 56th-day results is represented in graphical format in Figure 4 and Table 9. At the 56th day test, the FAC's compression strength is 48 MPa, the OPCC has the strength of compression as 52 MPa [33], and the RMC has a compressive strength of 47 MPa [34].

The proposed approach to carbonation concrete has compression strength of 59.31 MPa. The compression strength outcomes show that carbon curing concrete possesses higher strength than admixture replacement concretes and ordinary concrete.

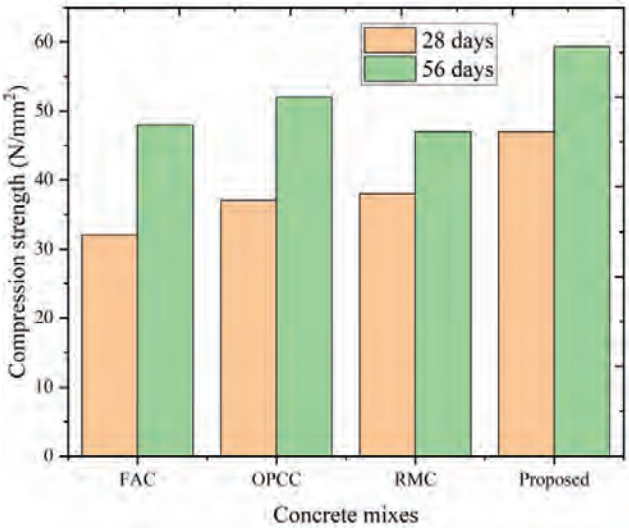


Figure 4. Comparison of compression strength in the proposed concrete with the recent concrete models.

Carbonation Depth

The CO_2 from the environment reacts with the hydration particles of the concrete and forms carbonate, leading to steel bar corrosion. Thus, the strength properties of concrete structures are altered. The depth at which the carbon particles penetrate inside the concrete member from the extreme fiber of concrete is termed carbonation depth. The depth of carbonation in the carbon-cured concrete is represented in graphical format in Figure 5.

On the 28th day, the depth of carbonation in the concrete is 4mm for the conventional mix, the mix with 20% GGBS has a carbonation depth of 5mm, the mix with 40% GGBS has a carbonation depth of 4mm, and the 60% GGBS mix has carbonation depth of 3mm. The statics of the carbonation depth of the carbonation concrete at the 28 days and 56 days of testing are illustrated in Table 10 and Figure 5.

On the 56th day of testing the conventional concrete had a depth of carbonation of 8mm, the concrete having 20% GGBS had a depth of carbonation of 9mm, the concrete

Table 9. Comparison of compression strength in the proposed concrete along with the recent concrete models

Testing day	Compression strength (MPa)			
	FAC	OPCC	RMC	Proposed
28th day	32	37	38	47
56th day	48	52	47	59.31

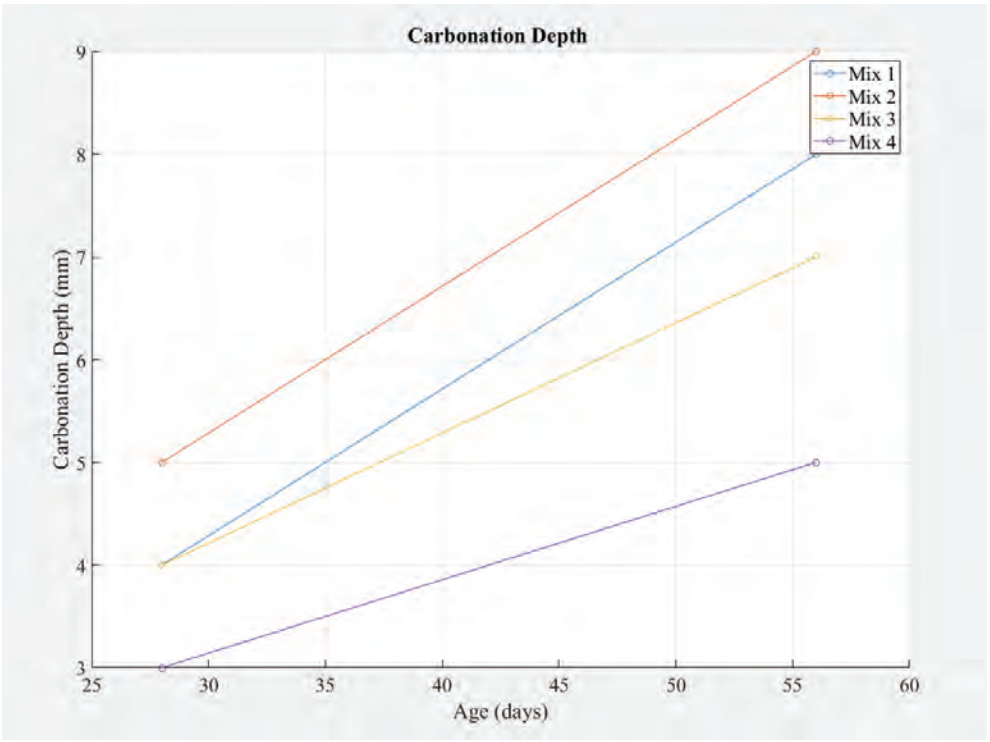


Figure 5. Carbonation depth of carbonation concrete at 28 and 56 days of testing.

Table 10. Carbonation depth of carbonated concrete at 28 and 56 days of testing

Proportions	Carbonation depth (mm)	
	28 days	56 days
0% SF 0% GGBS	4	8
15% SF 20% GGBS	5	9
15% SF 40% GGBS	4	7
15% SF 60% GGBS	3	5

having 40% GGBS had a depth of carbonation of 7mm, and the concrete having 60% GGBS has carbonation depth of 5mm. The mix containing 15% SF and 60% GGBS has a lower carbonation depth than the other mix proportions. The mineral admixture silica fume and GGBS will minimize the permeability in concrete, resulting in reduced carbonation depth in the admixture concrete.

The depth of carbonation in the concrete was represented in mm. The depth of carbonation in the proposed concrete was correlated with CC [35], SCC-MR [36], and GPC [37]. The carbonation depth in the suggested model is compared with the conventional carbonation concrete, eco-friendly self-compaction concrete, because the residual wastes in the environment are reduced in the form of concrete, and the full cement is replaced by admixture concrete, which is the GPC. The comparison statics of the carbonation depth of various concrete mixes with the suggested concrete are represented in Figure 6 and Table 11.

On the 28th day, the depth of carbonation in the existing carbonation concrete was 16.5 mm, the SCC-MR had a carbonation depth of 5 mm, and the GPC had a carbonation depth of 8.55 mm. The proposed carbonation concrete has a carbonation depth of 4mm.

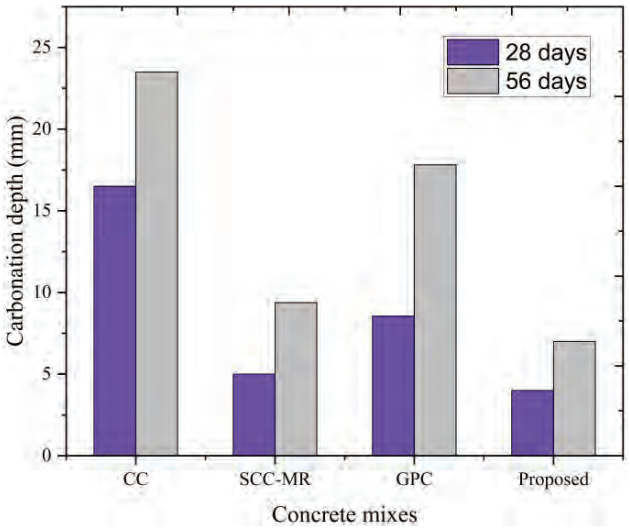


Figure 6. Comparison of carbonation depth of the proposed concrete with the recent concrete models.

Chloride Permeability

The permeability of chloride in the concrete defines the durability of the concrete at various exposure conditions. The ingression of chloride ions in concrete directly impacts its endurance. The chloride penetration in the concrete is expressed in terms of electrical conductance, which is concrete resistance to chloride ion penetration. The specimen size for the permeability test in chloride is a diameter of 100 mm and a height of 50 mm cylindrical specimen. The samples were tested on 3 and 28 days concerning ASTM 1202. One end of the sample is considered to be associated with the solution of sodium hydroxide, and the other end is associated with the solution of sodium chloride. The samples are tested for 6 hours. The potential difference maintained is 60V.

On 3rd day of testing the chloride permeability of the conventional concrete was 4276.28 coulombs, the mix containing GGBS of 20% had a chloride permeability of 3018.71 coulombs, the mix containing 40% GGBS had a chloride permeability of 1238.41 coulombs and the mix containing 60% GGBS has chloride permeability of 689.68 coulombs. The statistics of the chloride permeability test on the 3rd and the 28th day are represented in Table 12 and Figure 7. After curing for 28 days, the chloride permeability was reduced [44].

At the 56th day test the conventional mix has a chloride permeability of 2138.14 coulombs, the mix having 20% GGBS has a permeability by chloride of 1509.36 coulombs, the mix having 40% GGBS has a permeability of chloride as 619.21 coulombs. The mix with 60% GGBS has a chloride permeability of 344.84 coulombs. The mix containing 15% SF and 60% GGBS has lower chloride permeability than the other mix proportions because these mix proportions possess a higher admixture percentage than all the other

Table 11. Correlation in carbonation depth of the proposed concrete with the recent concrete models

Testing day	Carbonation depth (mm)			
	CC	SCC-MR	GPC	Proposed
28th day	16.5	5	8.55	4
56th day	23.5	9.38	17.82	7

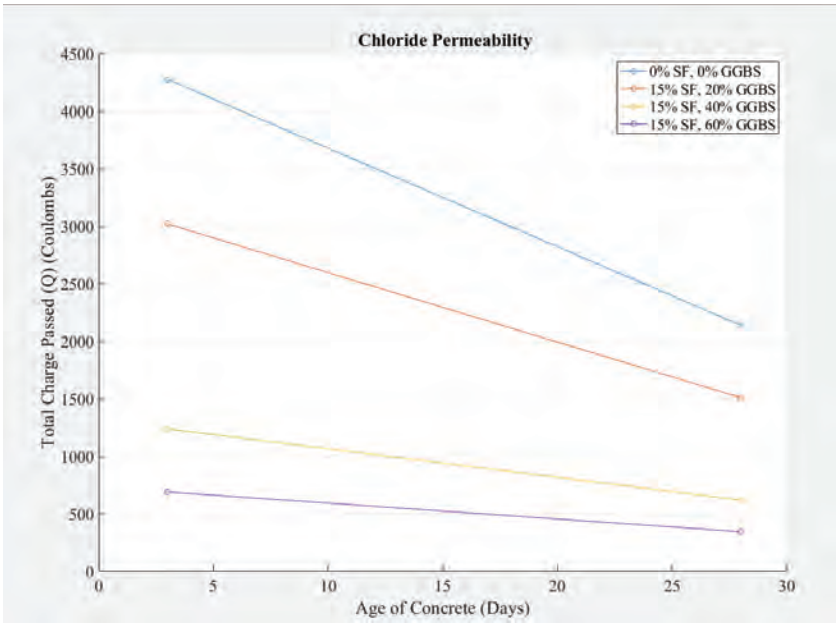


Figure 7. Chloride permeability of carbonation concrete at 3 and 28 days of testing.

Table 12. Chloride permeability of carbonation concrete at 3 and 28 days of testing

Proportions	Chloride permeability (Coulombs)	
	3 days	28 days
0% SF 0% GGBS	4276.28	2138.14
15% SF 20% GGBS	3018.71	1509.36
15% SF 40% GGBS	1238.41	619.21
15% SF 60% GGBS	689.68	344.84

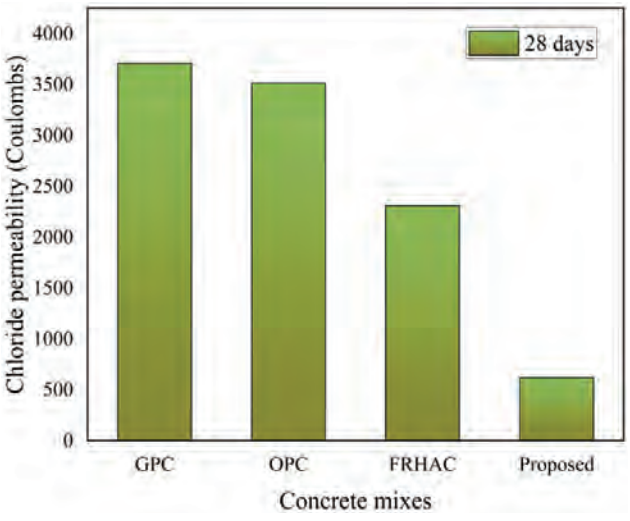


Figure 8. Comparison of chloride permeability of the proposed concrete with the recent concrete models.

mixes.All concrete mix mixtures saw a significant decrease in chloride permeability from 3 to 28 days since once the concrete cures, the cement particles keep continuing to hydrate, creating a dense microstructure using fewer overlapping pores that prevents chloride ions compared to permeating the concrete matrix.

The permeability of chloride ions in the concrete is represented in coulombs. The chloride permeability of the proposed carbonation concrete is correlated with GPC, OPCC [37], and FRHAC [38]. The OPCC is the traditional concrete, and GPC is the concrete where other admixtures replace the binder cement. The FRHAC concrete is made with fine rice husk with cement's partial replacement. The correlation of the permeability by the chloride ions of the carbonation concrete is illustrated in Figure 8 and Table 13. On the 28th day, the chloride permeability of GPC is 3712 coulombs. The OPCC has a chloride permeability of 3515 coulombs, and the FRHAC has a chloride permeability of 2314 coulombs. The proposed approach has a chloride permeability of 619.21 coulombs.

Table 13. Comparison of chloride permeability tests

Concrete	28th-day Chloride permeability (Coulombs)
GPC	3712
OPCC	3515
FRHAC	2314
Proposed	619.21

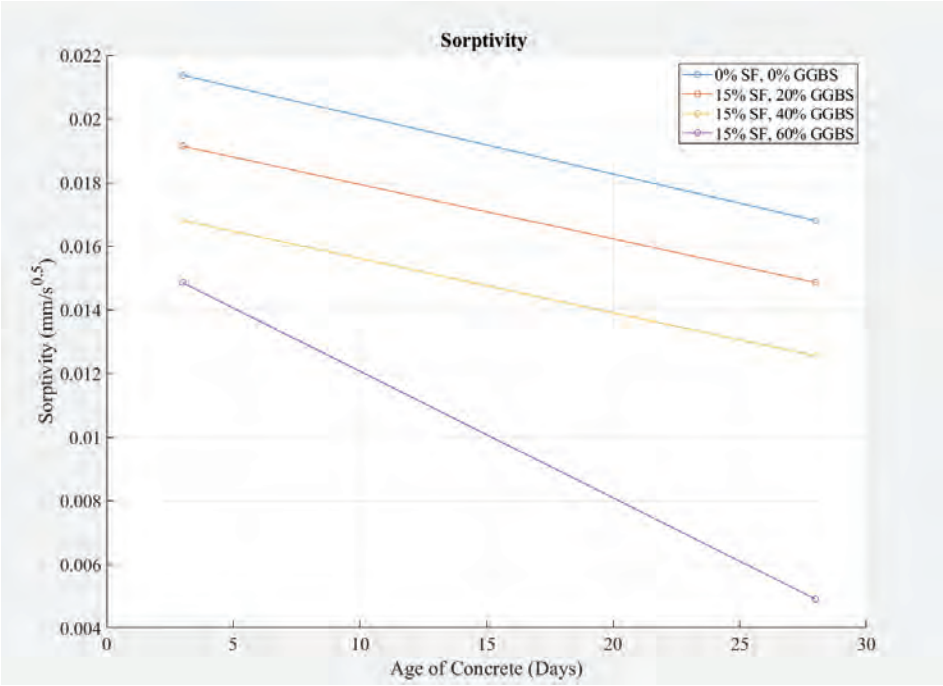


Figure 9. Water sorptivity of carbonation concrete at 3 and 28 days of testing.

Sorptivity

Sorptivity defines the concrete's ability to absorb and transmit water in the pore structure of concrete by the action of capillarity. The sorptivity test is conducted on 3 and 28 days as per ASTM C 1585 – 04. The specimen for testing is a cylinder with 100 mm diameter and 50 mm length. The curved surface in the specimen is considered to be coated with epoxy resin. The water level is kept 2-3 mm above the supporting frame. The test outcomes of the water sorptivity at 3 days and 28 days of carbon curing are represented in Figure 9 and Table 14.

On the 3rd day of testing the conventional mix had a water sorptivity of 0.0214 mm/s$^{0.5}$, the mix with 20% GGBS had a water sorptivity of 0.0191 mm/s$^{0.5}$, the mix with 40% GGBS had a water sorptivity of 0.0168 mm/s$^{0.5}$, and the mix with 60% GGBS has water sorptivity of 0.0149 mm/s$^{0.5}$. The statistics of the water sorptivity of the proposed carbonation concrete at 3 days and 28 days of testing are represented in Table 14.

Table 14. Water sorptivity of carbonation concrete at 3 and 28 days of testing

Proportions	Water sorptivity (mm/s$^{0.5}$)	
	3 days	28 days
0% SF 0% GGBS	0.0214	0.0168
15% SF 20% GGBS	0.0191	0.0149
15% SF 40% GGBS	0.0168	0.0125
15% SF 60% GGBS	0.0149	0.0049

On the 28th day, the water sorptivity of the conventional carbonation concrete is 0.0168 mm/s$^{0.5}$, the mix containing 20% GGBS has water sorptivity of 0.0149 mm/s$^{0.5}$, the mix containing 40% GGBS has water sorptivity of 0.0125 mm/s$^{0.5}$ and the mix containing 60% GGBS has water sorptivity of 0.0049 mm/s$^{0.5}$. The mix with 15% SF and 60% GGBS has lower water sorptivity than the other mixes. Due to the reason, that the mix with more amount of silica fumes and GGBS has lower water sorptivity due to the reason that rise in the admixture percentage of the concrete increases the permeability in concrete and reduces the water sorptivity.

The water sorptivity in the concrete was determined in terms of mm-sec$^{0.5.}$ The water sorptivity of the carbonation concrete is correlated with GPC [34], OPCC, SFC, and FAC [39]. The water sorptivity in the proposed concrete is correlated with the recent concrete models and is represented in Figure 10.

On the 28th day, the GPC has a water sorptivity of 0.0243 mm-sec$^{0.5}$, the OPCC has a water sorptivity of 0.0253 mm-sec$^{0.5}$, the SFC has a water sorptivity of 0.075 mm-sec$^{0.5}$ and the FAC has water sorptivity of 0.089 mm-sec$^{0.5}$. The proposed approach has a water sorptivity of 0.0125 mm-sec$^{0.5}$. The water sorptivity in the proposed concrete is correlated with the recent concrete model as represented in Table 15.

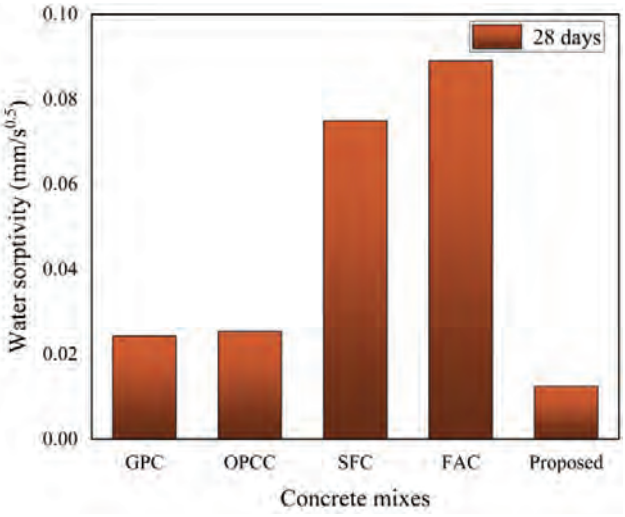


Figure 10. Correlation of water sorptivity in the proposed concrete with the recent concrete samples.

Table 15. Water sorptivity comparison of the proposed concrete with the recent models

Concrete	28th-day water sorptivity (mm/sec$^{0.5}$)
GPC	0.0243
OPCC	0.0253
SFC	0.075
FAC	0.089
Proposed	0.0125

Acid Attack

In chemically aggressive environments like sewage collection systems, concrete degradation occurs due to the presence of Hydrogen sulphide, moisture, and oxygen. Thus the design period of concrete structures is affected. An acid attack test is performed to determine the acid resistance in the concrete. The specimens in the acid attack test are of size 100 × 100 × 100 mm cube. Initially, for 10 days, the cubes are cured as usual. Afterward, the cubes are considered acid curing in a 2% sulphuric acid solution with pH maintained at 1 ± 0.1. Here, 0.1 is the fitness value adopted from the termite optimization. For the acid attack test, strength in compression of the acid-cured sample is tested at 28 and 56 days. The acid attack test outcome at 28 days and 56 days is represented in Figure 11.

Conventional carbonation concrete has a strength reduction of 9.98% on the 28th day of testing and an 18.04% strength reduction on the 56th day test. The mix containing 20% GGBS has a strength reduction of 8.73% on the 28th and 12.3% on the 56th day of testing. The mix containing 40% GGBS has a strength reduction of 6.26% on the 28th day of testing and a 10.93% strength reduction on the 56th day of testing. The mix containing 60% GGBS has a strength reduction of 7.27% on the 28th day of testing and a 14.53% strength reduction on the 56th day of testing. The statistics of the strength reduction in the acid attack test are represented in Table 16.

The mix containing 15% silica fume and 40% GGBS has a lower reduction in strength than the other models. This is because adding admixture in the proportions of concrete increases the concrete strength to an optimum level. After that, a reduction in strength occurs. Hence, compared with

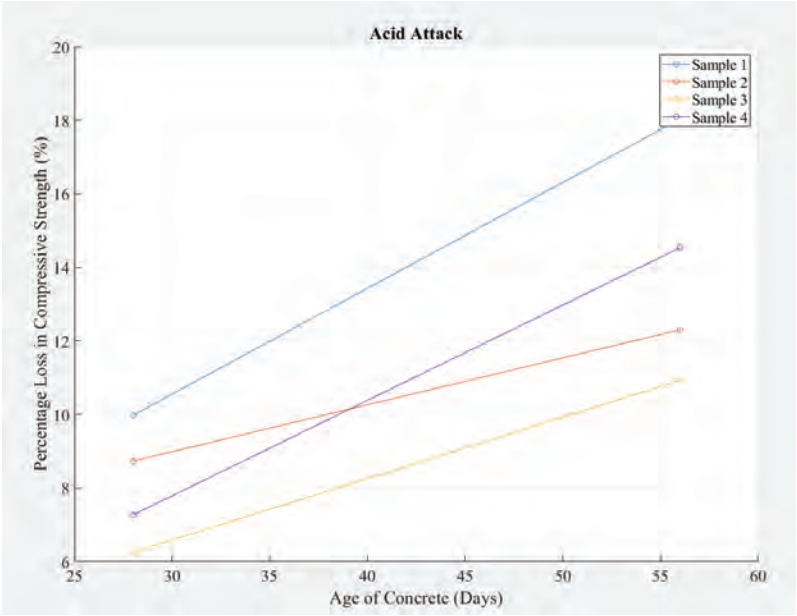


Figure 11. Percentage in strength loss of carbonation concrete at 28th, and 56th days of acid curing.

Table 16. Acid attack in carbonation concrete on the 28th, and 56th days of testing

Proportions	Reduction in strength (%)	
	28 days	56 days
0% SF 0% GGBS	9.98	18.04
15% SF 20% GGBS	8.73	12.30
15% SF 40% GGBS	6.26	10.93
15% SF 60% GGBS	7.27	14.53

all other mix proportions, the mix proportion of 15% silica fume and 60% GGBS has a lower reduction in strength in the acid attack test.

In an acid attack test, the curing is carried out by implementing a percentage of acid per amount of water. Thus, the strength in curing specimens is lowered through the acid percentage. This strength in terms of percentage is the outcome of the acid attack test. At the 28-day curing period, the GPC has a strength reduction of 7.124%, the FAC has a strength reduction of 9.51%, and the HFC has a strength reduction of 9.82%. The proposed approach has a strength reduction of 6.26%. The comparison graph of the acid attack in the 28th-day tests and the 56th-day tests is shown in Figure 12 and Table 17.

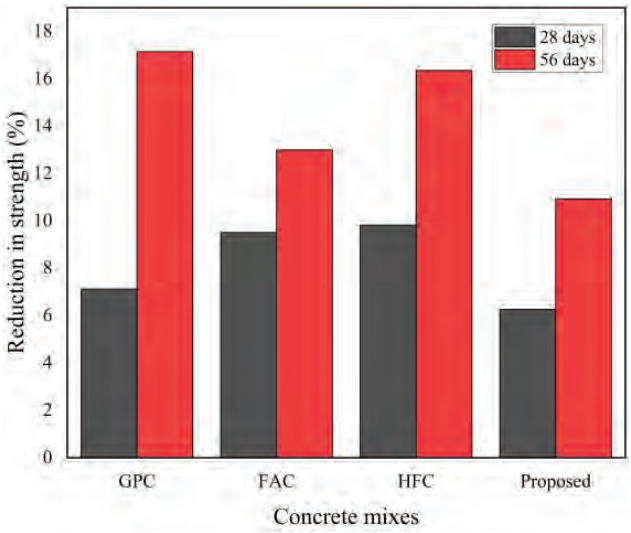


Figure 12. Comparison of the reduction in strength by acid attack of the proposed concrete with the recent concrete models.

Table 17. Comparison of acid attack test results

Testing day	Reduction in strength (%)			
	GPC	FAC	HFC	Proposed
28th day	7.124	9.51	9.82	6.26
56th day	17.14	13	16.35	10.93

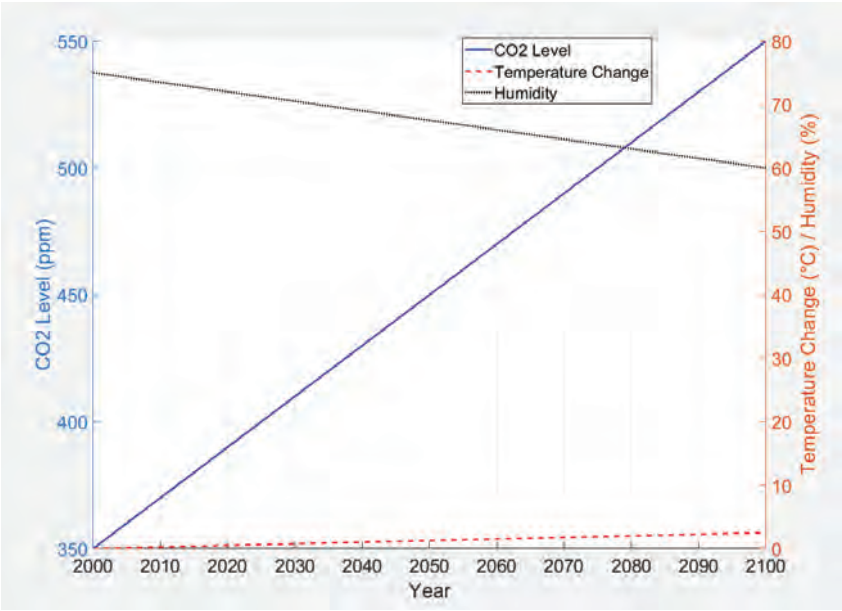


Figure 13. Environmental impact.

At the 56-day curing period, the GPC has a strength reduction of 17.14%, the FAC has a strength reduction of 13%, and the HFC has a strength reduction of 16.35%. The proposed approach has a strength reduction of 10.93%. The strength reduction is considered from the normal concrete strength.

The validated environmental impacts on three different parameters Co_2, temperature change, and humidity, are exposed in Figure 13.

CONCLUSION

This research aimed to deliver a novel Termite-based Taguchi Optimization to determine the concrete mix ratio. Initially, the material parameters in the concrete are assigned in the system. Based on the major, minor, and intermediate terms, the assigned parameters are optimized using the novel approach, and an optimal mix proportion for M30 grade concrete is obtained. Thus, mixed proportion concrete undergoes a carbonation curing training process for the required periods in definite sample sizes monitored under the fitness function. Afterward, the prediction system in the optimization is triggered to determine the durability parameters of the concrete. The durability parameters are determined regarding compression strength, carbonation depth, water sorptivity, chloride permeability, and acid attack.

- The mix containing 15% silica fume and 40% GGBS is considered the optimum mix because it possesses high compression strength and acid attack values. The carbonation depth, water sorptivity, and the permeability of chloride ions in the mix are also at the optimum level.
- The compression strength of the proposed carbonation concrete is 47 MPa on the 28th day test, which is 23.68% higher than other concrete, and on the 56th day, the compression strength is 59.31MPa, which is 14.1% higher than other compared models.
- The carbonation depth of the proposed concrete is 4 mm on the 28th day, which is 20 % lower than other compared models, and on the 56th day, the carbonation depth is 7 mm which is 25% lower than other compared models.
- The chloride permeability of the proposed concrete was 619.21 coulombs at the 28th day test, which is 73.24% lower than that of other models.
- The water sorptivity of the proposed concrete was 0.0125 mm-sec0.5 on the 28th day test, which is 48.55% lower than that of the other models.
- The acid attack of the proposed concrete showed a 6.26% strength reduction at the 28th-day test, which is 0.864% lower than other compared models. On the 56th day, the strength reduction was 10.93%, which is 2.07% lower than other compared models.

Considering the results, the proposed approach has considerably enhanced prediction ability with necessary features for building construction applications. Future studies in this section of concrete technology improve the process by including prediction accuracy, and behavior patterns of concrete, leading to the development of an accurate model in the upcoming days.

NOMENCLATURE

MPa	megapascal
sec	second
CO_2	carbon dioxide
pH	Potential of hydrogen
mm	millimeter
$Ca(OH)_2$.	Calcium hyroxide
µm	Micrometer
kg/m^3	Kilogram per meter cube
cm^2/g	Centimeter square per gram
SiO_2	Silicon dioxide
Al_2O_3	Aluminium oxide
Fe_2O_3	Ferric oxide
CaO	Calcium oxide
MgO	Magnesium oxide
Na_2O	Sodium oxide
K_2O	Potassium oxide
SO_3	Sulfur trioxide

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Song W, Yang W, Wang N, Wang C, Li Z, Ou Y, et al. How does industrial transformative process shape the spatial dynamics of urban development? Land Use Policy 2024;138:107015. [CrossRef]

[2] Ahmed I, Mehrez K, Socci C, Deriu S, Mathkur NM, Casasr IP. Disaggregated effect of construction investments on the Saudi economy: a dynamic computable general equilibrium model of Saudi Arabia. Financ Innov 2024;10:28. [CrossRef]
[3] Chen L, Hu Y, Wang R, Li X, Chen Z, Hua J. Green building practices to integrate renewable energy in the construction sector: a review. Environ Chem Lett 2024;22:751–784. [CrossRef]
[4] Santos G, Esmizadeh E, Riahinezhad M. Recycling construction, renovation, and demolition plastic waste: review of the status quo, challenges and opportunities. J Polym Environ 2024;32:479–509. [CrossRef]
[5] Ahmed O, Ahmad S, Adekunle SK. Carbon dioxide sequestration in cementitious materials: a review of techniques, material performance, and environmental impact. J CO2 Util 2024;83:102812. [CrossRef]
[6] Karadumpa CS, Pancharathi RK. Study on energy use and carbon emission from manufacturing of OPC and blended cements in India. Environ Sci Pollut Res 2024;31:5364–5383. [CrossRef]
[7] Pontikes Y, Snellings R. Cementitious binders incorporating residues. In: Handbook of recycling. Elsevier; 2024:429–444. [CrossRef]
[8] Yang Y, Lu P, Shao R, Zhao Q, Yang T, Wu C. A comprehensive review of multisource solid wastes in sustainable concrete: from material properties to engineering application. Const Build Mater 2024;435:136775. [CrossRef]
[9] Hamada HM, Al-Attar A, Abed F, Beddu S, Humada AM, Majdi A, Thomas B S. Enhancing sustainability in concrete construction: a comprehensive review of plastic waste as an aggregate material. Sustain Mater Technol 2024;40:e00877. [CrossRef]
[10] Xian X, Mahoutian M, Zhang D, Shao Y. Development of wet-cast Portland-cement-free concrete based on steel slag and ambient-pressure carbonation activation. Resour Conserv Recycl 2024;203:107455. [CrossRef]
[11] Wang J, Ren C, Huang T, Li X, Cao W, Zhu Y, Liu Z. Performances of concrete with binder and/or aggregates replacement by all-solid waste materials. J Clean Prod 2024;450:141929. [CrossRef]
[12] Niri AJ, Poelzer GA, Zhang SE, Rosenkranz J, Pettersson M, Ghorbani Y. Sustainability challenges throughout the electric vehicle battery value chain. Renew Sustain Energy Rev 2024;191:114176. [CrossRef]
[13] Maaze MR, Shrivastava S. Design optimization of a recycled concrete waste-based brick through alkali activation using Box-Behnken design methodology. J Build Eng 2023;75:106863. [CrossRef]
[14] Uratani JM, Griffiths S. A forward looking perspective on the cement and concrete industry: implications of growth and development in the Global South. Energy Res Soc Sci 2023;97:102972. [CrossRef]
[15] Hanifa M, Agarwal R, Sharma U, Thapliyal PC, Singh LP. A review on CO2 capture and sequestration in the construction industry: emerging approaches and commercialised technologies. J CO2 Util 2023;67:102292. [CrossRef]
[16] Schneider M, Hoenig V, Ruppert J, Rickert J. The cement plant of tomorrow. Cem Concr Res 2023;173:107290. [CrossRef]
[17] Knight KA, Cunningham PR, Miller SA. Optimizing supplementary cementitious material replacement to minimize the environmental impacts of concrete. Cem Concr Compos 2023;139:105049. [CrossRef]
[18] Singh RP, Vanapalli KR, Cheela VRS, Peddireddy SR, Sharma HB, Mohanty B. Fly ash, GGBS, and silica fume based geopolymer concrete with recycled aggregates: properties and environmental impacts. Const Build Mater 2023;378:131168. [CrossRef]
[19] Kanagaraj B, Lubloy E, Anand N, Hlavicka V, Kiran T. Investigation of physical, chemical, mechanical, and microstructural properties of cement-less concrete–state-of-the-art review. Const Build Mater 2023;365:130020. [CrossRef]
[20] Magesh B, Kumar VS, Raghunath PN, Suguna K. Structural performance of Nano alumina-based RC element with polypropylene fibre and GGBS: an experimental assessment. Sustain Energy Technol Assessments 2023;57:103290. [CrossRef]
[21] Amran M, Onaizi AM, Makul N, Abdelgader HS, Tang WC, Alsulami BT, Gamil Y. Shrinkage mitigation in alkali-activated composites: a comprehensive insight into the potential applications for sustainable construction. Results Eng 2023;20:101452. [CrossRef]
[22] García Del Ángel GDJ, Cabrera R, Rolón J, Pichardo R, Thomas García C. Systematic review on the use of waste foundry sand as a partial replacement of natural sand in concrete. Const Build Mater 2024;430:136460. [CrossRef]
[23] Qin J, Yi Y. Use of gasification fly ash, sodium carbonate, and ground granulated blast-furnace slag for soft clay stabilization. Const Build Mater 2024;426:136072. [CrossRef]
[24] Wang Z, Shui Z, Sun T, Hu T, Xiao X, Fan J. Reutilization of gangue wastes in phosphogypsum-based excess-sulphate cementitious materials: effects of wet co-milling on the rheology, hydration and strength development. Const Build Mater 2023;363:129778. [CrossRef]
[25] Reddy RRK, Yaragal SC, Sagar Srinivasa A. One-part eco-friendly alkali-activated concrete–An innovative sustainable alternative. Const Build Mater 2023;408:133741. [CrossRef]
[26] Xia P, Khan S, Hassam M, Sohaib M, Gong F, Zhao Y. Optimal utilization of low-quality construction waste and industrial byproducts in sustainable recycled concrete. Const Build Mater 2024;428:136362. [CrossRef]

[27] Ahmad J, Arbili MM, Alabduljabbar H, Deifalla AF. Concrete made with partially substitution corn cob ash: A review. Case Stud Constr Mater 2023;18:e02100. [CrossRef]

[28] Hamsashree Pandit P, Prashanth S, Katpady DN. Durability of alkali-activated fly ash-slag concrete-state of art. Innov Infrastruct Solut 2024;9:222. [CrossRef]

[29] Zhao C, Li Z, Peng S, Liu J, Wu Q, Xu X. State-of-the-art review of geopolymer concrete carbonation: from impact analysis to model establishment. Case Stud Constr Mater 2024:e03124. [CrossRef]

[30] Sun T, Li Z, Wang Z, Liu Y, Tang P, Deng Y. Optimization on hydration efficiency of an all-solid waste binder: carbide slag activated excess-sulphate phosphogypsum slag cement. J Build Eng 2024;86:108851. [CrossRef]

[31] Padavala SSA B, Kode VR, Dey S. Sustainable concrete development towards the eco-friendly construction: enhancing the strength and durability by using fly ash and silica fume. J Build Pathol Rehabil 2024;9:50. [CrossRef]

[32] Thorne J, Bompa DV, Funari MF, Garcia-Troncoso N. Environmental impact evaluation of low-carbon concrete incorporating fly ash and limestone. Clean Mater 2024;12:100242. [CrossRef]

[33] Pushpalal D, Danzandorj S, Bayarjavkhlan N, Nishiwaki T, Yamamoto K. Compressive strength development and durability properties of high-calcium fly ash incorporated concrete in extremely cold weather. Const Build Mater 2022;316:125801. [CrossRef]

[34] Silva YF, Delvasto S. Durability of self-compacting concrete with addition of residue of masonry when exposed to carbonation and chlorides mediums. Const Build Mater 2021;297:123817. [CrossRef]

[35] Shi X, Yao Y, Wang L, Zhang C, Ahmad I. A modified numerical model for predicting carbonation depth of concrete with stress damage. Const Build Mater 2021;304:124389. [CrossRef]

[36] Vediyappan S, Chinnaraj PK, Hanumantraya BB, Subramanian SK. An experimental investigation on geopolymer concrete utilising micronized biomass silica and GGBS. KSCE J Civ Eng 2021;25:2134–2142. [CrossRef]

[37] Zhao C, Li J, Zhu Z, Guo Q, Wu X, Wang Z, Zhao R. Research on the carbonation resistance and carbonation depth prediction model of fly ash-and slag-based geopolymer concrete. KSCE J Civ Eng 2024;28:2802–2817. [CrossRef]

[38] Lim YY, Pham TM. Influence of Portland cement on performance of fine rice husk ash geopolymer concrete: strength and permeability properties. Const Build Mater 2021;300:124321. [CrossRef]

[39] Çelik Z, Bingöl AF, Ağsu AS. Fresh, mechanical, sorptivity and rapid chloride permeability properties of self-compacting concrete with silica fume and fly ash. Iran J Sci Technol 2022;46:789–799. [CrossRef]

[40] Kumar R, Verma M, Dev N. Investigation on the effect of seawater condition, sulphate attack, acid attack, freeze–thaw condition, and wetting–drying on the geopolymer concrete. Iran J Sci Technol 2022;46: 2823–2853. [CrossRef]

[41] Somasekharaiah HM, Shobha MS, Mallikarjuna HM. Mechanical properties and acid attack test of hybrid fiber reinforced high performance concrete for fly ash based mineral admixture. Mater Today Proc 2022;51:742–749. [CrossRef]

[42] Sahoo S, Parhi PK, Panda BC. Durability properties of concrete with silica fume and rice husk ash. Clean Eng Technol 2021;2:100067. [CrossRef]

[43] Türköz M, Umu SU, Öztürk O. Effect of silica fume as a waste material for sustainable environment on the stabilization and dynamic behavior of dispersive soil. Sustainability 2021;13:4321. [CrossRef]

[44] El Ouni MH, Shah SHA, Ali A, Muhammad S, Mahmood MS, Ali B, Raza A. Mechanical performance, water and chloride permeability of hybrid steel-polypropylene fiber-reinforced recycled aggregate concrete. Case Stud Constr Mater 2022;16:e00831. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00141

Research Article

Investigation of correlation between standard penetration test blow counts and internal friction angles for loess soils

Murat GÜLEN[1,*]

[1]*Department of Civil Engineering, Siirt University, Siirt, 56100, Türkiye*

ARTICLE INFO

Article history
Received: 04 March 2025
Revised: 05 May 2025
Accepted: 22 July 2025

Keywords:
Direct Shear Box Test; Internal Friction Angle; Loess Soils; SPT Blow Count

ABSTRACT

The internal friction angle of soils is determined through laboratory tests. This parameter can also be estimated with high accuracy by referencing field test results. This approach enables the required soil parameters for design purposes to be determined with greater ease and reduces reliance on individual interpretation. A variety of engineering structures are constructed on rich loess deposits, and these loess soils differ significantly from other soil types. In contrast to the extensive research conducted on the strength properties of other soils, the field-test-based research on estimating the strength properties of loess soils remains quite limited. This study aims to address the gap in the literature on loess soils and contribute to this area of research. For this, standard penetration tests were conducted on Central Asian loess soils, known to be among the world's richest loess deposits. This study examined 86 samples exhibiting particle size distributions characteristic of loess soils. The samples were reconstituted in the laboratory to match their natural field water content and density, then subjected to direct shear box tests. Based on the resulting data, a correlation was proposed between standard penetration test blow counts and internal friction angles. The correlation coefficient R^2 = 0.732 indicates a moderately strong relationship. In addition, new equations were developed based on original data derived from loess soils and presented. These were then compared with the empirical relationships previously proposed for sandy soils, and the similarities and differences were interpreted in detail. These results provide practical guidance for estimating the internal friction angle of loess soils, particularly in the early stages of design when field and laboratory data may be limited.

Cite this article as: Gülen M. Investigation of correlation between standard penetration test blow counts and internal friction angles for loess soils. Sigma J Eng Nat Sci 2026;44(2):1172–1186.

INTRODUCTION

Loess soils have unique geotechnical properties that distinguish them from other soil types. Formed by wind-driven deposition, they have a loose, porous and metastable structure with variable grain sizes [1,2]. Their collapsibility under wetting and loading can lead to sudden settlement [3]. While partial cementation provides dry strength, saturation causes instability [4,5]. Its high void ratio, anisotropy

*Corresponding author.
*E-mail address: murat.gulen@siirt.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

and perturbation sensitivity reduce the reliability of standard empirical models [6]. The index and engineering properties of loess soils characterised by high bearing capacity in their dry state, high porosity, low water retention potential, and sudden settlement under saturated loading have been extensively investigated by various researchers [6-8]. Firstly, there is a clear divergence of opinion among researchers regarding the definition of loess soils. Sharifi Teshnizi et al. [4] characterised loess soils as loose to medium-dense silty soils that are partially cemented by clay particles and salts. Terzaghi et al. [9] and Liu [10] describe loess soils as wind-transported deposits of sedimentary origin. Meanwhile, Assadi-Langroudi et al. [5] define them as silt particles that have formed through the direct or indirect combination of silts, clays, and carbonates. Additionally, some researchers have classified loess soils based on their grain size distribution [11–14]. In particular, Swineford and Frye [11] defined loess soils as soils containing a mixture of fine sand and coarse silt, averaging between 0.02 and 0.05 mm in size. In their respective studies, Nouaouria et al. [8] and Li et al. [15] indicated that the predominant grain size of loess soils is silt, comprising 50–70% by weight, with a sand-to-clay ratio that is lower in weight.

Approximately 12% of the Earth's surface is covered by loess soils. The formation processes, index and engineering properties of these soils vary [1,16]. Figure 1 shows the global distribution of loess soils and indicates the areas in which loess of different geotechnical properties is commonly found. The areas indicated as Region 1 in Figure 1 include the Mississippi and Missouri Basins of North America and the glacial loess of the Carpathian Mountains in Eastern Europe. Region 2 comprises loess deposits originating from the foothills of the Alps in Central Europe and the high-altitude mountains of southern Russia. Region 3 consists of silt grains deposited by winds from deserts in Northern China and the Central Asian Republics [15, 17-19]. In particular, the origin of the desert loess in Region 3 is based on the grains transported by rivers from the surrounding mountains [20-22].

The mineral structure and natural water content play a significant role in the behaviour of loess soils. These soils, defined by a high void ratio and semi-unstable structure at low water content, exhibit high strength in their dry state. The weak clay and calcite bonds that hold loose soil particles together weaken or disappear completely when the soil becomes saturated, causing it to collapse suddenly (hydrocollapse) [23-26]. Oedometer tests were conducted by Nouaouria et al. [3] on loess soils in Algeria, which reported that hydrocollapse increases up to a specific level depending on the applied stress, reaching a maximum at 20-21% water content. In addition, the engineering properties of loess soils have been the focus of investigation by various researchers, taking into account the mineral structure, formation process and grain size distributions [6, 27-29]. Gibbs and Holland [30] reported that the liquid limit (LL) values of loess soils in the Kansas region of the USA, which exhibit cohesive properties, range from 25% to 60%. The plasticity index (PI) ranges from 5% to 37%. The maximum strength is realised at water contents of less than 10%, and the resistance to settlement is high. Xu et al. [31] investigated the critical state parameters of loess soils in North West China with LL ranging from 20% to 35.6% and PI ranging from 6% to 17.1%. They conducted triaxial compression tests to determine these values. Bandyopadhyay [32] examined the permeability of loess soils in the USA and stated that their vertical permeability was much higher

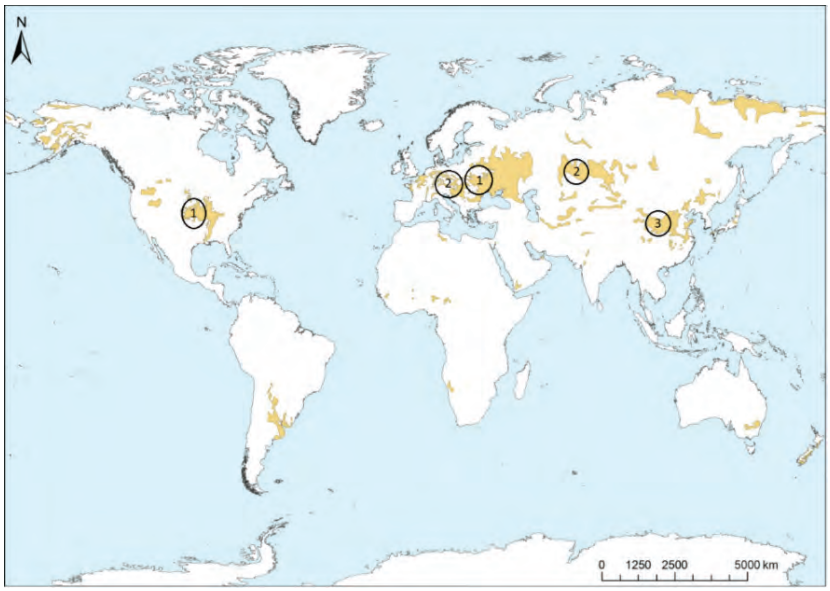


Figure 1. Rich loess soil deposits [22] [created by author].

than their horizontal permeability. However, Terzaghi et al. [9] noted that permeability varies due to changes in the structure of loess soils upon reaching saturation. Chindraprasirt et al. [33] prepared loess soil samples with varying clay contents from Northeast Thailand and subjected them to direct shear box tests. In these tests, where the maximum internal friction angle (ϕ) was found to be 25°, a decrease in the computed ϕ values was observed as the clay content increased. Ma et al. [34] performed CBR tests on 72 loess soil samples prepared with different compaction energies and water contents in China. They reported that samples on the dry side of the optimum density showed higher strength, but soaking the same samples in water resulted in a loss of bearing capacity of up to 97%. Wen and Yan [35] investigated the mechanical properties of unsaturated loess soils in Lanzhou, by conducting direct shear box tests. They highlighted that although significant decreases in measured shear strength and parameters (ϕ, c) were observed due to changes in soil structure, the changes in shear behavior were relatively small. Li et al. [15] carried out a study of the shear strength behaviour of Malan loess soils in China. They reported that the range of cohesion was 0–100 kPa, while the ϕ values varied between 6.5° and 38°. Both parameters were found to decrease with increasing water content and increase with increasing density. Sheeler [36] found that ϕ values varied between 28° and 36° in shear box tests on loess soil samples in the USA. Kong et al. [37] investigated the strength parameters of loess soils with varying grain diameter distributions in Northwest China using direct shear box tests. They reported that the ϕ values varied between 22.2° and 41.5° and that the cohesion changed between 8.3 kPa and 62.2 kPa under differing water contents and vertical stresses.

The analysis of rich loess soil deposits shown in Figure 1 highlights their widespread occurrence in densely populated areas. It emphasises the importance of the index and engineering properties of loess soils for the safe and effective design of structures in these regions. While the relationships between the index and engineering properties of cohesive and granular soils have been demonstrated in numerous studies, the applicability of these findings to silt and loess-type soils remains unclear. Although previous research has explored the characteristics of loess soils, few studies have systematically combined field and laboratory data. In particular, reliable empirical relationships that can be used to estimate the internal friction angle of loess soils from field test results are not widely accepted. To address this gap in the literature, this study investigated the empirical correlations between the internal friction angles obtained from direct shear tests on loess soils collected from Central Asia and Standard Penetration Test (SPT) blow counts. This will enable preliminary estimations in geotechnical design where laboratory testing is limited or unavailable [40-43].

FIELD AND LABORATORY STUDIES

General Site Characteristics and Properties of Loess Soils

The geological structure of the study area has been determined by drilling at pre-determined locations within the site. The findings show that the soils are primarily horizontally stratified, indicating a relatively consistent depositional pattern. Although the stratigraphic profile consists of three distinct soil layers, variations in thickness and composition are observed locally due to topographical and geological factors. Composed of sandy gravel, the upper layer extends to a depth of 3 m. Beneath this layer, between 3m and 6m, is a layer of clayey sand. The thickest layer consists of loess layers located at depths ranging from 3 to 30 metres, containing varying proportions of clay and sand. The variable clay and sand content of these layers has a significant effect on the soil's mechanical properties. In particular, an increase in clay content increases cohesion while reducing ϕ.

The sub-layer at depths between 30 and 36 metres consists of low-plasticity clays containing a certain proportion of loess, with LL% values ranging from 24 to 35%. It is important to emphasise that although this stratigraphic model provides a general framework, the depth and composition of the soil layers may vary locally depending on the geomorphological structure of the site and the depositional history. Detailed in-situ and laboratory investigations, including grain size distribution, Atterberg limits and direct shear box tests, would further improve the understanding of the geotechnical behaviour of these deposits.

Once the density values for the samples obtained from 20 boreholes at different depths had been determined, the samples extracted from the drill tubes were subjected to shear box tests. In consideration of the number of boreholes and their depths, the average depth of the samples tested was determined to be 16.0 m. Following the completion of the shear box tests, the samples were dried at 105°C for a period of 24 hours. The soil samples were then subjected to wet analysis tests. The samples that passed through the No. 200 (0.075 mm) sieve were analysed using a hydrometer, while those that were retained on the No. 200 sieve proceeded to sieve analysis. Consequently, only samples passing through the No. 40 (0.425 mm) sieve at a rate of at least 65% were considered in this study. Therefore, only 86 of the 135 shear box test samples met this criterion and were included in the analysis. The grain size distributions of the soil samples are illustrated in Figure 2.

Table 1 summarizes the specific gravity (Gs), water content (w), grain size percentages, liquid limit (LL), plastic limit (PL), and density (ρ) for 86 soil samples collected at depths ranging from 1.50 m to 36 m, based on ASTM D854-14, ASTM D6938-17, ASTM D6913/D6913M-17 and ASTM D7928-17, ASTM D4318-17, and ASTM D7263-21, respectively. The mean grain size (D_{50}) was determined for each sample, based on the obtained grain diameter distributions [44-49].

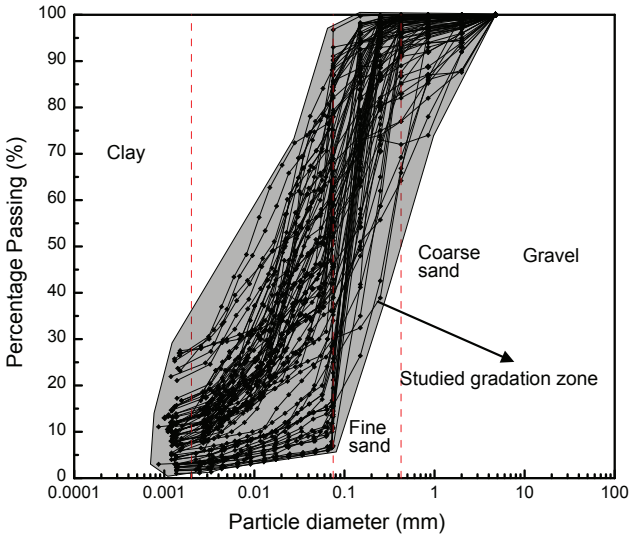


Figure 2. Grain size distribution curves of soils.

Sample Preparation and Direct Shear Test Setup

Soil internal friction angles are generally determined by direct shear box tests or triaxial compression tests, each of which has its own advantages and limitations. Although direct shear testing is less effective than triaxial testing at replicating true in-situ stress paths, it was preferred in this study due to methodological and practical considerations. A total of 405 direct shear tests were conducted on 135 loess samples at different normal stress levels, not including additional repeatability tests. Direct shear testing is faster and easier to handle. It allows samples to be tested at their in-situ water content and density. In contrast, triaxial tests involve highly time-consuming saturation and consolidation stages. Considering the large number of samples tested, direct shear testing was more suitable for achieving the empirical correlation objectives of this study. In this study, the SPT N and ϕ values required to investigate their correlations were determined by following the field and laboratory test procedures outlined below.

1) The SPT blow counts in the study area were determined using an MD300 auger-type water drilling machine equipped with an automatic hammer ram. The values were obtained from a total of different 20 boreholes, each with a diameter of 75 mm, and drilling depths ranging from 13.50 m to 36 m. SPTs were carried out by repeatedly dropping a 63.5 kg hammer from a height of 76.2 cm onto a standard sampler positioned at the upper level of the specified locations. A total of 45 cm of the sampler was allowed to penetrate the soil in three stages. The number of blows required for each 15 cm of penetration was recorded. The first 15 cm was excluded from the analysis as this was considered to be the stage at which the sampler settled in the soil. The total number of blows (N_{30}) was determined by summing the number of blows from the second and third 15 cm penetrations. For these boreholes, the energy rate was considered to be 60%, and the SPT guide rod length was set at 1 m. Following the drilling process, the drill tubes were extracted from the boreholes and sealed with paraffin to preserve the natural water content. The samples were then transferred to the laboratory for the next stage.
2) The density of the soil in the tubes was calculated by weighing the drill tubes, for which both the weight and volume were known. Subsequently, in order to test the homogeneity of the soil in the tubes, the tubes were cut with specialised cutting instruments, and density values were verified by taking samples with thin-walled steel samplers. In the laboratory, the water content of loess soils was found to range from 3.10% to 25.3%. In addition, wet analysis and hydrometer tests revealed that the clay content varied between 5% and 20%. The interplay between these two parameters and capillary effects was found to facilitate the sampling process (Fig. 3. Cylindrical samples were extracted from the drill tube for the sole purpose of calculating their density. It is important to note that these samples were not used in the shear box tests. The evaluation of the processes of taking samples from boreholes under field conditions, transporting them to the laboratory, and sampling in the laboratory environment revealed that the samples had been disturbed. It was therefore decided that the samples be reconstructed to match the water content and density values observed in the field. For this purpose, the soil samples from the tubes were removed and the water content was determined.

Table 1. Summary of the material properties of the soils within the scope of this study

	Gs	w (%)	Depth (m)	ρ (gr/cm³)	LL (%)	PL (%)	Coarse sand (%)	Fine sand (%)	Silt (%)	Clay (%)	D_{50} (mm)
Soil count	86	86	86	86	19	19	86	86	86	86	86
Max.	2.76	25.30	36.0	2.15	35.00	19.30	35.00	93.00	84.20	27.70	0.32
Min.	2.52	3.10	1.50	1.75	24.30	15.20	0.00	3.20	6.00	1.00	0.009
Mean	2.65	15.60	16.00	1.93	28.20	16.71	5.80	39.80	43.90	10.50	0.079

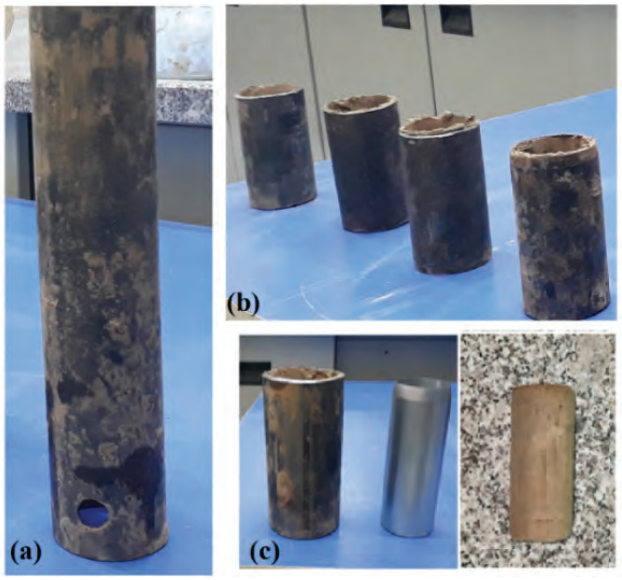


Figure 3. Determination of density a) sample in drill tube b) cutting the drill tube for sampling c) obtaining cylindrical samples.

3) Soil samples with known natural density (ρ) were compacted in a 6 cm x 6 cm x 5 cm shear box ring at a height of 4 cm. For a V=144 cm^3, the amount of soil (M) required for each sample was determined by ρ=M/V. Following the careful placement of all the soil in the ring, the compaction process was initiated. While it is not possible to preserve the microstructure of loess completely after sampling, the reconstituted samples were prepared using field-measured natural density and moisture contents to closely replicate the void ratios and conditions in situ. This methodological approach enables reliable assessment of macroscale geotechnical behaviour, particularly with regard to shear strength parameters. Similar procedures have been adopted in previous loess studies where undisturbed sampling is technically impractical and where controlled reconstitution is widely accepted as a valid alternative for maintaining representation [5, 12]. The reconstituted samples were assumed to represent in situ field conditions; however, no additional validation tests were conducted to quantitatively evaluate the accuracy of this assumption. A rigid plate was positioned to ensure that the loads applied during these compaction processes were distributed homogeneously across the upper level of the soil. As a result of the gradual application of loads to the loading head, the height of the soil in the ring continued until it reached a height of 4 cm. Details of the used for the shear box tests are presented in Figure 4. Displacement and force transducers were used to monitor and record shear strain and force during the test. Normal stress was applied to the top of the lid using a mechanism with a lever ratio of 1:9. The vertical displacement was allowed to complete after the application of the vertical load and then the shear phase started. The upper half of the shear box was fixed while the lower half was forced to move horizontally at a constant rate of 1 mm/min for 10% displacement. To minimize the frictional resistance caused by contact between the upper and lower sections during the shearing process, a 0.05 mm gap was left between the two halves of the direct shear box. In the direct shear test apparatus, adjustment screws are used to create a small gap and minimise friction between the lower and upper boxes. This procedure ensures the formation of a clear shear plane, enhancing reliability by making the measured shear strength parameters independent of box friction.

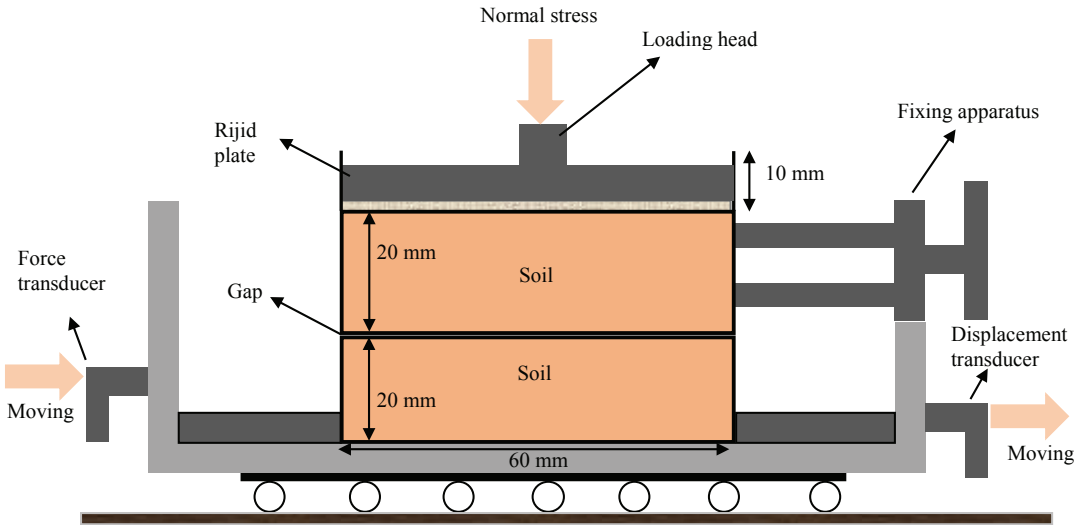


Figure 4. Direct shear test device.

4) The test samples were prepared by compacting and reconstituting them in a direct shear box to reflect the in situ density and moisture content values. The samples were then tested at a cutting speed of 1 mm/min under vertical stresses of 100 kPa, 200 kPa and 400 kPa up to a horizontal strain of 10%. The shearing process was initiated one minute after the application of the vertical normal stresses. Control tests were conducted to assess the repeatability of the results, and the findings were found to be consistent under repeated testing under similar conditions. The final tests were conducted to determine the shear strength parameters, based on 258 direct shear tests performed on 86 soil samples in accordance with ASTM D3080 [50].

RESULTS AND DISCUSSION

A review of the literature reveals a number of studies investigating SPT N–ϕ correlations in sandy soils, as summarized in Table 2. In the equations, N refers to the raw SPT blow count measured directly during field tests. N_1 is the corrected value obtained by applying an overburden correction and considering the effective vertical stress conditions in the soil profile. $N_{(60)}$ represents the SPT blow count standardised by applying energy correction and based on the differing energy transfer efficiencies of the field equipment used. $N_{1(60)}$ denotes the corrected SPT blow count obtained by applying overburden and energy corrections, as well as additional corrections based on the characteristics of the testing equipment used. In the first group of literature, no corrections are applied to the N values and the ϕ values are derived directly from these uncorrected values. The second group of literature applies various corrections to the N values, except for the overburden correction. Finally, the third group of literature applies all corrections to the N values. These different methods highlight the variability in data treatment and its potential impact on the accuracy of ϕ value estimates.

To correctly interpret SPT results under field conditions, it is important to understand the properties of the test equipment and the soil conditions. The measured SPT N values require correction for these factors. Incorrect determination of the N values can result in estimates of soil properties and bearing capacity that are both unsafe and unrealistic [64]. The application of the basic corrections given in Eq. 1 to the field to the N blow count gives the $N_{(60)}$.

$$N_{(60)} = \frac{C_R.C_S.C_B.E_R}{60} * N \quad (1)$$

In Eq. 1; E_R hammer efficiency, C_R rod length correction, C_B borehole diameter correction and C_S sampler correction, C_N overburden correction

In homogeneous soils, both vertical and horizontal effective stresses increase with depth, resulting in higher SPT blow counts. To account for this depth-related influence, an overburden correction factor (C_N) is applied and the corrected blow count is expressed as N_1. Considering the Turkish Building Earthquake Regulations [65] and Liao and Whitman's [66] proposed Equation 2, the C_N values were determined. In this equation, σ'_{v_0} represents the effective vertical stress acting on the test sample, which is calculated by multiplying the relevant sampling depth by the density. As the depth increases, the resulting

Table 2. SPT N versus f correlations recommended for sand soils in literature studies

Correlation type	Referance	Count	Proposed of equation	Limit
SPT N-ϕ	Ferdous [51]	8 samples	ϕ=0.28N + 22.72	-
	Kumar et al. [52]	-	ϕ= 0.2857N + 27.12	4<N<50
	Puri et al. [53]	1053 boreholes	ϕ=0.3128N + 26	1<N<52
	Yusof and Zabidi [54]	14 boreholes	ϕ = 0.481N +29.174	
SPT $N_{(60)}$-ϕ	Wolff [55]	-	$\phi' = 27.1 + 0.3N_{60} - 0.00054\,N^2_{60}$	$N_{60} < 60$
	Japan Road Assoc. [56]	-	$\phi=(15N_{(60)})^{0.5}+15 \leq 45$	N > 5
	Mujtaba et al. [38]	60 boreholes	$\phi=0.7N_{60} + 18$	
SPT $N_{1(60)}$-ϕ	Kuhawy and Mayne [57]	-	$\phi= \tan^{-1}\left(\frac{N_{60}}{12.2+20.3\frac{\sigma'_0}{P_a}}\right)^{0.34}$	$N_{60} < 60$; $\sigma'_v/Pa < 3$
	Hatanaka and Uchida [58]	12 samples	$\phi= (20N_{1(60)})^{0.5} +20$	$N_{1(60)} < 40$
	Mayne [59]	-	$\phi= (15.4N_{1(60)})^{0.5} +20$	-
	Chen [60]	48 samples	$\phi=7.1 \log_{10}[N_{1(60)}]_+ 29.8$	-
	Brown and Hettiarachchi [61]	36 samples	$\phi= 0.3818\tan^{-1}\left(\frac{0.25N_{(60)}xP_a}{\sigma'_0}\right)$	-
	Schnaid et al. [62]	-	$\phi=18\,[N_{1(60)}]^{0.234}$	-
	Ching et al. [63]	-	$\phi=9.2 \log_{10}[N_{1(60)}]_+ 27.5$	-

increase in vertical effective stress leads to a reduction in the C_N. Accordingly, the minimum C_N value of 0.52 was obtained at a depth of 33 m, while the maximum value of 1.70 was calculated at a depth of 1.5 m. Table 3 summarises the correction factors applied to the raw SPT N values obtained using an automated hammer system that provides a consistent energy input during testing. Despite this automation, adjustments are still required to account for variations in C_B, C_S, C_R, E_R, and C_N. These factors are crucial for standardising the number of counts and ensuring that the test results can be compared with those reported in the literature. In addition, Table 3 shows both the standard acceptable ranges and the specific correction values used in this study, which were selected based on relevant geotechnical standards and previous literature. These correction coefficients ensure that the characteristics of the experimental equipment and soil conditions are considered. Improving the accuracy of SPT data significantly enhances the reliability of the correlations between SPT N and ϕ.

$$C_N = 9.78 * \left(\frac{1}{\sigma'_{vo}}\right)^{0.5} < 1.70 \quad (2)$$

Terzaghi et al. (1996) proposed two equations that represent the correlations between SPT N and ϕ [9]. The results of the analysis demonstrate that grain size is a critical parameter that has a significant impact on the relationships under consideration. Salari et al. [67] performed SPT and direct shear tests on 50 gravel samples (25 GC and 25 GW). Based on all the data and the grain size distribution of the gravel, they proposed two equations to describe the relationship between the SPT N and ϕ values If the f value for SPT N-f correlations is determined under laboratory conditions, it is important to ensure that the samples have the same properties as they do under field conditions. Hatanaka and Uchida [58] obtained 12 undisturbed samples by freezing them to preserve their field conditions. They then conducted triaxial tests on these samples to determine their f values. In addition to SPT N-ϕ correlations derived from field and laboratory tests, this relationship has also been explored using various machine learning methods and statistical analyses. Kumar et al. [52] employed the random number generation technique to estimate the ϕ based on the SPT blow counts.

The correlation between the relative density (Dr) of cohesionless soils under field conditions and the N blow count has been investigated by various researchers, resulting in the proposal of several equations to express this relationship [68-70]. Furthermore, the density of the soils and the Dr (%) values are directly related. Therefore, an increase in the density of soils with the same properties will enhance soil resistance, resulting in higher SPT N values and strength. However, the relatively weak correlations shown in Figure 5 are mainly due to the variability in grain size distribution among the samples tested. The mechanical behaviour of soils is directly related to their particle size distribution. However, the variability of soil properties and particle size distributions, combined with the large number of variables, makes it difficult to identify clear statistical trends. When the drilling data is considered, it reveals that the upper layers contain loess soils with a high sand content, whereas the deeper layers consist of clayey loess. Soils with homogeneous grain size distributions generally become denser with increasing depth due to elevated geological stresses, resulting in higher SPT N values and internal friction angles [2, 26, 71]. In this study, however, soils with higher proportions of sand were mostly located at shallower depths. These samples exhibited SPT N values generally greater than 40 and internal friction angles exceeding 35°. Although an overall increase in the amount of fine-grained material is evident as depth increases, this trend is not consistent across all 20 boreholes. In other words, the distribution of grain sizes varied significantly with depth. In addition, increasing geological loads with depth led to higher soil density and compaction, while water content also varied considerably with depth. When considering these variables together, strong and consistent correlations could not be established for the SPT-N–ρ and ϕ–ρ relationships (Fig. 5).

Statistical summary of the SPT N and ϕ values obtained from direct shear box tests of 86 samples is shown in Table 4. The descriptive statistics such as mean ($\bar{x}$), minimum and maximum values (min–max), standard deviation (SD), number of samples (N), skewness coefficient (S) and kurtosis (K) were calculated to evaluate the samples. These parameters provide a solid foundation for assessing the scientific reliability of the correlation analysis and for gaining a comprehensive understanding of the dataset's distribution characteristics.

As seen in Table 4, the SPT N and ϕ values of the 86 samples exhibit a broad distribution range, with a mean SPT N value of 33.7 and a standard deviation of 12.47. In

Table 3. The correction values and the standard ranges

	Correction of the coefficients				
	C_B	**C_S**	**C_R**	**E_R**	**C_N**
Standard range	1.0-1.15	1.0	0.75-1.0	0.70	0-1.70
This study	1.0	1.0	0.75-1.0	0.70	0.52-1.70

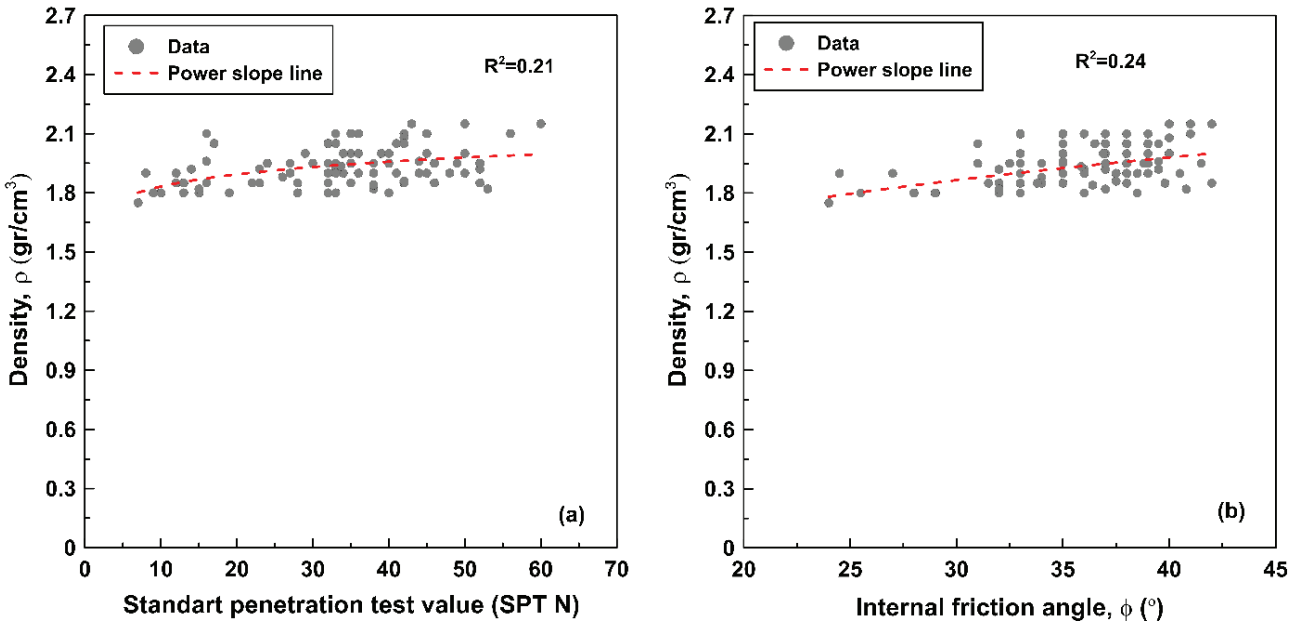


Figure 5. Effect of density on SPT N and ϕ a) SPT N-ρ b) ϕ vs. ρ.

Table 4. Summary of the statistical information from the field and laboratory data

Parameter	Samples (N)	Mean ($\bar{x}$)	Standard deviation (SD)	Minimum (Min.)	Maxiumum (Max.)	Skewness (S)	Kurtosis (K)
SPT N	86	33.71	12.47	7.00	60.0	-0.37	-0.54
SPT N_1	86	28.13	9.85	7.40	54.40	0.35	0.03
SPT $N_{(60)}$	86	37.55	15.38	6.13	68.00	-0.26	-0.76
SPT $N_{1(60)}$	86	30.74	10.79	6.47	58.61	0.18	-0.06
ϕ (°)	86	35.85	3.91	24.00	42.00	-0.95	0.85

contrast, the corrected values SPT N_1 and SPT $N_{1(60)}$ have respective means of 28.13 and 30.74. These results suggest that correction methods have a significant impact on the values. The average internal friction angle is 35.85°, and its skewness (-0.95) and kurtosis (0.85) suggest a left-skewed, peaked distribution. Overall, the limited skewness and kurtosis values suggest that the dataset is suitable for statistical analysis and meaningful interpretation of correlations.

Researchers have used SPT N values to estimate the relative density (Dr) or density of soils [38, 70]. Although this study does not focus directly on density estimation, Figure 6 presents soil density classifications based on the ranges proposed by Terzaghi & Peck [72]. Most of the 86 analysed samples fall within or just above the transition zone between medium-dense and dense soils, indicating that the soils generally have internal friction angles greater than 30°. Samples obtained from shallow depths and composed of coarser grains tend to exhibit higher ϕ values and are positioned near the upper boundary of this classification. However, samples taken from deeper layers with a higher clay content generally exhibit lower ϕ values and are located closer to the lower boundary of the dense soil category.

When field and laboratory test data a evaluated together, it becomes clear that increasing depth results in greater geological loading for samples with similar grain size distributions. This trend has also been observed in direct shear box tests. The samples, which have density ranging from 1.75 to 2.15 g/cm³ and similar grain size distributions, exhibit higher internal friction angles as the density increases. Furthermore, higher SPT N and ϕ values were obtained in samples with a higher sand content when samples were obtained from different depths and possessed similar grain size distributions. The SPT N values were only marginally affected by changes in moisture content. However, direct shear tests showed that an increase in water content, especially in higher

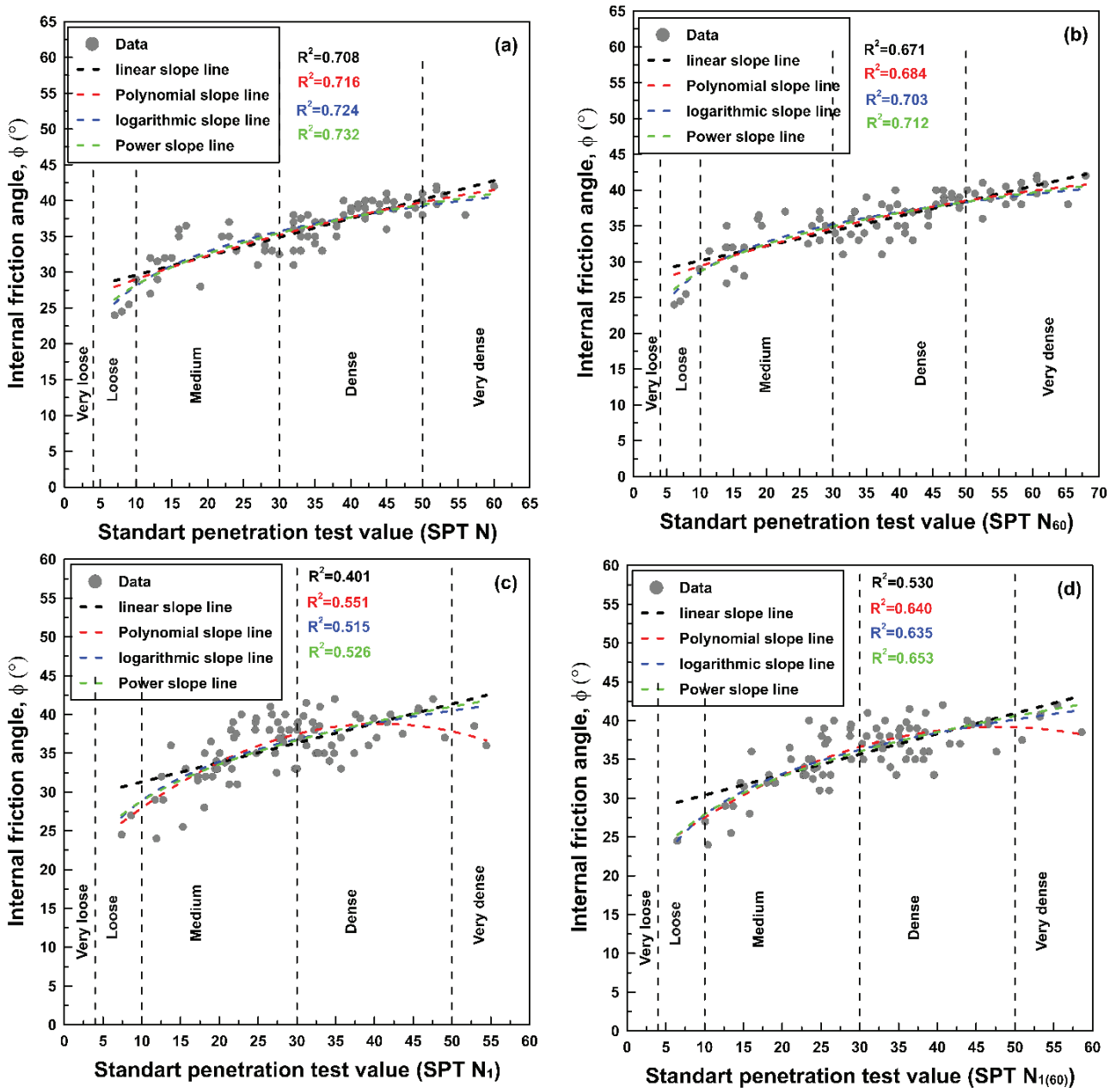


Figure 6. Correlation of SPT N-ϕ a) SPT N-ϕ b) SPT $N_{(60)}$-ϕ c) SPT N_1-ϕ d) SPT $N_{1(60)}$-ϕ.

clay content, reduced the ϕ and led to the development of cohesion. Considering all these findings together, it is clear that there is generally a positive and significant correlation between SPT N and ϕ values. This trend varies depending on the depth of the sample. Samples obtained from depths between 30 and 36 m with a high clay content generally exhibited internal friction angles below 20° and cohesion values of up to 60 kPa. In contrast, samples taken from shallow depths of between 0 and 6 m containing coarser grains showed higher internal friction angles of between 43° and 49°. For this reason, these samples were excluded from the correlation analysis, as their elevated clay or sand content means they do not reflect the typical characteristics of loess.

Table 5 indicates that a variety of regression curves, including linear, polynomial, exponential, and logarithmic, have been used by different researchers. In particular, when examining Figure 6, it is clear that some data points were outliers from the general trend. However, in order not to reduce the number of data points, these outliers were included in the analysis and not excluded. This caused the R^2 to remain at a maximum value of 0.732.

Table 5. SPT N versus f correlations recommended for sand soils in literature studies

Correlation type	Equation type	Equation	Equation ID	R^2
SPT N-ϕ	linear	$\phi=0.264N+26.97$	1a	0.708
	Polynomial	$\phi=-0.002N^2+0.39N+25.33$	1b	0.716
	Logarithmic	$\phi=6.93\ln(N)+12.13$	1c	0.724
	Power	$\phi=17.45N^{0.208}$	1d	0.732
SPT $N_{(60)}$-ϕ	linear	$\phi=0.208N_{(60)}+28.04$	2a	0.671
	Polynomial	$\phi=-0.0018N_{60}^2+0.34N_{(60)}+26.20$	2b	0.685
	Logarithmic	$\phi=6.05\ln(N_{(60)})+14.63$	2c	0.704
	Power	$\phi=18.81\ N_{(60)}^{0.182}$	2d	0.712
SPT N_1-ϕ	linear	$\phi=0.251N_1+28.79$	3a	0.401
	Polynomial	$\phi=-0.0115N_1^2+0.935\ N_1+19.74$	3b	0.550
	Logarithmic	$\phi=7.20\ln(N_1)+12.33$	3c	0.512
	Power	$\phi=17.44\ N_1^{0.218}$	3d	0.526
SPT $N_{1(60)}$-ϕ	linear	$\phi=0.263N_{1(60)}+27.76$	4a	0.530
	Polynomial	$\phi=-0.0082N_{160}^2+0.782N_{1(60)}+20.50$	4b	0.640
	Logarithmic	$\phi=7.65\ln(N_{1(60)})+10.20$	4c	0.635
	Power	$\phi=16.36\ N_{1(60)}^{0.232}$	4d	0.653

When evaluating the performance of the regression curves, it was observed that variations in grain size distribution, water content, and corrections applied to the SPT N values led to significant deviations from the general trend in some samples. However, all data were included in the analyses to improve the representativeness of the natural soil structure observed in the field. In particular, a 70% energy correction was applied for the automatic SPT testing machine used in field conditions for each sample shown in Figure 6b. As the fixed E_R correction affected the N values of all samples by the same proportion, similar R^2 values were obtained for the regression curves in Figures 6a and 6b. In contrast, the C_N correction factor varies between 0.52 and 1.70. For samples where C_N is greater than 1, N values increased, while for samples where C_N is less than 1, N values decreased. This effect increased the scatter of the data in Figures 6c and 6d and significantly reduced the R^2 values. In the analysis findings, the linear regression model showed the lowest correlation, whereas the power regression model generally provided the best fit. These results clearly demonstrate the effect of the correction factors applied to the N values on the selection and performance of the regression models.

The relationship between SPT N-ϕ was investigated in this study, with the results represented by four regression curves, and the corresponding formulations detailed in Table 5. The SPT N blow counts obtained under field conditions were then substituted into these equations to calculate the ϕ values for each sample. In Figure 7, the x-axis represents the ϕ values computed using the equations with the highest R^2 values from Table 5, while the y-axis represents the ϕ values obtained from the shear box tests. Since the R^2 values obtained from Figure 6 are similar (except for Fig. 6c), the variations shown in Figure 7 are closely matched.

The equations proposed by Ferdous [51], Kumar et al. [52], Puri et al. [53], and Yusof and Zabidi [54] for the SPT N-f relationship are applied to loess soils in this study. A comparison of the regression curves derived from these equations with the regression curve based on Equation 1d is shown in Figure 8. A detailed analysis indicates that the regression curve obtained for the soils in this study is aligned closely with the equation proposed by Yusof and Zahidi [54]. While the equation proposed by Kumar et al. [52] and Puri et al. [53] provides compatible results for SPT blow counts between 15 and 50. The equation proposed by Ferdous [51] demonstrates the least compatibility of the three.

The researchers analysing the SPT N-ϕ relationship prefer to utilise different approaches. The application of the SPT $N_{(60)}$-f equations proposed by the researchers to the loess soils considered in this study is presented in Figure 9. The regression curves reveal that the equations proposed by Wolff [55] and the Japan Road Association [56] are compatible with the regression curve developed for the soils in this study within the range of $15 < N_{60} < 50$. In contrast, the equation proposed by Mujtaba et al. [38] displayed the least compatibility.

Table 2 summarises the equations proposed by the researchers, considering the overburden and energy

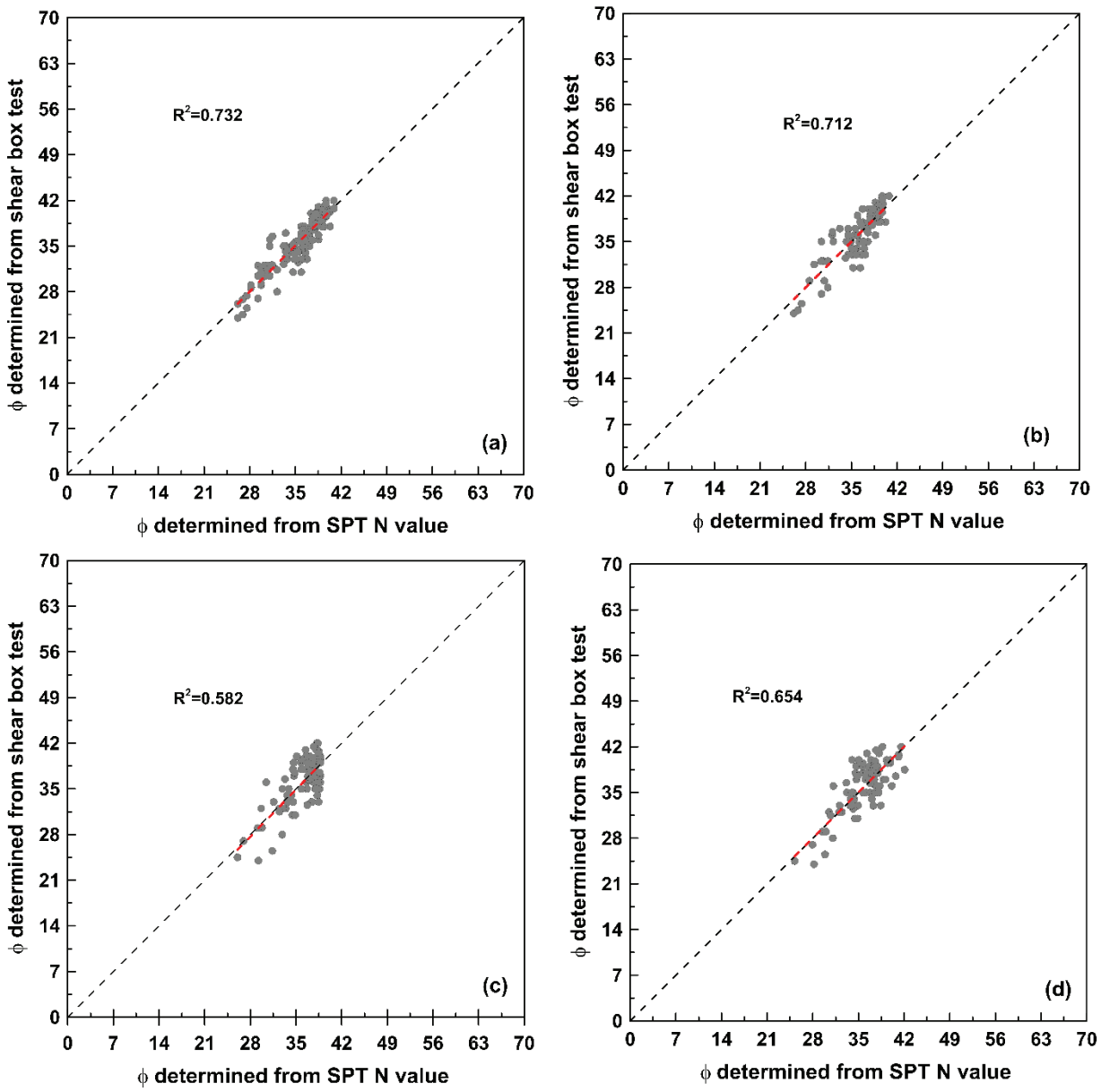


Figure 7. Comparison of ϕ values a) ϕ_{test}- $\phi_{SPT\text{-}N}$ b) ϕ_{test}- $\phi_{SPT\text{-}N(60)}$ c) ϕ_{test}- $\phi_{SPT\text{-}N1}$ d) ϕ_{test}- $\phi_{SPT\text{-}}N_{1(60)}$.

corrections in the field SPT N blow count. The regression curves derived from these equations are shown in Figure 10.

When the curves are analysed, the equations proposed by Kulhawy and Mayne [57], Hatanaka and Uchida [58], Mayne [59], and Chen [60] provide significantly higher f values, while the equation proposed by Brown and Hettiarachchi [61] provides the lowest f values. As a result, none of the equations proposed in the literature for the SPT $N_{1(60)}$-f relationship gave significant results for loess soils. Among these, the equation proposed by Brown and Hettiarachchi [61] seems to be appropriate for loess soils when safety is a priority.

A final discussion of Figures 8-10 indicates that while some of the equations proposed in the literature occasionally provide meaningful results for loess soils, their general applicability is limited. In particular, equations developed on the basis of sandy soils clearly produce inconsistent and unreliable results when applied to loess conditions. This is primarily due to the significant differences in the physical and engineering properties of loess compared to sand. These differences affect the ability of existing empirical equations to accurately describe the behaviour of loess soils, highlighting the need to develop correlations specifically tailored to this soil type.

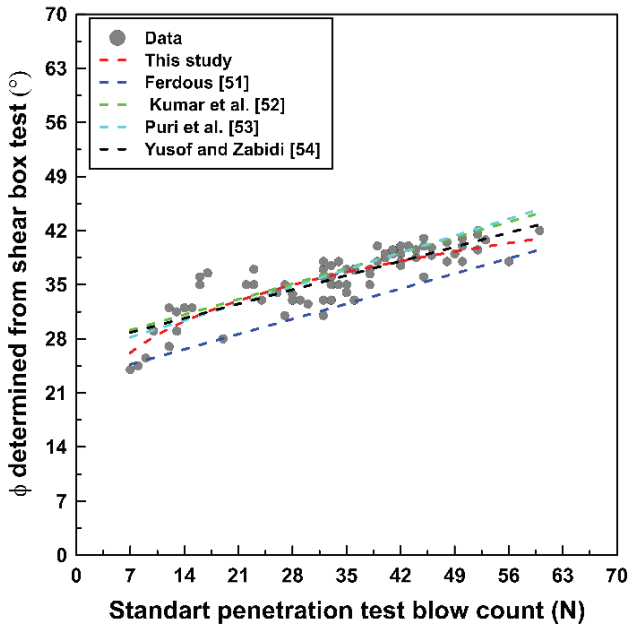


Figure 8. A comparison of regression curves for SPT N-ϕ relationship.

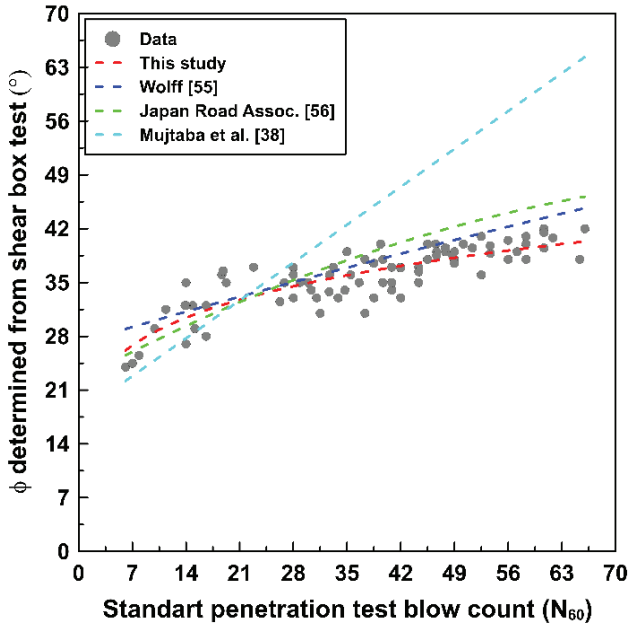


Figure 9. A comparison of regression curves for SPT $N_{(60)}$-ϕ relationship.

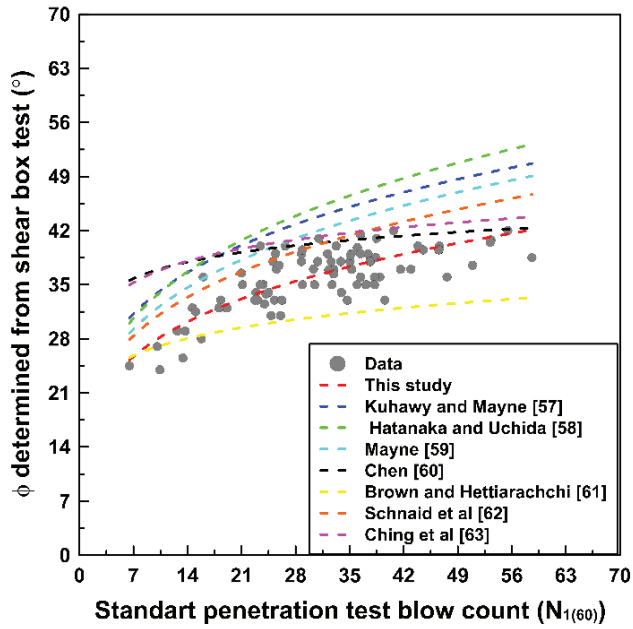


Figure 10. A comparison of regression curves for SPT $N_{1(60)}$-ϕ relationship.

CONCLUSION

The index and strength properties of loess soils are determined through a combination of laboratory and field tests. Although there are sufficient studies that focus on determining the engineering properties of cohesive and granular soils using field tests, research on loess soils remains limited. This study aims to contribute to this topic by determining the internal friction angles from standard penetration test blow counts values obtained in the field for loess soils in Central Asia. The results obtained are summarised below.

1) In consideration of the formation processes of the Central Asian loess, there is a notable variation in grain size distribution. While the majority of grain sizes fall within the range of 0.425 mm to 0.002 mm, there are instances where soils with grain sizes between 4.75 mm and 0.425 mm have been observed at specific depths.
2) Considering the standard penetration test blow counts, these values increased with depth for loess soils with similar grain size distributions. In addition, for loess soils at the same depth in different boreholes, higher standard penetration test blow counts were obtained for samples with larger D_{50} values.
3) Direct shear tests were carried out on samples that were reconstituted to reflect the density and water content of the in situ conditions. The results showed that the ϕ values increased as the density and D_{50} values increased.
4) The SPT-N-ϕ relationship for the loess soil properties considered in this study was examined for different regression models. The determination coefficients representing the different models range from 0.65 to 0.73. These values indicate that ϕ can be estimated from standard penetration test blow counts. Existing SPT N-ϕ correlations, originally established for clean sands, were evaluated for their applicability to loess soils. Although a limited number of studies have achieved satisfactory results, the findings indicate that improved correlations

based on loess soil properties are necessary to enhance the reliability of predictions.

5) These findings provide geotechnical engineers and applicators with practical empirical equations for estimating the internal friction angles of loess soils based on SPT N-values determined under field conditions. The equations contribute to a better understanding of loess soil behaviour in geotechnical site characterisation, foundation design and soil improvement applications during the preliminary design stages or in projects where laboratory testing is limited.

The internal friction angles of loess soils were determined in this study in a laboratory under conditions that reflect in situ density and water content. Meaningful correlations were obtained with standard penetration test blow counts. However, the degree of soil compaction is a critical factor influencing shear strength. Therefore, future studies should include tests representing different levels of compaction to improve the reliability of the correlations and provide a more comprehensive understanding of the engineering behaviour of loess soils.

This study is limited to loess soils in Central Asia. Therefore, the empirical correlations proposed herein are context-specific and unique to the soils studied in this region. They may not be directly applicable to loess soils from regions with different geological origins, mineral compositions, grain size distributions or post-depositional alterations. Therefore, caution should be used when applying these correlations outside the studied context, and validation at the site in question is strongly encouraged.

ACKNOWLEDGEMENT

The author would like to thank M.Sc. Ahmet Serdar Köşeli, M.Sc. Nazlı Ece Öz, Prof. Dr. Havvanur Kılıç, and the staff of the Geotechnical Laboratory at Yıldız Technical University for their contributions to this study.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Xie WL, Li P, Zhang MS, Cheng TE, Wang Y. Collapse behavior and microstructural evolution of loess soils from the Loess Plateau of China. J Mt Sci 2018;15:1642–1657. [CrossRef]

[2] Liu Z, Ma X, Zhou D, Lu L, Zhang H, Bai Y, et al. The geological origins and soil properties of loess-like silty clay: A case study in the Jinan area. Sci Rep 2024;14:12612. [CrossRef]

[3] Nouaouria MS, Guenfoud M, Lafifi B. Engineering properties of loess in Algeria. Eng Geol 2008;99:85–90. [CrossRef]

[4] Sharifi Teshnizi E, O'Kelly BC, Karimiazar J, Moosazadeh S, Arjmandzadeh R, Pani A, et al. Effects of cement kiln dust on physicochemical and geomechanical properties of loess soil, Semnan Province, Iran. Arab J Geosci 2022;15:1482. [CrossRef]

[5] Assadi-Langroudi A, Ng'ambi S, Smalley I. Loess as a collapsible soil: Some basic particle packing aspects. Quat Int 2018;469:20–29. [CrossRef]

[6] Hou Y, Li P, Wang J. Review of chemical stabilizing agents for improving the physical and mechanical properties of loess. Bull Eng Geol Environ 2021;80:9201–9215. [CrossRef]

[7] Lian B, Peng J, Zhan H, Wang X. Mechanical response of root-reinforced loess with various water contents. Soil Tillage Res 2019;193:85–94. [CrossRef]

[8] Yuan-jun J, Alam M, Li-Jun S, Umar M, Sadiq S, Jia LJ, et al. Effect of root orientation on the strength characteristics of loess in drained and undrained triaxial tests. Eng Geol 2022;296:106459. [CrossRef]

[9] Terzaghi K, Peck RB, Mesri G. Soil Mechanics in Engineering Practice. 3rd ed. New York: John Wiley & Sons; 1996.

[10] Liu TS. Loess and the Environment. Beijing: China Ocean Press; 1985.

[11] Swineford A, Frye JC. Petrography of the Peoria loess in Kansas. J Geol 1951;59:306–322. [CrossRef]

[12] Rogers CDF, Dijkstra TA, Smalley IJ. Hydroconsolidation and subsidence of loess: Studies from China, Russia, North America and Europe: In memory of Jan Sajgalik. Eng Geol 1994;37:83–113. [CrossRef]

[13] Nemecz E, Pécsi M, Hartyáni Z, Horváth T. The origin of the silt-size quartz grains and minerals in loess. Quat Int 2000;68-71:199–208. [CrossRef]

[14] Alam M, Jiang YJ, Umar M, Su LJ, Rahman M, Ullah F. Influence of drainage and root biomass on soil mechanical behavior in triaxial tests. Acta Geotech 2022;17:3151–3170. [CrossRef]

[15] Li Y, Zhao J, Li B. Loess and Loess Geohazards in China. 1st ed. CRC Press/Balkema; 2018. [CrossRef]
[16] Ajami M, Heidari A, Khormali F, Gorji M, Ayoubi S. Effects of environmental factors on classification of loess derived soils and clay minerals variations, northern Iran. J Mt Sci 2018;15:976–991. [CrossRef]
[17] Smalley I, Catt JA, Davy R, Thomas MF. Rivers and loess: The significance of long river transportation in the complex event-sequence approach to loess deposit formation. Quat Int 2008;181:88–98.
[18] Stevens T, Carter A, Watson TP, Vermeesch P, Andò S, Bird AF, et al. Genetic linkage between the Yellow River, the Mu Us desert and the Chinese Loess Plateau. Quat Sci Rev 2013;78:355–368. [CrossRef]
[19] Marković SB, Fitzsimmons KE, Sprafke T, Gavrilović D, Smalley IJ, Jović V, et al. The history of Danube loess research. Quat Int 2016;399:86–99. [CrossRef]
[20] Muhs DR. Loess and its geomorphic, stratigraphic, and paleoclimatic significance in the Quaternary. In: Shroder JF, editor. Treatise on Geomorphology. Academic Press; 2013. p. 149–183. [CrossRef]
[21] Smalley I, O'Hara-Dhand K, Kwong J. China: Materials for a loess landscape. Catena 2014;117:100–107. [CrossRef]
[22] Fenn K, Prud'Homme C. Dust deposits: Loess. In: Shroder JF, editor. Treatise on Geomorphology. 2nd ed. Academic Press; 2022. p. 320–365. [CrossRef]
[23] Chen ZY. Practical solutions to problems of collapsible loess in China. In: Proceedings of the 7th International Conference on Expansive Soils; 1992; Dallas. Vol 2, p. 1–12.
[24] Parsons RL, Johnson RM, Brown DA, Dapp S, Brennan JJ. Characterization of loess for deep foundations. DFI J 2009;3:14–24. [CrossRef]
[25] Rouaiguia A, Dahim MA. Hydrocollapse of semi-arid soils. Indian Geotech J 2016;46:25–33. [CrossRef]
[26] Hamdy DB. Origin and characteristics of collapsible soils: State-of-the-art report. Innov Infrastruct Solut 2024;9:401. [CrossRef]
[27] Drewnik M, Skiba M, Szymański W, Żyła M. Mineral composition vs. soil-forming processes in loess soils: A case study from Kraków (Southern Poland). Catena 2014;119:166–173. [CrossRef]
[28] Xu L, Coop MR. Influence of structure on the behaviour of a saturated clayey loess. Can Geotech J 2016;53:1026–1037. [CrossRef]
[29] Sun J, Chen R, Qin X, Lou B, Wang Y, Li W. Experimental study on mechanical properties of fissured loess under true triaxial conditions. Sci Rep 2025;15:9852. [CrossRef]
[30] Gibbs HJ, Holland WY. Petrographic and Engineering Properties of Loess (Engineer Monograph No. 28). U.S. Bureau of Reclamation; 1960.
[31] Xu L, Gao C, Zuo L, Liu K, Li L. The critical states of saturated loess soils. Eng Geol 2022;307:106776. [CrossRef]
[32] Chindraprasirt P, Kampala A, Arngbunta A. Engineering characteristics of Khon Kaen loess as construction material. Lowland Technol Int 2020;22.
[33] Bandyopadhyay SS. Geotechnical evaluation of loessial soils in Kansas. Transp Res Rec 1983;945:29–36.
[34] Ma F, Yang J, Bai X. Water sensitivity and microstructure of compacted loess. Transp Geotech 2017;11:41–56. [CrossRef]
[35] Wen B, Yan Y. Influence of structure on shear characteristics of the unsaturated loess in Lanzhou, China. Eng Geol 2014;168:46–58. [CrossRef]
[36] Sheeler JB. Summarization and comparison of engineering properties of loess in the United States. In: Conference on Loess: Design and Construction. Highway Research Board; 1968. p. 1–9.
[37] Kong D, Wan R, Zhao C, Dai J, Dong T, et al. Effect of conglomeration gradation on loess shear strength with different water content. Sci Prog 2021;104(2):00368504211010581. [CrossRef]
[38] Mujtaba H, Farooq K, Sivakugan N, Das BM. Evaluation of relative density and friction angle based on SPT-N values. KSCE J Civ Eng 2018;22:572–581. [CrossRef]
[39] Ghali M, Chekired M, Karray M. Framework to improve the correlation of SPT-N and geotechnical parameters in sand. Acta Geotech 2020;15:735–759. [CrossRef]
[40] Hegde A, Anand A. Resistivity correlations with SPT-N and shear wave velocity for Patna soil in India. Indian Geotech J 2022;52:161–173. [CrossRef]
[41] Ekmen AB. Evaluation of SPT-N values and internal friction angle correlation using artificial intelligence methods in granular soils. Soil Res 2023;61:495–509. [CrossRef]
[42] Alshkane YM. Correlations between geotechnical design parameters and index properties for low-plasticity clays. Transp Infrastruct Geotech 2024;11:3714–3725. [CrossRef]
[43] Guo F, Gu W. Study on dynamic resilient modulus prediction model of subgrade fine-grained soil based on physical property parameters. Sci Rep 2025;15:17605. [CrossRef]
[44] ASTM International. ASTM D854-14: Standard Test Methods for Specific Gravity of Soil Solids by Water Pycnometer. ASTM International; 2023.
[45] ASTM International. ASTM D6938-17: Standard Test Method for In-Place Density and Water Content of Soil and Soil-Aggregate by Nuclear Methods (Shallow Depth). ASTM International; 2023.
[46] ASTM International. ASTM D6913/D6913M-17: Standard Test Methods for Particle-Size Distribution (Gradation) of Soils Using Sieve Analysis. ASTM International; 2023.
[47] ASTM International. ASTM D7928-17: Standard Test Method for Particle-Size Distribution (Gradation) of Fine-Grained Soils Using the Sedimentation (Hydrometer) Analysis. ASTM International; 2023.

[48] ASTM International. ASTM D4318-17: Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils. ASTM International; 2023.
[49] ASTM International. ASTM D7263-21: Standard Test Methods for Laboratory Determination of Density (Unit Weight) of Soil Specimens. ASTM International; 2023.
[50] ASTM International. ASTM D3080/D3080M–11: Standard test method for direct shear test of soils under consolidated drained conditions. West Conshohocken, PA: ASTM International; 2011.
[51] Ferdous S. Geotechnical Characterization of the Subsoil in Khulna City Corporation (KCC) Area [MSc thesis]. Bangladesh University; 2007.
[52] Kumar R, Bhargava K, Choudhury D. Estimation of engineering properties of soils from field SPT using random number generation. INAE Lett 2016;1:77–84. [CrossRef]
[53] Puri N, Prasad HD, Jain A. Prediction of geotechnical parameters using machine learning techniques. Procedia Comput Sci 2018;125:509–517. [CrossRef]
[54] Yusof NQAM, Zabidi H. Reliability of using standard penetration test (SPT) in predicting properties of soil. J Phys Conf Ser 2018;1074:012086. [CrossRef]
[55] Wolff TF. Pile capacity prediction using parameter functions. In: Predicted and Observed Axial Behavior of Piles: Results of a Pile Prediction Symposium. ASCE; 1989. p. 96–106.
[56] Japan Road Association. Specifications for Highway Bridges, Part IV. Japan Road Association; 1990.
[57] Kulhawy FH, Mayne PW. Manual on Estimating Soil Properties for Foundation Design. Report No. EPRI-EL-6800. Palo Alto, CA: Electric Power Research Institute; 1990.
[58] Hatanaka M, Uchida A. Empirical correlation between penetration resistance and internal friction angle of sandy soils. Soils Found 1996;36:1–9. [CrossRef]
[59] Mayne PW. Stress-strain-strength-flow parameters from enhanced in-situ tests. In: Proceedings of the International Conference on In-Situ Measurements of Soil Properties and Case Histories; 2001; Bali, Indonesia. p. 27–48.
[60] Chen JR. Axial Behavior of Drilled Shafts in Gravelly Soils [PhD thesis]. Cornell University; 2004.
[61] Brown T, Hettiarachchi H. Estimating shear strength properties of soils using SPT blow counts: An energy balance approach. ASCE Geotech Spec Publ 1998;179:16–26.
[62] Schnaid F, Odebrecht E, Rocha MM, de Paula Bernardes G. Prediction of soil properties from the concepts of energy transfer in dynamic penetration tests. J Geotech Geoenviron Eng 2009;135:1092–1100. [CrossRef]
[63] Ching J, Lin GH, Chen JR, Phoon KK. Transformation models for effective friction angle and relative density calibrated based on generic database of coarse-grained soils. Can Geotech J 2016;54:481–501. [CrossRef]
[64] Sivrikaya O, Toğrol E. Relationships between SPT-N value and undrained shear strength of fine-grained soils in Turkey. Teknik Dergi 2007;18:4229–4246.
[65] Disaster and Emergency Management Authority (AFAD). Turkish Building Earthquake Code (TBEC). Off Gaz Repub Turkey 2018;30364.
[66] Liao S, Whitman RV. Overburden correction factor for SPT in sand. J Geotech Eng 1986;112:373–377. [CrossRef]
[67] Salari P, Lashkaripour G, Ghafoori M. Presentation of empirical equations for estimating internal friction angle of SP and SC soils in Mashhad, Iran using standard penetration and direct shear tests and comparison with previous equations. Int J Geogr Geol 2015;4:89–102. [CrossRef]
[68] Meyerhof GG. Penetration tests and bearing capacity of cohesionless soils. J Geotech Eng Div ASCE 1956;82:866/1–866/19. [CrossRef]
[69] Kibria S, Masood T. SPT, relative density and PHI relationships for Indus sands at Chashma. In: Proceedings of VII National Conference of Pakistan National Society for Soil Mechanics and Foundation Engineering; 1998. p. 169–188.
[70] Cubrinovski M, Ishihara K. Empirical correlation between SPT N-value and relative density for sandy soils. Soils Found 1999;39:61–71. [CrossRef]
[71] Kabir ME, Sabab SR, Bondhon MMM, Shuvo NA. Determining Relations Between SPT-N & Shear Strength Parameters of Dhaka and Sylhet Soils Using Machine Learning Approach [BSc thesis]. Islamic University of Technology; 2022.
[72] Terzaghi K, Peck RB. Soil Mechanics in Engineering Practice. 2nd ed. New York: Wiley; 1967.

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1187–1218, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00062

Research Article

Analyzing challenges to maximize the efficiency of industrial processes using carbon capture and hydrogen energy production from methane

Kareem A. ABDELKAREEM[1,2,*], Ahmed El-AWADY[1]

[1]*Department of Environmental Engineering Sciences, Faculty of Graduate Studies and Environmental Research, Ain Shams University, Cairo, 11566, Egypt*
[2]*Sustainability Department, Ministry of Petroleum and Mineral Resources, 11191, Egypt*

ARTICLE INFO

Article history
Received: 14 October 2024
Revised: 11 November 2024
Accepted: 07 March 2025

Keywords:
Carbon Capture and Storage (CCS); Energy Efficiency; Hydrogen; Methane

ABSTRACT

Greenhouse gases (GHGs) significantly impact on our environment, negatively affecting living conditions. Therefore, we strive to harness these GHGs for humanity's benefit. Carbon dioxide is viewed as a valuable resource for generating new products, including concrete treatments and chemicals. In the food industry, carbon dioxide is used for enzyme inhibition, protein precipitation, maintaining food quality, and also being used in the soft drink sector. This paper presents a novel approach to utilizing greenhouse gas emissions for beneficial purposes. Methane and carbon dioxide are two primary gases contributing to climate change, which present significant environmental, social, and economic challenges. We also aim to explore enhancing industrial operations by implementing advanced carbon capture technologies and leveraging methane as a source for hydrogen production. Focusing on these two key areas can significantly reduce emissions and promote sustainable energy practices. By addressing the research gaps, it is possible to analyze most of the challenges in order to significantly enhance the efficiency and sustainability of industrial processes, contributing to a low-carbon future while supporting research and development for future applications of CO_2 that could contribute to a CO_2-neutral economy. And this may have a massive impact on the carbon markets as carbon credits are emerging globally. A legislative and regulatory framework for trading hydrogen and carbon in various new uses and applications should be established. In this paper, different methods for producing hydrogen from methane have proven an energy efficiency of 65% to 75%. Other methane applications are still in the experimental phase with an estimated energy efficiency of about 58%, marking a significant advancement in the production of clean hydrogen and solid carbon for various applications, all while generating minimal emissions.

Cite this article as: Abdelkareem KA, El-Awady A. Analyzing challenges to maximize the efficiency of industrial processes using carbon capture and hydrogen energy production from methane. Sigma J Eng Nat Sci 2026;44(2):1187–1218.

*Corresponding author.
*E-mail address: kmadel@live.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

A Brief History of Efforts to Reduce Carbon Dioxıde and Methane Emissions

The recognition of the need to address greenhouse gas emissions, primarily carbon dioxide (CO_2) and methane (CH_4), is a relatively recent development, dating back to the mid-20th century. However, significant efforts to mitigate these emissions have gained momentum over the past few decades, driven by growing scientific evidence of climate change and its potential impacts, as shown in Figure 1 [1]:

❖ **Early awareness and initial actions [2]:**
- 1950s and 1960s: Scientists began to understand the greenhouse effect and the potential role of human activities in altering the Earth's climate. Scientists such as Roger Revelle and Hans Suess released the revolutionary findings of the growing amount of CO_2 in the atmosphere because of the burning of fossil fuel.
- 1970s: A turning point came in 1979 as the first international conference on climate change was held. Instead, the conference was acknowledged in the risks of climate change and the need to further study the issue and collaborate internationally.

❖ **International cooperation and climate treaties [3]:**
- 1980s: Intergovernmental panel on climate change (IPCC) was formed in 1988 to evaluate the science of climate change. The reports by the IPCC have furnished very important evidence that can guide the policy makers and other individuals on the dangers of climate change.
- 1990s: The framework convention on climate change of the United Nations (UNFCCC) was adopted in 1992 and it paved way to global action on climate change. The Kyoto Protocol was signed up in 1997 and it enacted legally binding emissions reduction objectives to the developed countries.
- 2000s: An important step was made in 2015 with the Paris Agreement. It seeks to reduce global warming by at least two degrees Celsius (or less) beyond the pre-industrial levels and pursue an aim of reduction of at least 1.5 degrees Celsius.1 The agreement introduced the principle of Nationally Determined Contributions (NDCs) whereby the countries establish their emission reduction targets.
- 2010s and 2020s: Over the past several years, attention to the contribution of methane emissions to climate change has grown. Bridging the gap between Global warming and the effects of global warming, the Global Methane Pledge was introduced in 2021, destination being to lower the total amount of methane emission on the planet by 30 percent by 2030, as illustrated in Figure 2.

It is worth mentioning that the history of the struggle to decrease emission of greenhouse gases has not finished yet. The rest of the world is striving to achieve sustainable future as the scientific community deepens their understanding on climate change and the world is opening up to more technological developments.

Greenhouse gases (GHGs) trap heat in the atmosphere of the earth. The ultraviolet radiation of the sun heats up the surface of the earth during the day which cools down and emits heat back to the atmosphere when it is dark. But, greenhouse gases cause some of that heat to be trapped. One of these greenhouse gasses is carbon dioxide (CO_2); the gas is not flammable and is colorless at standard temperature and pressure. Carbon dioxide is far less prevalent in the environment of the Earth in comparison to nitrogen and oxygen but it is important in controlling climate on our planet. The release of carbons into the atmosphere

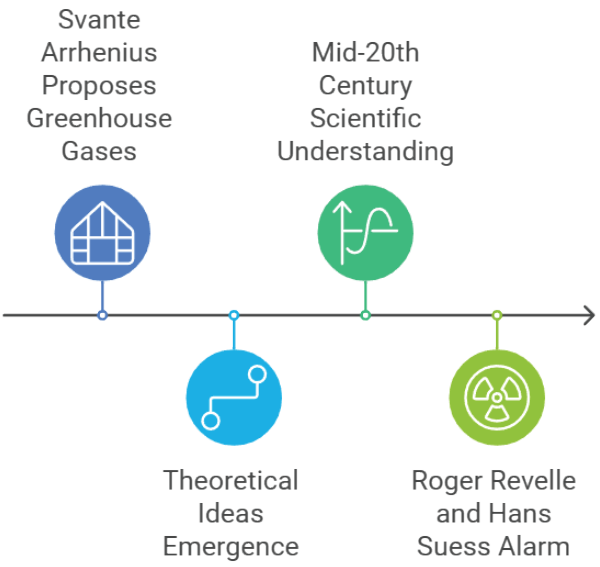


Figure 1. Evolution of climate change awareness.

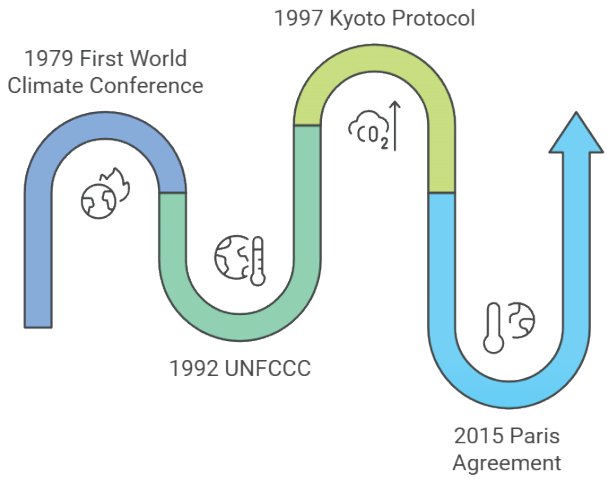


Figure 2. Key milestones in global climate agreements.

due to human activities is increasing and this is caused by the burning of fossil fuels, the increase in industrial establishment, and deforestation. It can take hundreds of years to dissipate and unlike other green house gases, this gas helps in fueling climate change by trapping heat stored within the solar radiation. At this point, one of the recent projects and an effective method to reduce climate change impacts is carbon capture. It particularly focuses on significant contributors to the emissions e.g., power plants and energy-intensive industrieses [4].

Three common ways of capturing carbon are available, post-combustion systems, oxy-fuel combustion systems and pre-combustion systems. The extracted carbon could be further stored in underground geologic storage sites or be recycled to produce new products and uses, including cement, chemicals, food, and drinks. The risks, tracking, and the legal issues of CO_2 capture projects are not new because these areas are included in the existing health, safety and environmental monitoring actions in this industry. But these projects are unprofitable to the companies and stakeholders at this time. As such, it is important to put up a binding legal and regulatory system among these parties. Figure 3 illustrates the carbon dioxide capture cycle that comprises of emission sources, channels of capture, storage or reuse, and monitoring and regulations [5].

Conversely, hydrogen is another clean type of energy to be produced and it is the most common chemical element with an estimated 75 percent universe mass. There are very large numbers of hydrogen atoms in water, plants, animals and humans on planet Earth. The most important benefit of hydrogen is that hydrogen burns up with oxygen and gives no CO_2 reaction because hydrogen has no carbon atom in the hydrogen molecule. When the pure oxygen is burnt with hydrogen, the pure water is the only product that is formed [6].

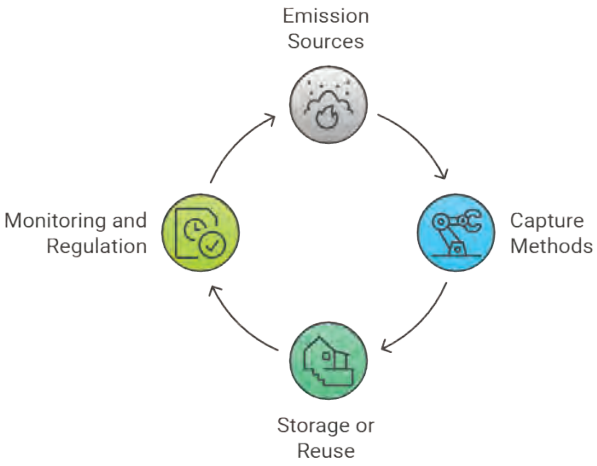


Figure 3. Carbon dioxide capture cycle.

Extraction of hydrogen needs energy and the energy requirement is different and depends on the production process. This process is dependent on the amount of hydrogen that can be derived out of an energy source, determining its efficiency. This efficiency ought to be maximized ideally to make the energy originating the source efficiently converted to a form of hydrogen as an energy carrier. There are various resources that can be used to produce hydrogen. Today, hydrogen is produced primarily based on fossil fuels, especially, natural gas, by means of chemical reactions, among which are steam methane reforming (SMR), auto-thermal reforming (ATR), and methane pyrolysis. Also, hydrogen can be produced with electricity (again either using supplies of renewable energy, such as wind, sun, geothermal, and biomass, or fed by the grid, as illustrated in Figure 4) [7].

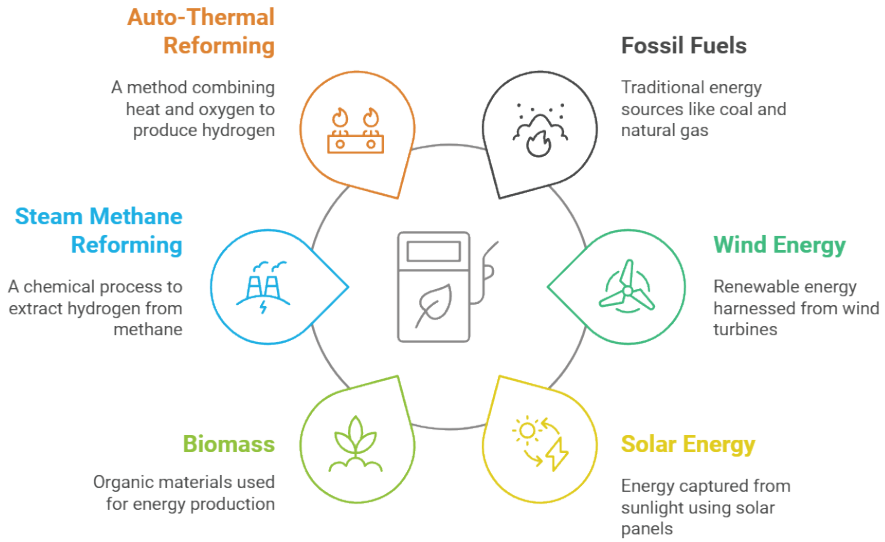


Figure 4. Hydrogen production from different resources.

When talking about carbon and hydrogen, it becomes impossible not to mention such a powerful greenhouse gas as methane, which does not exist in the environment long enough. Methane has an average atmospheric lifetime of approximately 12 years; this is a short period compared to hundreds of years that it takes carbon dioxide to exit the atmosphere. This implies that cutting down the number of methane cases may be a simple and effective measure of curbing part of the climate change in a matter of several decades. Agriculture, oil and gas, landfills, wastewater management, and a number of different industrial processes are the primary contributors to methane emissions [6]. Figure 5 shows how the ability to capture the CO_2 and CH_4 can be useful in reducing climate change via positive pathways, including the storage or re-use of CO_2 and the separation of methane to create alternative products.

The Sustainable Development Goals (SDG) are meant to reduce the effects of climate change and environmental degradation. This can be done by capturing the methane emissions and converting them into both carbon and hydrogen. This does not only improve the efficiency of energy consumed in industrial activities but also it makes it possible to develop new goods and services that bring into support the principle of the closed-loop. A more in-depth discussion of this will follow in this paper.

CARBON CAPTURE, HYDROGEN AND METHANE

We examine and discuss several methods of capturing carbon based on carbon dioxide and methane in this paper. We aim to discover the possibilities of using carbon in the industrial processes and harvesting the hydrogen that is extracted form the methane and used in other industries, energy production and in energy saving activities. We will start by defining carbon capture, hydrogen and methane.

Carbon Capture

Carbon capture refers to a collection of technologies that capture carbon dioxide (CO_2) emissions from industrial processes and power plants, preventing them from staying in the atmosphere. As shown in Figure 6, the captured CO_2 can then be either utilized for various purposes or stored permanently undergrounds [8].

The different CO_2 capture methods (introduced in Figure 7) are:

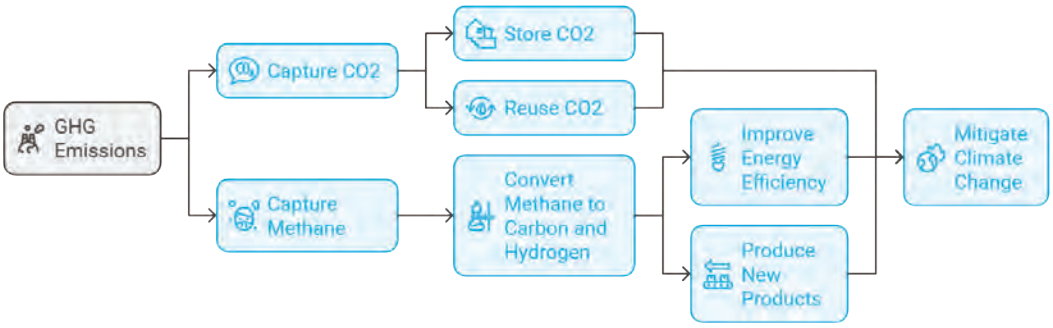


Figure 5. Mitigation pathway from capturing CO_2 & CH_4.

Figure 6. Carbon capture concept.

1. Carbon Capture and Utilization /Storage (CCUS/CCS)
2. Carbon Capture at Power Plants
3. Direct Air Capture (DAC)
4. Enhanced Rock Weathering (ERW)
5. Aqueous Amine-Based CO_2 Capture
6. Membrane Gas Separation
7. Carbon Capture and Conversion
8. Bioenergy with Carbon Capture and Storage (BECCS)
9. Chemical Looping
10. Cryogenic Carbon Capture (CCC)
11. Carbon Capture Using Nanotechnology

Table 1 below shows a comparison of various CO_2 capture technologies with their description, advantages, and challenges [8-12].

The process of CO_2 capture technologies is complicated, and can frequently be misconstrued as a step to a more sustainable future. Although these technologies cannot be used as an excuse to carry on consuming fossil fuels, they are very important in curbing emissions in a manner that cannot be achieved by other technologies. More efficient alternatives exist in most situations (such renewable energy sources in the power sector), but in some areas of industry and at the system level (to offset emissions), capturing and trapping CO_2 is critical.

In order to capture CO_2, a separation of carbon dioxide by a number of physical and chemical techniques, which include: absorption, adsorption and membrane technology, are used. Alternative methods of capturing carbon

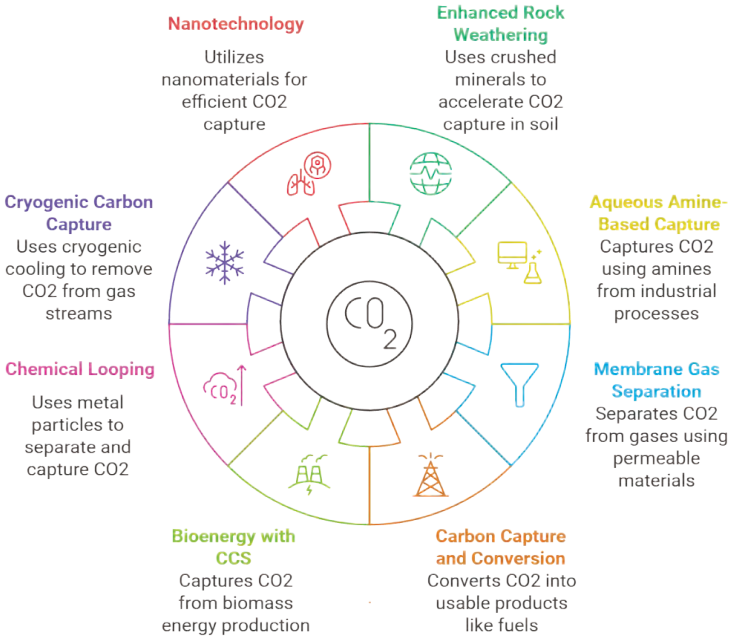


Figure 7. Carbon capture technologies.

Table 1. Comparison of CO_2 Capture Technologies [created by author]

No.	Capture Technology	Description	Advantages	Challenges
1	Carbon Capture and Utilisation /Storage (CCUS/CCS)	It is a set of technologies that capture carbon dioxide (CO_2) emissions from industrial processes and either stores it underground or utilize it for other purposes.	▪ Mitigation of Climate Change ▪ Enhanced Oil Recovery (EOR) ▪ Production of Synthetic Fuels ▪ Industrial Raw Material	▪ High Cost ▪ Energy Intensive ▪ Leakage Problem ▪ Public Perception Issues
2	Carbon Capture at Power Plants (CCPP)	It is a technology designed for the removal of carbon dioxide (CO_2) emissions by fossil-fuel power plants..	▪ Low Greenhouse Gas Emissions ▪ Increased Operational Lifetime of Power Plants ▪ Capacity to Generate Negative Emissions	▪ High Energy Intake ▪ High Operating Expenses ▪ Technological Complexity ▪ CO_2 Storage and Transport

Table 1. Comparison of CO_2 Capture Technologies [created by author] (continued)

No.	Capture Technology	Description	Advantages	Challenges
3	Direct Air Capture (DAC)	The technology involved here entails capturing carbon dioxide (CO_2) emissions directly from the atmosphere. The technology involves the intake of air, CO_2 filtration, and its subsequent storage underground or utilization.	▪ Carbon Negative Emissions ▪ Global Implementation Capacity ▪ Potential for CO_2 Utilization	▪ High Energy Requirement ▪ High Cost ▪ Limited Capacity ▪ CO_2 Storage and Transport
4	Enhanced Rock Weathering (ERW)	The method involved here entails enhancing the rate of rock weathering to remove CO_2 emissions from the atmosphere through a natural process. This process entails the crushing of rocks, which is subsequently scattered on the ground to enhance exposure to atmospheric CO_2.	▪ Natural Process ▪ Long-Term Storage ▪ Soil Fertility	▪ Large Scale Land Requirements ▪ Limited Capacity ▪ Environmental Impacts of Massive Rock Crushing and Transport
5	Aqueous Amine-Based CO_2 Capture	It is a mature technology involving the use of a liquid medium (typically aqueous amine solutions) for absorbing CO_2 emissions from flue gases.	▪ Mature Technology ▪ High CO_2 Capture Rate ▪ Flexibility	▪ High Energy Requirement ▪ Corrosion ▪ Solvent Degradation ▪ Environmental Impact of Amine Solvent
6	Membrane Gas Separation	The technology involves using semi-permeable membranes to separate gases according to their molecular weights. In carbon capture applications, the membrane technology can be used to separate CO_2 emissions from other gases within the flue gas stream.	▪ Low Energy Intake ▪ No Chemical Reagents ▪ Modularity	▪ Lower CO_2 Capture Rate ▪ Membrane Fouling ▪ Constraining Operating Conditions
7	Carbon Capture and Conversion	Consists of the capture and conversion of CO_2 emissions to valuable products such as fuels, chemicals, or construction materials.	▪ GHG Emission Reduction ▪ Value-added Products Creation	▪ High Energy Requirement ▪ Technological Maturity ▪ Economic Viability ▪ Scalability
8	Bioenergy with Carbon Capture and Storage (BECCS)	This is a technology involving the capture and storage of CO_2 emissions from biomass combustion. In this case, the CO_2 emissions from biomass burning to produce energy are captured and stored underground.	▪ Negative Emissions ▪ Energy Production ▪ Sustainable Land Utilization	▪ Land Utilization Competition ▪ Necessity for Water Sources ▪ Sustainability Challenges
9	Chemical Looping	The technology entails an efficient and cost-effective approach to the capturing of CO_2 emissions from power plants and industrial processes. The process involves a cyclical operation using a metal oxide (also referred to as an oxygen carrier) to move oxygen from an air reactor to a fuel reactor.	▪ Automatic CO_2 Separation ▪ High Thermal Efficiency ▪ Low Environmental Impact ▪ Flexibility	▪ Oxygen Carrier Development ▪ Reactors' Design and Operation ▪ High Capital Costs
10	Cryogenic Carbon Capture (CCC)	It is a modern technology involving cooling of flue gases to cryogenic temperatures, thereby allowing the solidification of CO_2 and separation from other gases.	▪ High Efficiency of Capture ▪ Efficient Process ▪ Removal of Other Pollutants ▪ Opportunity for Product Utilization	▪ High Capital Cost ▪ High Energy Intake ▪ Complex Process ▪ Scalability
11	Carbon Capture Using Nanotechnology	Nanotechnology is employed to create advanced materials to be applied in efficient CO_2 capture due to manipulation of materials at the nanoscale.	▪ High Surface Area ▪ Selectivity ▪ Low Energy Requirement ▪ Potential for DAC	▪ Scalability ▪ Material Stability ▪ Environmental Impact Issues

Figure 8. Membrane technology applications.

dioxide through new alternative technologies are proving to be effective and membrane technology is gaining more and more appreciation. It is currently used in other industries to large-scale separate a variety of mixtures of gases. The increasing use of membrane technology is due to its low cost of operation and eco-friendliness. Gas separation using membranes is an excellent alternative to the conventional unit operations because it is relatively easy to run, economically viable, it requires little maintenance, it deals in a small space and because it consumes less energy. In gas separation, hollow fiber or spiral wound module is used and that one makes use of selective non porous membranes. Membrane gas separation is considered an alternative mechanism to the more conventional processes such as adsorption and absorption. Oxygen recovery, natural gas sweetening, recovery of hydrogen in steam and removal of carbon dioxide in methane processes using polymeric membranes are some of the applications of membrane technology (Fig. 8). Hollow fiber membranes are advantageous in that they have a good area to volume ratio, can be densely packed, and can be desired to not only withstand mechanical strength but also to be mechanically stable which as well as resistant to fouling. Moreover, hollow fiber membranes are some of the cheapest membranes in the market since they do not need supporting substance. [13].

On the other hand, adsorption is one of the most promising methods for reducing costs associated with the hold step in carbon capture. It typically has high CO_2 capture capacity, high selectivity, low-cost loot material, and adsorption capacity remained constant after several cycles.

Layered double hydroxide (LDHs) are inorganic compounds and belong to the group of two-dimensional clay minerals. They are made up of several cations (metals in the inner layers) and interlayers made of anions. the constitution of LDHs consists of metal hydroxide metals with a positive charges, like in brushite [$Mg(OH)_2$], and anions and water molecule that counterbalance the charges between the layers. The possibility of using LDHs in high temperature carbon capture and storage is very promising. They do not need a lot of energy to regenerate during adsorption

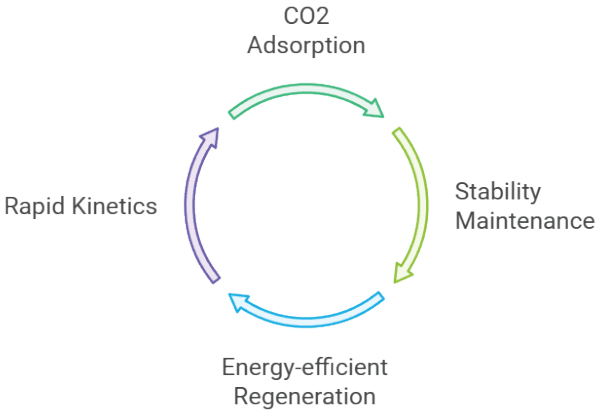


Figure 9. LDHs in carbon capture cycle.

of CO_2 and are more stable than certain solid adsorbents. Moreover, LDHs have fast adsorption and desorption kinetics, especially under the influence of water, and thus would be a promising choice to use in pre-combustion CO_2 capture processes (Fig. 9) [14].

So far, the adoption rate of CO_2 capture technologies has been slow. Though positive signals are that there are changes that are likely to take place, the industry is not beginning as a complete one. It will not be realized that these technologies will make any real difference in reducing emissions until many years down the line when capture, transport and storage infrastructure can be built [15,16].

As outlined by the net-zero emissions (NZE) scenarios of the International Energy Agency (IEA) on future stage (2050), the carbon capture and storage have a considerable potential. The aggregate ability to capture carbon dioxide (CO_2) is even estimated to rise to 1.2 gigatons (Gt) of CO_2/year by the year 2030 and beyond at 7.6 $GtCO_2$/year by 2050. It has been reported by Intergovernmental panel on climate change (IPCC) that a total of 15 $GtCO_2$ of carbon reduction will be required by the year 2050 in case average temperature increase is to be maintained at 1.5 C. Likewise, the Energy Transformation Commission (ETC) estimates

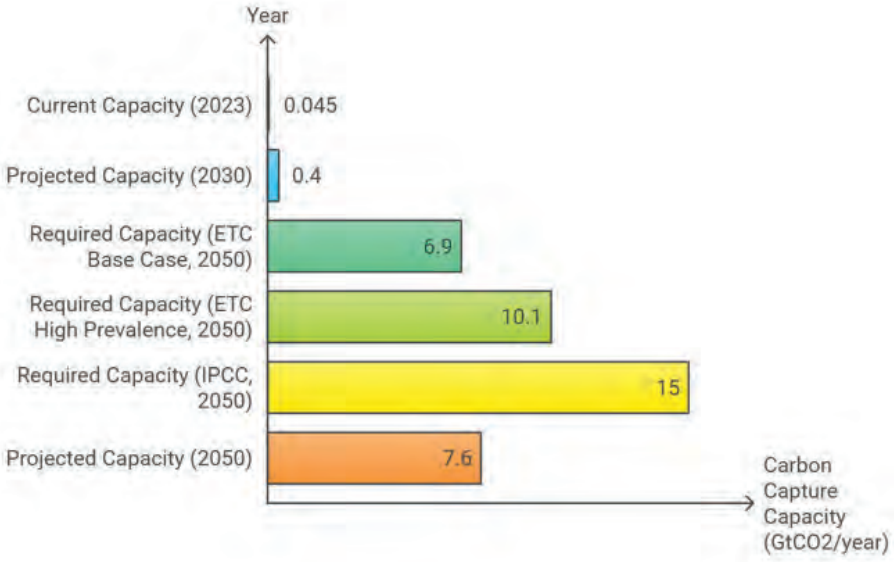


Figure 10. Carbon capture capacity and emissions reduction targets.

that up to 6.9 Gt (base case) to 10.1 Gt (high prevalence) of carbon dioxide emissions have to be captured per year to achieve net-zero by 2050 (Fig. 10). According to ref. [17], indicates the total amount of carbon that was captured in 2023 was about 45 million tons because of an IEA report. Although deployment momentum has continued to improve and approximately 200 new capture facilities have been announced as becoming operational by 2030, with all the developments, the overall annual capacity is estimated to reach almost 400 million tons of CO_2 by 2030. This is still very small against the amounts needed to achieve the net-zero goal of 2050 period [17].

The current buzzword in climate change mitigation is carbon capture and storage (CCS). The goal of CCS is to capture emissions produced during the production and processing of fuel. Nevertheless, it fails to represent the majority of the emissions, which take place down the line. Moreover, CCS consumes a lot of energy to run and therefore in the case that the energy is obtained through fossil fuels, there is a risk of increased emission. This critical review shows that in certain situations, CCS may result into a net increase in emissions. A lot of studies that seek to illustrate the possibility of CCS do not incorporate a more elaborate life cycle analysis (LCA) to cloud the actual ineffectiveness of technology. Although CCS is among the few solutions that can be employed in the reduction of emissions in sectors that are hard to decarbonize, like cement industries, renewable sources of energy are eclipsing it.. Furthermore, CCS is sometimes proposed to expand fossil fuel production [18].

The operation of a CO_2 capture system reduces the overall efficiency of power generation and other processes. This leads to increased fuel requirements, more solid waste, and greater environmental impacts compared to an equivalent facility without capture. Therefore, minimizing the energy needed for CO_2 capture and enhancing the efficiency of energy conversion processes will be key priorities for future technological development aimed at reducing environmental impact and costs. Carbon dioxide is currently separated into large industrial facilities, such as those involved in natural gas processing and ammonia production. However, these plants primarily remove carbon dioxide to meet their production requirements rather than for storage purposes. While power plants are significant sources of carbon dioxide emissions, carbon capture technology has yet to be implemented in these facilities [8].

There are three main approaches for capturing CO_2 in industrial and power plant applications (as introduced in Figure 11) [8]:

- **Post-Combustion Systems**: Captures CO_2 from flue gases after combustion. This is the most common method currently used.
- **Oxy-fuel Combustion** uses oxygen instead of air for combustion, producing flue gas that is mainly H_2O and CO_2 and is readily captured; this option is still under research and development.
- **Pre-Combustion Systems**: Captures CO_2 before combustion through gasification or oxy-fuel combustion. This method is more energy-intensive but can capture a higher percentage of CO_2.

Monitoring, risk, and legal aspects associated with CO_2 capture systems present no new challenges, as they are all elements of the industry's long-standing health, safety, and environmental control practices [8].

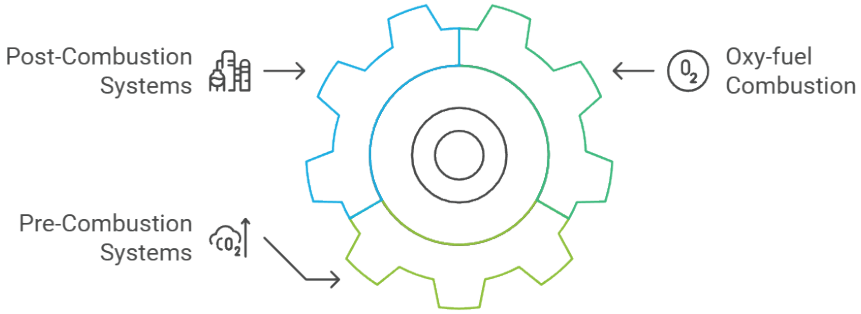


Figure 11. Carbon capture methods industrial and power plant.

Hydrogen

Hydrogen (H_2) is the most abundant element in the universe and has recently gained recognition as a crucial component of decarbonization strategies in many countries [19].

Hydrogen can be produced from both biological and non-biological sources, using various production pathways categorized by color, which indicates the type of feedstock, power source, and technique employed [20].

Hydrogen can have numerous applications, including the possibility of hydrogen being a raw material (energy storage) or a fuel source. Notably, hydrogen does not release carbon dioxide when it burns up [19].

In its normal state, hydrogen is a flammable, colorless, and odorless gas, comprising approximately 75% of the material in the universe, by mass. It is capable of holding a lot of energy and it is in its gaseous state at ambient pressure, although it can be converted to liquid at very low temperatures -253°C. To facilitate storage, hydrogen is typically stored in cylinders under high pressure, usually at 350 or 700 bar, and in some cases up to 1,000 bar. This high-pressure storage also allows for the transportation of hydrogen in copious quantities over significant distances [21].

It's important to note that hydrogen acts as an energy carrier rather than an energy source. The method of hydrogen production can be identified through its "color," which corresponds to the source used. Additionally, the price of each type of hydrogen is influenced by the source and efficiency of the production technology. Often, the cost of hydrogen production is intricately linked to electricity prices [21].

The main colors of hydrogen can be explained as shown in Figure 12, but other colors, such as yellow, black, and brown, are also associated with hydrogen production [20].

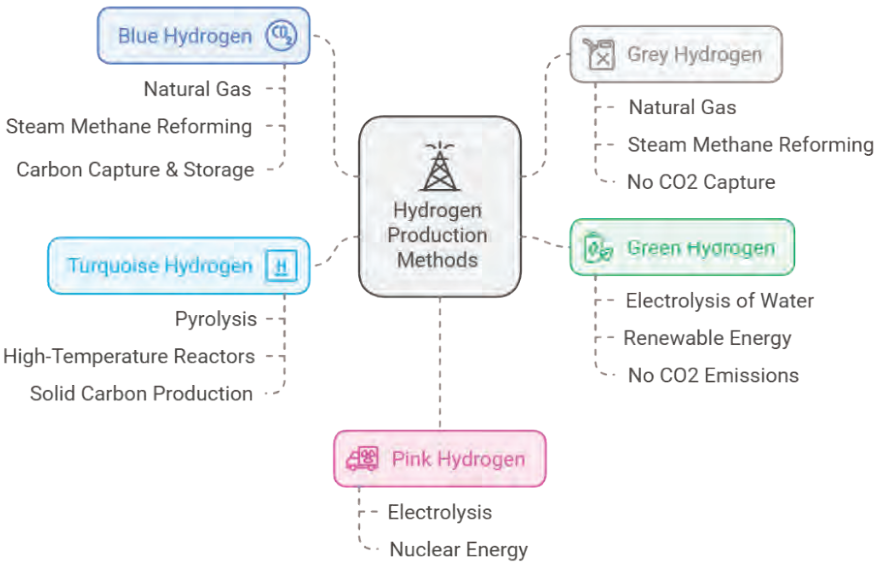


Figure 12. Main hydrogen colors production pathways.

Methane

Methane is a colorless and odorless gas that forms abundantly in nature and is also produced by certain human activities. As the simplest member of the paraffin series of hydrocarbons, methane is one of the most potent greenhouse gases [22].

Although it constitutes only 3% of greenhouse gas emissions by mass, methane has the potential to cause global warming that is 84 times more powerful than that of carbon dioxide over a period of 100 years. Therefore, methane is often used as a benchmark in climate change projections. It accounts for approximately 23% of the overall impact of global warming [23,24].

There is a unique opportunity to mitigate methane emissions, as the most cost-effective technologies and practices to reduce emissions from the main human contributors are already widely available and in use around the world. Reducing methane emissions not only helps mitigate climate change but also offers various energy, health, and safety benefits, along with positive local environmental impacts (Fig. 13). Many methane abatement technologies and practices also lower emissions of volatile organic compounds and hazardous air pollutants, improving local air quality.

Projects aimed at reducing methane from landfills and wastewater treatment plants can eliminate unpleasant odors. In agriculture, such practices manage manure to safeguard the local water bodies and environment. Coal mine-gas is a gas that is very safe to capture because it helps in reducing chances of explosions that could result to industrial accidents. Low emitting equipment and enhanced management techniques in oil and gas systems can reduce the emission of methane gases, which has health and safety advantages but adds to the quantity of products released to the market and adds to the revenues. In addition, the recovered methane can be used as a local source of clean energy and as one that fosters economic development. It can replace energy sources that have higher CO_2 emissions and pollutants, such as wood, coal, and oil. Additionally, recovered methane can become a sustainable and abundant energy source in developing countries [25].

The major sources of atmospheric methane may include the following:

1. Agricultural Activities
2. Oil and Gas Operations
3. Coal Mining
4. Landfills
5. Wastewater Treatment

The following is an introduction and explanation of each one of them:

Agricultural Activities

Agriculture accounts for 40% to 46% of global methane emissions, and with the increasing demand for food production, these emissions are projected to rise by approximately 40% by 2050. Enteric methane from ruminant livestock contributes to about two-thirds of these emissions, while rice production accounts for 20%, and managed manure adds another 7% (Fig. 14). Several scenarios have been developed to mitigate agricultural methane emissions by 54% compared to potential emissions in 2050, and by 36% compared to current levels.

The mitigation opportunities include [26]:

- Improving feed efficiency for ruminants
- Promoting alternative sources of animal protein
- Implementing effective water management in rice fields
- Establishing a unified mechanism to reduce methane emissions from managed manure
- Raising awareness and addressing food waste globally.

Oil And Gas Operations

In 2020, oil and gas operations worldwide emitted more than 70 million tons of methane, according to the International Energy Agency (IEA). Most of this data was

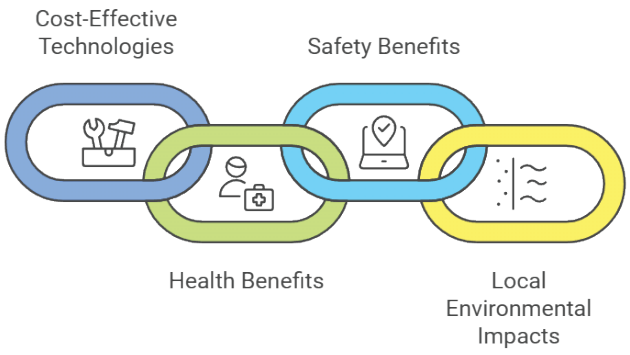


Figure 13. Methane emissions mitigation.

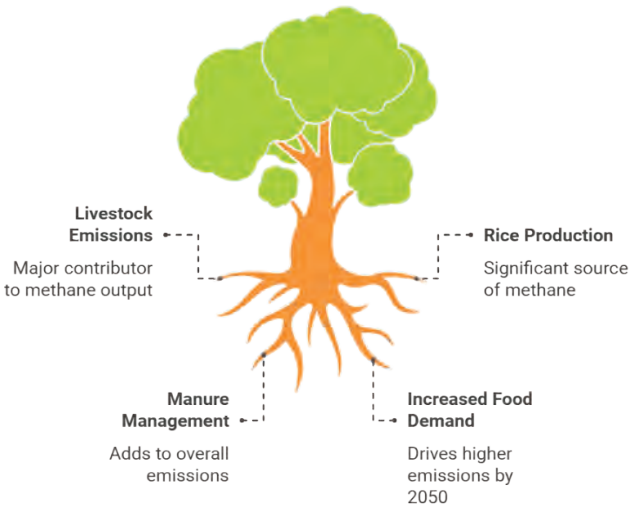


Figure 14. Methane emissions sources from agriculture.

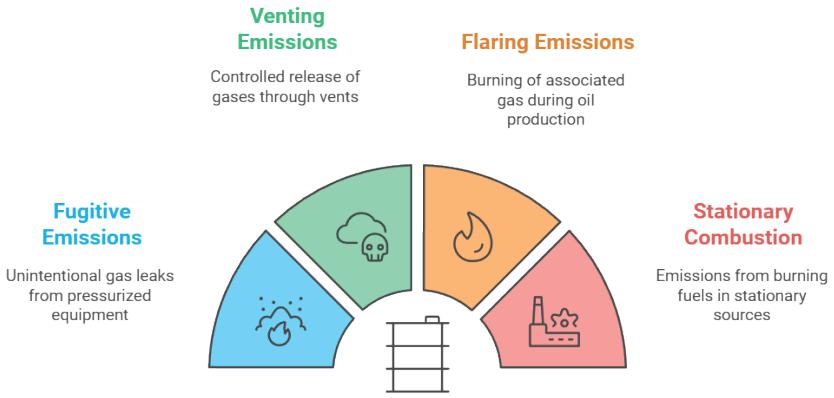


Figure 15. Methane emissions from oil & gas operations.

derived from upstream operations, which include drilling, production, and treatment.

Methane emissions, in oil and gas industry are classified into four categories based on the GHG emission classification and international standards, as follows (Fig. 15) [27]:

- Fugitive: emissions resulting from unintentional loss of the gas phase from equipment containing over-pressurized fluid.
- Venting: emissions related to the intentional control release of gases directly into the atmosphere resulting from the process design, most typically through a vent pipe, seal, or duct.
- Flaring: emissions resulting from burning the "associated" gas that accompanies oil production or simply releasing it to the atmosphere.
- Stationary Combustion: Emissions from stationary combustion include boilers, heaters, furnaces, kilns, ovens, flares, thermal oxidizers, dryers, and any other equipment or machinery that combusts carbon-bearing fuels or waste stream materials.

Coal Mining

Methane is produced during the process of coalification, which is the formation of coal. A small portion of this methane remains trapped in the surrounding rock layers and coal seams (Fig. 16). This trapped methane is released during the fracturing of the coal seam.

Generally, deeper seams contain a higher concentration of gas compared to those accessed by surface or open-pit mining. As a result, underground coal mining tends to release more methane into the atmosphere [28].

Landfills

Landfills release mostly methane when organic waste is decomposed by bacteria. The landfill gas primarily consists of methane and carbon dioxide (90% to 98%), while the remaining 2% to 10% includes nitrogen, oxygen, ammonia, hydrogen, and other gases. Figure 17 explained the factors contributing to methane emissions from landfills as fish bone diagram.

In landfills, any increase in temperature or moisture can lead to a higher gas output. Although gas production typically peaks within five to seven years, a landfill can continue to generate gases for over 50 years [29].

Wastewater Treatment

In 2010, methane emissions from wastewater were estimated to produce 512 million metric tons of CO_2 equivalent ($MMTCO_2E$) globally, accounting for nearly 7% of all methane emissions worldwide. Methane is released during the handling and treatment of municipal wastewater

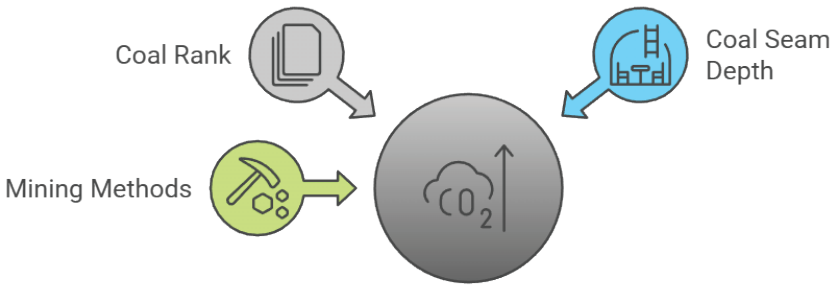


Figure 16. Factors contributing to methane emissions from coal minig.

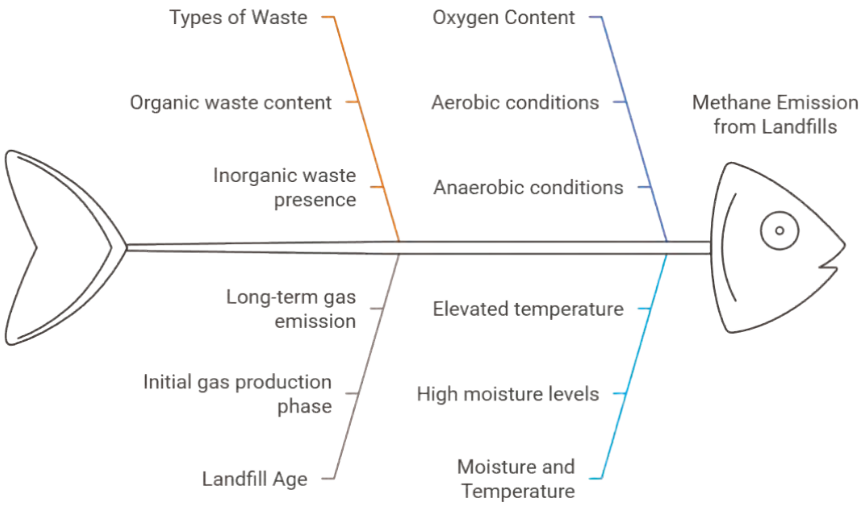


Figure 17. Factors contributing to methane emissions from landfills.

because of the anaerobic decomposition of organic materials. Most developed countries use centralized aerobic wastewater treatment systems to collect and process municipal wastewater. While these systems generate lesser amounts of methane, they also produce significant quantities of biosolids, which can lead to higher rates of methane emissions (Fig. 18) [30].

In conclusion, Table 2 compares the major sources of methane emissions and their description and challenges for each source [26-30]. From these five sources, methane can be used in some industrial applications such as hydrogen for ammonia synthesis, acetylene by exposure to an electric arc, methyl chloride by chlorination, hydrogen sulfide by reaction with sulfur, oxidation products such as methanol, formaldehyde, formic acid, and nitromethan [31].

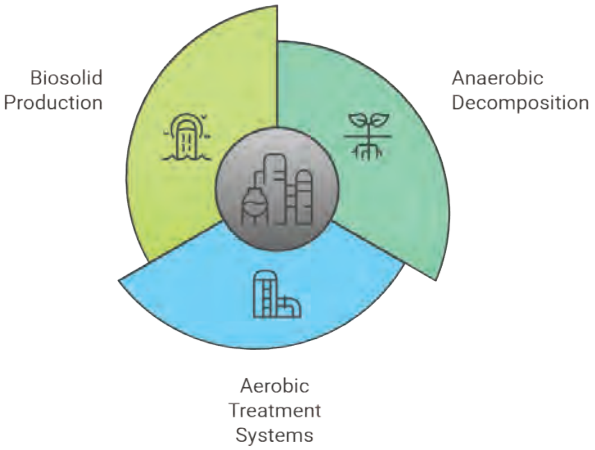


Figure 18. Methane emissions from wastewater treatment.

Extracting Carbon and Hydrogen From Methane

Methane can be used to make hydrogen and carbon by engaging in several chemical reactions. A closed-loop system can be employed by using hydrogen and carbon dioxide that can be used in many industrial applications and processes to increase energy efficiency and decrease emissions [32]. In this section, the various processes along with chemical reactions that use methane to form hydrogen and carbon are introduced as follows:

Steam Methane Reforming (SMR)

One of the ways of producing syngas is the reaction of methane commonly obtained through natural gas with water steam. Methane is reactive with water steam to form carbon monoxide and hydrogen in this process. Also, more water-steam could be used with carbon monoxide to produce carbon dioxide and additional hydrogen. A chemical equilibrium can be represented as this series of reactions::

$CH_4 + H_2O \rightarrow CO + 3H_2\ \Delta H_{SR} = 206$ kJ/mol

$CO + H_2O \rightarrow CO_2 + H_2\ \Delta H_{SR} = -41.2$ kJ/mol

The process has about 65-75 efficiency that is one of the highest recorded rates of the existing commercial approaches to production [33].

Autothermal Reforming (ATR)

ATR reacts oxygen with methane producing syngas in one chamber, where methane is partially oxidized. An equilibrium can be described as follows:

$CH_4 + O_2 \rightarrow CO_2 + 2H_2\ \Delta H_R = -902.3$ kJ/mol

The small to mid-scale ATR processes using oxygen instead of air being preferred to use about 40% is an acceptable reforming technology [34,35].

Methane Pyrolysis

The term "pyrolysis" refers to the process of breaking down molecules through the application of heat. In the case

Table 2. Comparison of major sources of methane emissions [created by author]

No	Major Source	Description	Challenges
1	Agricultural Activities	Agriculture is a major source of methane (CH_4), a potent greenhouse gas that contributes significantly to climate change.	▪ Complex Systems ▪ Require Significant Investments ▪ Technological Limitations ▪ Behavioral Changes ▪ Policy and Regulatory Framework
2	Oil and Gas Operations	The oil and gas industry are a significant contributor to global methane emissions. Methane leaks and venting occur throughout the natural gas supply chain, from production to distribution.	▪ Age and Maintenance of infrastructure. ▪ Remote Operations ▪ Regulative and İmplementing Barriers ▪ Significant Upfront Costs
3	Coal Mining	It is an important origin of methane. Coal seams contain a natural supply of methane which is emitted when disturbances occur to the seams during mining processes.	▪ Technical Challenges ▪ Regulatory Challenges ▪ Safety Concerns ▪ Infrastructure Limitations
4	Landfills	Methane which is a powerful greenhouse gas emits in landfills. An anaerobic (the absence of oxygen) breakdown of the organic material in the landfill emits methane gas.	▪ Waste Composition ▪ Landfill Design and Operation ▪ Regulatory Framework ▪ Public Awareness and Behavior
5	Wastewater Treatment	Methane may be a major source of wastewater treatment plants, especially those that take part in anaerobic digestion. Anaerobic digestion is a natural process that organic matter is dismantled by microorganisms without oxygen, resulting in the formation of biogas, which are mainly made up of the matter of methane and carbon dioxide.	▪ Process Optimization ▪ Energy Efficiency ▪ Methane Capture and Utilization ▪ Public Perception and Regulatory Framework

of methane, heat is used to decompose methane molecules, resulting in the production of hydrogen, often referred to as "turquoise hydrogen." Additionally, the pyrolysis of methane generates solid carbon as a byproduct, which can be easily stored and repurposed in various industries, such as tire and rubber production. Methane pyrolysis has an energy efficiency of 58%. [36].

$CH_4 \rightarrow C_s + 2H_2\ \Delta H_\circ = 74.91$ kJ/mol

Using Hydrogen In Energy Processes

Hydrogen storage is a crucial aspect of the hydrogen value chain that has not been thoroughly addressed until an integrated hydrogen economy is established. The absence of a viable business model for hydrogen as either an internationally or locally traded commodity overlooks the fact that, in many cases, hydrogen needs to be stored immediately after production and before it reaches the end user. Currently, hydrogen is primarily consumed close to where it is produced, mainly for industrial purposes such as oil refining and ammonia fertilizer production. As a result, the significance of storage in the hydrogen value chain has often been underestimated. To enhance hydrogen circulation, it is essential to consider the characteristics and challenges related to storing hydrogen in various quantities for different durations [19].

As shown in Figure 19, hydrogen can be stored in three combined cases (gaseous, liquid and solid), where it can be divided into six main types according to different conditions:

1. Compressed Hydrogen
2. Liquefied Hydrogen
3. Liquid Organic Hydrogen Carriers (LOHCs)
4. Chemical Hydrides
5. Metal Hydrides
6. Porous Materials

Pure Hydrogen Storage

Hydrogen can be stored in a highly compressed form at 700 bar. It can also be kept in a cryogenic liquefied state at -253 °C or in a hybrid form. At 700 bar, hydrogen has a physical limit for volumetric density of 1.3 kg/m³, which provides an acceptable level of energy density at a reasonable cost. Gaseous hydrogen can be stored in closed systems without losses for extended periods. Liquid hydrogen has higher storage densities compared to gaseous hydrogen. At 2 bar, liquid hydrogen achieves a volumetric limit of 2.3 kg/m³, although its boiling point is exceptionally low at -253 °C (Fig. 20). Through greater effort in production and storage, larger energy densities can be reached using liquid hydrogen storage particularly when large amounts of

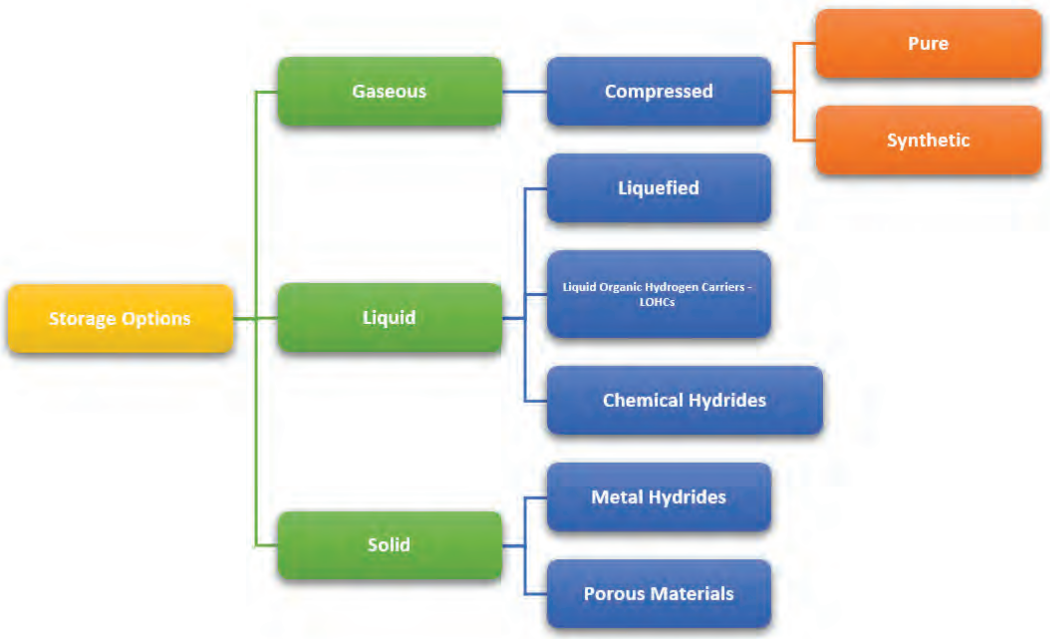


Figure 19. Different hydrogen storage options.

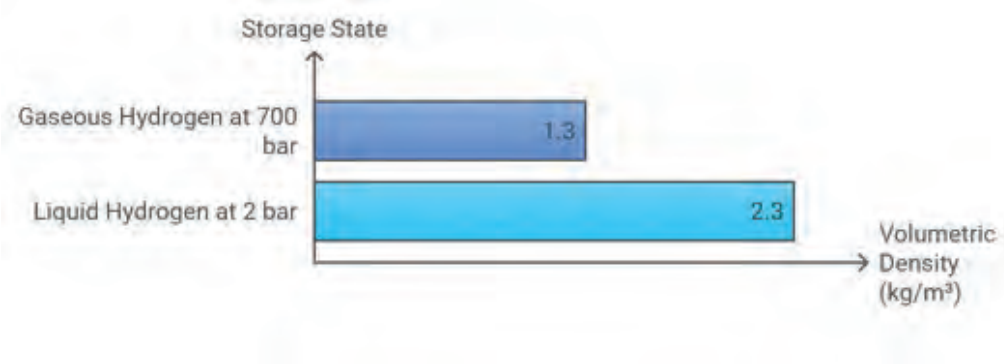


Figure 20. Volumetric density of hydrogen in different storage states.

energy are required, like in centralized production, delivery or space applications such as rocket fuel [37].

Pure hydrogen storage can be utilized centralized; however, it is imperative to adhere to all safety requirements due to its classification as an explosive material.

Synthetic Hydrocarbons

Synthetic hydrocarbons are fuel types produced out of renewable energy sources rather than fossil fuels that are not renewable. They contain equal quantities of chemicals to conventional fuels such as gasoline. These hydrocarbons offer an easy substitute to fossil fuels, such as oil, coal and natural gas as well as all their derivatives. Recent technology demonstrations have demonstrated that one can use a combination of renewable energy with waste carbon sources- such as biomass or recycled CO_2 to generate synthetic hydrocarbons like methanol, dimethyl ether and synthetic methane (Fig. 21). The fuels can be used as an alternative to fossil fuels in a number of industries such as transportation, chemicals, electricity production, and heating. This transition has the potential to decrease greenhouse gas emissions by a large margin and also reduce the dependence on diminishing stocks of fossil supplies. Finally, by capturing and recycling carbon and sending it back to the origin of synthetic hydrocarbon manufacture, the whole process could theoretically be carbon-neutral, a process that is known as closing the loop [38-40].

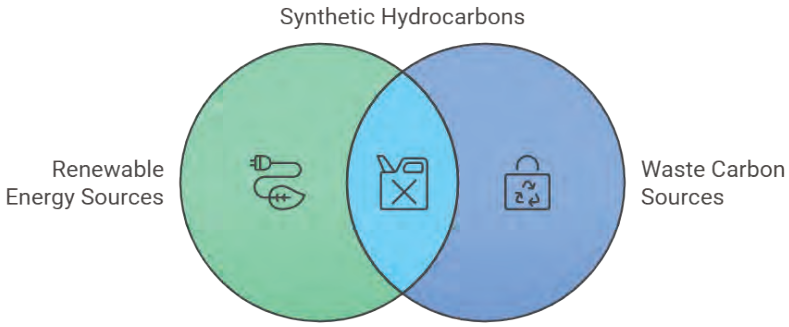


Figure 21. Pathway to carbon-neutral energy.

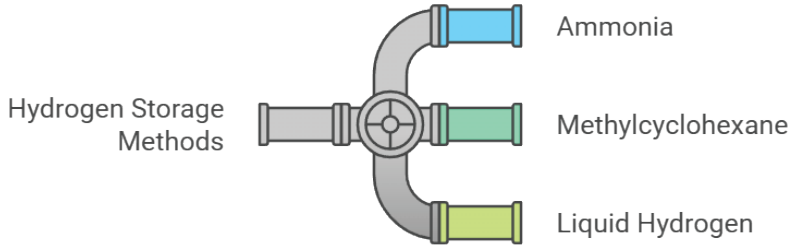


Figure 22. Different chemical hydrides pathway for hydrogen storage.

Chemical Hydrides

Chemical hydride storage methods are one of the many ways to store hydrogen. This category encompasses a wider range of chemicals than synthetic hydrides, yet they generally share similar storage characteristics. The primary advantages of chemical hydride storage include low energy consumption and the absence of a need for infrastructure development. Additionally, their production and dehydration processes tend to require less energy, resulting in lower overall costs. Ammonia, methylcyclohexane, and liquid hydrogen are being researched globally in conjunction with hydrogen storage methods, particularly for the transportation of massive quantities of hydrogen over long distances (Fig. 22). While practical applications of these techniques are promising, they remain in the experimental phase and are awaiting future results [41].

Liquid Organic Hydrogen Carriers (LOHCS)

Liquid organic hydrogen carriers (LOHCs) operate on the principle that organic molecules can chemically bond with hydrogen. Unsaturated hydrocarbon molecules, such as ammonia and methanol, contain high-density double bonds. Storing and transporting hydrogen using LOHCs has a low hazard profile. While a high flashpoint does not eliminate all risks, it, combined with storage and transport at ambient temperature and pressure, significantly reduces the likelihood of leaks, sprays, ignition, and fires. However, the primary concern associated with LOHCs is their aquatic toxicity, see Figure 23 for the pros and cons of using LOHC [42] .

Currently, some of the most promising LOHCs that are attracting the most attention from researchers are as follows:

- Toluene/Methyl-Cyclohexane MCH (C_7H_8/C_7H_{14}),
- Naphthalene/Decalin ($C_{10}H_8/C_{10}H_{18}$)
- Benzene/Cyclohexane (C_6H_6/C_6H_{12})

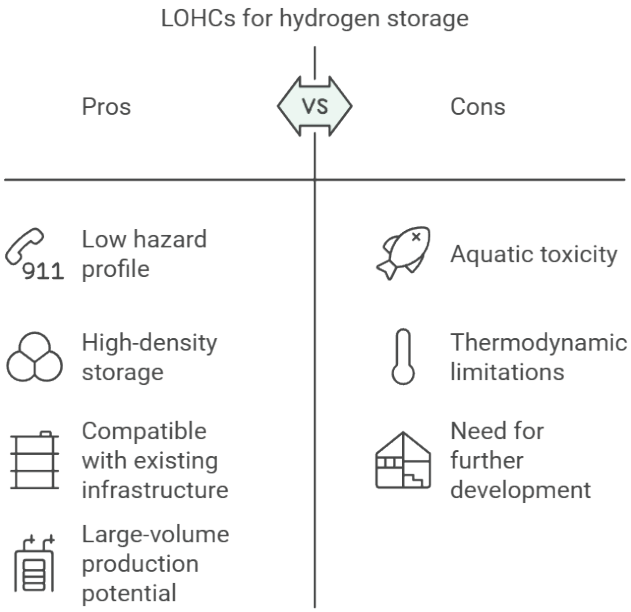


Figure 23. LOHCs pros & cons.

- Dibenzyl/Toluene DBT
- Perhydro-dicenzyl toluene PDBT ($C_{21}H_{20}/C_{21}H_{33}$)

LOHC systems have several attractive characteristics that are like those of crude oil, making them well-suited for integration with existing energy infrastructure. Among these systems, those based on toluene and dibenzyl toluene are particularly suitable for large-scale production. However, further development of LOHC technology is necessary, particularly from a thermodynamic standpoint. Lower dehydrogenation enthalpies would lead to reduced heat requirements and enhance both the molecular stability and recyclability of the candidate molecules over multiple cycles [43,44]

Metal Hydrides

Metal hydrides represent a diverse range of materials that can stabilize hydrogen in a concentrated solid form. In this process, hydrogen bonds with metallic elements and alloys, allowing for some of the highest volumetric densities of hydrogen storage available. In simple terms, hydride is the annoyance of hydrogen. It is a chemical compound in which hydrogen atoms display nucleophilic, basic, or reducing properties.

Some applications of metal hydrides include hydrogen storage and compression, heat pumps and refrigerators, battery electrodes, temperature sensors, and actuators [45].

The boron-nitrogen-hydrogen-containing compounds (BNHC) hold significant promise for enhancing hydrogen storage systems. Among these compounds, ammonium borane (NH3BH3, AB) stands out as one of the most attractive options due to its high hydrogen content (19.6% by weight) and favorable properties such as stability, transportability, and non-toxicity, making it an environmentally friendly hydrogen carrier.

Hydrogen can be efficiently generated from an aqueous solution of AB at a pH value higher than 7 in the presence of catalytic materials through a hydrolysis process under ambient conditions [46].

Overall, the main advantages of metal hydrides are their high volumetric density, low operational pressure, and endothermic reaction during desorption. However, there are also drawbacks, such as the required operating temperature, hydrogen sorption kinetics, cost, and potential pyrophoricity (Fig. 24).

Currently, this option is still under research and development before metal hydrides can become a viable solution for large-scale hydrogen storage [45].

These metal hydrides include the following:

- Ionic or Saline Hydrides: NaH - CaH_2
- Covalent Hydrides: SiH_4 - Al $(BH_4)_3$
- Metallic Hydrides: TiH_2 - $LaNi_5H_6$ - $LiAlH_4$ - $NaAlH_4$

Porous Materials

Porous materials are found throughout the environment and are integral to many aspects of daily life. They are utilized in various applications, including energy management, vibration dampening, thermal insulation, sound absorption, and filtration [47] .

The amount of hydrogen that can be adsorbed by porous substances at normal temperatures and elevated pressures is often quite low, typically falling below 0.5 wt.% (wt% means weight percent). The effect of temperature on hydrogen adsorption in these materials presents challenges for their use in hydrogen storage, particularly when large storage capacities are required [48].

There are two primary types of porous materials capable of storing significant amounts of hydrogen, (Fig. 25 showing the capacity vs. readiness) [49]:

- Carbon-based Hydrogen such as activated carbon, carbon nanotubes, and graphene, are promising candidates for hydrogen storage due to their high surface area, tunable pore structure, and chemical stability.

Figure 24. Weighing the pros and cons of metal hydrides.

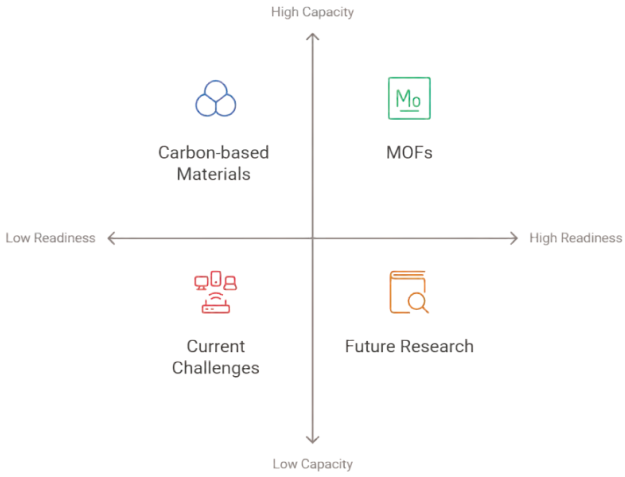


Figure 25. Porous materials for hydrogen storage.

- Metal Organic Frameworks (MOFs) are a class of crystalline, porous materials composed of metal ions or clusters connected by organic linkers. Their highly porous structure, tunable pore size, and large surface area make them promising candidates for hydrogen storage.

Despite numerous efforts to bring these solutions to the market, no one has yet achieved the level of technological readiness necessary for full commercialization, due to the following:

- Require complex synthesis processes.
- It can degrade under harsh conditions, such as elevated temperature, humidity, or exposure to specific chemicals.
- High cost and energy-intensive processes

The following table 3 introduces a comparison of the hydrogen storage options while showing description, advantages, and challenges [19].

Table 3. Comparison of hydrogen storage options

No.	Hydrogen Storage Option	Description	Advantages	Challenges
1	Compressed Hydrogen	It is one of the most common methods for storing hydrogen gas. It involves compressing hydrogen gas to high pressures, typically around 700 bar, and storing it in high-pressure tanks.	▪ Mature technology ▪ High energy density ▪ Flexibility	▪ Energy ıntensive ▪ Safety concerns ▪ Storage tanks are too large and heavy ▪ Efficiency losses
2	Liquefied Hydrogen	Making Hydrogen gas cool to very low temperatures (-253°C) converting it to liquid form. This greatly makes it smaller in volume and therefore easier to store and carry around.	▪ High energy density ▪ Efficient storage and transport	▪ Needs Special Cryogenic Storage Tanks and Packing ▪ Energy ıntensive ▪ Boil-off losses. ▪ Safety concerns
3	Liquid Organic Hydrogen Carriers (LOHCs)	Are the organic compounds that can store/release hydrogen reversibly exist.These substances might exist as liquid at ambient temperature and pressure and hence easier to process and ship compared to gaseous or liquid hydrogen. Safe and convenient to work with.	▪ Safe and Easy to Handle ▪ High Energy Density ▪ Improving with Existing Fuel Infrastructure	▪ Lower Energy Density by Volume ▪ Slow Hydrogen Release ▪ Material Cost
4	Chemical Hydrides	Can work with solid materials able to take up and give off hydrogen gas by way of chemical reactions. The materials provide a possible solution to store large quantities of hydrogen in a small size.	▪ High Hydrogen Storage Capacity ▪ Safe and stable	▪ Slow hydrogen release ▪ Energy-ıntensive ▪ Not as Well Suited to Transportation
5	Metal Hydrides	Are the solid-state materials exist that are capable of adsorbing and desorbing hydrogen gas. This property renders them a possible solution in storage of large quantities of hydrogen in a condensed form.	▪ High hydrogen storage capacity ▪ Safe and stable ▪ Moderate operating conditions	▪ Low Hydrogen Absorption/ Desorption rates ▪ Sensitivity to ımpurities ▪ Limited cycle life
6	Porous Materials	They have high surface areas and tunable pore structures which makes them ideal in adsorbing and storing hydrogen gas.	▪ High surface area ▪ Tunable pore structure ▪ Environmental friendliness	▪ Low storage capacity ▪ Sensitivity to environmental factors ▪ Expensive cost

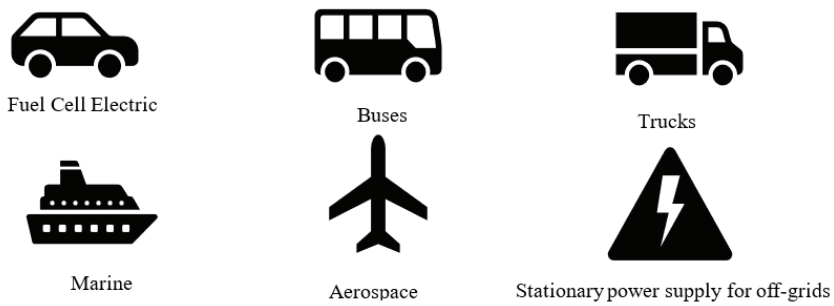


Figure 26. Hydrogen fuel cell applications.

Using Hydrogen Fuel Cells for Transportation and Electricity Generation

Figure 26 shows how the hydrogen is directly used in the transportation sector via fuel cells (FCs). A fuel cell is an electrochemical apparatus that reacts hydrogen and oxygen to create electricity and water and heat is the byproduct. Simply, a fuel cell is composed of two electrodes; the anode and the cathode with an electrolyte in the middle. At the anode, hydrogen is reacted with a catalyst, which produces a positively charged ion, and a negatively charged electron. The electric current then flows through an external circuit as the positively charged ion (proton) flows through the electrolyte and the electron flows through the electron circuit. At the cathode the oxygen reacts with the ions and electrons to produce water and liberate productive heat [50].

Aviation

Decarbonization is one of the largest problems of the aviation industry. The current level of CO_2 emitted by the aviation is more than 900 million tons/year. When the industry keeps growing with an average of 3 to 4 percent / year and the industry becomes more efficient with an average of 2 percent every year, then the CO_2 emissions would be much higher in 2050. In the same timeframe, the Air Transport Action Group (ATAG) has also established an aim to cut down on CO_2 emissions by half of the current situation in 2005 and the European Union has also made a target to be a carbon neutral nation by its Green Deal. Besides the production of CO_2 emissions, aircrafts also produce NOx emissions, soot and water vapor that contribute to the formation of contrails and the formation of cirrus clouds. As a result, the net effect of the aviation to global warming is much higher as compared to the isolated CO_2 emissions. Hydrogen (H_2) propulsion can fundamentally reduce such climate effects. It can break down CO_2 during flights and can be produced in a carbon-free manner. Recent estimates suggest that H_2 combustion alone could cut climate impacts during flight by 50-75%, and by 75-90% when using fuel-cell propulsion, compared to a reduction of about 30-60% with synthetic fuels [51,52].

To scale H_2-powered aviation, several actions need to be taken [53]:

- Lighter tanks (aiming for 12 kWh/kg/ gravimetric index 35%) to improve overall efficiency and fuel cell systems (aiming for 2 kW/kg including cooling).
- Liquid Hydrogen (LH_2) distribution inside the aircraft and Turbines can burn hydrogen with low NO_x emissions.
- The development of high efficiency refueling technologies that can provide the same flow rates as kerosene.

Maritime Applications

The marine sector is responsible for transporting over 80% of global goods through international shipping routes, consuming more than 330 million tons of fuel each year. An important part of transport is also played by ferries; says the United States Bureau of Transportation Statistics of 2015, almost 119 million passengers and 25 million vehicles were transported by ferries in the United States. U.S ferries used 48.6 million gallons of diesel fuel in 2018 and the subsequent 494,000 metric tons of CO_2 emissions each year. The maritime sector is considering fuels like liquefied natural gas (LNG), biofuels, and electric propulsion to cut emissions by using batteries and/or fuel cells using hydrogen (Fig. 27). Renewable hydrogen power can be used to greatly reduce or even eliminate the emissions in marine operations.

Nonetheless, due to the prevailing low cost of low-sulfur marine gasoil (LSMGO), competing of hydrogen fuel in container ship business is not easy. LSMGO is more densely volumetric and that is more preferable in longer voyages. As much space to hold the quantity of hydrogen necessary to supply the energy requirements during such voyages is larger than what a similar lengthy distance by diesel-powered ships would have consumed in terms of cargo carrying and revenue generation. Fuel costs account for approximately 87% of the total cost of ownership (TCO) in the maritime sector. The lower efficiency of fuel cells compared to diesel engines, in terms of energy output during regular operation, does not offer sufficient energy savings to justify the higher cost of hydrogen fuel, which is currently around $2 per kilogram compared to LSMGO [54].

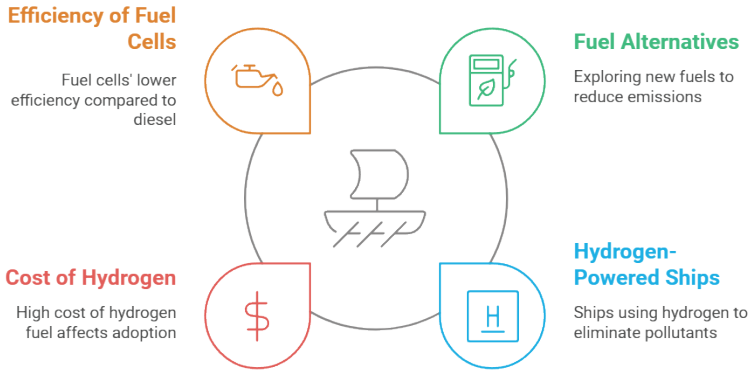


Figure 27. Factors influencing maritime emissions reduction.

Ground Vehicles

There are two primary alternatives to conventional internal combustion engine (ICE) vehicles for reducing greenhouse gas (GHG) emissions:

- Hydrogen Fuel Cell Vehicles (HFCVs).
- Battery Electric Vehicles (BEVs).

Both types of vehicles produce zero tailpipe emissions and have the potential to significantly lower emissions in the transportation sector by utilizing electricity or low-emission hydrogen, (Fig. 28). The current increase in the commercialization of hydrogen-fueled vehicles (HFCVs) is a result of substantial technological advancements. HFCVs have accumulated over 20 million kilometers in real-world conditions using Polymer Electrolyte Membrane Fuel Cells (PEMFC). Since 2002, the costs of fuel cells have decreased by 80 percent, and the durability of HFCVs has improved fourfold, now exceeding 120,000 miles. As of 2018, there were 11,200 HFCV cars on the road worldwide [55,56].

Electricity Generation

Since most of the energy is generated by burning fossil fuels, power generation accounts about for 25% of all carbon dioxide emissions. Hydrogen will play a crucial role in the future energy mix and can help reduce the carbon footprint of power generation.

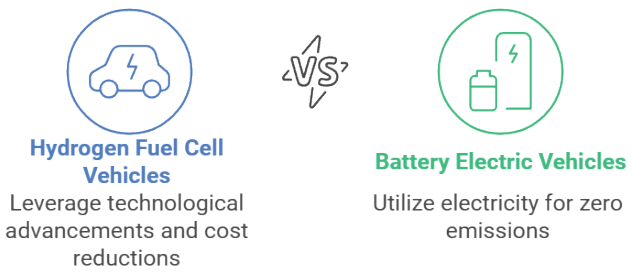


Figure 28. New alternatives to lower-emission vehicles.

Depending on the fuel cell technology, the practical electrical efficiency of fuel cells can range from 40% to 65% at the beginning of their lifecycle. When combined with Combined Heat and Power (CHP) systems, the overall efficiency can exceed 90%, making it an ideal option for hydrogen applications. In the short to medium term, clean hydrogen is likely to remain expensive, which highlights the importance of power generation efficiency in the project's economic model. Since hydrogen prices can differ based on the production process, transportation methods, geographical location, and other site-specific conditions, electrical efficiency plays a significant role in the economics of projects involving hydrogen. A new generation of solid oxide fuel cells (SOFCs) presents several inherent advantages over traditional fuel cell technologies. These advantages include high efficiency, stability, reliability, and fuel flexibility. SOFCs can utilize either natural gas (methane) or hydrogen as fuel for the reaction [57,58] Figure 29 presents the role of hydrogen in energy generation.

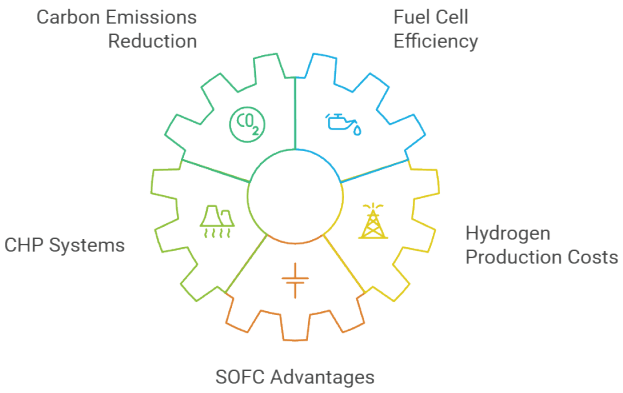


Figure 29. Hydrogen's role in energy efficiency.

DOMESTIC ENERGY

Figure 30 illustrates the domestic use of fuel cells. Along with generating electricity, the heat produced can also be utilized. This process is known as combining heat and power (CHP) transformation. When these systems are applied in the domestic heating sector, they are commonly referred to as micro/mini-CHP plants due to their smaller output capacities [50].

The heat generated by the combination of heat and power is used to partially meet a building's heating needs. Fuel cell heating systems produce low thermal output mainly because they operate primarily on electricity. To address the remaining heating requirements of the building, an additional heating system, such as a condensing boiler, is utilized. Fuel cells are particularly well-suited for structures that require minimal space heating, like low-energy or nearly zero-energy buildings. In contrast, buildings with higher space heating demands use hybrid heating systems that combine a fuel cell with a boiler to satisfy peak heating needs [50,59].

Using Carbon Dioxide In Energy Efficiency & Industrial Processes

Climate change mitigation is a key factor driving interest from governments and stakeholders in leveraging new opportunities to utilize CO_2 for developing products and services. This also aims to promote technological leadership and support the circular economy [60]. Figure 31 introduced some pathways to a circular carbon economy related to carbon dioxide. The use of CO_2 has the potential to significantly contribute to achieving net-zero greenhouse gas emissions in the future. CO_2, a greenhouse gas, can be captured from the Earth's atmosphere, water, or waste gas streams and chemically transformed into commercially viable carbon-based products. The U. S. Congress, as part of the Energy Act of 2020, teamed up with various national academies to examine the prospects and obstacles of moving forward with improvements in CO_2 capture technology, infrastructure, and CO_2-based product markets. The goal is to integrate carbon waste streams in a circular carbon economy. The application of carbon is a core element to

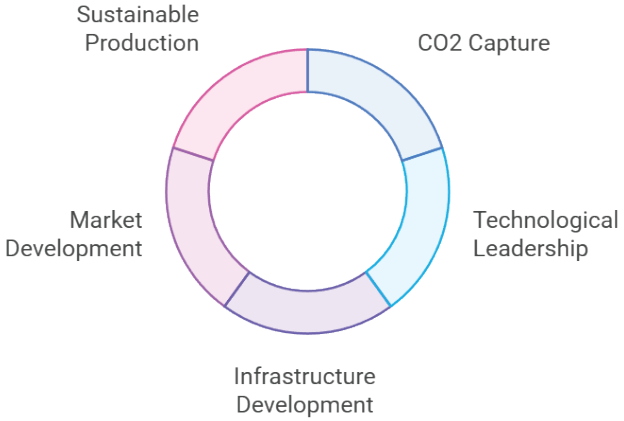


Figure 31. Pathways to a circular carbon economy.

several products, namely plastics and commodity chemicals. Nevertheless, the existing production technologies are based on the extensive use of fossil fuels and carbon input based on heat, power, and transport powered by fossil fuels. A net-zero-carbon economy will require that the carbon-based products be produced in a sustained manner that no net CO_2 emissions are emitted into the atmosphere.

Today, industrial-scale CO_2-based products include urea, organic carbonates, methanol, salicylic acid, and CO_2-treated concrete. However, these products generally do not achieve net-zero emissions at any point in their lifecycle. Pilot-scale initiatives are being developed to create additional commercial products, but they face challenges such as a lack of low-carbon product incentives, the inherently higher costs and energy requirements of using CO_2 for hydrocarbon products compared to readily available fossil carbon, and the need for clean hydrogen, electricity, and heat as inputs. CO_2 is subject to a chemical change that creates long- and short-lived products. Indicatively, the concrete and aggregates may last over 100 years since they are capable of storing carbon. These materials are sustainable to be manufactured by fossil, atmospheric or bio-based CO_2 as in Figure 32. In contrast, short-lived products, such as chemical intermediates and many plastics, have shorter

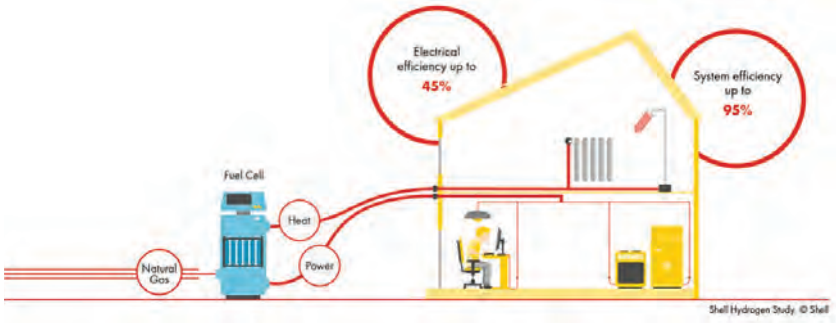


Figure 30. Hydrogen fuel cells in domestic energy usage [created by author].

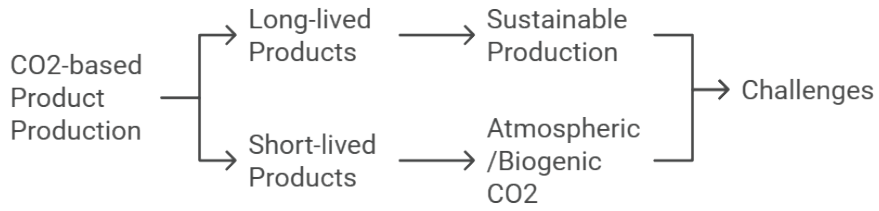


Figure 32. CO_2-based products pathway.

lifespans and only temporarily remove CO_2 from the atmosphere because they eventually degrade back into CO_2 at the end of their life cycle. Therefore, the production of short-lived products requires CO_2 derived from either the atmosphere or biogenic sources in a way that allows the embodied carbon to contribute to the circular flow of CO_2 in the atmosphere [61].

The key categories of products and services derived from CO_2 that are anticipated to see new developments in the short term include the following:

- Fuels
- Chemicals
- Mineral building materials
- Building materials from waste
- Enhancing the yields of biological processes

Each of these categories has the potential to be scaled up to a market size of at least 10 million tons of CO_2 per year, which is comparable to the current CO_2 demand in the food and beverage industry. At present, the technologies for utilizing CO_2 in products and services are still in the development stage, as illustrated in Figure 33 [60].

When an application is scalable, utilizes low-carbon energy, and replaces a product with higher life cycle emissions, the use of CO_2 can contribute to achieving climate goals. In the short term, the market for CO_2 utilization will likely remain small, particularly for early opportunities related to building materials. Additionally, public purchases of low-carbon products can help develop technological standards and create an initial market for these goods. Furthermore, CO_2 sourced from biomass, or the air can play a significant role in ensuring a net-zero economy in the long run, as it can be used as a carbon source for jet fuel and chemicals. [60]. Carbon dioxide (CO_2) can be utilized in several ways. It can be used directly, meaning it retains its chemical state, or indirectly, where it participates in multiple chemical and biological processes. These processes include the production of fuels, chemicals, and building materials.

The interest in CO_2 stems from its potential to help achieve climate objectives. This interest is combined with technological innovation, energy security, and expectation of low-cost, abundant renewable energy sources, which may reduce CO_2 conversion technologies to be more economical. Moreover, the application of CO_2 can be a low-volume source or a low-volume option [62].

The Emerging Market Opportunities for CO_2

Recently, there has been an increased government and industry interest in using CO_2 in other applications mainly due to the need to reduce climate change. The 2015 Paris Agreement established a framework governing the process of intensifying climate programs with the objective of ensuring that the increase in the global average temperatures is below 2°C beyond preindustrial levels. Power plants are responsible for 80% of global CO_2 emissions from large stationary sources, while the remaining 20% originates

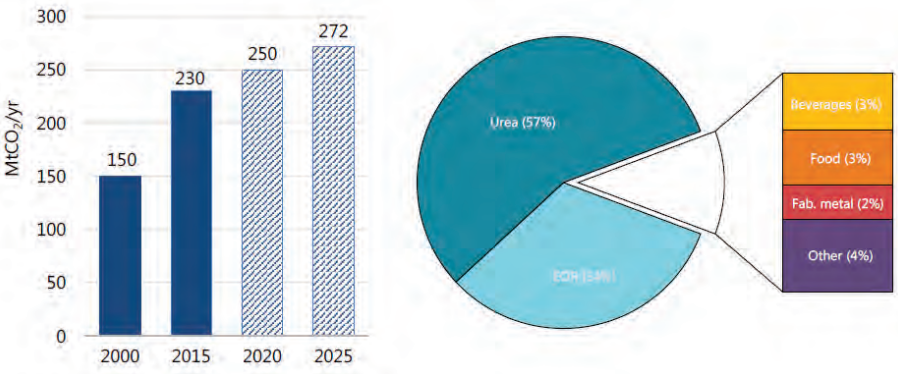


Figure 33. Growth in the global demand for CO_2 over time [created by author].

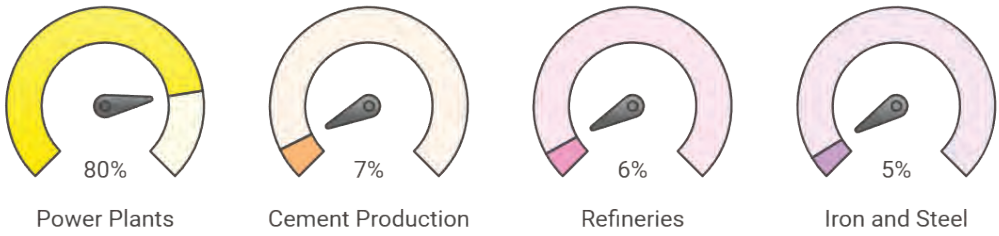


Figure 34. Global CO_2 emissions from different industrial applications.

from industrial sources. The major contributors in this category include cement production plants (7% of total emissions), refineries (6%), and the iron and steel industry (5%), as shown in Figure 34. While modern technologies contribute a smaller overall percentage of emissions, the concentration of CO_2 in flue gas from these industrial sources can be higher than that from power plants. As a result, CO_2 capture processes in industrial settings may be more efficient and can lead to competitive costs for the CO_2 that is avoided compared to power plants. It is also essential to recognize that industrial sources can represent a sizable portion of total emissions, especially in countries where low-carbon energy sources such as nuclear power as in France, and renewable energy (for example, in Scandinavian countries) play a key role in reducing overall emissions. The potential for CO_2 utilization may help in lowering overall emissions and is one of many strategies for mitigating CO_2. Even in a carbon-free economy, carbon will still be necessary for various applications [63,64].

Some examples of emerging applications of CO_2 in industry include the following:

The Cement Industry

CO_2 can be utilized as an input in the concrete production process. In principle, concrete produced with CO_2 can be used for the same applications as conventional concrete, provided that its material properties are similar to or better than those of traditional concrete (Fig. 35). CO_2 curing concrete technology is an innovative method that enhances the performance of cured concrete while being cost-effective and environmentally friendly. This green, low-carbon technology enables the carbonatable components in fresh concrete and the calcium silicate in cement clinker to react with CO_2, resulting in the formation of calcium carbonate and silica gel. The raw materials used in CO_2 curing concrete include aggregates, a specific water-to-binder ratio, CO_2 concentration, pressure, and pre-treatment processes. When concrete undergoes CO_2 curing, its micromorphology changes significantly. The surfaces of cement particles and the slurry are coated with calcium carbonate crystals and silica gel. This combination of silica gel and calcium carbonate interlocks, reducing the overall porosity of the concrete and increasing the density of its microstructure.

Figure 35. CO_2 curing concrete process.

Subsequent CO_2 curing leads to a reduction in the inner porosity of the concrete, resulting in a more compact structure that boasts increased hardness, density, and improved resistance to freeze-thaw cycles, weathering, and sulfate exposure. The transition from conventional concrete to CO_2-treated concrete is projected to create a demand for approximately 1,000 $MtCO_2$ worldwide today, increasing to 1,200 $MtCO_2$ by 2030. The annual utilization of 10 $MtCO_2$ corresponds to about 300 million tons of CO_2-cured concrete [65,66]

Chemical Industry

The utilization of CO_2 in chemical and biotechnological processes involves using it as a reactant in various chemical and biochemical reactions. This concept of employing CO_2 in chemical reactions is not new; several established reactions are already based on this idea. Products derived from CO_2 include basic chemicals, polymers, specialty chemicals, and synthetic chemicals. However, it can be challenging to categorize CO_2-based products strictly as chemicals or fuels. For example, some substances, like methanol, fit into both categories, as illustrated in Figure 36. Currently, approximately 110 million tons of carbon dioxide are used as a raw material in the chemical industry. The primary applications of CO_2 as a raw material include

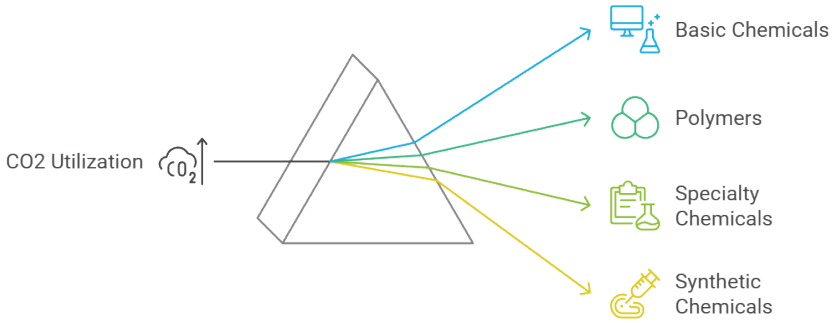


Figure 36. Diverse pathways of CO_2 utilization.

the production of synthesizers and products such as urea, methanol, polymers, and formic acid [67].

However, the following are some chemical industries using CO_2 in their processes:

Methanol Synthesis

The industrial process that utilizes CO_2 as a reactant involves adding small amounts of it to hydrogen-rich synthesis gas to produce methanol. Currently, the world produces approximately 80 million metric tons (Mt) of methanol annually, using about 2 million tons (Mt) of CO_2 in the process. While the direct synthesis of CO_2 and hydrogen from methanol has not yet been implemented on a large industrial scale, it is already in use at a few small commercial plants, Figure 37, show the taken steps in synthesis methanol production [67].

Synthesis of Urea

The urea synthesis reaction involves the combination of carbon dioxide (CO_2) and ammonia to produce urea. Urea serves as a nitrogen fertilizer and as a raw material in the chemical industry to produce various chemical products, including urea resins. These resins can be used for adhesives, impregnation, and insulation purposes. Approximately 150 million metric tons of urea are produced annually, using about 110 million tons of CO_2 in the synthesis process. Typically, the synthesis units for ammonia and urea are integrated, meaning that the CO_2 used in the urea synthesis is sourced directly from the ammonia production process [68].

Polymers

Polymers are chemical substances made up of many repeating units of the same groups of atoms, organized into linear or branched chains. The term "plastic" typically refers to synthetic polymers. The production of plastics involves linking together chemicals known as monomers, which form the polymer.

Carbon dioxide (CO_2) can be used as one of the monomers and can be incorporated directly into the polymer chain. Alternatively, the polymer can consist of monomers that have been synthesized beforehand using CO_2. The chemical industry foresees that plastic products that are based on CO_2 will result in innovative products that possess new or enhanced material properties. Moreover, the use of CO_2 in plastic production could be more sustainable,

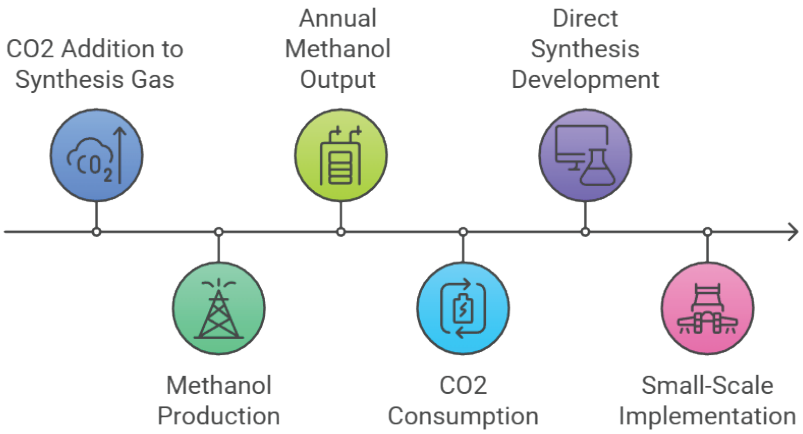


Figure 37. CO_2 Utilization in Methanol Production.

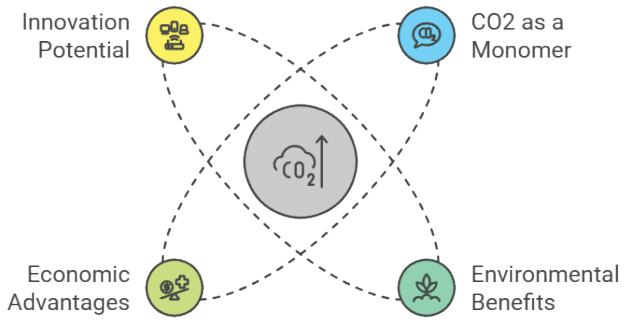


Figure 38. The role of CO_2 in plastic production.

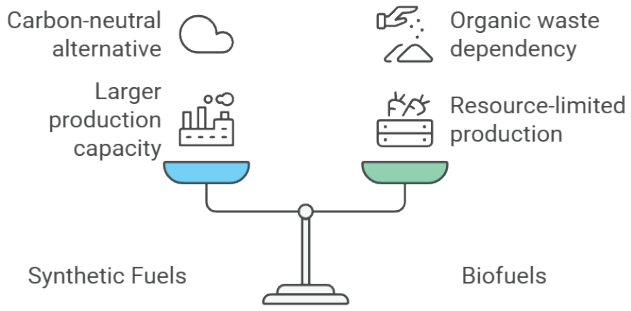


Figure 39. Comparing synthetic & biofuels: capacity & resources.

leading to a decrease in the processing of CO_2 emissions and the dependence on fossil fuel.

One of the key advantages of CO_2-derived compounds is that they provide a carbon-containing molecule that is often less expensive and has a smaller environmental impact than its fossil fuel-based counterparts, as introduced in Figure 38. Some chemical feedstocks cannot be decarbonized because carbon is essential for their structure and properties. Along with biomass and waste, carbon dioxide (CO_2) is one of the few carbon-building blocks that can serve as a raw material for carbon-containing chemicals [69].

Synthetic Fuel

Synthetic fuel is not a new concept. In general terms, it refers to any fuel that is not derived from fossil fuels. Synthetic fuels can be used to prevent the green house gas emissions which would arise due to the combustion of fossil fuels like petrol and kerosene yet they offer the same energy density and meet the international standards of fuel. Synthetic fuels can be used as an alternative to fossil fuels since they are a carbon-neutral source of energy and can be used to curb the frequency of fossil consumption among us. Synthetic fuels do not require limited resources like the use of carbon and hydrogen through the input of organic waste feed into biofuels and thus can be reproduced in large volumes.

The production of synthetic fuels can utilize carbon dioxide from various sources, including point sources like industrial emissions, or through carbon removal techniques such as direct air capture. Figure 39 presents a comparing between synthetic & biofuels according to capacity and resources. With supportive government policies emerging in many countries to unlock the potential of synthetic fuel production, now is an ideal time to establish future markets and regulatory frameworks that can stimulate demand [70].

FOOD & BEVERAGE INDUSTRY

Carbon dioxide is commonly used in the beverage industry for carbonation. It also plays a crucial role in controlling atmosphere storage and modified atmospheric packaging, which helps reduce spoilage, eliminate insect pests, and extend the shelf life of packaged foods such as grains, fruits, vegetables, meats, and processed foods. Additionally, liquid, and supercritical CO_2 are utilized for extracting flavors, oils, and other chemical compounds, such as in the decaffeination of coffee and tea. CO_2 also finds applications in food processing within the baking, dairy, and other industries. Furthermore, dry ice and liquid CO_2 are employed in freezing and refrigeration processes, as shown in Figure 40 [71].

Inactivation of Enzymes

Lipolytic microorganisms, such as psychrotrophic bacteria, are commonly found in fatty foods. These bacteria produce enzyme lipases during storage, which leads to rancidity and negatively impacts food quality. Research has shown that carbon dioxide gas can effectively inactivate the lipase enzyme; approximately 84% of the enzyme's initial activity is lost within just five minutes when exposed to carbon dioxide at an initial pH of 7.15 and a temperature of 50°C. Furthermore, carbon dioxide treatment, applied under atmospheric pressure, presents promising

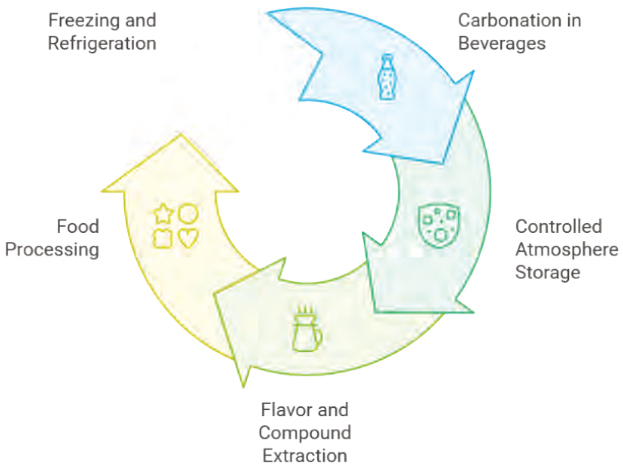


Figure 40. CO_2 utilization cycle in food industries.

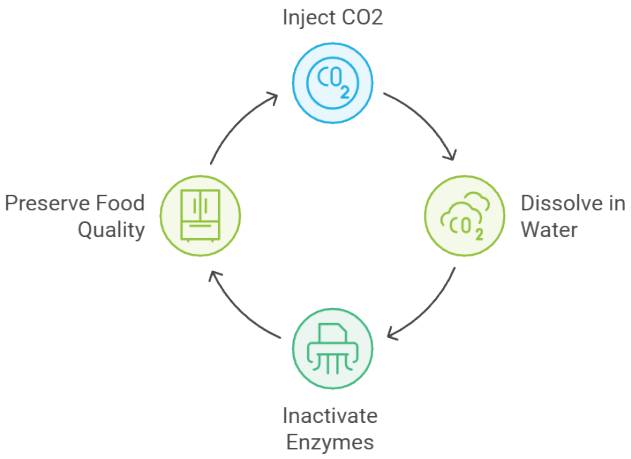


Figure 41. Carbon dioxide treatment cycle.

non-thermal technology for preserving liquid foods. This method can also be effective in inactivating enzymes that require high pressure. By injecting carbon dioxide through submerged water, the oxygen content is reduced, which inhibits enzymatic browning in pre-peeled potatoes without adversely affecting their hardening, thanks to the increased acidity of the water.

Figure 41 shows the carbon dioxide treatment cycle in inactivation of enzymes. In modified atmosphere packaging or controlled atmosphere storage, carbon dioxide inactivates enzymes by dissolving in the water present in the product [72].

Protein Precipitation

A continuous process has been developed for the precipitation of casein, the primary protein extracted from milk and the main ingredient in cheese, using carbon dioxide at a pressure of less than 14 MPa in a tubular reactor. The optimal casein product is obtained at temperatures of 38°C and a pressure of 5.52 MPa. Instead of lactic or mineral acids, casein precipitation with CO_2 is advantageous since the area remaining after precipitation is relatively high, i.e. 6.0, thus there is less pretreatment in comparison to other conventional methods. Also, it was possible to precipitate soy protein by the use of high-pressure carbon dioxide with sizes attaining approximately 68.3% (by weight) of precipitation owing to the use of 3.0 MPa pressure, pH 5.6 and constant temperature of 22°C.In contrast to traditional mineral oil-based precipitation methods, high-pressure CO_2 precipitation is a cleaner process. It produces highly purified food protein without the need for further purification before use in food and non-food applications, as introduced in Figure 42 [72].

Monitoring Food Quality

The use of carbon dioxide extends beyond preserving and processing food products; it also serves as an indicator of changes in the quality of food materials during processing and storage. This is done by observing changes in the concentration of carbon dioxide in the grain storage bins where growth and infestation of insects and microorganisms like fungi are observed. After determining the impact of humidity, temperature and the mechanical damages on the carbon dioxide that is generated by grains such as corn and wheat during storage, we can measure the dry matter loss which is a sign that the grain is deteriorating. Consequently, linking carbon dioxide levels with grain quality will help determine the optimal conditions—such as humidity and temperature—for storing grains effectively and maintaining their quality, as shown in Figure 43 [72].

Carbonated Beverages

The most common and oldest use of carbon dioxide (CO_2) is in the production of carbonated soft drinks and soda water. CO_2 enhances the distinctive sharp taste or

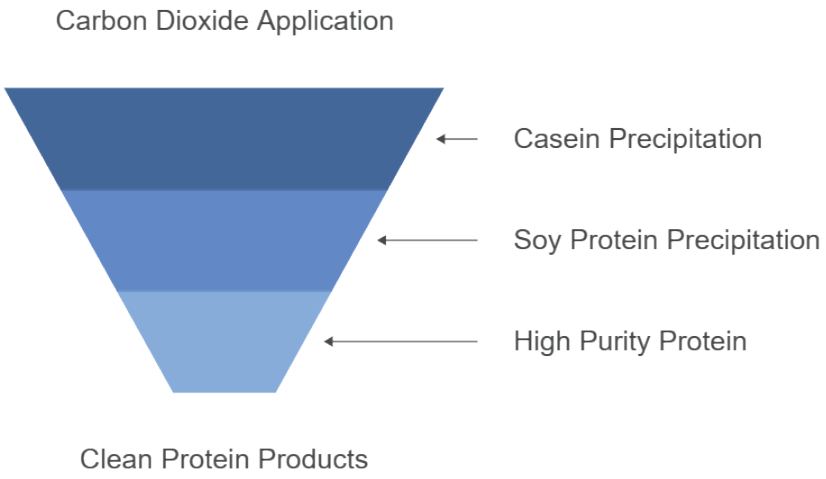


Figure 42. Efficient protein precipitation process.

Figure 43. Carbon dioxide monitoring in grain storage.

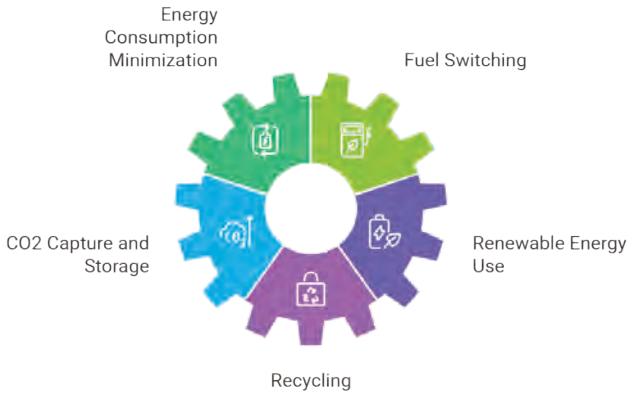


Figure 44. Strategies for reducing steel industry CO_2 emissions.

"bite" of soft drinks by stimulating the olfactory and taste nerves. Also, the carbonation contained in these drinks prevents mould growth and the growth of bacteria. In other instances, carbonation might even kill some bacteria, which is based on the amount of carbonation employed. Together with the acidity present in the beverage, the CO_2 content effectively preserves carbonated drinks for an extended period [72].

Inerting

Inerting the production of fruit juice with CO_2 helps minimize the reaction between the juice and the oxygen present in the containers. This reduces undesired fermentation later in the shelf life of the product thus prolonging the high quality sensory in the product [73].

STEEL INDUSTRY

With the increasing worry over climate change, manufacture of steel is now challenged to regulate carbon dioxide emission by providing a solution with minimal effect on its ability to perform its processes and without raising the cost. The largest industrial source of CO_2 emissions lies in both the iron and steel sectors, which are mainly because of the energy consumption required to make steel, the use of carbon based raw materials, and its production of 1,414 million tonnes and above of steel in the year 2010. The entire world steel production has been on a steady increase within the last 40 years and projected to continue increasing even more in the future, particularly in the developing economies. By providing all steel plants with electricity of hydroelectric generation stations, nuclear power generation stations or other environmentally friendly sources, they would greatly lower their indirect carbon dioxide emission as in Figure 44. The resulting shift-off would result in an almost zero carbon dioxide emissions of scrap-based electric arc furnace (EAF) mini-mills. Direct and indirect CO_2 emissions can be decreased with several technologies and processes that include:

- Reduction of energy consumption, and enhance process efficiency.
- Replacing a fuel or an agent with a lower CO_2.
- Sequestration of CO_2 and its storage in the ground.

In the future, indirect carbon dioxide emissions of the steel and iron industries are predicted to be reduced with the passage of time due to the creation of carbon sinks. Recycling scrap steel is also important towards saving energy and demanding carbon dioxide emissions [74].

Let's Capture The Future

Picarro, an Italian technology firm deals in advanced laser technology in gas network monitoring. Their creativity assists the network operators to detect gas leakages and minimise fugitive gas emissions. The laser system of Picarro makes it possible to detect areas much more accurately, in record time and covering much more area than conventional systems using cavity ring-down spectroscopy (CRDS).

The technology can be deployed in vehicles, boats, or backpacks, as shown in Figure 45. The detection equipment

Figure 45. Vehicles gas leak detector during the activities in COP 29 Baku, Azerbaijan.

includes an anemometer, GPS antenna, tablet, and smartphone. Its sensitivity is 1,000 times higher than that of conventional systems, allowing it to detect gas leaks from distances of about 150-200 meters. Notably, this technology does not require the vehicle to follow the route of the underground pipeline and can differentiate between methane (CH_4) and ethane (C_2H_6). The benefits of this surveying technology include are Greater Safety, is a tenfold increase in network safety and sustainability, Lower Costs, its optimization of operating costs through the reduction of emergency calls, it also More Efficiency by Three times faster detection procedures, leading to greater network coverage, and will be More Reliability. Overall, Picarro's technology significantly advances gas network monitoring, making it safer and more efficient.

Another application is using drones to detect fugitive emissions through conducting methane detection and measurement campaigns using drones equipped with AUSEA (Airborne Ultra Lightweight Spectrometer for Environmental Applications) technology, as shown in Figure 46. This technology, developed by TotalEnergies and its research and development partners, enables highly accurate detection and quantification of methane emissions from oil and gas facilities.

AUSEA features a miniaturized dual sensor mounted on a drone that can detect methane and carbon dioxide emissions while simultaneously identifying their sources. Developed by TotalEnergies, the French National Centre for Scientific Research (CNRS), and the University of Reims Champagne-Ardenne, AUSEA is now recognized as the most accurate technology in the world for detecting and measuring methane emissions.

Single-atom solution in capturing carbon and creating energy is presented in a recent study that presents, for the first time, the electrochemical conversion of carbamate species to methane (CH_4). In this study, they demonstrate that a single-atom nickel (Ni) catalyst, when dispersed on various conductive electrode materials, is capable of directly

Figure 46. Drone leak detector to reducing methane emissions.

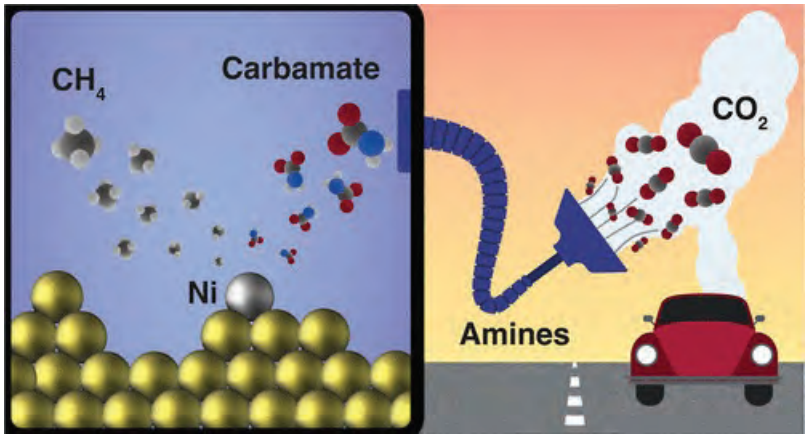


Figure 47. Convert the capture CO_2 using amines into methane by Ni catalyst.

reducing an amine capture solution to methane (CH_4). This catalyst exhibits a high selectivity for methane (CH_4) over other carbon-based products (Fig. 47) [75].

This integrated approach facilitates the conversion of carbon dioxide (CO_2) and the regeneration of the capture species in a single step, thereby completely eliminating the energy loss typically associated with the release of CO_2 from capture solutions. Despite the significance of completing the carbon (C) cycle, there are currently no known examples of an electrocatalyst that can directly convert the CO_2 capture species, carbamate, into a hydrocarbon.This study identifies a novel class of catalysts that can directly convert carbamate to hydrocarbons (CH) and offers significant mechanistic insights that may inform future applications involving the electrochemical conversion of CO_2 capture species into hydrocarbon products [75].

RESULTS AND DISCUSSION

It is necessary to promote the better quantification of CO_2 applications and their climate benefits. Exploring and facilitating early market opportunity to utilize CO_2 that is scalable, commercially viable, and potentially resulting in cuts in emissions must be a priority.

Although tremendous progress has over the past few years been achieved in carbon capture, utilisation, and storage (CCU/S) technologies and hydrogen energy production using methane, the listed several of questions in research will help to optimize the efficiency of industrial processes:

- **Hydrogen Production + Ccu/S Integration:**
  - Synergistic Optimization: Creating combined systems that receive maximum optimization in carbon capture and in hydrogen production to minimize energy losses and greatly enhance efficiency.
  - Co-product Utilization: The utilisation of by-products of each process, like carbon and value-added chemicals, to enhance the overall economic value.

- **More Efficient Carbon Capture Technologies:**
  - High-Performance Sorbents: Producing ultra-efficient and selective sorbents to capture CO_2 in industrial flue gases, particularly to difficult situations, such as steel and cement manufacturing.
  - Membrane-Based Separation: How can the performance and the life of membranes used in the separation of CO_2 be enhanced, especially in high-temperature and pressure setting.
  - Energy-Efficient Capture Processes: Advancing new methods to decrease energy usage related to carbon capture, including solvent regeneration and pressure swing adsorption.

- **Efficient Hydrogen Generation of Methane:**
  - Catalytic Processes: Hyperefficient and stable catalysts in other processes such as reforming of methane and pyrolysis to enhance process efficiency and hike the hydrogen production.
  - Designing Revolutionary Reactor Designs: Native reactor design: coming up with new reactor designs that promote efficiency and decrease emissions as a result of improved heat and mass transfer.
  - Renewable Energy Integration: Investigating the possibility of integrating renewable energy, the source of solar and wind power, to serve as the needed energy source to produce hydrogen, which will decrease fossil fuel consumption.

- **Environmental Impact Assessment & Economic Feasibility:**
  - Life Cycle Assessment: Carrying out of thorough life cycle assessments to analyze the environmental and economic effects of various carbon capture and hydrogen generation technologies, including energy use aspect, carbon emissions and water use.
  - Cost-Benefit Analysis: Building strong cost-benefit analyses to determine the most cost-effective and eco-friendly solutions to industrial processes.

- **Policy And Regulatory Frameworks:**
  - Incentive Mechanisms: Delineation into effective policies and incentives to spur adaptation of CCU/S and hydrogen technology, like carbon tax credential, carbon certification, and carbon pricing/trading.
  - Standardization and Certification: Developing universal guidelines and certification of carbon capture and hydrogen technologies that achieve quality, safety and environment friendliness.

These gaps in research can be bridged to potentially realize a high degree of efficiency and sustainability of the industrial process, leading to a low-carbon future.

Therefore, it is important to support research and development for future applications of CO_2 that could contribute to a CO_2-neutral economy. A legislative and regulatory framework for trading hydrogen and carbon in various new uses and applications should also be established.

CONCLUSION

This paper reviews and analyzes the potential and the challenges of carbon capture and storage, hydrogen energy storage technologies, and their energy efficiency in various applications, as well as the benefits of properly and sustainably exploiting both methane and carbon emissions.

The most common method for producing hydrogen from methane is steam reforming, which has an energy efficiency of 65% to 75%. This process can produce carbon dioxide as a byproduct, which can be integrated into new services and products aimed at achieving sustainable development goals. Although methane pyrolysis is still in the experimental phase and has an estimated energy efficiency of about 58%, it marks a significant advancement in

the production of clean hydrogen and solid carbon for various applications, all while generating minimal emissions.

Hydrogen is an emerging energy carrier that exists naturally and can be produced in various forms for electricity generation and as a key raw material in several important industries. Additionally, hydrogen serves as a clean fuel for transportation, much like its historical use in spacecraft.

One of the primary sources of hydrogen that contributes to the reduction of carbon emissions in the atmosphere is methane. Hydrogen can be extracted from methane through various chemical reaction mechanisms, enabling its use in a range of applications.

Carbon dioxide can facilitate new industrial development in a sustainable manner, contributing to the production of low-emission concrete, advancements in the chemical industry, and innovations in the food industry. For example, it can play a role in enzyme immobilization for food quality control and in protein precipitation processes, as well as in the production of soft drinks.

ABBREVIATIONS

ATAG	Air Transport Action Group
ATR	Auto-Thermal Reform
BEVs	Battery Electric Vehicles
CDR	Carbon Direct Removal
CHP	Combined Heat & Power
CCS	Carbon Capture & Storage
CCU	Carbon Capture & Utilisation
EAF	Electric Arc Furnace
ETC	Energy Transformation Commissions
EU	European Union
FCs	Fuel Cells
GHGs	Greenhouse Gases
GT	Giga Tone
HFCVs	Hydrogen Fuel Cell Vehicles
HSE	Health, Safety & Environment
IEA	International Energy Agency
IPCC	Intergovernmental Panel on Climate Change
LCA	Life Cycle Analysis
LH2	Liquid Hydrogen
LNG	Liquified Natural Gas
LOHCs	Liquid Organic Hydrogen Carriers
LSMGO	Low Sulfur Marine Gas Oil
$MMTCO_2$	Million Tons of Carbon Dioxide
NZE	Net-Zero Emissions
O&G	Oil & Gas
PEMFC	Polymer Electrolyte Membrane Fuel Cell
SMR	Steam Methane Reform
SOFCs	Solid Oxide Fuel Cells

ACKNOWLEDGEMENTS

I extend my sincere thanks to the Faculty of Graduate Studies and Environmental Research, Department of Engineering, Ain Shams University, for their efforts to help us understand the effort that I could review, revise, and analyze in preparing this paper.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Wallace-Wells D. The Uninhabitable Earth: Life After Warming. New York: Tim Duggan Books; 2019.

[2] Pahuja G, Hawken P. Drawdown: The Most Comprehensive Plan Ever Proposed to Reverse Global Warming. London, New York: Penguin Books, 2016.

[3] Kolbert E. The Sixth Extinction: An Unnatural History. London: Bloomsbury Publishing; 2014.

[4] Hansen J. Climate Change in a Nutshell: The Gathering Storm, Bloomsbury USA, 2009.

[5] Doll JE, Baranski M. Greenhouse Gas Basics. Michigan: Michigan State University; 2011.

[6] Dentener F, Derwent R, Dlugokencky E, Holland E, Isaksen I, Katima J, et al. Atmospheric Chemistry and Greenhouse Gases. Washington: Intergovernmental Panel on Climate Change; 2008.

[7] Zohuri B. Hydrogen Energy: Challenges and Solutions for a Cleaner Future. New York: Springer International Publishing; 2019. [CrossRef]

[8] Metz B, Davidson O, de Coninck H, Loos M, Meyer L. Carbon Dioxide Capture and Storage, Intergovernmental Panel on Climate Change (IPCC), 2005.

[9] Madejski P, Chmiel K, Subramanian N, Kuś T. Methods and Techniques for CO2 Capture: Review of Potential Solutions and Applications in Modern Energy Technologies, Energies 2022;15:3. [CrossRef]
[10] Levy CR, Almaraz M, Beerling DJ, Raymond P, Reinhard CT, Suhrhoff TJ, Taylor L. Enhanced rock weathering for carbon removal: Monitoring and mitigating potential environmental ımpacts on agricultural land. Am Chem Soc 2024;58:17215–17226. [CrossRef]
[11] Nagireddi S, Agarwal JR, Vedapuri D. Carbon dioxide capture, utilization, and sequestration: current status, challenges, and future prospects for global decarbonization. ACS Eng Au 2024;4:22–48. [CrossRef]
[12] Silveira BH, Costa HKM, Santos EM. Bioenergy with carbon capture and storage (BECCS) in Brazil: A review. Energies 2023;16:4. [CrossRef]
[13] Qadir S, Hussain A, Ahsan M. Numerical analysis of a gas separation of CH4/CO2 using hollow fiber membrane modüle. Sigma J Eng Nat Sci 2018;36:2:501-510.
[14] Eren D, Sari Yilmaz M. Synthesis of silica xeorogel@ mg-al layered double hydroxide composite for CO2 capture, Sigma J Eng Nat Sci 2024;42:1101–1107. [CrossRef]
[15] Lyons M, Durrant P, Kochhar K. Reaching Zero with Renewables: Capturing Carbon. Abu Dhabi: International Renewable Energy Agency; 2021.
[16] Mannion P, Parry E, Patel M, Ringvold E, Scott J. Carbon Removals: How to Scale a New Gigaton Industry. New York: McKinsey & Company; 2023.
[17] Jewad R. Capture Carbon, Capture Value: An Overview of CCS Business Models. Oxford: Oxford Institute for Energy Studies OIES; 2024.
[18] Rempel Z, Cameron L, Bois von Kursk O. Unpacking Carbon Capture and Storage: The Technology Behind the Promise. 28 Nov 2023. Retrieved Available at: https://www.iisd.org/ Accessed Apr 29, 2026.
[19] Patonia A, Poudineh R. Hydrogen Storage for a Net-Zero Carbon Future, Oxford Institute for Energy Studies. Oxford: Oxford Institute for Energy Studies; 2023.
[20] Pasteris M, Lorin A, Decorte M. Decarbonising Europe's Hydrogen Production with Biohydrogen, European Biogas Association EBA, 2023. Available at: https://www.europeanbiogas.eu/publication/decarbonising-europes-hydrogen-production-with-biohydrogen/ Accessed Apr 29, 2026.
[21] Helmi T, Lorentz B, Schlaak T, Trüby J. Hydrogen: Pathways to Decarbonization, Deloitte, 2023. Available at: https://www.deloitte.com/global/en/issues/climate/pathways-to-decarbonization-hydrogen.html Accessed Apr 29, 2026.
[22] Methane Emissions from Natural Gas Infrastructure, Deutscher Verein des Gas- und Wasserfaches e.V. DVGW, 2020. Available at: https://www.dvgw.de/english-pages/dvgw-services/publications/methane-emissions Accessed on Apr 29, 2026.
[23] Heilig GK. The Greenhouse Gas Methane (CH4): Sources and sinks, the ımpact of population growth, possible ınterventions. Int Inst Appl Syst Anal 1994;16:109-137. [CrossRef]
[24] International Methane Emissions Observatory IMEO. An Eye on Methane, United Nations Environment Programme UNEP, 2022.
[25] Global Methane Emissions and Mitigation Opportunities, Global Methane Initiative GMI, 2021. Available at: https://www.globalmethane.org/documents/analysis_fs_en.pdf Accessed on Apr 29, 2026.
[26] Searchinger T, Yan X, Wang J, Dumas P, Beauchemin K, Kebreab E. Opportunities to Reduce Methane Emissions from Global Agriculture. Discussion Paper. Princeton: Princeton University & Cornell University; 2021.
[27] Progetti E. Fugitive Emission Activities. ENI, 2023. Available at: https://www.eni.com/static/en-IT/infographics/eni-for-2023/carbon-neutrality/ Accessed on Apr 29, 2026.
[28] Tailakov W, Irving I, Oleg. CH4 Emissions: Coal Mining and Handling, IPCC National Greenhouse Gas Inventories Programme, 2000. Available at: https://www.ipcc-nggip.iges.or.jp/public/gp/bgp/2_7_Coal_Mining_Handling.pdf Accessed on Apr 29, 2026.
[29] Krause M, Kenn S, Stephenson J, Singleton A. Quantifying Methane Emissions from Landfilled Food Waste. New York: Environmental Protection Agency; 2023.
[30] Municipal Wastewater Methane: Reducing Emissions, Advancing Recovery and Use Opportunities. Washington: The Global Methane Initiative; 2013.
[31] Chenier PJ. Chemicals from Methane. New York: Springer; 2002, p. 205–215. [CrossRef]
[32] Timmerberg S, Kaltschmitt M, Finkbeiner M. Hydrogen and hydrogen-derived fuels through methane decomposition of natural gas– GHG emissions and costs, Energy Convers Manag 2020;7:100043. [CrossRef]
[33] New York State Energy Research and Development Authority. Hydrogen Production – Steam Methane Reforming (SMR). New York: New York State Energy Research and Development Authority; 2014.
[34] Blumberg T, Morosuk T, Tsatsaronis G. A Comparative Exergoeconomic Evaluation of the Synthesis Routes for Methanol Production from Natural Gas. Appl Sci 2017;7:1213. [CrossRef]
[35] Liu K, Song C, Subramani V. Hydrogen and Syngas Production and Purification Technologies. Washington: American Institute of Chemical Engineers;Ü 2010. [CrossRef]
[36] Patlolla SR, Katsu K, Sharafian A, Wei K, Herrera OE, Mérida W. A Review of Methane Pyrolysis Technologies for Hydrogen Production, Renew Sustain Energy Rev 2023;181:113323. [CrossRef]

[37] Hirscher M. Handbook of Hydrogen Storage. Weinheim: Wiley-VCH GmbH & Co; 2010.
[38] Ferrari M, Varone A, Stückrad S, White RJ. Sustainable Synthetic Fuels. RIFS Policy Paper and Policy Briefs. Potsdam, Germany: Institute for Advanced Sustainability Studies; 2014.
[39] Treyer K, Sacchi R, Bauer C. Life Cycle Assessment of Synthetic Hydrocarbons for Use As Jet Fuel. Switzerland: Paul Scherrer Institute; 2022.
[40] Matuszewska A. Review of Technologies of Synthetic Hydrocarbons Production from Biomass. Poland: Institute for Fuels and Renewable Energy; 2010.
[41] Takahashi K. Hydrogen Storage by Chemical Hydrides. Energy Carriers Convers Syst 2009;2:13.
[42] Melcher BU, George M, Paetz C. Liquid Organic Hydrogen Carriers – A Technology to Overcome Common Risks of Hydrogen Storage, Hydrogenious LOHC Technologies GmbH. 2021. Available at: https://hysafe.info/uploads/papers/2021/5.pdf Accessed on Apr 29, 2026.
[43] Rao PC, Yoon M. Potential Liquid-Organic Hydrogen Carrier (LOHC) Systems: A Review on Recent Progress, Energies 2020;13:6040. [CrossRef]
[44] Carvalho MCS de. Liquid Organic Hydrogen Carriers (LOHCs): A Review and techno-economic analysis (Master thesis). Lisbon: University of Lisbon; 2021.
[45] Huot J. Basics of Metal Hydrides, International Centre for Theoretical Physics (ICTP). ICTP-IAEA Advanced School on the Role of Nuclear Technology in Hydrogen-Based Energy Systems Trieste Italy, 13-18 June 2011. Available at: https://indico.ictp.it/event/a10148/session/25/contribution/ Accessed on Apr 29, 2026.
[46] Coşkuner Filiz B, Pişkin S. Investigation on Activities of Metal Oxides Catalyzed Ammonia Borane Hydrolysis, Sigma J Eng Nat Sci 2016;2:159-173.
[47] Van Der Voort P, Leus K, De Cank E. Introduction to Porous Materials. New York: Wiley; 2019.
[48] Thomas KM. Hydrogen adsorption and storage on porous materials. Catal Today 2007;120:389-398. [CrossRef]
[49] Chen Z, Kirlikovali KO, Idrees KB, Wasson MC, Farha OK. Porous Materials for Hydrogen Storage, Chem 2022;8:693-716. [CrossRef]
[50] Adolf J. Shell Hydrogen Study: Energy of the Future?, Shell Deutschland Oil. Hamburg, 2017. Available at: https://epub.wupperinst.org/frontdoor/deliver/index/docId/6786/file/6786_Hydrogen_Study.pdf Accessed on Apr 29, 2026.
[51] Clean Hydrogen Partnership. Hydrogen-Powered Aviation, Clean Sky 2 & Fuel Cells and Hydrogen 2 Joint Undertaking FCH 2 JU, 2020. Available at: https://www.clean-hydrogen.europa.eu/media/publications/hydrogen-powered-aviation_en Accessed on Apr 29, 2026.
[52] Hydrogen in Aviation. Hydrogen in Aviation UK.2024. Available at: https://hydrogeninaviation.co.uk/ Accessed on Apr 29,
[53] Hydrogen in Aviation Alliance (HIA). Launching Hydrogen-Powered Aviation, Hydrogen in Aviation UK, 2024.
[54] Kopasz J, Krause T, Ahluwalia R, Papadias D, Wang X. Hydrogen for Maritime Applications, Argonne National Laboratory, 2022. [CrossRef]
[55] Üçok MD. Hydrogen Fuel Cell Vehicles. Istanbul: Istanbul International Center for Energy and Climate IICEC, 2019.
[56] Brooker A, Birky A, Reznicek E, Gonder J, Hunter C, Lustbader J, et al. Vehicle Technologies and Hydrogen and Fuel Cell Technologies, National Renewable Energy Laboratory NREL, 2020.
[57] Cigolotti V, Genovese M. Stationary Fuel Cell Applications: Current and Future Technologies - Costs, Performances, and Potential, IEA - Technology Collaboration Programme, Italy, 2021.
[58] Hydrogen Power Generation and the Significance of Efficiency, Bloom Energy, 2023. Available at: https://www.scribd.com/document/1001205437/Bloom-Energy-Electrolyzer-Datasheet-December-2023 Accessed on Apr 29, 2026.
[59] Rongé J, François I. Use of Hydrogen in Buildings, WatersoftNet & KU Leuven, 2021. Available at: https://www.waterstofnet.eu/_asset/_public/BatHyBuild/Hydrogen-use-in-builings-BatHyBuild-29042021.pdf Accessed on Apr 29, 2026.
[60] Putting CO2 to Use: Creating Value from Emissions. New York: International Energy Agency; 2019.
[61] National Academies Press. Carbon dioxide utilization markets and infrastructure: Status and opportunities. 2022. Available at: https://www.nationalacademies.org/projects/DEPS-BEES-21-02/publication/26703 Accessed on Apr 29, 2026.
[62] Global CO2 Initiative. Global roadmap for implementing CO2 utilization. 2016. Available at: https://www.law.uh.edu/faculty/thester/courses/Climate%20Intervention%20Law%202019/Class%20readings/DAC%20Fuels%20Oxy/CO2U_ICEF_Roadmap_FINAL_2016_12_07.pdf Accessed on Apr 29, 2026.
[63] Romano MC, Anantharaman R, Arasto A, Ozcan DC, Ahn H, Dijkstra JW, Carbo M, Boavida D. Application of advanced technologies for CO2 capture from industrial sources. Energy Procedia. 2013;37:7176–7185. [CrossRef]
[64] Baruch S. Opportunities in the emerging carbon trading markets. Washington: U.S. Environmental Protection Agency; 2015.
[65] Zheng W, Chen H, Zhao J. Summary of CO_2 curing concrete technology. Highlights Sci Eng Technol 2022. Preprint. DOI:10.54097/hset.v18i.2655 [CrossRef]

[66] Kumar TS, Balaji KVGD, Nath Reddy ST, Rao GS. Mechanical properties of concrete when cured with carbon dioxide. Int J Eng Adv Technol 2019;8:2544–2549. [CrossRef]

[67] Bazzanella A, Krämer D. Technologies for sustainability and climate protection chemical processes and use of CO2. DECHEMA Gesellschaft für Chemische Technik und Biotechnologie, 2017.

[68] Röper M. Carbon dioxide: Feedstock to the chemical industry. Toulouse, France: Interuniversity Center for Materials Research and Engineering; 2013.

[69] Ayash AA. Co-polymerization. Mustansiriyah University;2020.Avialableat:https://uomustansiriyah.edu.iq/media/lectures/5/5_2025_02_20!08_44_58_AM.pdf Accessed on Apr 29, 2026.

[70] Nuclear Industry Association. Synthetic fuels. London: Nuclear Industry Association; 2023.

[71] Middle East Gases Association. Carbon dioxide usage in beverage industry. MEGA; 2018.

[72] Kaliyan N, Gayathri P, Alagusundaram K, Morey RV, Wilcke WF. Applications of carbon dioxide in food and processing industries: Current status and future thrusts. Am Soc Agric Biol Eng 2007. Preprint. doi:10.13031/2013.23553.

[73] Reliable CO_2 monitoring in the beverage industry. Dräger Safety AG & Co.; 2017. Available at: https://www.draeger.com/Content/Documents/Content/carbon-dioxide-food-and-beverage.pdf Accessed on Apr 29, 2026.

[74] Carpenter A. CO2 abatement in the iron and steel industry. London: IEA Clean Coal Centre; 2012.

[75] Neves-Garcia T, Hasan M, Zhu Q, Li J, Jiang Z, Liang Y, Wang H, Rossi LM, Warburton RE, Baker LR. Integrated carbon dioxide capture by amines and conversion to methane on single-atom nickel catalysts. J Am Chem Soc 2024;146:31633–31646. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2033

Research Article

Hyperparameter optimization web tool: Hyperopt

Fatma ALTINSOY[1,*], Muhammed Maruf ÖZTÜRK[2]

[1]*Graduate School of Natural and Applied Sciences, Suleyman Demirel University, Isparta, Türkiye*
[2]*Department of Computer Engineering Faculty of Engineering and Natural Sciences, Suleyman Demirel University, Isparta, Türkiye*

ARTICLE INFO

Article history
Received: 04 September 2024
Revised: 08 November 2024
Accepted: 27 February

Keywords:
Hyperparameter Optimization; Machine Learning; Optimization Techniques Stack Overflow Dataset; Web-Based Tool

ABSTRACT

This study introduces a web-based application designed to facilitate hyperparameter optimization for machine learning models, leveraging data from the stack overflow dataset. A primary contribution of this research is the development of a novel hyperparameter optimization method, integrated alongside established techniques such as Grid Search, Random Search, Nelder-Mead, and Bayesian Optimization. This integration provides users with the flexibility to explore various optimization strategies and identify the most suitable approach for their specific datasets and models.
The proposed web application enables users to select datasets, define optimization methods, and fine-tune hyperparameters through an intuitive and user-friendly interface. Empirical results demonstrate that the optimized models achieved a 15% improvement in prediction accuracy, attaining approximately 70% accuracy in predicting coding expertise and programming languages. Furthermore, the Proposed Method enhanced memory efficiency by 20% in SVM-based optimizations, with only a modest 10% increase in computational time. These findings underscore the method's effectiveness in balancing accuracy with resource efficiency.

Cite this article as: Altınsoy F, Öztürk MM. Hyperparameter optimization web tool: Hyperopt. Sigma J Eng Nat Sci 2026;44(2):1219–1239.

INTRODUCTION

Maintenance phases in software engineering like debugging, refactoring, and performance improvement are vital for maintaining the trustworthiness, extendibility, and productivity of software systems [1]. Through these methods, the functionalities of the current systems are not only upgraded but also the systems become more capable of adapting to changing requirements and technologies [2]. Among these, one critical performance optimization technique that has gained prominence, particularly in machine learning, is hyperparameter optimization [3].

Hyperparameter optimization is probably one of the most critical steps in further improving model performance in any machine learning project, considering the big data era we live in. In a nutshell, hyperparameters are important settings or options that are not directly learned during the model's training but significantly impact the model's success. Selecting the right hyperparameters leads to enhancing the overall performance and flexibility of the model to

*Corresponding author.
*E-mail address: fatmaaltinsoy@sdu.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

new data [4]. However, in most cases, the hyperparameter search space is so big and complex that it is very time-consuming and expensive to find the optimal combinations for a particular problem or dataset [5]. Hence, an effective strategy for hyperparameter optimization is highly desired to improve model accuracy while keeping computational resources inefficient.

Among various hyperparameter optimization strategies, widely used methods in machine learning models are Grid Search, Random Search, and Bayesian Optimization [6]. However, such approaches cannot be handled more straightforwardly in big search spaces. The curse of dimensionality, a phenomenon where the computational complexity and data sparsity grow exponentially with the number of hyperparameters, exacerbates the difficulty of searching large spaces efficiently [7]. This challenge requires deep technical expertise, which is cumbersome to provide for big datasets or model architectures where the associated computational costs are truly substantial. Current solutions of hyperparameter optimization are, most of the time, not user-friendly and tend to be problematic for novices to use [8]. This again highlights the demand for platforms that can make hyperparameter optimization processes more accessible with time and resource efficiency.

Web-based tools have become popular for handling such challenges, largely because they are easy to reach, simple to use, and cut down on repetitive setup work. Platforms like Dlib-ml [9], Teachable Machine [10], and H2O.ai [11] let users run data analysis, train models, and perform optimization with little or no coding. Building on this direction, we introduce a web-based hyperparameter optimization platform aimed at a wide audience, from beginners to experienced practitioners. The tool pairs a clean interface with strong optimization methods, so users can spend their time on model performance rather than setup.

Two goals drive this study: designing a new hyperparameter optimization method and testing it on a real-world dataset, the Stack Overflow (SO) platform. We also built a web-based tool to make the method easier to use in practice. SO holds a large body of developer-focused content and offers clues about programming languages, coding habits, and expertise levels [12]. That said, its sheer size and frequent mislabeling make finding the right information difficult [13]. We propose a rule-based labeling method to estimate developer experience levels, which then feeds into better content sorting and recommendation.

The platform is built on the ASP.NET Core MVC framework, with R handling the backend computation. Users can pick a machine learning algorithm, set hyperparameters, and choose an optimization strategy from the interface. Results appear in both visual and numerical form, keeping the process open and easy to follow. Taken together, the study connects optimization theory with day-to-day use, offering a practical tool for hyperparameter tuning in both research and industry settings.

This study is devised to develop an easy-to-use, web-based hyperparameter optimization platform. Specifically, the main contributions are as follows:

1. Enabling users to select datasets, specify optimization strategies, and perform hyperparameter tuning through a web interface.
2. Proposing a heuristic optimization method that improves the model's accuracy and efficiency.
3. A rule-based labeling of coding experience and preferred languages applied to the SO dataset, which makes the data more useful for future work.

The paper continues as follows. Section 2 covers related work and situates this study among existing approaches. Section 3 presents the preliminaries, and Section 4 describes the method in detail. Results and discussion, along with their implications, are presented in Section 5. Finally, Section 6 concludes the study and outlines potential avenues for future research.

RELATED WORK

SO is a large online community in which many types of information are shared with developers, ranging from programming languages to data analysis [14]. The study of the impact of SO on social media has indicated that some features, such as the transparency of information sharing, are likely to cause complex emotional outcomes for developers [15]. Other works have focused more on the reliability of SO as a source of information and the dissemination of knowledge across software development communities [16]. However, due to the huge amount of information available on SO and the number of users mislabeling, access to valid information is rather complicated, which ultimately reduces the overall efficiency [17].

Various data mining and deep learning techniques have been proposed to answer the labeling issues in a large platform like SO: for example, Sonam et al. [18] proposed the TagStack framework to enhance the accuracy of tag prediction, while Harrag and Khamliche [19] proposed the deep learning-based system intended for the correction of programming errors. Among the recent tag prediction methods, the one suggested in a work by He et al. [20], relying heavily on the significance of the title and demonstrating state-of-the-art language models applied in this domain such as CodeBert, is excellent.

High-value answer platforms represent indispensable tools that solve development problems and are the repository of development techniques [21]. Chen et al. [22], in their work related to code reuse, put the accent on the interactions between platforms like GitHub and SO and showed how these platforms have gained an essential place in the software development process.

Hyperparameter optimization is a crucial process for improving the performance of machine learning models [23]. The classical methods have ranged from Grid and Random Search; even the more advanced ones like Bayesian

Optimization have been employed in recent literature [24]. Grid Search is costly as it tests all parameters, while Random Search tries random combinations, and Bayesian Optimization optimizes the search process through inferential reasoning based on previous predictions [25]. The limitations of these methods make them computationally expensive for large datasets. While other alternative methods have been developed, searching for effective solutions in large hyperparameter spaces is still an ongoing process [26].

Hybrid optimization approaches have drawn attention in recent work [27, 28]. Uslu et al. [29], for example, combined Ant Colony Optimization (ACO) with Genetic Algorithms (GA) and reported strong results on tasks with limited computing resources. The Continuous Genetic Algorithm (CGA) has likewise been used on complex systems such as Troesch's and Bratu's problems, which points to its usefulness for machine learning tasks including hyperparameter tuning [30, 31]. Work of this kind helps frame how combinatorial optimization problems can be tackled and opens new directions for hyperparameter optimization [32].

A number of libraries and AutoML tools aim to simplify hyperparameter tuning. Widely used options include Optuna [33], Google Vizier [34], Ray Tune [35], and Microsoft NNI [36], each supporting a range of algorithms and optimization strategies. The catch is that most of them are built as libraries, meaning users have to write code to use them. That limits their reach among people without much programming background.

FairerML [37] is another notable contribution. It is an extensible platform for studying, visualizing, and reducing bias in machine learning models, with a simple upload-and-analyze workflow for both data and models. It also supports joint modeling for multi-target learning, letting users weigh accuracy and fairness at the same time.

DebiAI [38] takes a similar direction. As an open-source tool, it covers analysis, visualization, evaluation, and comparison of machine learning models. Users can work with their data and metadata in several ways, from selection and editing to annotation, and the tool also supports contextual evaluation and bias detection.

LAMDA-SSL [39] is a user-friendly and powerful open-source toolkit offering comprehensive functionality,

Table 1. Comparison of the proposed platform with other AutoML tools

Criterion	AutoGluon	Scikit-learn	Auto-sklearn	TPOT	H2O.ai	Proposed platform
Ease of Use	Easy to use, offers automatic optimization.	Simple API, manual optimization.	Automatic model and hyperparameter selection.	Moderate ease, some settings require effort.	Easy to use with API and web support.	Simplified web interface, accessible to novices.
Performance	Strong across various data types.	Varies by data/model selection.	Strong performance but resource-intensive.	Good but longer training times.	High speed on large datasets.	High accuracy on SO-specific data, optimized models.
Flexibility	Broad model and data support.	Flexible for diverse applications.	Extensive algorithm support.	Automatic feature engineering.	Supports machine and deep learning models.	Supports SO tagging, and multi-label classification.
Resource Usage	High resource usage is possible.	Low resource usage.	Resource-intensive in long processes.	Computation-intensive.	Efficient with distributed computing.	Efficient SO focused optimizations with low memory.
Hyperparameter Optimization	Fully automatic.	Manual, user-controlled.	Automatic and effective.	Genetic algorithms for optimization.	Automatic and efficient.	Heuristic optimization integrated with user controls.
Community and Support	Active community, extensive docs.	Very large, comprehensive docs.	Benefits from scikit-learn ecosystem.	Active, good documentation.	Large community, commercial support available.	Community friendly, SO data focus.
Data Processing	Automatic processing, some features.	Tools are available but manual tweaks are.	Automatic processing/ selection.	Automatic feature engineering.	Automatic feature engineering.	Rule-based SO tagging, TDM matrix generation.
Use Case Relevance	General AutoML.	General AutoML.	General AutoML.	General AutoML.	General AutoML and distributed computing.	Specific to SO, improves data labeling and recommendations.

intuitive interfaces, detailed documentation, and extensive support for algorithms, data types, and tasks compared to similar tools. The goal of LAMDA-SSL research is to make SSL research more feasible and widely applicable, particularly given the scarcity of labeled data.

Additionally, another notable work in the literature is MLweb [40], an open-source software toolkit for web-based machine learning. MLweb's distinguishing feature is that all computations are performed client-side without sending data to a third-party server. Key components of MLweb include a JavaScript API for scientific computations (LALOLib), extensions of this library with machine learning tools (ML.js), and an online development environment with numerous examples (LALOLab).

Another noteworthy project is WebDraw [41], an automation tool developed to enhance the efficiency of web designers. It produces mock-up designs and working websites from screenshots of web pages. The process runs in three stages: GUI element detection, classification, and code generation. WebDraw is designed with a user-friendly interface and is used to improve industrial web design and development processes.

Work in this area shows a mix of approaches to hyperparameter optimization and related machine learning problems. Tools and platforms built for hyperparameter optimization sit at the center of this space. That said, most are shipped as libraries with a narrow set of optimization methods. Table 1 sets HyperOpt, our proposed tool, against popular AutoML options: AutoGluon [42], scikit-learn [43], Auto-sklearn [44], TPOT [45], and H2O.ai [11]. The AutoML tools lean toward automation and ease of setup, whereas our platform is built for flexibility and customization, especially on domain-heavy data like the SO dataset. HyperOpt brings several practical gains: lighter resource use, optimization strategies suited to the task, and an interface that handles complex multi-label datasets without friction.

We address this gap with an interactive web-based platform that handles hyperparameter optimization through the interface itself, so users do not have to touch scripts. The platform runs on ASP.NET Core MVC and lets users pick a machine learning algorithm, set hyperparameters, and choose an optimization strategy. Backend optimization is handled by R, and results come out in both visual and numerical form, which supports qualitative and quantitative review side by side.

BACKGROUND

Web Platform Development

In this research, the developed web application was created based on the ASP.NET Core MVC framework. The main benefit of such an MVC architecture is a clear line of separation between UI and business logic, which itself interacts with data. The above-mentioned configuration makes this suitable for handling machine learning algorithm hyperparameter optimization processes through inputs from the user and offers the ability to run R scripts that work in the background.

Users can select datasets, hyperparameters, machine learning algorithms, and optimization methods from the web interface. Based on the selected parameters, respective R scripts are selected to be executed by passing the arguments to the R interpreter from the operating system command line using the Process class in the System.Diagnostics namespace of .NET. These scripts are run, their outputs captured, error-checked, and presented to the user as results of the analysis: graphical and numerical. Further, this makes the hyperparameter optimization process more intuitive and accessible to both machine learning experts and non-expert users with lesser programming backgrounds.

This paper has used an extensive optimization process to determine the hyperparameter settings of various machine learning algorithms, including SVM, GBM, EBR, EPS, and Random Forest. In order to maximize the performance of each algorithm, different optimization techniques have been used, including Grid Search, Random Search, Bayesian Optimization, Nelder-Mead, and a newly developed optimization method.

The goal, therefore, is to find the θi associated with a model M_i that maximizes the accuracy performance of such a model, evaluated as $\hat{\theta}_i$ assessed as a vector of hyperparameters. This can be done through the following generalized optimization problem:

$$\hat{\theta}_i = \underset{\theta i \in \ominus i}{\operatorname{argmax}} \text{ Accuracy } (\text{Mi } (\theta i), \text{Dtrain}) \quad (1)$$

Here, $\hat{\theta}_i$ represents the set of hyperparameters that maximizes the accuracy performance of the model. The argmax expression aims to find the value of θi that provides the highest accuracy. The hyperparameter search space is denoted as $\theta i \in \ominus i$, which specifies the range within which each hyperparameter is evaluated. Additionally, the expression Accuracy (Mi (θi), D_{train}), represents the accuracy of model Mi trained with a specific set of hyperparameters on the training dataset D_{train}.

The optimization processes have been carried out using the R programming language, and interfaces have been created particularly for optimized models using caret [46], rBayesOptimization [47], gbm [48], e1071 [49], randomForest [50], and other relevant R packages. Numerous experiments have been designed to evaluate the performance of each algorithm. The optimization process followed these steps:

1. We split the data into two subsets, Dtrain (training set) and Dtest (test set), to keep training and evaluation separate.
2. The search space θsearch was set to define where hyperparameters could be tuned.

3. Each θ (hyperparameter) went through cross-validation during training, so we could check how stable the results were.
4. Performance metrics were then computed for every model that came out of this process.
5. Execution time and memory usage (Ptime, Pmemory) were tracked throughout, since working within resource limits was part of the problem.
6. Finally, performance metrics and the impact of hyperparameters were plotted using the ggplot2 [51] package.

Another side of this work is the careful use of computing resources. This matters for large datasets and for scaling up machine learning models. Tracking CPU and memory usage gives a clearer picture of what each algorithm actually costs to run. Algorithms that hit the CPU hard can drag down performance and scalability, and memory-heavy operations tend to choke on big data.

Looking at these metrics lets us tune implementations, clean up workflows, and place computing resources where they are needed most. The upshot is a setup that scales, holds up under load, and remains workable for machine learning tasks, even when the data volume and computing demands are large. For each hyperparameter setting, we also generate performance curves, which make it easier to see how parameter changes shape model behavior and to guide choices during optimization.

Optimization Techniques

Grid Search consists of hyperparameter group H and all possible values of the hyperparameters ξ. Then a comprehensive search algorithm is employed, and by using the Cartesian product of ξ, a parameter grid is made. This grid covers all possible combinations of hyperparameters [52]. The complexity of Grid Search is $O(k^n)$, where n is the number of hyperparameters and k represents the search range defined by the user. The complexity of Grid Search increases as the number of hyperparameter combinations grows. Consequently, it may become costly when dealing with a relatively large search space. In Equations 2-4, the hyperparameter group H, the set of possible values ξ, and the grid structure of the algorithm are defined, respectively.

$$H = h_1, h_2, \ldots, h_3 \tag{2}$$

$$\xi = \xi_1, \xi_2, \ldots, \xi_m \tag{3}$$

$$\text{Grid} = h_1 * \xi, h_2 * \xi, h_3 * \xi \ldots, h_n * \xi \tag{4}$$

In Random Search, a probability function f takes a hyperparameter h as an argument. f(h) is expressed as random sampling.

$$\phi = (h_1, \xi_1), (h_2, \xi_2), \ldots, (h_n, \xi_n) \tag{5}$$

In Equation 5, $\phi \sim f(h_i)$ is defined for i=1,2,…,n [53]. This random sampling determines the hyperparameter range $H = h_1, h_2, \ldots, h_n$ for tuning. Random Search skips the exhaustive sweep that Grid Search does. It works on a user-defined subdomain of H, where values are drawn randomly from H itself.

Its time complexity is O(p), with p being the number of iterations. Note that p_h is not necessarily equal to p, since some iterations can land on values already tried earlier. This makes Random Search faster than Grid Search and lighter on memory, which helps most on tuning problems of modest size.

Bayesian Optimization looks for the minimum or maximum of an objective function, aiming to get as close to the optimum as it can [54]. Its complexity is $O(f * n^2)$, where n is the number of hyperparameters and f is the cost of evaluating them. It belongs to the family of black-box optimization methods and picks candidate values from the search space through a probability model.

This model compares previous results with new ones to explore the parameter space effectively, enabling the achievement of the global optimum with fewer trials. However, tuning the cost depends on the number of values defined in the search space, which could result in a high cost due to the curse of dimensionality. The acquisition function f is used to find the global optimum x∗ as shown in Equation 6 where $A \subseteq R$ [55].

$$x^* = \text{argmax}_{x \in A} f(x) \tag{6}$$

Here, x* represents the best solution value that corresponds to the maximum value of the objective function. This value is the optimal hyperparameter combination found within the scope of the optimization problem. A stands for the solution space under search and holds all possible ranges that the hyperparameter values can fall into. The objective function f(x) measures how well each hyperparameter combination performs on the model, and it is this function that guides the search toward better-performing hyperparameters. What Bayesian Optimization really seeks is the point x in A where f(x) reaches its maximum, which in practice means finding the best hyperparameter combination for the problem at hand.

On the other hand, Nelder-Mead [56] is specially designed for stochastic responses. In this method, candidate points form the set *S* and are provided to a function as shown in Equation 7.

$$f : S^n \rightarrow S \tag{7}$$

Here, S^n represents the set of candidate points in n-dimensional space, while the function f maps these points to another set S. This mapping implies that the function takes a point from the set S^n as input and returns a different point in the set S as output. The Nelder-Mead method is not classified as a direct search method because it does not utilize gradient information. The complexity of Nelder-Mead is $O(n^3)$, where n represents the number of variables used for the optimization. It updates the objective function

by comparing the previous result with the next one, gradually modifying an optimization triangle throughout the remaining optimization steps. Nelder-Mead is well-suited for small-scale optimization problems, although it performs more slowly compared to its counterparts.

The Proposed Method and proofs provide a comprehensive mathematical foundation for optimization methods. The Proposed Method operates through a unique mechanism: the execution cost of tuning steps is closely related to the number of iterations, where the execution cost is expected to be equal to or greater than one-fifth of the iteration count.

One strength of this approach is how it handles different hyperparameters within a Grid Search through a linear relationship, so every possibility gets covered without having to split the space.

Another feature is cost control through the number of words used in labeling functions. We found that as the word count grows, the total cost climbs quadratically, which tells us something useful about where the cost comes from.

There is also an accuracy threshold that caps how many iterations run. This cutoff does not tie directly to the iteration count, which gives the algorithm more room to breathe during the search.

Algorithm 1 handles the tuning of an arbitrary value, which we call optimizationValue. Its aim is to find the best setting for a given hyperparameter when building a model on the training data with a machine learning algorithm. In Step 3, the value range is defined as a vector bounded by an upper value x and lower value y, and this vector is assigned to optimizationValue.

In Step 4, the first entry of optimizationValue is assigned to j, which stands for a hyperparameter. Steps 5 through 7 take care of setting initial values for flag, accuracy, and accuracyOld. The flag acts as a control switch: it decides whether j should be increased on the next iteration. Accuracy is evaluated at each iteration of the algorithm to modify the hyperparameter value accordingly, while accuracyOld stores the previous accuracy obtained in the previous iteration.

Steps 9-17 replace the better hyperparameter and accuracy with their respective old values. Steps 19 through 26 bring the loop to a close, leaving us with the hyperparameter that gave the highest accuracy in the algorithm. The variable v controls how much the hyperparameter value changes from one iteration to the next. For instance, if the hyperparameter value fluctuates between 0.1 and 0.9, v is set to 0.1 to observe the impact on accuracy. The codes devised to experiment are publicly available for replication purposes.

Algorithm 1. Heuristic tuning algorithm

1: **Input:** dataset, threshold, iteration, v
2: **Output:** maxAccuracy, optimalHyperparameter
3: optimizationValue ← c(x : y)
4: j ← optimizationValue[1]
5: flag ← 0
6: accuracyOld ← 0.6
7: thresholdAccuracy ← 0.8, optimalHyperparameter ← −1
8: count ← 0
9: **while** (accuracy < thresholdAccuracy||count <= iteration) **do**
10: accuracy ← ϕ(j)
11: **if** accuracy >= accuracyOld **then**
12: maxAccuracy ← accuracy
13: optimalHyperparameter ← j
14: j ← j + v
15: **else**
16: j ← j − v
17: **end if**
18: accuracyOld ← accuracy
19: **if** accuracy > thresholdAccuracy **then**
20: flag ← 1
21: optimalHyperparameter ← j
22: break
23: **end if**
24: count ← count + 1
25: **end while**
26: **return** (optimalHyperparameter, maxAccuracy) ▷ result includes optimal configurations found by the algorithm

Machine Learning Algorithms

Support Vector Machines (SVM) are robust algorithms widely used in classification tasks [57]. SVM gives us room to tune several hyperparameters, including kernel, cost, degree, and gamma, which lets it take on both linear and non-linear problems [58]. With kernel functions such as linear, polynomial, sigmoid, and Radial Basis Function (RBF), SVM pushes the data into higher-dimensional spaces, where complex decision boundaries become tractable [59].

We used SVM to predict developer expertise and spot programming languages. Optimization techniques were applied to tune its hyperparameters, pushing up accuracy without letting the computational load grow out of hand.

Random Forest is another ensemble learning method, and it works for both classification and regression tasks. It merges many decision trees, avoiding overfitting of this combined model and improving generalization performance [60]. It has two major hyperparameters: mtry, defining the number of features available for each split, and ntree, fixing the number of trees in a forest [61]. The performance of Random Forest in this study showed a better trend for large and high-dimensional datasets. However, it is the balance between robust accuracy and computational cost. An optimized mtry and ntree often yield the best results.

Gradient Boosting Machine (GBM) encompasses an ensemble method that iterates and improves weak learners by using their errors to correct them [62]. Some of the key

hyperparameters of the said model are shrinkage (learning rate), n.trees (number of trees), and n.minobsinnode (minimum number of observations in each node). Proper tuning of these parameters can significantly improve the model's performance. While this approach maximized accuracy in predicting developer expertise, it required a careful balance of computational resources.

Ensemble of Probabilistic Subclassifiers (EPS) is a probabilistic ensemble method designed for multi-label classification; it is particularly effective on sparse datasets [63]. Other important hyperparameters are subsample (the ratio of data used in training), strategy (the approach for handling sparse label input), b (the number of subsets used in the strategy), and p (pruning rate to reduce error) [64]. Performance on sparse data conditions made it best suited for recommendation systems relying on a few labeled examples.

Ensemble of Binary Relevance (EBR) is a multi-label classification method in which every label is independently treated by training one model for each [65]. Key hyperparameters in this regard include subsample, which defines the data sample ratio used during training, and attr.space, which indicates the proportion of features used in training [66]. Tuning these parameters gave us balanced generalization and kept overfitting in check.

Table 2 lays out the machine learning algorithms we used along with their hyperparameter settings. Tuning these hyperparameters mattered a lot for how well the algorithms performed, especially on the harder multi-label classification tasks.

METHOD

Dataset

Within this study, a new dataset has been created to predict a programmer's experience using tag and question datasets obtained from the SO website [67]. The dataset has been enriched with features extracted from the content of the questions, alongside tag information corresponding to each question.

The resulting dataset follows a structural format where each record has a unique identifier. In addition to features, the dataset contains tags representing different programming languages (JavaScript, SQL, Java, C#, Python, C++, C, PHP, Ruby, Swift, Objective- C, VB.NET, Perl, Bash, CSS, Scala, HTML, Lua, Haskell, Markdown, and R) and experience levels. This dataset has been created based on

Table 3. The features of the dataset

Dataset	Feature	Description
Question	Id	Unique number
	OwnerUserId	User id
	CreationDate	The date input is entered
	Score	The reputation of the input
	Title	Title of the input
	Body	Context of the input
	ClosedDate	The date input is closed
Label	Id	Id of the input
	Tag	Tags are given by the user

Table 2. Summary of machine learning algorithms and hyperparameter settings

Algorithm	Parameter	Description	Strengths	Range	Step
SVM	kernel	Kernel type for SVM	Effective for non-linear problems	-	-
	degree	Polynomial kernel exponent		1-10	1
	gamma	Kernel coefficient		2^{-10}-2^{10}	0.1
	cost	Penalty of the error term		2^{-10}-2^{10}	1
RF	mtry	Features for each split	Generalization, scalability	1-8	1
	ntree	Trees in the forest		100-500	1
GBM	shrinkage	Impact of additional base-learner	High accuracy, iterative learning	0.001-1	0.1
	n.trees	Trees to fit		100-500	1
	n.minobsinnode	Min. observations in terminal nodes	Sparse data, multi-label tasks	5-15	1
EPS	subsample	Training instance rate		0.1-1	0.1
		Strategy for sparse clusters			
	strategy	A Number used in strategy		A-B	-
	b	Instances pruned		1-4	1
	p			1-4	1
EBR	subsample	Training instance rate		0.1-1	0.1
	attr.space	Attributes rate	Label independence, scalability	0.1-1	1

tags expressing relationships between the skill levels of programmers and programming languages, representing their experiences. For detailed information about the corpus, refer to Table 3.

Text Mining Operations

To achieve optimal results, the dataset size was incrementally increased during the experiments. Each experiment was run at least 20 times to ensure reliable results, allowing the observation of the impact of different dataset sizes on the inputs. It was determined that 20,000 records appeared to be a good fit. The maximizing of model parameters through text processing techniques set the stage for correctly tuning the data that was used to maximize the suggestion system, overall, endowing it to provide more successful outcomes. We also created subsets of records containing 500, 1000, and 3000 records to test certain scalability and performance dimensions on several sizes of datasets. These subsamples enabled a thorough analysis of the effect of dataset size on accuracy, memory consumption, and computational complexity metrics.

Data Cleaning and Preprocessing

Text mining techniques were used to transform the titles and contents of questions in the SO dataset into vectors. In the dataset, every word was taken as the feature, and occurrences of the same word in a document were quantified. Data were cleaned by removing punctuation, numbers, and spaces in this step. Subsequently, a Term Document Matrix (TDM) was generated using the processed data [68]. This resulted in the addition of two new features, word count and character count, to the dataset, allowing for more granular insight into the text.

After applying data mining techniques to the datasheet, the words extracted from the dataset were graphically represented using a visualization technique referred to as a "word cloud." Word clouds are a visualization tool to highlight unique words from textual data; words are represented in varying sizes according to their frequency [69]. This visualization has word size corresponding to occurrence frequency in the text. More frequent or more robust words in the dataset appear larger. Thus, the transmission of the essence of data can be understood more rapidly and effectively. The visual analysis illustrated in Figure 1 played a role in dissecting important pieces of the data, generating keywords for functional programming, and rendering outcomes in a way that's easier to digest.

Experience Tagging and User Level

To create the labels, 21 prominent programming languages from the past 12 years were selected and labeled from y1 to y21. These languages include JavaScript, SQL, Java, C#, Python, C++, C, PHP, Ruby, Swift, Objective-C, VB.NET, Perl, Bash, CSS, Scala, HTML, Lua, Haskell, Markdown, and R [70]. Stack Overflow's feature of publicly sharing datasets every two years enabled meaningful insights to be extracted from the entries. Labeling for these languages was performed using rule-based functions developed in the R programming language. Each function assigns a value of "1" when keywords related to the respective language are found in the question titles or contents.

The y22 label was used to determine the experience levels of SO users. User experience was classified into three

using
can
data
way
file
aspnet
server
application
net
sql
use
java
get
windows
function
table
text
visual
best
class
html
mvc
studio
possible
javascript
jquery
database
php
iphone
object
string
value
custom
user
open
create
problem
search
array
python
multiple
query
xml
files
set
select
without
method
framework
return
design
time
flash
inside
dynamic
system
methods
good
call
event
line
another
code
error
ajax
linq
web
control
variable
perl
version
single
load
item
objects
validation
help
work
number
based
button
software
record
document
app
site
null
model
need
date
getting
hibernate
show
simple
properties
directory
first
search
array
pass
tools
excel
source
type
two
wcf
forms
iis
key
like
access
add
update
testing
issue
ruby
linux
wpf
api
column
classe
conten
detect
adding
inpl
usedbuild
find
visual
field
fields
function
view
mvc
studio
windows
best
class
another
line
event
get
java
object
iphone
database
framework
number
based
button
software
record
document
app
site
null
model
need
date
getting
method
hibernate
show
return
design
time
flash
inside
know
dynamic
system
methods
good
call
program
oracle
text
table
possible

levels: beginner (1), intermediate (2), and advanced (3). Figure 2 illustrates the categorization of coding experience into these three categories. If the questions lacked words or phrases indicating programming experience, the y22 label was assigned a value of "0." For the classification of user experience, the identified words or patterns in the questions were taken into account.

Table 4 presents an example of the labeled dataset after preprocessing and tagging. The dataset includes 21 binary labels for programming languages (y1 to y21) and a multiclass label (y22) for user experience.

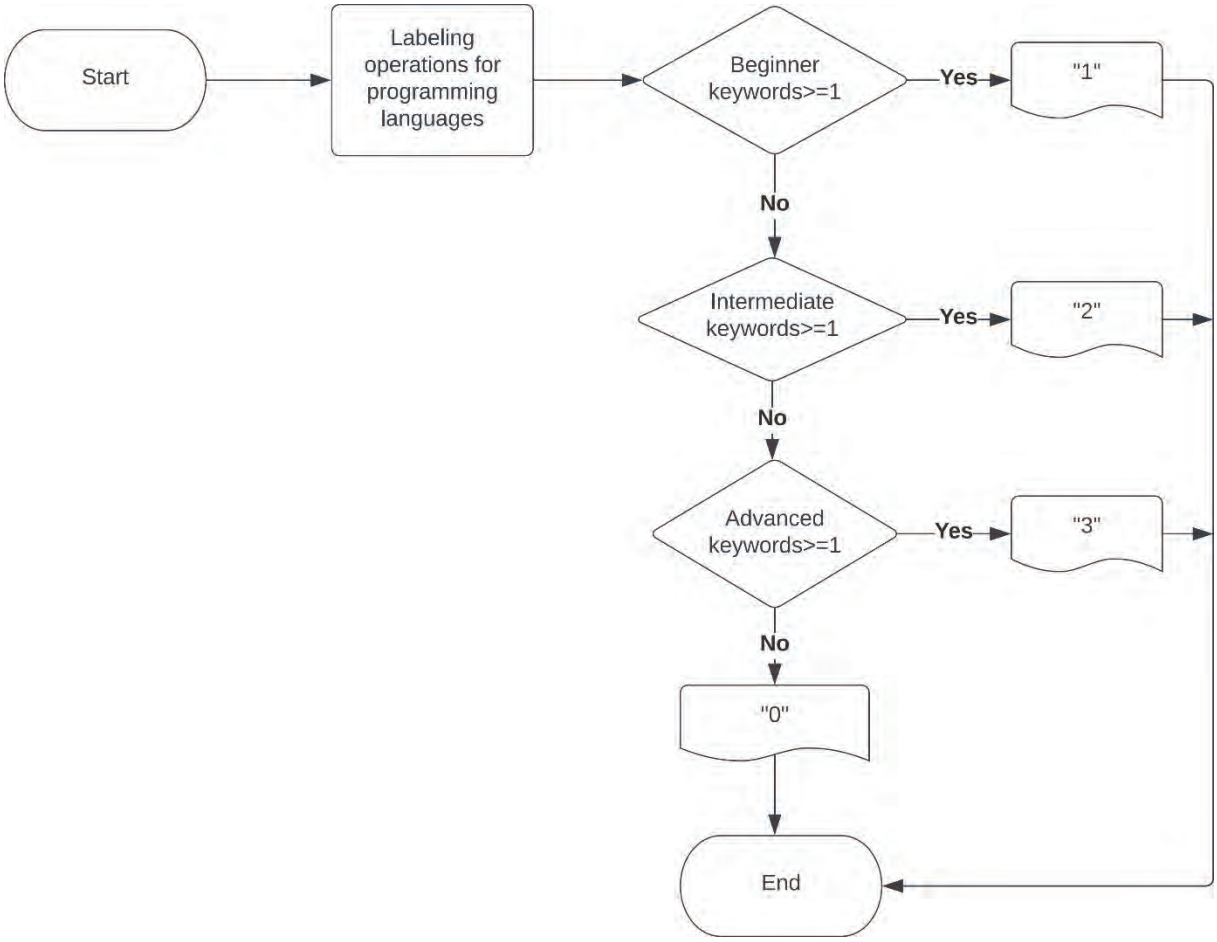


Figure 2. Steps of the labeling process.

Table 4. Processed dataset

y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12	y13	y14	y15	y16	y17	y18	y19	y20	y21	y22
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1

Software Framework (HyperOpt)

HyperOpt was built to make hyperparameter optimization straightforward and to let R code run as part of the same workflow. Figure 3 shows the architecture, which pulls together a few core components into one flexible and efficient platform. Things start at the data input stage, where users upload their datasets and the system pushes them through to the optimization pipeline. From there, the optimization module tests hyperparameters using Grid Search, Random Search, Bayesian Optimization, Nelder-Mead, and our proposed heuristic method, all with the goal of getting the model to perform as well as it can. Each method iteratively refines hyperparameters based on user-defined thresholds and evaluation metrics like accuracy, resource usage, and computation time.

The HyperOpt web platform is divided into four main sections, two key components being: (a) hyperparameter optimization tools and (b) result visualization. These tools together allow the user to set up optimization processes, monitor the progress, and effectively analyze the results.

Some of the techniques that can be used (some are shown in Figure 4(a)) are Grid Search, Random Search, Bayesian Optimization, Nelder-Mead, and a heuristic approach proposed in this paper. All of these methods systematically go through all the possible ranges of hyperparameters to search for the best model performance. The threshold value and iteration count are examples of parameters that can be defined by the user to drive the optimization. In each iteration, the platform logs the accuracy along with its associated hyperparameter combination. Also, to facilitate decision-making, relevant indicators like elapsed time in seconds and memory consumption in MB are computed and displayed.

Figure 4(a) shows a plot of the range across different values of one specific hyperparameter, illustrated in this example by using the mtry hyperparameter. Thus, on the horizontal axis, it shows the range for a certain mtry, while on the vertical axis, the scores of accuracy are reflected, and recorded for the optimizations through this range. Though mtry is used as the example here, any other hyperparameter is open to be analyzed with the help of visualization for assessing a particular value that may be most influential toward improving the model.

Figure 4(b) provides a detailed representation of the relationships between hyperparameter settings and model performance. This graph illustrates both the mean accuracy and its standard deviation for various hyperparameter configurations. The horizontal axis carries the values of the chosen hyperparameter, and the vertical axis shows the accuracy metrics that go with them. A smooth curve

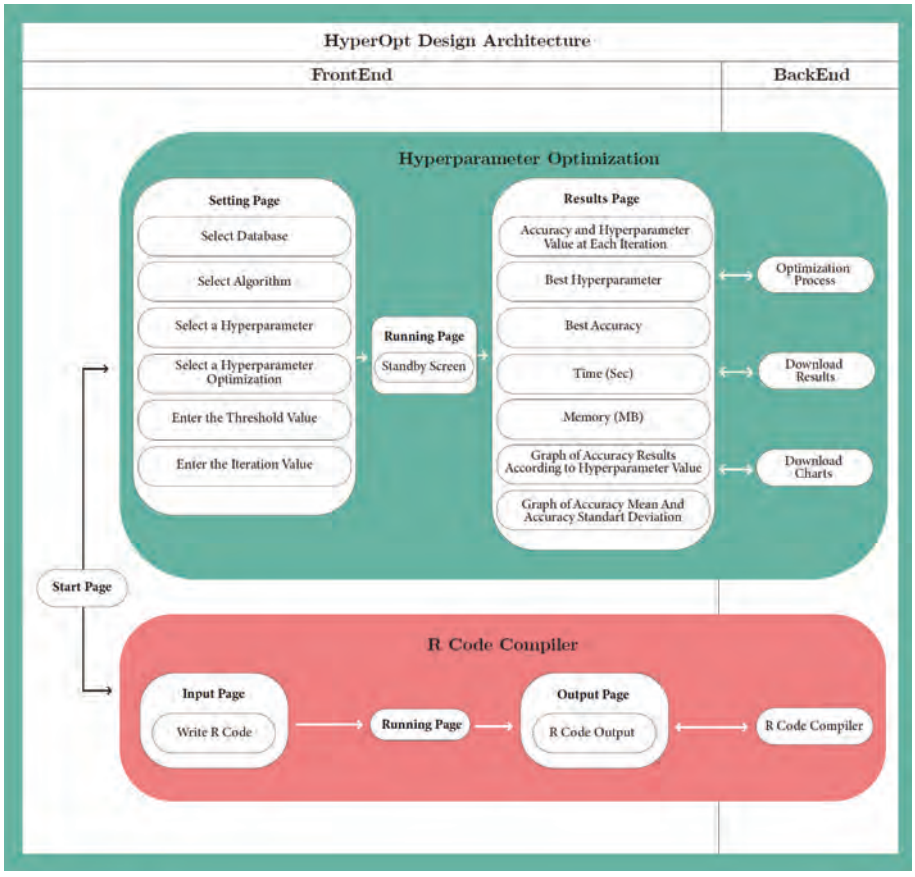


Figure 3. Architecture of HyperOpt.

traces the mean accuracy, while a shaded region or error bars mark the standard deviation. Taken together, these give users a quick read on performance trends, how much results vary, and how consistent the model is across different configurations.

The mean accuracy is simply the average accuracy across several runs for a given hyperparameter value. Formula 8 below shows how we compute it:

$$\text{Accuracy Mean} = \frac{1}{N}\sum_{i=1}^{N} \text{Accuracy}_i \quad (8)$$

In this formula, N is the total number of iterations, and Accuracy_i is the accuracy obtained at the i-th iteration for the hyperparameter value in question. This gives us a single number to anchor on when we want to judge how the model performs under a particular hyperparameter setting.

The standard deviation reflects the variability or spread of accuracy scores. Its formula is given as 9:

$$\text{Accuracy Standard Deviation} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{Accuracy}_i - \text{Accuracy Mean})^2} \quad (9)$$

In this calculation, (Accuracyi−Accuracy Mean) refers to the deviation of an individual accuracy score from the mean, squared to emphasize outliers. When the standard deviation is small, the model performs consistently across iterations; when it is larger, the results vary more.

HyperOpt also makes life easier on the saving side: users can export the resulting graphs as JPG, PNG, or PDF files, which keeps storing and sharing simple. This feature supports collaboration and thorough documentation, making the platform invaluable for researchers and practitioners focused on model optimization.

Additionally, an integrated R code compiler (illustrated in Figure 5) enables users to execute R scripts directly within the HyperOpt environment. The user can enhance the workflow efficiency due to advanced dynamic adjustment of parameters: in real-time, they see all the results and can immediately validate changes to the model. As discussed, HyperOpt supports iterative model refinement without requiring any additional application; hence, it is very powerful and practical.

HyperOpt embeds a tailor-made R code compiler, which is an interesting approach to problems in a certain field. Modifications of scripts can be easily done to fit either with the specific dataset or the model, which makes this platform useful for more complex analyses. This thus makes HyperOpt a general-purpose tool that can be used for many research directions and more sophisticated modeling activities.

Figure 4. HyperOpt web interface.

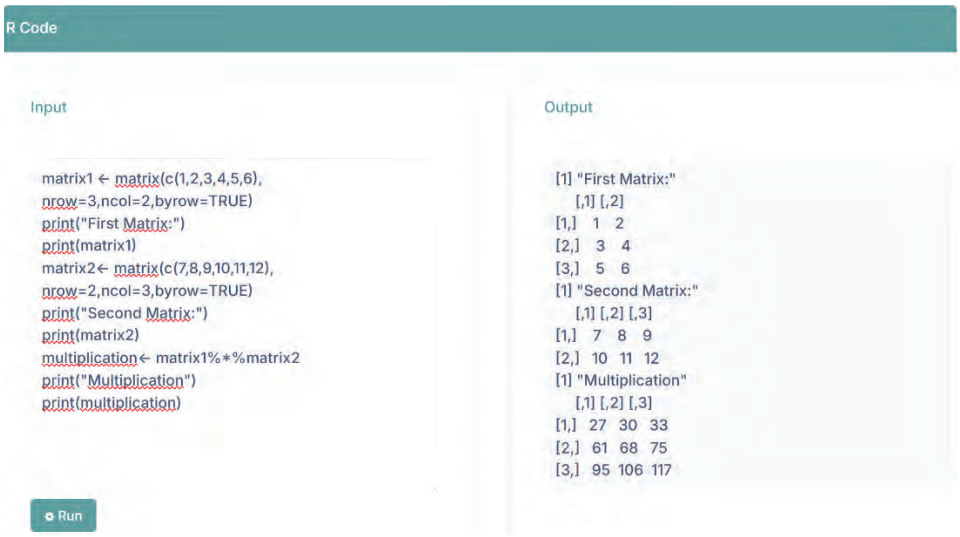


Figure 5. R code compiler.

With a persistent focus on design, HyperOpt aims to make state-of-the-art optimization tools simple to use. It is designed to be an easy-to-use tool regardless of experience but also serves as a powerful tool for users who are experienced.

The usable and functional interface of HyperOpt is designed with the end user in mind. This allows them the opportunity to test a variety of optimization techniques, track trends in accuracy and how they are affected by different hypers and allows them to gain an insight into the model better. This delineation empowers users to fine-tune the models and other elements with ease, therefore achieving the best possible outcome with minimal effort.

RESULTS AND DISCUSSION

This research presents a detailed performance comparison of machine learning models with different dataset sizes in terms of optimization techniques and algorithms used. Each method was performed in detail using different configurations of hyperparameters and each of the results was met with a broad comparison.

In Figure 6, the authors provide the results in a bar graph comparing the accuracy of five algorithms which are Random Forest, GBM, SVM, EBR and EPS across six optimization techniques including: Before Optimization, Grid Search, Random Search, Bayesian Optimization, Nelder Mead and Proposed Method. The results were based on sample datasets of 500, 1000, and 3000 while averages were obtained from 10 runs per sample and sample method.

The status Before Optimization are as a comparison where one can truly see the effects that hyperparameter tuning would have. It was observed that for every dataset size used, the baseline accuracy figures were on the lower side. For instance, for the dataset with 3000 records, the baseline accuracy of both Random Forest and GBM stood at 0.62, this indicates that there is so much room for further enhancement. SVM, EBR, and EPS had baseline accuracies of 0.66, 0.51, and 0.46, respectively and similarly better results were expected.

In general, the Proposed Method outperformed others in post-optimization accuracy across datasets and algorithms, hence proving efficiency in navigating hyperparameter spaces. For Random Forest, the accuracy increased to 0.66, 0.65, and 0.66 for datasets with 500, 1000, and 3000 entries, respectively, which again proved the scalability and robustness of the method. Similarly, GBM reached the highest values of accuracy with the Proposed Method, outperforming all other alternative optimization approaches across dataset sizes.

While the accuracy for SVM was steadily at 0.66 for all dataset sizes, surprisingly, some methods like Bayesian Optimization and Nelder-Mead have shown a drop in accuracy with larger datasets. On the other hand, the Proposed Method acted to negate such effects, thus giving back the baseline accuracy of these methods and further supporting their reliability.

The Proposed Method had considerable improvements, such as EBR and EPS, for models with low baseline accuracies. For instance, EBR had an increase in accuracy to 0.58, 0.57, and 0.58 for the three dataset sizes, while EPS had a constant accuracy of 0.56 across the datasets. This reflects how well the method generalizes to models with complex hyperparameter landscapes.

On the other hand, alternative methods for Bayesian Optimization and Nelder-Mead have mostly failed on datasets of larger sizes or with more complex models. Except for these, Grid Search and Random Search could give a decent improvement in accuracy but are much more

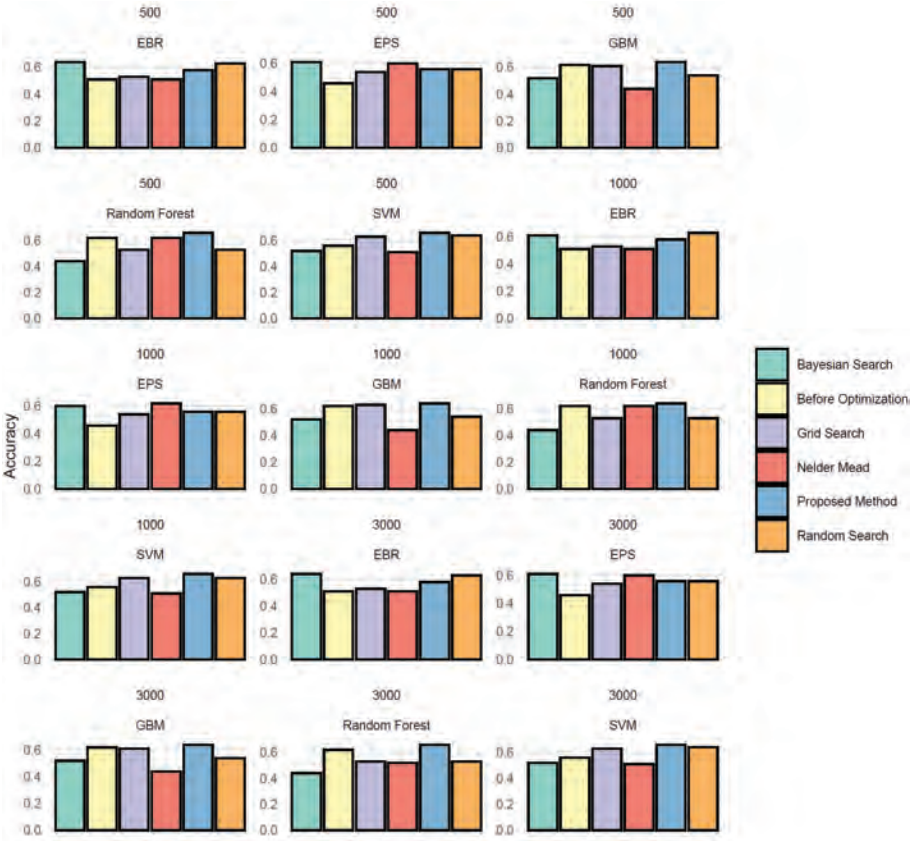


Figure 6. Accuracy comparison of machine learning algorithms under different optimization techniques.

computationally expensive reflected by longer execution time and higher memory usage.

Figure 6 summarizes, in general, that the Proposed Method does well in every algorithm and dataset scale concerning accuracy, while it provides a clear indication of different optimization techniques that perform badly across different conditions.

Figure 7, the performance comparison of five algorithms, namely, Random Forest, GBM, SVM, EBR, and EPS, for different optimization methods, namely, Before Optimization, Grid Search, Random Search, Bayesian Optimization, Nelder-Mead, and the Proposed Method. These results are averaged over several runs on datasets with 500, 1000, and 3000 entries.

This analysis puts into perspective the computational cost of each of these optimization techniques. It is good to refer to baseline times, such as from the Before Optimization scenario, for additional overhead brought in by the hyperparameter tuning process. Taking the example of the 3000-entry dataset, we observe that the baseline time taken for both Random Forest and GBM were well below times for Grid Search and Random Search. The numbers do not consider the performance improvement achieved by these optimizations.

Among all the algorithms and datasets, Grid Search and Random Search were the most time-consuming. That is in line with their exhaustive and stochastic approaches since they have to try a load of combinations of hyperparameters. For instance, for SVM on the 3000-entry dataset, Grid Search had to spend a significantly longer amount of time compared to the Proposed Method. It reflected the trade-off between time and performance: one invests time and enjoys superior improvements.

In most applications, the Proposed Method presented almost consistent low time consumption in contrast to competitive optimization output. It has shown competitive matches or outperformed those methods, like Random Forest and GBM, with significantly lower required time, underlining efficiency in performing hyperparameter exploration.

Bayesian Optimization and Nelder-Mead showed mixed performance. Bayesian Optimization, on the one hand, is faster compared to Grid Search and Random Search but often returns less significant improvements in accuracy for EPS and EBR. The results of Nelder-Mead were moderately consumptive with time but inferior to the Proposed Method in terms of performance and scalability, especially on large datasets.

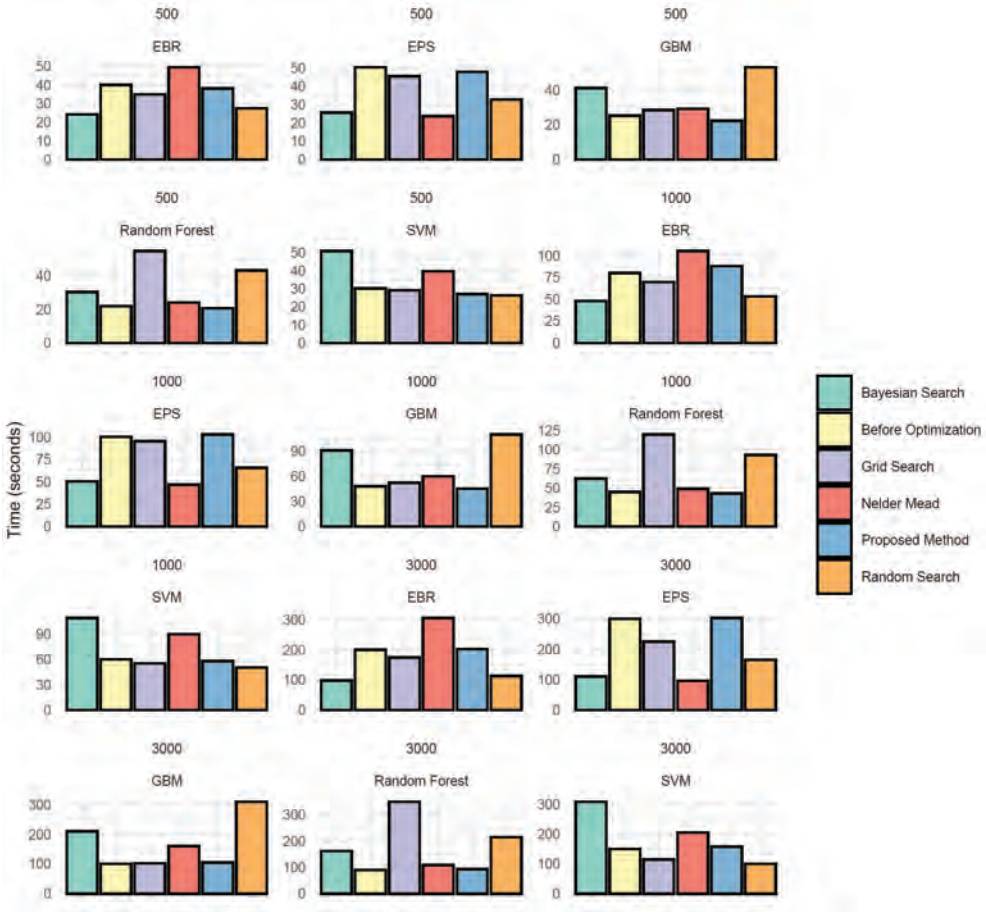


Figure 7. Time comparison of machine learning algorithms under different optimization techniques.

Among these, time consumption analysis underlined again that the Proposed Method would be effective in balancing between optimization effectiveness and computational efficiency, establishing a stronger option of hyperparameter tuning across diverse settings.

Figure 8 presents the memory consumption of the five algorithms, namely Random Forest, GBM, SVM, EBR, and EPS using different optimization methods: Before Optimization, Grid Search, Random Search, Bayesian Optimization, Nelder-Mead, and the Proposed Method. The memory consumptions were averaged over multiple runs while the datasets contained 500, 1000, and 3000 entries.

This analysis gives an indication of the differing memory requirements for hyperparameter tuning techniques. The Before Optimization scenario acts as our baseline. It shows what memory usage looks like without any iterative tuning. Not surprisingly, baseline memory stayed lower than any of the optimization methods across every dataset size we tried, since no extra computation is happening at this stage.

Grid Search and Random Search turned out to be the heaviest on memory in every scenario. Both have to go through a large number of hyperparameter combinations (Grid Search exhaustively, Random Search stochastically), and that pushes up the resource demand. On the 3000-entry dataset, for example, Grid Search on GBM and EPS used noticeably more memory than our Proposed Method did.

The Proposed Method came out ahead on memory efficiency. For Random Forest, GBM, and SVM, it produced results that were competitive with or better than the alternatives, but with a much lighter memory footprint. That makes it a good fit for setups where computing resources are tight, or where the work needs to scale up to bigger datasets.

Bayesian Optimization landed somewhere in the middle on memory use, doing a better job of balancing efficiency and computational demand than Grid Search or Random Search. The catch is that it did not always deliver large enough performance gains to justify itself. Nelder-Mead used the least memory of all the methods, but its results were often weak, especially on algorithms with complicated hyperparameter spaces like EBR and EPS.

To wrap up, the memory analysis points to the Proposed Method as the one that best balances resource use and

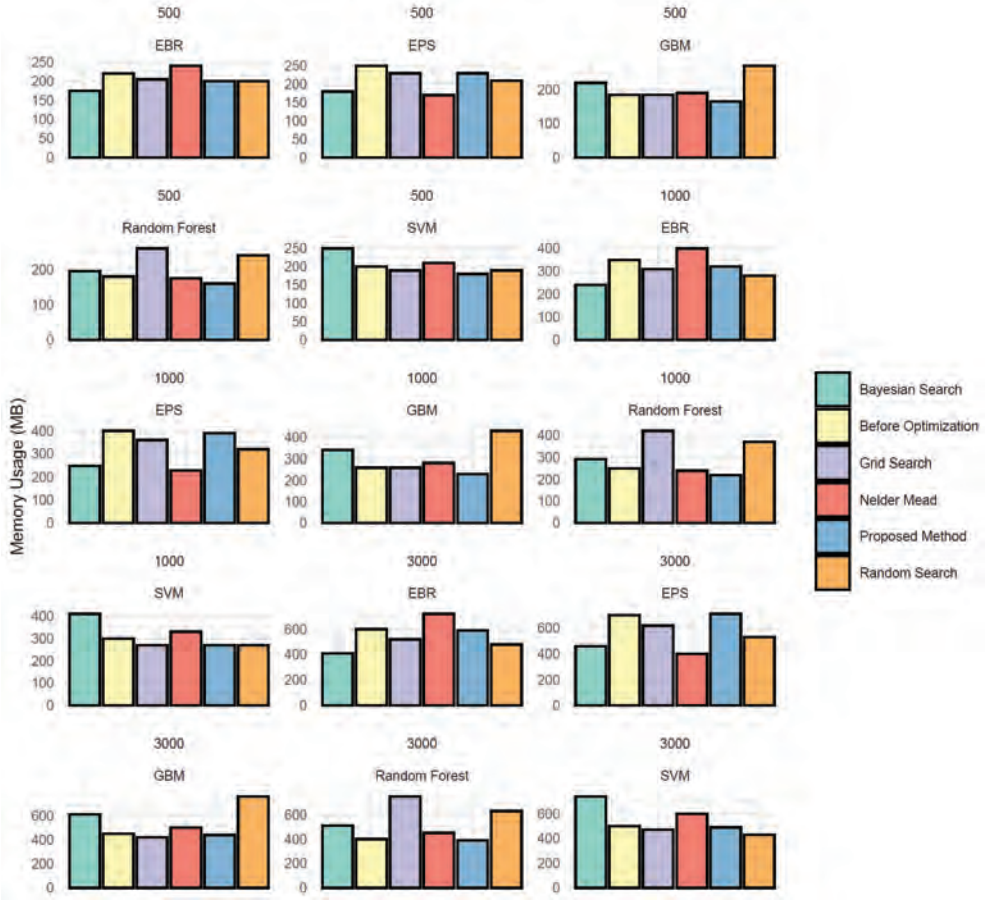


Figure 8. Memory comparison of machine learning algorithms under different optimization techniques.

optimization quality. That makes it a solid option across a range of applications, whether you are working in a tight computing environment or handling large-scale datasets.

The findings in Table 5 highlight the performance of various hyperparameters across optimization methods, dataset sizes, accuracy, processing time, and memory usage. Each hyperparameter's best-performing method is identified, and the effect of dataset size on overall performance is analyzed comprehensively.

As illustrated in Figure 9, the hyperparameter optimization results are further visualized across the three metrics (accuracy, time, and memory) for five optimization methods: Proposed Method, Grid Search, Random Search, Bayesian Optimization, and Nelder-Mead. The figure offers a detailed breakdown of the performance of these methods for each hyperparameter and dataset size (500, 1000, and 3000 entries).

mtry: The Proposed Method achieved the best performance for the mtry parameter across all dataset sizes. While the accuracy for the small dataset (500 samples) was 0.6174, it increased to 0.6644 for the large dataset (3000 samples). Processing time grew with dataset size, from 5.83 seconds for the small dataset to 60.27 seconds for the large dataset. Similarly, memory usage increased with larger datasets.

ntree: The ntree hyperparameter performed best with the Grid Search method, achieving the highest accuracy of 0.6857 on the medium-sized dataset (1000 samples). However, this performance came with significant costs in processing time and memory usage. For the large dataset, processing time rose to 540 seconds, with memory usage reaching 226 MB.

degree: The Proposed Method provided high accuracy for the degree parameter on small and medium-sized datasets but showed a decline in accuracy (0.5851) for the large dataset (3000 samples). Processing time significantly increased for larger datasets, reaching 20.43 seconds.

gamma: The Proposed Method outperformed other methods for small datasets, while Random Search achieved the best accuracy on medium-sized datasets. For large datasets, the Proposed Method maintained shorter processing times than other methods, though its accuracy was limited to 0.5608.

cost: For the cost parameter, the Proposed Method demonstrated shorter processing times and lower memory

Table 5. Hyperparameter optimization results across different dataset sizes and methods

Hyperparameter	Best Method	Dataset Size	Best Accuracy	Time (s)	Memory (MB)
mtry	Proposed Method	500	0.6174497	5.83135	181.3848
		1000	0.65	16.0182	189.792
		3000	**0.664451**	**60.2661**	**227.0672**
ntree	Grid Search	500	0.671428	60.1103	189.615
		1000	0.685714	120.629	198.834
		3000	0.682577	540.857	226.311
degree	Proposed Method	500	0.614864	1.43987	158.584
		1000	**0.634551**	**3.36655**	**159.090**
		3000	0.585177	20.4345	160.967
gamma	Proposed Method	**500**	**0.590604**	**0.7777**	**161.242**
	Random Search	1000	0.63815	16.1707	176.018
	Proposed Method	**3000**	**0.560840**	**22.3856**	**160.930**
cost	Proposed Method	500	0.53333	60.2175	158.583
		1000	0.51986	300.907	158.890
		3000	**0.5608**	**23.8777**	**161.124**
ntree	Proposed Method	**500**	**0.664429**	**26.146**	**161.801**
		1000	0.6369	26.294	162.451
		3000	0.64269	32.2069	165.083
minobsinnode	Proposed Method	500	0.65100	25.295	161.797
		1000	**0.67109**	**27.135**	**162.448**
		3000	0.651548	32.2606	165.078
shrinkage	Bayesian Optimization	500	0.68	25.6277	160.619
		1000	0.6345	35.2656	162.121
	Proposed Method	**3000**	**0.63676**	**33.465**	**165.079**
eps-subsample	Bayesian Optimization	500	0.2234	6.7714	180.346
	Proposed Method	**1000**	**0.5619**	**51.800**	**227.340**
	Grid Search	3000	0.503	900.151	291.569
eps-b	Proposed Method	**500**	**0.482**	**240.811**	**360.983**
	Random Search	1000	0.3660	300.320	266.552
	Bayesian Optimization	3000	0.5009	420.680	182.422
eps-p	Random Search	500	0.2391	60.457	211.74
		1000	0.3604	180.436	240.547
	Bayesian Optimization	3000	0.4885	420.122	182.422
ebr-subsample	Grid Search	500	0.3678	360.503	395.283
	Random Search	1000	0.3769	1380.72	467.708
	Grid Search	3000	0.5409	3600.05	733.654
ebr-attrspace	Random Search	500	0.3395	660.852	387.531
		1000	0.4126	1860.81	468.678
		3000	0.540	3600.92	181.742

usage. That said, on accuracy, it trailed behind Grid Search and Bayesian Optimization.

minobsinnode and ntree: For these two hyperparameters, the Proposed Method held up well across dataset sizes of every scale, staying efficient on both runtime and memory use relative to the other methods.

shrinkage: On small datasets, Bayesian Optimization came out on top with an accuracy of 0.68. As the dataset grew, though, the Proposed Method generalized better, which speaks to how well it adapts under different conditions.

eps-subsample and eps-p: While Grid Search and Bayesian Optimization had better accuracy, these methods

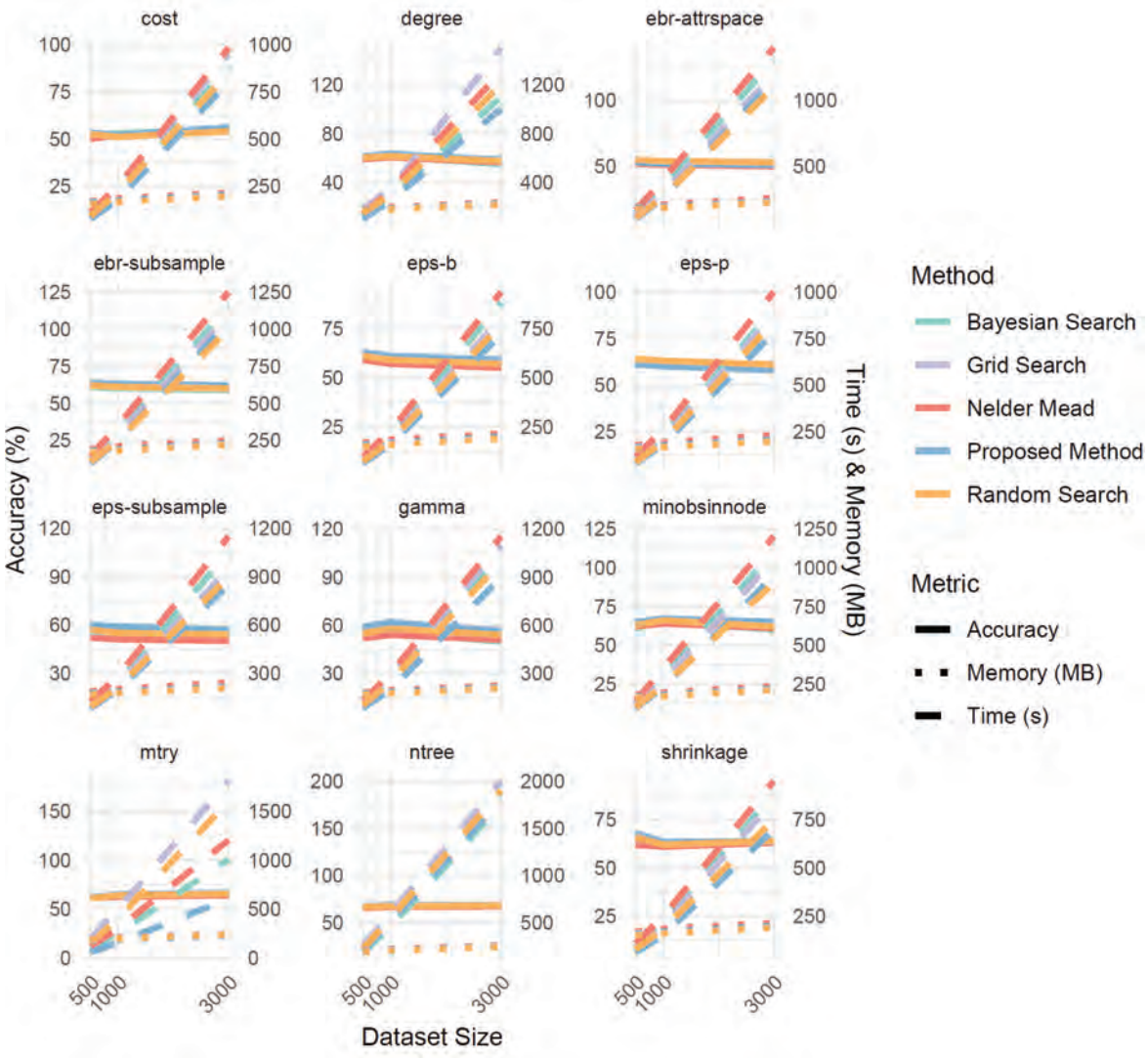


Figure 9. Hyperparameter optimization results for different methods and dataset sizes across various metrics.

consumed quite a lot of processing time and memory, especially for large datasets. On the contrary, the Proposed Method yielded balanced results, performing exceptionally well on medium-sized datasets, hence quite feasible for application.

ebr-subsample and ebr-attrspace: The Grid Search produced the best accuracy with 0.5409 for big datasets but at the price of very high processing time, 3600 seconds, and memory, 733 MB, hence it becomes computationally infeasible in resource-constrained settings.

In a nutshell, the Proposed Method has established itself as a robust and efficient optimization method for a variety of hyperparameters and dataset sizes. While traditional methods like Grid Search and Random Search are highly accurate, their computational cost is prohibitively high for larger datasets. Bayesian Optimization and Nelder-Mead gave a moderate performance but failed to be consistent. The Proposed Method strikes a good balance between accuracy, processing time, and memory efficiency; hence, it is the most practical choice for hyperparameter optimization in diverse contexts.

CONCLUSION

This paper proposes a holistic framework for predicting developers' experiences with programming languages using the SO dataset. The researchers have prepared three datasets of different sizes: 500, 1000, and 3000 entries. They applied different machine learning algorithms along with different optimization techniques. Results show the interrelation between dataset size, optimization methods, and important performance metrics such as accuracy, memory usage, and computational efficiency.

The Proposed Method came out to be the best among them, as it always outperformed traditional optimization techniques in the prediction of developer experiences from

the SO dataset. It achieves better performance and efficiency due to enhanced hyperparameter tuning and higher-order machine-learning strategies. This approach marks a leap in the development of hyperparameter optimization by incorporating easy-to-use tools with advanced optimization techniques.

Additionally, in many cases where the lighter datasets were indeed produced, larger datasets are usually associated with less precise results and impose a cost on higher computation. Some of the above considerations denote a delicate balance that existed between dataset size and algorithmic choice toward optimal performance. The Proposed Method presents excellent adaptability among different optimization techniques that include Grid Search, Random Search, and Bayesian Optimization. Coupled with Bayesian Optimization, the approach grants an excellent balance between exploration and exploitation for improved performance with low computational overhead.

One of the key contributions of this study is the development of HyperOpt, a web application designed to automate hyperparameter optimization and support model building. Beyond the new optimization method we propose here, HyperOpt gives users an R-based platform they can actually work with, letting them predict developer experience from the programming languages they prefer. It takes much of the pain out of hyperparameter tuning on large datasets, which is where the real bottleneck tends to sit, and that usefulness carries across both research and industry settings.

In the future, many avenues are open to explore for improving both the methodology and its application. Much more can be done to extend the range of supported machine learning algorithms, introducing more hyperparameters for an even more flexible and fine-grained insight into model performances. The testing of methodology on various datasets and research into the adaptation of this approach within a real-time system would further assert robustness.

Apart from these, a comparative study incorporating techniques like Hyperband, TPE, and genetic algorithms may also be presented to establish the efficacy of the Proposed Method against those in the state of the art. Also, besides providing accuracy, performance metrics like precision, recall, F1 score, and ROC curves should not go unmentioned. At the backend, a much more scalable and responsive HyperOpt is achieved through some additional steps of optimization regarding parallel computation, the use of distributed systems, and feedback mechanisms in real time.

The interface of the web application, including result visualization and integration of user feedback, requires further improvements to enhance its usability. Such enhancements are expected to make the application significantly more accessible to users ranging from beginners to expert practitioners. Working on these aspects pushes the Proposed Method further toward being a method that scales well, holds up under real use, and stays approachable for practitioners looking at hyperparameter optimization.

This work sets a base to build on, both for optimization techniques themselves and for the web-based systems that put them to use. Sharpening the optimization methods, bringing in more algorithms, and folding in a wider set of evaluation metrics are the natural next steps if the goal is to stay close to the state of the art. What we have learned along the way feeds into building machine learning tools that run leaner and reach further.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The dataset used in the manuscript is taken from one source: https://www.kaggle.com/datasets/stackoverflow/stacksample.

Code availability: The codes for the experiment are available at

https://github.com/fatmaaltinsoy/a-heuristic-technique-for-hyperpameter-tuning.git.

Web interface: The web interface is available at https://hypermlopt.sdu.edu.tr.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Mejía J. The software engineering evolution: A look through 11 years. In: 11th International Conference on Software Process Improvement (CIMPS); 2022. p. 245–248. [CrossRef]

[2] Glass R. A study about software maintenance. Inf Syst Manag 2012;29:338–339. [CrossRef]

[3] Karl F, Pielok T, Moosbauer J, Pfisterer F, Coors S, Binder M, et al. Multi-objective hyperparameter optimization in machine learning—An overview. ACM Trans Evol Learn Optim 2022;3:1–50. [CrossRef]

[4] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020;415:295–316. [CrossRef]
[5] Yu S, Ma P, Singh B, Silva S, Pritchard M. Two-step hyperparameter optimization method: Accelerating hyperparameter search by using a fraction of a training dataset. Artif Intell Earth Syst 2024;3:e230013. [CrossRef]
[6] Bartz E, Bartz-Beielstein T, Zaefferer M, Mersmann O. Hyperparameter tuning for machine and deep learning with R: A practical guide. Singapore: Springer Nature; 2023. [CrossRef]
[7] Chen S, Bolufé-Röhler A, Montgomery J, Zhang W, Hendtlass T. Using average-fitness based selection to combat the curse of dimensionality. In: IEEE Congress on Evolutionary Computation (CEC); 2022. p. 1–8. [CrossRef]
[8] Andonie R. Hyperparameter optimization in learning systems. J Membr Comput 2019;1:279–291.
[9] King D. Dlib-ml: A machine learning toolkit. J Mach Learn Res 2009;10:1755–1758. [CrossRef]
[10] Carney M, Webster B, Alvarado I, Phillips K, Howell N, Griffith J, et al. Teachable machine: Approachable web-based tool for exploring machine learning classification. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems; 2020. CHI 2020, April 25–30, 2020, Honolulu, HI, USA. [CrossRef]
[11] LeDell E, Poirier S. H2O AutoML: Scalable automatic machine learning. In: 7th ICML Workshop on Automated Machine Learning (AutoML); 2020.
[12] Tambad S, Nandwani R, McIntosh SK. Analyzing programming languages by community characteristics on GitHub and StackOverflow. arXiv 2020;2006.01351. doi:10.48550/arXiv.2006.01351.
[13] Sengupta S, Haythornthwaite C. Learning with comments: An analysis of comments and community on Stack Overflow. In: Proceedings of the 53rd Hawaii International Conference on System Sciences; 2020; Hawaii, USA. p. 2898–2907. [CrossRef]
[14] Tehrani MJ, Arjomand P. Finding the potential accepted answer on Stack Overflow: A text mining approach. Research Square 2022. Preprint. doi: 10.21203/rs.3.rs-2067571/v1 [Epub ahead of print] [CrossRef]
[15] Storey M-A, Treude C, Van Deursen A, Cheng LT. The impact of social media on software engineering practices and tools. In: Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research; 2010; New York, USA. p. 359–364. [CrossRef]
[16] Nasehi SM, Sillito J, Maurer F, Burns C. What makes a good code example? A study of programming Q&A in StackOverflow. In: Proceedings of the 28th IEEE International Conference on Software Maintenance; 2012; Trento, Italy. p. 25–34. [CrossRef]
[17] Vasilescu B, Filkov V, Serebrenik A. StackOverflow and GitHub: Associations between software development and crowdsourced knowledge. In: Proceedings of the 2013 International Conference on Social Computing; 2013; Alexandria, VA, USA. p. 188–195. [CrossRef]
[18] Sonam S, Verma A, Lal S, Sardana N. TagStack: Automated system for predicting tags in Stack Overflow. In: Proceedings of the 2019 International Conference on Signal Processing and Communication; 2019; India. p. 223–228. [CrossRef]
[19] Harrag F, Khamliche M. Mining Stack Overflow: A recommender systems-based model. Preprints 2020;2020080265. [CrossRef]
[20] He J, Xu B, Yang Z, Han D, Yang C, Lo D. PTM4Tag: Sharpening tag recommendation of Stack Overflow posts with pre-trained models. In: Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension; 2022. p. 1–11. [CrossRef]
[21] Zhu W, Zhang H, Hassan AE, Godfrey MW. An empirical study of question discussions on Stack Overflow. Empir Softw Eng 2022;27:148. [CrossRef]
[22] Chen X, Xu F, Huang Y, Zhou X, Zheng Z. An empirical study of code reuse between GitHub and Stack Overflow during software development. J Syst Softw 2024;210:111964. [CrossRef]
[23] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281–305.
[24] Liashchynskyi P, Liashchynskyi P. Grid search, random search, genetic algorithm: A big comparison for NAS. arXiv 2019;1912.06059.
[25] Bischl B, Binder M, Lang M, Pielok T, Richter J, Coors S, et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. Wiley Interdiscip Rev Data Min Knowl Discov 2023;13:e1484. [CrossRef]
[26] Hossain S, Islam F, Tayeb NT, Aslam M, Kim JH. Performance enhancement of the micromixer by the multiobjective genetic algorithm and surrogate model based on a Navier–Stokes analysis using trade-off objective functions. Math Probl Eng 2021;2021:9924849. [CrossRef]
[27] Wang C, Wilhelm M, Stuber M. Semi-infinite optimization with hybrid models. Ind Eng Chem Res 2022. [CrossRef]
[28] Azevedo B, Rocha A, Pereira A. Hybrid approaches to optimization and machine learning methods. In: 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA); 2023. p. 1–2. [CrossRef]
[29] Uslu MF, Uslu S, Bulut F. An adaptive hybrid approach: Combining genetic algorithm and ant colony optimization for integrated process planning and scheduling. Appl Comput Informatics 2022;18:101–112. [CrossRef]

[30] Abu Arqub O, Abo-Hammour Z. Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm. Inf Sci 2014;279:396–415. [CrossRef]
[31] Abo-Hammour Z, Abu Arqub O, Momani S, Shawagfeh N. Optimization solution of Troesch's and Bratu's problems of ordinary type using novel continuous genetic algorithm. Discrete Dyn Nat Soc 2014;2014:401696. [CrossRef]
[32] Parouha R, Verma P. Design and applications of an advanced hybrid meta-heuristic algorithm for optimization problems. Artif Intell Rev 2021;54:5931–6010. [CrossRef]
[33] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2019; New York, USA. p. 2623–2631. [CrossRef]
[34] Golovin D, Solnik B, Moitra S, Kochanski G, Karro J, Sculley D. Google Vizier: A service for black-box optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017; Halifax, Canada. p. 1487–1495. [CrossRef]
[35] Liaw R, Liang E, Nishihara R, Moritz P, Gonzalez JE, Stoica I. Tune: A research platform for distributed model selection and training. arXiv 2018;1807.05118.
[36] Microsoft. Neural Network Intelligence. GitHub; 2018. Available at: https://github.com/microsoft/nni. Accessed on Oct 07, 2024.
[37] Yuan B, Gui S, Zhang Q, Wang Z, Wen J, Mao B, et al. FairerML: An extensible platform for analysing, visualising, and mitigating biases in machine learning. IEEE Comput Intell Mag 2024;19:129–141. [CrossRef]
[38] Mansion T, Braud R, Amrani A, Chaouche S, Adjed F, Cantat L. Debiai: Open-source toolkit for data analysis, visualisation and evaluation in machine learning. In: 20th International Conference on Autonomic and Autonomous Systems; 2024; Athens, Greece.
[39] Jia L, Guo L, Zhou Z, Li Y. LAMDA-SSL: A comprehensive semi-supervised learning toolkit. Sci China Inf Sci 2023;67:1. [CrossRef]
[40] Lauer F. MLWeb: A toolkit for machine learning on the web. Neurocomputing 2018;282:74–77. [CrossRef]
[41] Kaluarachchi T, Wickramasinghe M. WebDraw: A machine learning-driven tool for automatic website prototyping. Sci Comput Program 2024;233:103056. [CrossRef]
[42] Erickson N, Mueller J, Shirkov A, Zhang H, Larroy P, Li M, Smola A. AutoGluon-Tabular: Robust and accurate AutoML for structured data. arXiv 2020;2003.06505.
[43] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–2830.
[44] Feurer M, Klein A, Eggensperger K, Springenberg J, Blum M, Hutter F. Auto-sklearn: Efficient and robust automated machine learning. In: Advances in Neural Information Processing Systems; 2015. p. 2962–2970.
[45] Olson RS, Bartley N, Urbanowicz RJ, Moore JH. Evaluation of a tree-based pipeline optimization tool for automating data science. In: Proceedings of GECCO; 2016. p. 485–492. [CrossRef]
[46] Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26. [CrossRef]
[47] Yan Y. rBayesOptimization: Bayesian optimization of hyperparameters. CRAN; 2016. [CrossRef]
[48] Ridgeway G. Generalized boosted models: A guide to the GBM package. 2007.
[49] Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. E1071: Misc functions of the Department of Statistics (E1071), TU Wien; 2009.
[50] Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18–22.
[51] Wickham H. ggplot2: Elegant graphics for data analysis. Springer; 2016. [CrossRef]
[52] Huang Q, Mao J, Liu Y. An improved grid search algorithm of SVR parameters optimization. In: 2012 IEEE 14th International Conference on Communication Technology; 2012; Chengdu, China. p. 1022–1026. [CrossRef]
[53] Larochelle H, Erhan D, Courville A, Bergstra J, Bengio Y. An empirical evaluation of deep architectures on problems with many factors of variation. In: Proceedings of ICML; 2007. p. 473–480. [CrossRef]
[54] Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH. Hyperparameter optimization for machine learning models based on Bayes optimization. J Electron Sci Technol 2019;17:26–40.
[55] Marinov D, Karapetyan D. Hyperparameter optimisation with early termination of poor performers. In: 2019 11th Computer Science and Electronic Engineering Conference (CEEC); 2019; Colchester, UK. p. 160–163. [CrossRef]
[56] Kolda TG, Lewis RM, Torczon V. Optimization by direct search: New perspectives on some classical and modern methods. SIAM Rev 2003;45:385–482. [CrossRef]
[57] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–297. [CrossRef]
[58] Feurer M, Hutter F. Hyperparameter optimization. In: Automated Machine Learning: Methods, Systems, Challenges. Springer; 2019. p. 3–33. [CrossRef]
[59] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. In: Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory; 1992; New York, USA. p. 144–152. [CrossRef]

[60] Breiman L. Random forests. Mach Learn 2001;45:5–32. [CrossRef]
[61] Probst P, Bischl B, Boulesteix AL. Tunability: Importance of hyperparameters of machine learning algorithms. arXiv 2018;1802.09596.
[62] Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29:1189–1232. [CrossRef]
[63] Read J, Pfahringer B, Holmes G. Multi-label classification using ensembles of pruned sets. In: 2008 IEEE International Conference on Data Mining; 2008; Pisa, Italy. p. 995–1000. [CrossRef]
[64] Chen H, Tiňo P, Yao X. Predictive ensemble pruning by expectation propagation. IEEE Trans Knowl Data Eng 2009;21:999–1013. [CrossRef]
[65] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Mach Learn 2011;85:333–359. [CrossRef]
[66] Rivolli A, de Carvalho ACPLF. The utiml package: Multi-label classification in R. R J 2019;10:24–37. [CrossRef]
[67] Kaggle. Stack Overflow dataset. Available at: https://www.kaggle.com/datasets/stackoverflow/stacksample. Accessed on Oct 17, 2023.
[68] Feinerer I. wordnet package. CRAN; 2024. Available from: https://cran.r-project.org/web/packages/wordnet/.
[69] Fellows I. wordcloud package. CRAN; 2024. Available at: https://cran.r-project.org/web/packages/wordcloud/
[70] Peruma A, Simmons S, AlOmar EA, Newman CD, Mkaouer MW, Ouni A. How do I refactor this? An empirical study on refactoring trends and topics in Stack Overflow. Empir Softw Eng 2022;27:1. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1240–1250, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2034

Research Article

Numerical simulation of magnetohydrodynamic casson Williamson hybrid nanofluid flow accounting thermal stratification

Jintu Mani NATH[1,2,*], Ashish PAUL[2], Tusar Kanti DAS[2,3]

[1]*Department of Mathematics, Mangaldai College (Autonomous), Mangaldai, 784125, India*
[2]*Department of Mathematics, Cotton University, Guwahati, 781001, India*
[3]*Department of Mathematics, Dudhnoi College (Autonomous), Dudhnoi, 783124, India*

ARTICLE INFO

Article history
Received: 13 June 2024
Revised: 21 August 2024
Accepted: 11 November 2024

Keywords:
Heat Stratification; Hybrid Nanofluid; Vertically Stretching Cylinder; Volume Concentration; Williamson Fluid

ABSTRACT

This study investigates the thermally stratified magnetohydrodynamic flow of Casson-Williamson hybrid nano liquid around a linearly stretched vertical cylinder in a porous region. The unique aspect of this research is the inclusion of thermal stratification effects on non-Newtonian hybrid nanofluid flow. Computational solutions are derived by employing MATLAB's Bvp4c algorithm, with velocity and thermal profiles illustrated graphically, and distinct non-dimensional factors. Shear and thermal transmission rates are also calculated and displayed in tables. Results expose that thermal stratification reduces the thermal profile, with notable heat stratification, even causing negative temperature values. It's also shown that the Williamson hybrid nanofluid exhibits a 16.5% increase in heat transmission rate over the Williamson nanofluid and a 30.7% higher shear stress rate. The heat transmission rate is enhanced by increasing the thermal buoyancy factor but decreases with higher values of the Casson factor, thermal stratification factor, porosity factor, and Weissenberg number. Additionally, the thermal distribution increases with a higher Weissenberg number and curvature factor. These findings offer substantial potential for enhancing thermal management in various industrial applications, marking a significant contribution to fluid mechanics and nanofluid research. The outcomes align well with prior studies.

Cite this article as: Nath JM, Paul A, Das TK. Numerical simulation of magnetohydrodynamic casson Williamson hybrid nanofluid flow accounting thermal stratification. Sigma J Eng Nat Sci 2026;44(2):1240–1250.

INTRODUCTION

Nanofluids are engineered fluids containing nanoparticles that augment thermal conductivity and thermal transmission. They are broadly operated in solar energy applications, thermal exchangers, and electronics cooling structures. Nanofluids boost thermal performance, which increases the efficacy of heating and cooling systems in a variety of industries, such as energy production, automotive, and aerospace. Kataria and Mittal [1] investigated the velocity and temperature of gravity-based convective nanofluid

***Corresponding author.**
*E-mail address: nathjintu67@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

movement via a vertical plate in the existence of a magnetic field. Moreover, Sheikholeslami et al. [2] explored the radiation influences on thermal transmission of nanofluid flow incorporating heat interfacial resistance. Furthermore, many researchers [3-9] explored distinct nanofluid flow phenomena incorporating various conditions. Moreover, a new sort of fluid that can offer noteworthy heat transport for a diverse range of industrial sectors is essential due to the significance of the improvement of nanotechnology. Thus, a unique fluid known as a hybrid nanofluid was created. The recent discovery of hybrid nanofluid has prompted multiple researchers to expand their numerical and experimental work. A hybrid nano-liquid may be formed by scattering several sorts of nano-particles in one or more fundamental fluids. In comparison to conventional nanofluids, hybrid nanofluids exhibit more thermal traits. These fluids are used in freezing, thermoelectric refrigeration, electronic cooling processes, medical applications, and more.

The heat transportation augmentation in hydro-magnetic Al_2O_3 - Cu hybrid nano-liquid movement via an extending cylinder was designated by Maskeen et al. [10]. Their findings indicate that related to conventional fluids or nanofluids, hybrid nanofluids are more effective at transmission of heat. Additionally, Khashi'ie et al. [11] used a porous longitudinal cylinder to track the flow of a hybrid nanofluid near its stagnation point, showing an effect on temperature stratification. Throughout their examination in a contracting cylinder, Al_2O_3-water nanofluid performed at a lower heat transfer rate than Cu-Al_2O_3/water hybrid nanofluids. Additionally, Elsaid et al. [12] conducted a computational study on the impact of a hybrid nanofluid's performance across a rotating cylinder. Also, Khan et al. [13] investigated the radiative movement created by the hybrid nano-liquid in a porous vertical cylinder. Recently, Nath et al. [14] performed computational comparative research on magnetohydrodynamics stagnation point movement of hybrid nanofluids flow with different fundamental fluids.

The utmost stimulating area of investigation is the inquiry of fluid movement via stretched cylinders. Applications for flow throughout a stretching cylinder encompass wire drawing, fiber spinning, and polymer extrusion operations, where regular stretching advances material features. In refrigeration systems, chemical reactors, and filtering units, it is vital to optimize heat and mass transmission. Furthermore, controlled stretching impacts thermal conductivity and viscosity, and also enhances system efficiency. In these scenarios, this is crucial for the development of biomedical devices and thermal exchangers. Crane [15] first studied how a stretched cylinder caused the boundary layers to flow. Wang [16] then investigated the flow of a viscous fluid across a stretched, hollow cylinder. Waqas et al. [17] recently used a vertically elongating cylinder to study the process of thermal conduction in a magnetised motion of hybrid nanoliquids. Additionally, to ascertain the procedure of thermal transmission integrating the movement of hybrid nanofluids via an extended cylinder, Sreenivasa et al. [18] conducted numerical research. Recently, Nath et al. [19] investigated the Newtonian and non-Newtonian hybrid nanofluid movement in cylindrical geometry.

The flow and thermal transmission features of non-Newtonian fluids are noteworthy to diverse systems in bioengineering and pharmacological medicine. Therefore, despite its intricacy, applied mathematician's quintessence on it. Hamid et al. [20] investigated the effect of thermal radiation on Williamson nanofluid movement. After that, the study of Khan et al. [21] emphasized how heat stratification influence with the viscosity of Williamson nano-liquid movement. Also, Mittal et al. [22] investigated the combined convection micropolar ferrofluid movement incorporating viscous dissipation, joule heating, and convective boundary limitations. Furthermore, Mittal and Patel [23] examined the consequences of thermophoresis and Brownian motion on combined convection MHD Casson fluid movement. Also, Nath et al. [24] addressed the thermal transport features of magnetohydrodynamic stratified Casson hybrid nanofluid movement via a porous extending cylinder. Additionally, Hussain et al. [25] considered the Williamson fluid with homogeneous-heterogeneous interactions in an assessment of sheet and cylinder on MHD convective flow. The results demonstrate that fluid flow and concentration are more noticeable by cylinder than by sheet. Williamson nanofluid movement in a convective cylinder with heat radiation and variable conductivity was investigated by Bilal et al. [26]. A. Almaneea [27] quantitatively studied the enhancement of heat and mass transfer in MHD Williamson fluid using hybrid nanoparticles. Recently a few researchers [28-33] have investigated in the field of non-Newtonian fluid incorporating distinct conditions.

Researchers have thoroughly investigated the role that thermal and solute stratification plays in the transmission of heat and mass. Stratification in flow domains may result from variations in fluid temperature, density, or concentration. Salinity, seismic fluxes, lake thermo-hydraulics, and underground water supplies are all affected by the process of heat stratification. Srinivasacharya and Surender [34-37] examined the stratified nanofluid movement via a vertical plate. Also, Srinivasacharya et al. [38] computationally studied the doubly stratified free convection in a non-darcy porous medium. Through the Keller box approach, Bilal et al. [39] numerically computed the thermally stratiform Williamson liquid movement. The findings illustrate that the thermal transmission rate exhibits prominent development for ascending magnitudes of Prandtl number for cylindrical geometry relatively for plane geometry. Williamson nanofluid flow's variation in viscosity as a consequence of heat and solute stratification was addressed by Khan et al. [40]. Furthermore, radiative bi-nano-liquid movement on a curved outward under temperature stratification circumstances was investigated computationally by Ramzan et al. [41]. Khan et al. investigated the flow and heat transport of a bio-convective hybrid nanofluid that included triple stratification impacts [42]. Masood et al. investigated the effects

of radiation and heat stratification on a hybrid nanofluid [43]. Heat stratification and unequal thermal radiation in the hybrid nanofluid movement were studied by Manzoor et al. [44]. Kairi et al. [45] studied the layered thermo-solutal Marangoni bioconvective movement of gyrotactic microbes in Williamson nanofluid. Moreover, an array of investigations [46-54] explored into different flow phenomena observed in the conventional fluid or hybrid nanofluids over numerous constraints.

The vitality of the Cassion Williamson hybrid nanofluid flow through solid geometries is brought out by the literature assessment, which additionally stimulates further research in this discipline. Casson Williamson hybrid nanofluid flow problems over vertical cylindrical bodies are also infrequently discussed in the literature. Nobody has inspected the flow of Casson Williamson hybrid nanofluid integrating the thermal stratification impact over a linearly stretched vertical cylinder. Which provides the required novelty for the current investigation. So, motivated by this, the present exploration investigates how heat stratification influences the MHD Casson Williamson Ethylene glycol driven MoS_2 - *Cu* hybrid nano-liquid movement via a stretched vertical cylinder in a permeable region incorporating the utilization of an inclined magnetic field. In this study integrating *Cu* and MoS_2 nanoparticles augment the lubricating, thermal transmission, and heat conductivity of nanofluids. MoS_2 has superior lubrication, stability, and anti-friction features, whereas *Cu* nanoparticles have prodigious heat conductivity, improving cooling efficiency. Also choosing Ethylene glycol offers a low-freezing, greater-boiling point base fluid. This amalgamation advances thermal transmission, diminishes friction, and escalates system stability, making it perfect for lubrication, cooling, and energy-efficient applications in the electronics and automotive industries. The fundamental objective of this investigation is to address the following queries:

- What are the outcomes and features of the flow under the specific conditions?
- What impact does the existence of MoS_2, and Cu nanoparticles influence the flow behavior and heat transport?
- How does the strong thermal stratification impact influence the Casson Williamson hybrid nanofluid?
- How does the Weissenberg number and curvature factor affect the thermal curve and the rate of thermal transmission?
- What insights can be gained from this investigation on the non-Newtonian hybrid nanofluid flow via a stretching cylinder?

Bvp4c [55] approach is employed to solve the non-linear ODE arrangement. Tabulated data are used to signify and compare several parameters visually. This study has an extensive range of applications. It is indispensable for maximizing thermal transmission in industrial operations, including energy storage, chemical and nuclear reactors, and cooling mechanisms. It also supports greater magnetic fluid-based technologies in engineering and medicine, improved oil recovery, heat insulation frameworks, and biomedical fluid dynamics.

Mathematical Formulation

We examine a stable Casson Williamson hybrid nanofluid that passes via a vertical cylinder that is stretched and linearly permeable. In this flow issue, the coordinate system used is cylindrical polar (x,r), where x and r indicate the axial and radial directions, correspondingly. Also, the magnetic field of strength B_0 is applied vertically at an angle of ξ. The momentum equation also incorporates the impact of heat buoyancy. The cylinder expands linearly, while integrating the surface velocity, $w_w = a\left(\frac{x}{l}\right)$, employing with its axis. In which a and l denote the velocity and characteristic-length of the cylinder. $T_w(x) = T_0 + E\left(\frac{x}{l}\right)$ is considered to be wall temperature and $T_\infty(x) = T_0 + F\left(\frac{x}{l}\right)$ is the surrounding

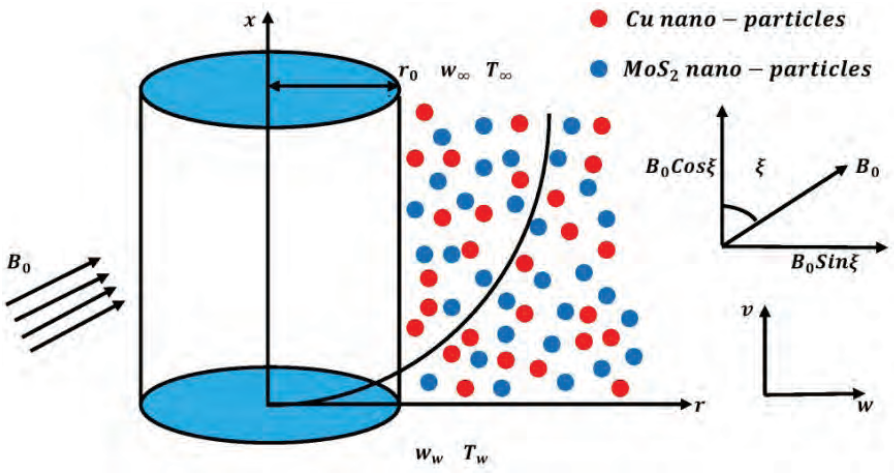


Figure 1. Representation of the flow model.

temperature; in which E, F and T_0 symbolize positive constants ($E > F$)and initial temperature, respectively.

Figure 1 represents the flow model of this study.

The governing equations are as follows (Ref. [27, 56-61]):

$$\frac{\partial(rw)}{\partial x} + \frac{\partial(rv)}{\partial r} = 0 \tag{1}$$

$$\begin{aligned} w\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial r} = {} & (1+\tfrac{1}{B})\frac{\mu_{hnf}}{\rho_{hnf}}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right) \\ & - (1+\tfrac{1}{B})\frac{\mu_{hnf}}{\rho_{hnf}}\left(\sqrt{2}\Gamma\frac{\partial w}{\partial r}\frac{\partial^2 w}{\partial r^2} + \frac{\Gamma}{\sqrt{2}}\frac{1}{r}\left(\frac{\partial w}{\partial r}\right)^2\right) \\ & + \frac{(\rho\beta)_{hnf}}{\rho_{hnf}}g(T-T_\infty) - \frac{\sigma_{hnf}}{\rho_{hnf}}{B_0}^2.Sin^2\xi.w \\ & - \frac{\mu_{hnf}}{\rho_{hnf}}\frac{w}{k}(1+\tfrac{1}{B}) \end{aligned} \tag{2}$$

$$w\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial r} = \frac{k_{hnf}}{(\rho c_p)_{hnf}}\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right) \tag{3}$$

The boundary constraints are:

$$\begin{aligned} & w = a\frac{x}{l},\ \ v = 0,\ \ \ T = T_w(x) \ \ \ while\ r = r_0 \\ & w = 0\ ,\ \ T \to T_\infty(x) \qquad\qquad\ \ while\ r \to \infty \end{aligned} \tag{4}$$

The similarity transformations [61] are as follows:

$$\eta = \frac{r^2 - {r_0}^2}{2r_0}\sqrt{\frac{a}{\upsilon_f l}},\ \psi = \sqrt{\frac{a\upsilon_f}{l}}xr_0 f(\eta),\ \theta = \frac{T-T_\infty(x)}{T_w(x)-T_0} \tag{5}$$

Where $w = \frac{1}{r}\frac{\partial\psi}{\partial r}$, $v = -\frac{1}{r}\frac{\partial\psi}{\partial x}$

The transformed dimensionless nonlinear equations, taking into account the previously indicated transformation, are described as follows:

$$\begin{aligned} f'^2 - f.f'' = {} & (1+\tfrac{1}{B})\left(\frac{\rho_f}{\rho_{hnf}}\frac{\mu_{hnf}}{\mu_f}\right)\{2\gamma f'' + (1+2\gamma\eta)f'''\} \\ & + \left(\frac{\rho_f}{\rho_{hnf}}\frac{\mu_{hnf}}{\mu_f}\right)(1+\tfrac{1}{B})\{3.We.\gamma(1+2\gamma\eta)^{\frac{1}{2}}f''^2 \\ & + 2We(1+2\gamma\eta)^{\frac{3}{2}}f''.f'''\} + (\frac{\rho_f}{\rho_{hnf}}\frac{(\rho\beta)_{hnf}}{(\rho\beta)_f})\,\lambda.\theta \\ & - \{Sin^2\xi.\left(\frac{\rho_f}{\rho_{hnf}}\frac{\sigma_{hnf}}{\sigma_f}\right)M + \left(\frac{\rho_f}{\rho_{hnf}}\frac{\mu_{hnf}}{\mu_f}\right)P(1+\tfrac{1}{B})\}f' \end{aligned} \tag{6}$$

$$\begin{aligned} f'(\theta+\delta) - f\theta' = {} & \left(\frac{k_{hnf}}{k_f}\frac{(\rho c_p)_f}{(\rho c_p)_{hnf}}\right)\left(\frac{1}{Pr}\right)\{2\gamma\theta' \\ & + (2\gamma\eta+1)\theta''\} \end{aligned} \tag{7}$$

The transformed boundary restrictions are represented as:

$$\begin{aligned} & f'(0) = 1,\ \ \ f(0) = 0 \\ & \theta(0) = 1-\delta,\ \ \ f'(\infty) \to 0,\ \ \ \theta(\infty) \to 0 \end{aligned} \tag{8}$$

In which,

$$M = \frac{{B_0}^2 l\sigma_f}{a\rho_f},\ P = \frac{l}{a}\frac{\vartheta_f}{k},\ \gamma = \sqrt{\frac{l\vartheta_f}{a{r_0}^2}},$$

$$\lambda = \frac{Gr_x}{{Re_x}^2},\ \delta = \frac{F}{E}\ ,Pr = \frac{\vartheta_f(\rho c_p)_f}{k_f},We = \frac{\Gamma a^{3/2}x}{\sqrt{2\nu}l^{3/2}}$$

The density ρ_{hnf}, thermal conductivity k_{hnf}, heat expansion $(\rho\beta)_{hnf}$, heat capacity $(\rho C_p)_{hnf}$, dynamic viscosity μ_{hnf} of hybrid nano-fluid is specified as follows (Ref. [56, 61]):

$$\rho_{hnf} = (1-\phi_2)\{(1-\phi_1)\rho_f + \phi_1\rho_{s1}\} + \phi_2\rho_{s2}\ ,$$

$$k_{hnf} = k_{bf}\left\{\frac{k_{s2}+2k_{bf}-2\phi_2(k_{bf}-k_{s2})}{k_{s2}+2k_{bf}+\phi_2(k_{bf}-k_{s2})}\right\}$$

$$\text{Where } k_{bf} = k_f\left\{\frac{k_{s1}+2k_f-2\phi_1(k_f-k_{s1})}{k_{s1}+2k_f+\phi_1(k_f-k_{s1})}\right\},$$

$$(\rho\beta)_{hnf} = (1-\phi_2)\{(1-\phi_1)(\rho\beta)_f + \phi_1(\rho\beta)_{s1}\} + \phi_2(\rho\beta)_{s2}\ ,$$

$$\mu_{hnf} = \frac{\mu_f}{(1-\phi_1)^{2.5}(1-\phi_2)^{2.5}}\ ,$$

$$(\rho C_p)_{hnf} = (1-\phi_2)\left\{(1-\phi_1)\left(\rho C_p\right)_f + \phi_1\left(\rho C_p\right)_{s1}\right\} + \phi_2\left(\rho C_p\right)_{s2}$$

Moreover, , the skin friction and the thermal transport rate are indicated by:

$$C_f {Re_x}^{1/2} = \frac{1}{(1-\phi_1)^{2.5}(1-\phi_2)^{2.5}}(1+\tfrac{1}{B})\left(f''(0) + \frac{We}{2}f''(0)^2\right)$$

$$Nu_x {Re_x}^{-1/2} = -\frac{k_{hnf}}{k_f}\theta'(0)$$

METHODOLOGY

The dimensionless higher order ODEs (6)– (7) are transformed into first-order boundary value problem integrating with the associated boundary conditions (8). The formerly mentioned non-dimensional ODEs are numerically included using MATLAB's bvp4c method in conjunction with a shooting strategy. Even for complex and disordered formulations, the bvp4c solver produces correct results by more precisely addressing boundary value issues. Its powerful mesh augmentation mechanism reduces computational burden and increases computation efficiency. In the meantime, using the bvp4c solver requires a preliminary projection that satisfies its spatial constraints. With a convergence threshold of 10^{-6}, Bvp4c is by far the superlative method for computing the outcomes.

RESULTS AND DISCUSSION

The thermally stratified Casson Williamson Ethylene glycol driven MoS_2 - Cu hybrid nanofluid movement via a stretched vertical cylinder is investigated in this research. To produce a hybrid nanofluid, molybdenum disulfide

Table 1. Thermo physical appearances of base fluid and nano particles [56,60] [created by author]

Thermo-physical attributes	$\rho(kg/m^3)$	$C_p(J/kgK)$	$k(W/mK)$	σ (S/m)	β (K^{-1})
MoS_2	5060	397.21	904.4	2.09 x 10^4	2.8424 x 10^{-5}
Cu	8933	385	400	59.6 x 10^6	1.67 x 10^{-5}
$C_6H_2O_6$	1109	2400	0.258	5.5 x 10^{-6}	5.7 x 10^{-1}

Table 2. The related results of $-\theta'(0)$ incorporating $M = P = \gamma = \delta = \lambda = We = \phi_1 = \phi_2 = 0$, $\xi = \pi/2$, $B \to \infty$ and different values of Prandtl number [56-58] [created by author]

Pr	Ishak and Nazar [57]	Elbashbeshy et al. [58]	Paul et al. [56]	Current study
1	1	1	1.0005	1
10	3.7207	3.7207	3.72	3.7207

(MoS_2) with a solid volume percentage of 0.1 is integrated to the base fluid, ethylene glycol, followed by copper (Cu) nanoparticles with a solid volume concentration of 0.02. This volume ratio doesn't change while the issue is being investigated. As a base fluid, ethylene glycol is used. The thermo-physical features of the hybrid nanoparticle and the basic fluid are summarised in Table 1. The MATLAB Bvp4c method is used to evaluate the numerical results for the velocity and heat distribution. Compared to other studies, this one demonstrates a greater level of data accuracy. Table 2 lists many results for the rate of thermal transport in contrast to other studies showed by Ishak and Nazar [57], Elbashbeshy et al. [58], and Paul et al. [56]. It demonstrates that the Bvp4c approach is successful in producing computational results that are correct and reliable with the results of the other methods.

In Figure 2, how the heat stratification factor influenced the heat profile is displayed. The heat profile diminishes as heat stratification hikes. Substantial temperature variances between the layers caused by strong thermal stratification preserve the denser, colder fluid at the bottom of the fluid and the lighter, warmer fluid at the top. The fluid's inability to mix convectively due to this stratified structure stops thermal energy from being dispersed throughout the fluid volume. Also, the fluid surrounding the cylinder can be colder than the ambient whenever there is appropriate thermal stratification, generating the temperature to lessen. Deka and Paul [59] established that a greater stratification causes fluid flow to opposing across an infinite vertical cylinder. Besides, in our investigation, we have attained the same outcome for Cu/MoS_2 -$C_6H_2O_6$ Casson Williamson hybrid nanofluid movement.

Figure 3 illustrates that as the Weissenberg number (We) upsurges, the velocity profile declines, as advancing Weissenberg number growths viscosity, which upsurges resistance to fluid movement and shrinkages velocity. Figure 4 demonstrates how the Weissenberg number affects the heat profile. Temperature has a direct relationship with the Weissenberg number. Therefore, when the Weissenberg number increases, so does the thermal curve. Enhancing the Weissenberg number emphasizes greater stretching and alignment of fluid elements by exhibiting greater elastic

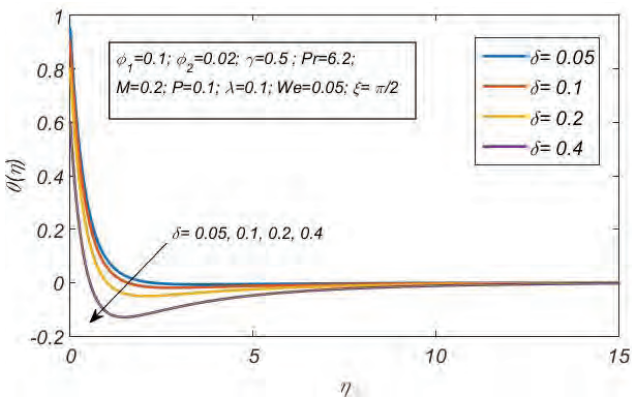


Figure 2. Illustration of temperature with stratification term.

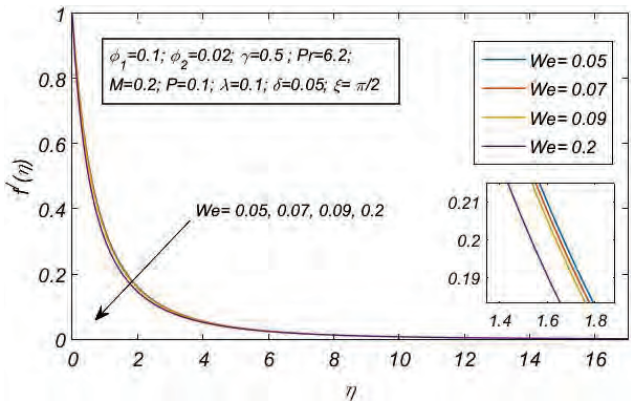


Figure 3. Illustration of velocity with Weissenberg number.

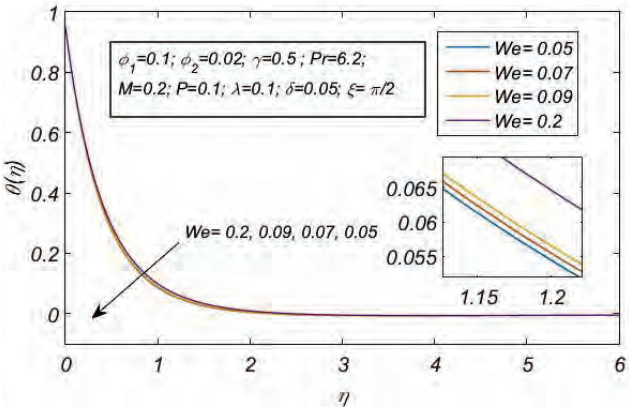


Figure 4. Influence on temperature with Weissenberg number.

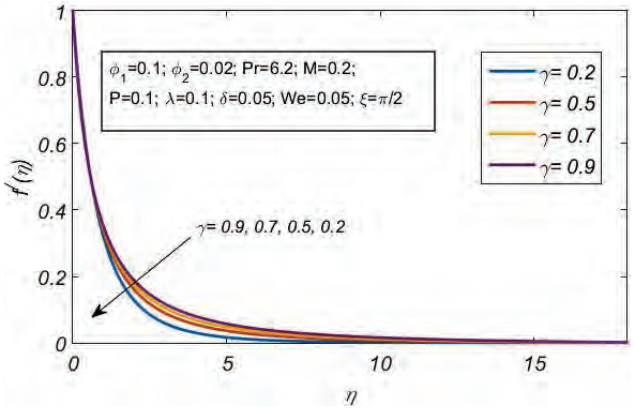


Figure 5. Illustration of velocity with curvature factor.

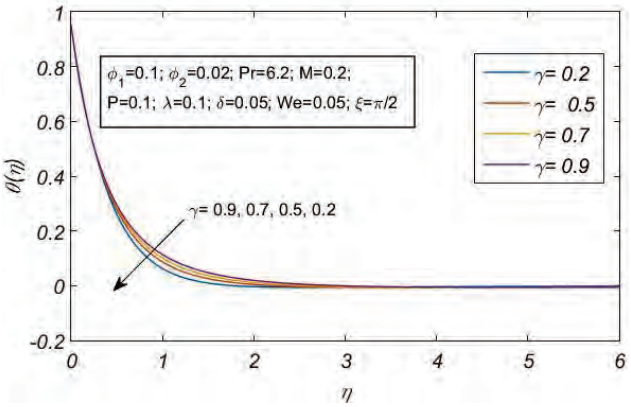


Figure 6. Illustration of temperature with curvature factor.

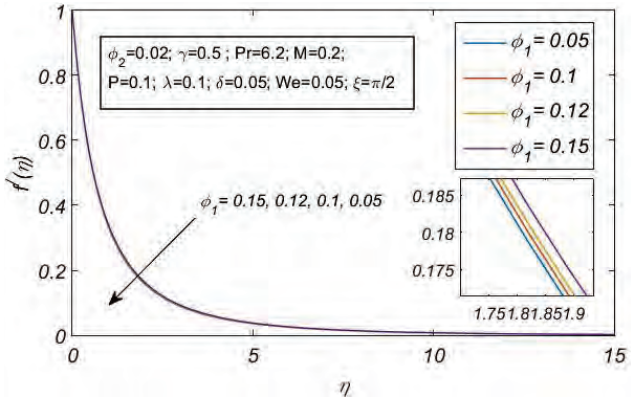


Figure 7. Demonstration of volume proportion (ϕ_1) of MoS_2 on velocity profile.

effects in non-Newtonian fluids. This optimizes heat transport inside the fluid and improves internal energy dissipation. Paul and Nath [32] obtained the same findings for both velocity and thermal curves without thermal stratification influence.

Figures 5 and 6 display how flow and the thermal profile are affected by the curvature factor γ. It appears that when γ broadens, fluid velocity increases. The circumference of the cylinder reduces as γ grows because the radius of curvature decreases. As a result, fluid particle impedance inside the cylinder decreases and fluid velocity increases. Additionally, the temperature curves increase as the size of γ increases. The thickness of the boundary layer increases with an increase in the curvature parameter, which improves thermal diffusion in the fluid. As a result of the enhanced heat flow throughout the layer, temperature curves are raised.

The influence of the volume fraction of MoS_2 on both the velocity and temperature gradient curves are exemplified in Figures 7 and 8, while the volume fraction of Cu remains constant. It is detected that both the flow velocity and heat distribution upsurge with advancement in the volume fraction of MoS_2 nanoliquids.

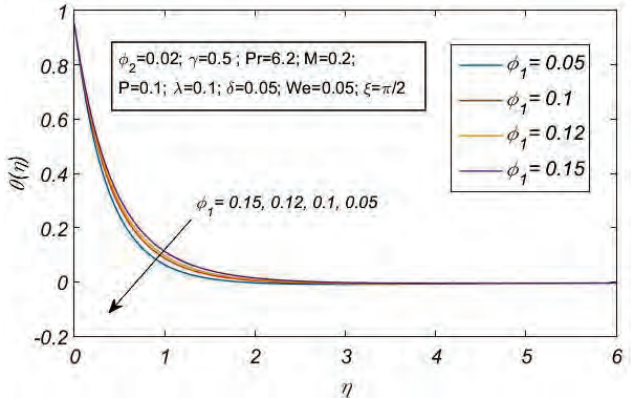


Figure 8. Demonstration of volume proportion (ϕ_1) of MoS_2 on heat distribution.

The skin friction coefficient and thermal transport rate are computed via the distinct noteworthy factors like the We, δ, and γ while ϕ_1= 0.1 and ϕ_2= 0.02 in Table 3. An enhancement in the skin friction, and the rate of heat

Table 3. Skin-friction and Nusselt number for changing value of *We, δ, γ* when ϕ_1= 0.1 and ϕ_2= 0.02

We	*δ*	*γ*	*Skin-Friction Coefficient*	*Nusselt Number*
0.05	0.01	0.1	2.4776	2.5218
0.08			2.4888	2.5186
0.1			2.5016	2.5122
0.3			2.7276	2.4768
0.05	0.03	0.1	2.4796	2.5198
	0.08		2.4845	2.5148
	0.2		2.4964	2.5029
	0.4		2.5163	2,4827
	0.6		2.5362	2.4624
0.05	0.01	0.1	2.4776	2.5218
		0.5	2.9948	2.7508
		0.7	3.2487	2.8622
		1.2	3.8707	3.1332

Table 4. Skin-friction and Nusselt number for varying magnitude of *B, P, λ* when ϕ_1= 0.1 and ϕ_2= 0.02

B	*P*	*γ*	*Skin-Friction Coefficient*	*Nusselt Number*
0.5	0.1	0.1	3.7819	2.7702
0.7			3.2840	2.7539
0.9			2.9984	2.7417
1.2			2.7412	2.7284
0.9	0.1	0.1	2.9984	2.7417
	0.4		3.2348	2.7216
	0.8		3.5265	2.6966
	1.2		3.7962	2.6734
0.9	0.1	0.1	2.9984	2.7417
		0.3	2.9557	2.7441
		0.5	2.9132	2.7464
		0.7	2.8708	2.7486

Table 5. Contrast of the Skin friction and Nusselt numbers for $Cu/C_6H_2O_6$ Casson Williamson-nanoliquid and $Cu\text{-}MoS_2$ / $C_6H_2O_6$ Casson Williamson-hybrid- nanoliquid

***Cu*/Ethylene Glycol**				***Cu-MoS*$_2$ /Ethylene Glycol**			
$\phi_1(MoS_2)$	$(\phi_2(Cu)$	**Skin Friction**	**Nusselt Number**	$\phi_1(MoS_2)$	$(\phi_2(Cu)$	**Skin Friction**	**Nusselt Number**
0	0.02	2.2925	2.3521	0.1	0.02	2.9984	2.7417
	0.03	2.3759	2.3924		0.03	3.0932	2.7907
	0.04	2.4606	2.4333		0.04	3.1899	2.8403
	0.05	2.5468	2.4747		0.05	3.2887	2.8905
	0.09	2.9083	2.6462		0.09	3.7067	3.0987

transference is reported with intensifying *γ*. Moreover, increasing magnitude of *We*, and *δ*, the rate of shear stress is also escalated.

The friction drag coefficient and Nusselt number are computationally simulated against numerous factors, such as the Casson factor, Porosity term, and Thermal buoyancy

parameter in Table 4. An enhancement in the absolute value of friction drag is noticed for boosting the Porosity factor (P), but the trend is the opposite for escalating the Casson (B) and Thermal buoyancy factor (γ). Moreover, the thermal transmission rate is improved for enlarging the heat buoyancy factor but decreases for boosting the value of the Casson and porosity factor.

Table 5 shows that enhancing the volume fraction of *Cu* nano-particles, keeping $\phi_1 = 0$ leads to advances in skin friction and the Nusselt number. Also, the heat transmits rate advances by more than 16.5% while the volume fraction of MoS_2 is 0.1. Moreover, it has been visible that the MoS_2 - *Cu*/ Ethylene glycol Casson Williamson hybrid nanofluid's absolute skin friction augmented by more than 30.7%.

CONCLUSION

The movement of MHD Casson Williamson Ethylene glycol based MoS_2 - *Cu* Hybrid nano-fluid via a linearly stretch vertical cylinder integrating the heat stratification impact and inclined magnetic field have been inspected numerically by employing the Bvp4c approach. The is also taken into consideration in this investigation. The impressions of governing factors on flow velocity and temperature distribution are demonstrated pictorially, along with their impacts on the Nusselt number and skin friction. The following are the main conclusions of this investigation:

- While there is adequate heat stratification, the temperature tends to be negative for ethylene glycol-based MoS_2 - *Cu* Casson Williamson hybrid nanoliquid.
- The rate of thermal transmission is decreasing for the augmenting thermal stratification, Casson and Williamson factors. Moreover, higher as the curvature and thermal buoyancy increase.
- The heat profile escalates with the augmented values of the Weissenberg number and curvature factor.
- Noteworthy expansion in the velocity profile is observed with the improving curvature term, but the opposite pattern is noticed for the Weissenberg number.
- The absolute value of friction drag is enhanced for rising Weissenberg number, thermal stratification, and curvature factor.
- The heat transfer speed goes up by more than 16.5% for MoS_2 - *Cu*/ Ethylene glycol Williamson hybrid nano-liquid in contrast with *Cu*/ Ethylene glycol Williamson nanoliquid.
- Casson Williamson hybrid nano-liquid has a thermal transport rate that is clearly advanced than that of Casson Williamson nanofluid.
- Enhancing the volume fraction of MoS_2 nanoparticle, while the volume fraction of *Cu* nanoparticle is constant, with both the velocity and thermal profile escalating.

Future directions for this research include investigating non-Newtonian fluid practices across multiple geometries, combining experimental validation with numerical simulations, and examining more intricate hybrid nanofluids with varied nanoparticle combinations for boosted thermal properties.

NOMENCLATURE

$(C_p)_f$	Specific heat capacity of fluid ($J\,kg^{-1}\,K^{-1}$)
$(C_p)_{hnf}$	Hybrid nanofluid's specific thermal capacity ($J\,kg^{-1}\,K^{-1}$)
$(C_p)_{s1}$	Specific thermal capacity of the MoS_2 ($J\,kg^{-1}\,K^{-1}$)
$(C_p)_{s2}$	Specific thermal capacity of the Cu ($J\,kg^{-1}\,K^{-1}$)
Gr_x	Grashof term
k_f	Heat conductivity of fluid ($Wm^{-1}\,K^{-1}$)
k_{bf}	Heat conductivity of nano-fluid ($Wm^{-1}\,K^{-1}$)
k_{hnf}	Heat conductivity of hybrid nano-fluid ($Wm^{-1}\,K^{-1}$)
k_{s1},k_{s2}	Heat conductivity of solid nanoparticles ($Wm^{-1}\,K^{-1}$)
k	Permeability (m^2)
M	Magnetic term
P	Porosity term
Pr	Prandtl Number
B	Casson factor
Re_x	Local Reynolds factor
T	Fluid temperature (K)
T_∞	Fluid's surroundings temperature (K)
T_w	Temperature at the wall (K)
w	The component of velocity in the r-direction(ms^{-1})
v	The component of velocity in the x direction (ms^{-1})
B_0	Strength of Magnetic field (NmA^{-1})
$(\beta_T)_f$	Fluid's thermal expansion (K^{-1})
$(\beta_T)_{s1}$	Thermal expansion of the MoS_2 (K^{-1})
$(\beta_T)_{s2}$	Thermal expansion Cu (K^{-1})
μ_f	Fluid's dynamic viscosity (mPa)
μ_{hnf}	Dynamic viscosity of hybrid nanofluid (mPa)
σ_{hnf}	Electric conductivity of the hybrid nanofluid ($Ohm^{-1}\,m^{-1}$)
ρ_f	Density of the fluid (kgm^{-3})
ρ_{hnf}	Density of the hybrid nanofluid (kgm^{-3})
ρ_{s1}	Density of MoS_2 (kgm^{-3})
ρ_{s2}	Density of Cu (kgm^{-3})
We	Weissenberg number
ϕ_1,ϕ_2	Volume fraction of MoS_2 and *Cu*
γ	The curvature parameter
δ	Thermal stratification factor
η	Similarity variable
λ	Heat buoyancy factor
ψ	Stream function

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Kataria HR, Mittal AS. Mathematical model for velocity and temperature of gravity-driven convective optically thick nanofluid flow past an oscillating vertical plate in presence of magnetic field and radiation. J Niger Math Soc 2015;34:303–317. [CrossRef]

[2] Sheikholeslami M, Kataria HR, Mittal AS. Radiation effects on heat transfer of three dimensional nanofluid flow considering thermal interfacial resistance and micro mixing in suspensions. Chin J Phys 2017;55:2254–2272. [CrossRef]

[3] Kataria HR, Mittal AS. Velocity, mass and temperature analysis of gravity-driven convection nanofluid flow past an oscillating vertical plate in the presence of magnetic field in a porous medium. Appl Therm Eng 2017;110:864–874. [CrossRef]

[4] Sheikholeslami M, Kataria HR, Mittal AS. Effect of thermal diffusion and heat-generation on MHD nanofluid flow past an oscillating vertical plate through porous medium. J Mol Liq 2018;257:12–25. [CrossRef]

[5] Mittal AS, Kataria HR. Three dimensional CuO–Water nanofluid flow considering Brownian motion in presence of radiation. Karbala Int J Mod Sci 2018;4:275–286. [CrossRef]

[6] Akbarı N, Gholınıa M, Gholınıa S, Dabbaghıan S, Ganjı DD. Analytical and numerical study of hydrodynamic nano fluid flow in a two–dimensional semi-porous channel with transverse magnetic field. Sigma J Eng Nat Sci 2018;36:587–608.

[7] Li Z, Sheikholeslami M, Mittal AS, Shafee A, Haq RU. Nanofluid heat transfer in a porous duct in the presence of Lorentz forces using the lattice Boltzmann method. Eur Phys J Plus 2019;134:1–10. [CrossRef]

[8] Patel HR, Mittal AS, Darji RR. MHD flow of micropolar nanofluid over a stretching/shrinking sheet considering radiation. Int Commun Heat Mass Transf 2019;108:104322. [CrossRef]

[9] El Hattab M, Boumhaout M, Oukach S. MHD natural convection in a square enclosure using carbon nanotube-water nanofluid with two isothermal fins. Sigma J Eng Nat Sci 2024;42:1075–1087. [CrossRef]

[10] Maskeen MM, Zeeshan A, Mehmood OU, Hassan M. Heat transfer enhancement in hydromagnetic alumina–copper/water hybrid nanofluid flow over a stretching cylinder. J Therm Anal Calorim 2019;138:1127–1136. [CrossRef]

[11] Khashi'ie NS, Hafidzuddin EH, Arifin NM, Wahi N. Stagnation point flow of hybrid nanofluid over a permeable vertical stretching/shrinking cylinder with thermal stratification effect. CFD Lett 2020;12:80–94.

[12] Elsaid EM, Abdel-wahed MS. Impact of hybrid nanofluid coolant on the boundary layer behavior over a moving cylinder: Numerical case study. Case Stud Therm Eng 2021;25:100951. [CrossRef]

[13] Khan U, Zaib A, Ishak A, Sherif ESM, Waini I, Chu YM, et al. Radiative mixed convective flow induced by hybrid nanofluid over a porous vertical cylinder in a porous media with irregular heat sink/source. Case Stud Therm Eng 2022;30:101711. [CrossRef]

[14] Nath JM, Paul A, Das TK. Magnetohydrodynamics Stagnation Point Flow of Hybrid Nanofluids with Distinct Base Fluid over an Exponentially Extending Cylinder: A Numerical Comparative Analysis.Nano 2024;2450123. [CrossRef]

[15] Crane LJ. Boundary layer flow due to a stretching cylinder. Z Angew Math Phys 1975;26:619–622. [CrossRef]

[16] Wang CY. Fluid flow due to a stretching cylinder. Phys Fluids 1988;31:466–468. [CrossRef]

[17] Waqas H, Naqvi SMRS, Alqarni MS, Muhammad T. Thermal transport in magnetized flow of hybrid nanofluids over a vertical stretching cylinder. Case Stud Therm Eng 2021;27:101219. [CrossRef]

[18] Sreenivasa BR, Faqeeh JA, Alsaiari A, Alzahrani HA, Malik MY. Numerical study of heat transfer mechanism in the flow of ferromagnetic hybrid nanofluid over a stretching cylinder. Waves Random Complex Media 2022;1–17. [CrossRef]

[19] Nath JM, Das TK, Dey B. Flow of Newtonian and non-Newtonian hybrid nanofluid in cylindrical geometry. Nanofluids Technol Therm Sci Eng 2024;190–208. [CrossRef]

[20] Hamid A, Khan M, Khan U. Thermal radiation effects on Williamson fluid flow due to an expanding/contracting cylinder with nanomaterials: Dual solutions. Phys Lett A 2018;382:1982–1991. [CrossRef]

[21] Khan M, Salahuddin T, Malik MY, Mallawi FO. Change in viscosity of Williamson nanofluid flow due to thermal and solutal stratification. Int J Heat Mass Transf 2018;126:941–948. [CrossRef]

[22] Mittal AS, Patel HR, Darji RR. Mixed convection micropolar ferrofluid flow with viscous dissipation, joule heating and convective boundary conditions. Int Commun Heat Mass Transf 2019;108:104320. [CrossRef]

[23] Mittal AS, Patel HR. Influence of thermophoresis and Brownian motion on mixed convection two dimensional MHD Casson fluid flow with non-linear radiation and heat generation. Physica A: Stat Mech Appl 2020;537:122710. [CrossRef]
[24] Nath JM, Paul A, Das TK. Heat transfer characteristics of magnetohydrodynamic Casson stratified hybrid nanofluid flow past a porous stretching cylinder. J Thermal Eng 2021;10:1137–1148. [CrossRef]
[25] Hussain Z, Hayat T, Alsaedi A, Ullah I. On MHD convective flow of Williamson fluid with homogeneous-heterogeneous reactions: A comparative study of sheet and cylinder. Int Commun Heat Mass Transf 2021;120:105060. [CrossRef]
[26] Bilal M, Siddique I, Borawski A, Raza A, Nadeem M, Sallah M. Williamson magneto nanofluid flow over partially slip and convective cylinder with thermal radiation and variable conductivity. Sci Rep 2022;12:1–15. [CrossRef]
[27] Almaneea A. Numerical study on heat and mass transport enhancement in MHD Williamson fluid via hybrid nanoparticles. Alexandria Eng J 2022;61:8343–8354. [CrossRef]
[28] Kataria HR, Mistry M, Mittal A. Influence of non-linear radiation on MHD micropolar fluid flow with viscous dissipation. Heat Transf 2022;51:1449–1467. [CrossRef]
[29] Rashad AM, Nafe MA, Eisa DA. Heat variation on MHD Williamson hybrid nanofluid flow with convective boundary condition and Ohmic heating in a porous material. Sci Rep 2023;13:6071. [CrossRef]
[30] Das TK, Paul A, Nath JM. Endo/exothermic analysis of Casson and Maxwell quadra hybrid nanofluid flow configured by pollutant concentration with thermal and solute jump. J Taibah Univ Sci 2024;18:2382940. [CrossRef]
[31] Nabwey HA, EL-Hakiem AMA, Khan WA, Abdelrahman ZM, Rashad AM, Hawsah MA. Magnetic williamson hybrid nanofluid flow around an inclined stretching cylinder with joule heating in a porous medium. Chem Eng J Adv 2024;18:100604. [CrossRef]
[32] Paul A, Nath JM. Magneto-Hydrodynamic Stagnation Point Flow of Casson Williamson Hybrid Nanofluid Incorporating Viscous Dissipation and Suction/Injection Effect Past an Exponentially Stretching Cylinder. J Nanofluids 2024;13:710–720. [CrossRef]
[33] Das TK, Paul A, Nath JM. Thermo-convection driven Sodium Alginate-based Darcy–Forchheimer EMHD Williamson hybrid nanofluid flow with varying thermal distribution. Mod Phys Lett B 2024;2450405. [CrossRef]
[34] Srinivasacharya D, Surender O. Non-similar solution for natural convective boundary layer flow of a nanofluid past a vertical plate embedded in a doubly stratified porous medium. Int J Heat Mass Transf 2014;71:431–438. [CrossRef]
[35] Srinivasacharya D, Surender O. Non-darcy mixed convection induced by a vertical plate in a doubly stratified porous medium. J Porous Media 2014;17. [CrossRef]
[36] Srinivasacharya D, Surender O. Double Stratification effects on mixed convection along a vertical plate in a non-Darcy porous medium. Procedia Eng 2015;127:986–993. [CrossRef]
[37] Srinivasacharya D, Surender O. Effect of double stratification on mixed convection boundary layer flow of a nanofluid past a vertical plate in a porous medium. Appl Nanosci 2015;5:29–38. [CrossRef]
[38] Srinivasacharya D, Motsa SS, Surender O. Numerical study of free convection in a doubly stratified non-darcy porous medium using spectral quasi-linearization method. Int J Nonlinear Sci Numer Simul 2015;16:173–183. [CrossRef]
[39] Bilal S, Rehman KU, Malik MY. Numerical investigation of thermally stratified Williamson fluid flow over a cylindrical surface via Keller box method. Results Phys 2017;7:690–696. [CrossRef]
[40] Khan M, Salahuddin T, Malik MY, Mallawi FO. Change in viscosity of Williamson nanofluid flow due to thermal and solutal stratification. Int J Heat Mass Transf 2018;126:941–948. [CrossRef]
[41] Ramzan M, Gul N, Chung JD, Kadry S, Chu YM. Numerical treatment of radiative Nickel–Zinc ferrite-Ethylene glycol nanofluid flow past a curved surface with thermal stratification and slip conditions. Sci Rep 2020;10:1–14. [CrossRef]
[42] Khan MN, Ahmad S, Nadeem S. Flow and heat transfer investigation of bio–convective hybrid nanofluid with triple stratification effects. Phys Scr 2021;96:065210. [CrossRef]
[43] Masood S, Farooq M. Influence of thermal stratification and thermal radiation on graphene oxide-Ag/H2O hybrid nanofluid. J Therm Anal Calorim 2021;143:1361–1370. [CrossRef]
[44] Manzoor U, Muhammad T, Farooq U, Waqas H. Investigation of thermal stratification and nonlinear thermal radiation in Darcy-Forchheimer transport of hybrid nanofluid by rotating disk with Marangoni convection. Int J Ambient Energy 2022;1–8. [CrossRef]
[45] Kairi RR, Roy S, Raut S. Stratified thermosolutal Marangoni bioconvective flow of gyrotactic microorganisms in Williamson nanofluid. Eur J Mech B Fluids 2023;97:40–52. [CrossRef]
[46] Mohammadi H, Kumar S, Rezapour S, Etemad S. A theoretical study of the Caputo–Fabrizio fractional modeling for hearing loss due to Mumps virus with optimal control. Chaos Solitons Fract 2021;144:110668. [CrossRef]
[47] Khan H, Alzabut J, Shah A, He ZY, Etemad S, Rezapour S, et al. On fractal-fractional waterborne disease model: A study on theoretical and numerical aspects of solutions via simulations. Fractals 2023;31:2340055. [CrossRef]

[48] Chávez-Vázquez S, Lavín-Delgado JE, Gómez-Aguilar JF, Razo-Hernández JR, Etemad S, Rezapour S. Trajectory tracking of Stanford robot manipulator by fractional-order sliding mode control. Appl Math Modelling 2023;120:436–462. [CrossRef]

[49] Salahuddin T, Iqbal MA, Bano A, Awais M, Muhammad S. Cattaneo-Christov heat and mass transmission of dissipated Williamson fluid with double stratification. Alexandria Eng J 2023;80:553–558. [CrossRef]

[50] Farooq U, Waqas H, Makki R, Ali MR, Alhushaybari A, Muhammad T, Imran M. Computation of Cattaneo-Christov heat and mass flux model in Williamson nanofluid flow with bioconvection and thermal radiation through a vertical slender cylinder. Case Stud Therm Eng 2023;42:102736. [CrossRef]

[51] Alsallami SA, Abbas T, Al-Zubaidi A, Khan SU, Saleem S. Analytical assessment of heat transfer due to Williamson hybrid nanofluid (MoS2+ ZnO) with engine oil base material due to stretched sheet. Case Stud Therm Eng 2023;51:103593. [CrossRef]

[52] Salahuddin T, Fatima G, Awais M, Khan M, Al Awan B. Adaptation of nanofluids with magnetohydrodynamic Williamson fluid to enhance the thermal and solutal flow analysis with viscous dissipation: A numerical study. Results Eng 2024;21:101798. [CrossRef]

[53] Farooq U, Safeer M, Cui J, Hussain M, Naheed N. Forced convection analysis of Williamson-based magnetized hybrid nanofluid flow through a porous medium: Nonsimilar modeling. Numer Heat Transf B Fundam 2024;1–17. [CrossRef]

[54] You X, Wang Y. Series Solutions of Three-Dimensional Magnetohydrodynamic Hybrid Nanofluid Flow Heat Transf Nanomater 2024;14:316. [CrossRef]

[55] Shampine LF, Kierzenka J, Reichelt MW. Solving boundary value problems for ordinary differential equations in MATLAB with bvp4c. Tutorial Notes 2000;1–27.

[56] Paul A, Nath JM, Das TK. An investigation of the MHD Cu-Al2O3/H2O hybrid-nanofluid in a porous medium across a vertically stretching cylinder incorporating thermal stratification impact. J Therm Eng 2023;9:1300847. [CrossRef]

[57] Ishak A, Nazar R. Laminar boundary layer flow along a stretching cylinder. Eur J Sci Res 2009;36:22–29.

[58] Elbashbeshy EMA, Emam TG, El-Azab MS, Abdelgaber KM. Laminar boundary layer flow along a stretching cylinder embedded in a porous medium. Int J Phys Sci 2012;7:3067–3072. [CrossRef]

[59] Deka RK, Paul A. Transient free convection flow past an infinite moving vertical cylinder in a stably stratified fluid. Trans ASME J Heat Transfer 2012;134:0425031. [CrossRef]

[60] Unyong B, Vadivel R, Govindaraju M, Anbuvithya R, Gunasekaran N. Entropy analysis for ethylene glycol hybrid nanofluid flow with elastic deformation, radiation, non-uniform heat generation/absorption, and inclined Lorentz force effects. Case Stud Therm Eng 2022;30:101639. [CrossRef]

[61] Paul A, Nath JM, Das TK. Thermally stratified Cu–Al2O3/water hybrid nanofluid flow with the impact of an inclined magnetic field, viscous dissipation and heat source/sink across a vertically stretching cylinder. ZAMM J Appl Math Mech 2023;e202300084. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.1956

Research Article

Evaluation and effectiveness of student's awareness on outcome-based education perception and implementation in teaching/learning activities a cross-sectional study

Vennila J.[1], Basker P.[2], Venkatachalam M.[3,*], Prakash V. S.[4], Dafik[5], Devi P.[6]

[1]*Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal 576104, India*
[2]*Department of Mathematics, Chandigarh University, Punjab, 140413, India*
[3]*Department of Mathematics, Kongunadu Arts and Science College, Coimbatore, Tamil Nadu, 641029, India*
[4]*Department of Computer Science, Kristu Jayanti College (Autonomous), Bengaluru, 560077, India*
[5]*Department of Mathematics Education, University of Jember, Jember, 68121, Indonesia*
[6]*Department of Mathematics, Kongunadu Arts and Science College, Coimbatore, Tamil Nadu, India*

ARTICLE INFO

Article history
Received: 16 February 2024
Revised: 18 April 2024
Accepted: 20 December 2024

Keywords:
Assessment; Attainment Level; Course outcome; Program Outcome; Teaching/learning

Mathematics Subject Classification
62P10, 62P20, 62P25

Computing Classification System
A20, C10

ABSTRACT

Outcome Based Education evaluates students' knowledge, behavior, and practical abilities. It is crucial for determining learning goals, changing courses, pinpointing learning gaps, boosting motivation and engagement, and fostering ongoing educational progress. Outcome-Based Education implementation is enhanced and the learning experience is fostered by the use of data from student evaluations. This research investigates students' knowledge of Outcome-Based Education and analyzes how demographic factors like age and gender affect their understanding. The data was analyzed using descriptive statistics and the chi-square test (version 4.2.3) in R software after 250 pupils responded to the survey. 65.86% of pupils are acquainted with outcome-based education. Students who were older or had completed NPTEL courses were more informed. According to these findings, age and further coursework have an impact on one's understanding of outcome-based education, which may inform methods for putting Outcome-Based Education into practice. This is one of the first studies to examine how demographic variables, such as age and NPTEL course completion, affect the awareness of Outcome-based education. Student understanding of Outcome-Based Education can be evaluated to gain insightful knowledge about its practical efficacy. Student viewpoints have a significant influence on the development of learning strategies that are student-centered, relevant, and successful. Including pupils in the evaluation process helps to improve the overall quality of education as well as promote ongoing development.

Cite this article as: Venila J, Basker P, Venkatachalam M, Prakash VS, Dafik, Devi P. Evaluation and effectiveness of student's awareness on outcome-based education perception and implementation in teaching/learning activities a cross-sectional study. Sigma J Eng Nat Sci 2026;44(2):1251–1260.

***Corresponding author.**
*E-mail address: venkatmaths@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Conventional education system, with its focus on standardized teaching, Rote recitation and repeated testing may not effectively educate pupils for the difficulties of the twenty-first century. To deal with this, India has adopted outcome-based education (OBE), which emphasizes the results students should achieve after completing their course. Problem-solving, practical knowledge, career preparation, and critical thinking are given priority in OBE. India aims to improve its global competitiveness by aligning academic achievement with the line of business needs and international norms. The method of giving students the essential and required skills, accreditation, and flexible opportunities to succeed in today's competitive world. Implementation of this method seeks to overcome the shortcomings of traditional education and educational methods and foster economic development. OBE is mainly focused on initiating and putting forward learning objectives and developing curricula, pedagogy, and assessments to fulfill those objectives. As the technology is advancing expeditiously, not only developing just technical proficiency help graduates, but also effective communication, problem-solving and other skills. Traditional teaching methods are insufficient for developing these skills, making OBE a crucial strategy for preparing students for the future. In recent decades, scholars worldwide have placed significant emphasis on a system known as outcome-based education. In this framework, the facilitator first defines the desired objectives or competencies that students must evince, which generally cover the combinations of skills, and enhances the mindset related to the topic and develops expertise. Once when these outcomes are established, the curriculum designed aligns to help students to achieve them with the instructional strategies, regular practices and evaluation methods accordingly. Unlike the traditional educational systems, which emphasize the teacher-led guidance, self-directed learning in outcome-based education. The new approach is set in such a way that clear goals are set and are also accomplished by the students by the end of each course upon graduation. In this system, professors function as facilitators who enhance students' learning experiences. Outcome-Based Education is implemented by including Course Outcomes (COs), Program Outcomes (POs), and Program Educational Objectives (PEOs) into the curriculum. The technique of Outcome-Based Education (OBE) is constantly changing. OBE has two primary goals: first, to set precise learning outcomes around which the entire academic system is built, and second, to provide conditions and opportunities that encourage students to accomplish these planned results. The current paper intends to emphasize and analyze the many writers' contributions to OBE, as well as to suggest a theoretical framework for its application. With universities increasingly embracing OBE to align with international standards, recent educational reforms have sought to depart from conventional teaching methods. The implementation of this transition in Nigeria's educational system has drawn significant academic attention and debate. Key elements of Outcome-Based Education (OBE) have been explored by Borsoto et al. [1], Hansen et al. [2], and Towers et al. [3], that includes the method of assessing students' learning and altering educational practices to stimulate outcome-based approaches.

Harden et al. [4], studied the efficacy of Outcome-Based Education (OBE), while Mukhopadhyay and Smith [5] examined its broader effects. Tan et al. [6] discussed how novel instructional strategies have to be integrated into OBE, while Asis [7] focused on specific student learning difficulties in the OBE framework. Taken together, these research offers helpful suggestions and instructions to the educators and students on how the OBE principles can be implemented into practice. Additionally, studies by Caguimbal et al. [8], Deneen et al. [9], Cook et al. [10], and Laguador et al. [11] looked at the knowledge and interests that students had in OBE. Their reseach was also focused on principles, instructional strategies, methodological frameworks and the opportunities it offers, including the challenges faced. Collectively, these studies highlight the importance of OBE in contemporary education, underscoring its effects, implementation challenges, and strategies for intensifying teaching and learning results. By 2014, Borsoto et al. [1] and Laguador et al. [12] investigated the gap between knowledge and practice in Outcome-Based Education (OBE) implementation among engineering educators at LPU, highlighting the extent of implementation and its utility within the engineering and technology department. The theoretical perspectives and current tasks within OBE frameworks were provided by Shaheen [13], it offers insights into ongoing discussions and practical outcomes. Vennila J et al. [14] studied how COVID-19 pandemic affected academic achievement, which also highlighted how teachers and students felt about modern pedagogical methods, particularly the use of R programming,. This research underscores the ongoing debate around OBE, the challenges of implementing OBE and the prerequisite to modify the educational practices to keep up with changing conditions and technological breakthroughs.

While OBE is implemented broadly, there is still research going on on how aware the students are of it. Hardly many studies have looked at how variables like age, interest and involvement in additional courses like NPTEL affect students' level of understanding and improvements, while the majority of them have concentrated on its implementation. This study aims to close the gap by looking at the effect of demographic factors on students' awareness and understanding of OBE. Based on this literature review, potential research gaps may be identified in ongoing trends and challenges within education.

1. It can be less sufficient research to determine how students retain and learn knowledge based on their OBE paradigm. It can provide useful information

on the long-term applicability of OBE in the development of learning and skill mastery.

2. Despite the fact that OBE has the potential to foster equity and inclusion in that it caters to the different learning needs of all learners, there has been very little research on its impact on marginalized and underrepresented groups, including special needs learners, ELL learners, and socio-economically disadvantaged learners. More research should be done to investigate how OBE can be used to cater to the needs of all kids in order to eradicate inequalities in education.
3. Despite growing recognition of the significance of teachers' professional development in the implementation of OBE, there has been very little research on effective strategies for teachers in the process. Further research should be conducted to examine the forms of teachers' training that can be most helpful in assisting teachers to comprehend the concepts of OBE.
4. As evaluation forms an integral part of the process in OBE, there has been continuous debate and confusion about the best methodologies for evaluating student learning outcomes in an OBE system. Further research should be conducted to examine the validity, reliability, and fairness of different forms of evaluation in achieving the desired results in student learning.
5. The effectiveness of OBE can differ substantially depending on the institutional culture, policy context, and socio-cultural setting. There is a need for research on these variables to assess their impact on the formulation and implementation of OBE.
6. In view of the increasing use of technology for various educational pursuits, there is a need for study on its possible integration with OBE for more effective teaching and learning experiences, evaluation and feedback mechanisms, and individualized learning paths.

Recent developments and advancements in Microgrid (MG) and Optimal Power Flow (OPF) have greatly improved the efficiency and reliability of modern energy infrastructure. Similarly, Optimal Power Flow calculations have improved in terms of precision and efficiency with the inclusion of recent advances in optimization techniques, i.e., Convex Relaxation and Machine Learning. These advances are important in improving energy efficiency and are an example of how technology is helping to solve the challenges posed by modern issues in power management. Future developments in this area are expected to improve these systems even further, resulting in improved energy solutions.

By addressing these gaps in research, we hope to improve our exploration of the intricacies and complexities of Outcome-Based Education (OBE) and improve our efforts in ensuring its effectiveness and its positive impacts on students' learning outcomes. To achieve this objective, this research aims to generate findings that will be useful in developing strategies on how to effectively promote OBE and improve students' engagement and learning outcomes and experiences. The objectives of this study are as follows: First, it aims to develop and validate a questionnaire that will measure students' awareness of and perceptions about OBE implementation in teaching and learning activities; and second, it aims to investigate students' knowledge and attitudes about OBE. It should allow professionals to take well-informed decisions about the best way of integrating OBE and establishing an environment of continuous improvement throughout the process of teaching and learning.

MATERIALS AND METHODS

Data for this research was acquired using a survey questionnaire designed for descriptive analysis. The primary objectives of the questionnaire were to assess students' knowledge and attitudes towards outcome-based education (OBE) and their commitment to implementing OBE in teaching and learning activities Rhaffor et al., [15].

Sample

The survey included 250 pupils in all, and G*Power 3.1 was used to calculate the sample size. Using the convenience sampling method, the participants were chosen for the study. To identify the significant effects, the statistical validity was preserved thanks to this sample size's ability. A questionnaire was meticulously designed and validated to measure the Students' understanding, views and awareness of OBE. Validation involved expert evaluation, focus group discussions and a pilot study. From this 5% to 10% of the data collected and used for finding reliability analysis. Data for the pilot survey and primary analysis were collected using convenience sampling techniques. Descriptive statistics and Chi-square tests were used to analyze the data. R software(version 4.2.3) was used to perform all the statistical analyses.

Phase of Study

Phase I: Baselines identify the sample size.

Phase II: Questionnaire developed, validated, and circulated to respondents with their consent using focus group discussions.

Phase III: Data analyzed using R open sources (4.2.3. version)

Phase IV: Result and importance discussed.

Phase V: Conclusion.

DATA ANALYSIS

A systematic approach was adopted in the development and validation of the questionnaire, which includes steps like literature study, FGD, expert evaluation, and

pilot study, among others. Informed consent was obtained from all the participants before they were involved in the study, and the study was approved by the institutional ethics committee of Kongunadu Arts and Science College in Coimbatore, Tamil Nadu, India. In order to validate the questionnaire, a study was done using a survey research design with a convenience sampling approach. There were two main parts to the study: Section A and Section B. Section A was used to obtain background information about the participants, while Section B was further divided into three parts: dedication to OBE implementation, teaching and learning, and awareness of OBE, as shown in Table 1. The questionnaire attempted to measure the awareness and attitude of students towards OBE using a 5-point Likert scale ranging from 1: Strongly Disagree to 5: Strongly Agree. The data was then statistically analyzed using R programming. The questionnaire was also administered in English language.

Reliability Assessment

The questionnaire was also validated using a cross-sectional study, and data was collected from 20 participants using a convenience sampling technique. The internal consistency of the entire questionnaire was also verified using Cronbach alpha coefficient, and it was found to be 0.85, indicating a very high level of internal consistency. The independent alpha coefficients for the components of the questionnaire were 0.92, 0.85, and 0.83, indicating an excellent level of internal consistency. According to Sekaran (2000), internal consistency reliability greater than 0.6 is good, greater than 0.7 is very good, and greater than 0.9 is exceptional.

Content and face validity were assessed through expert judgment and focus group discussions (FGD). Construct validity was carried out through factor analysis, and item analysis was done to assess construct validity based on the filled-out forms. A correlation matrix was constructed to measure the level of association among the items. Sampling adequacy was assessed using Kaiser-Meyer-Olkin (KMO), and it scored 0.947. Bartlett's sphericity test was carried out and gave a chi-square of 23.56 with a p-value of 0.001. Factor analysis was done based on Kaiser's criteria and varimax rotation to assess domain validity. Twenty-one items were analyzed, and they were confirmed through factor analysis.

Table 1. Data verbal interpretation guide

Weight	Verbal Interpretation
5	Strongly agree
4	Agree
3	Neutral
2	Disagree
1	Strongly disagree

RESULTS AND DISCUSSION

Frequency and percentage were used to assess students' knowledge of Outcome-Based Education (OBE) and their attitudes toward its implementation. Using the well-structed questonnaire the data were collected, analyzed and interpreted based on the guidelines outlined in Table 2.

Descriptive statistical analysis were analysed and mentioned in Table 2 and Figure 1, From the results, it observed that, majority of the respondents are lies in between the age group of 20-23 (52%), form that most of them are female respondents (60.6%) to compare the male respondents. Most of participatents are educated in PG level (55.9%) with having science background (54.7%).

Measures of Association Between Demographic Data And Awareness of OBE (Knowledge and Based Questions)

The results in Table 3 and Figure 2 show how students know about their colleges Outcome-Based Education strategy. This varies based on a lot of things. We used the Chi-square test to see how students answers relate to things like gender, age what they are studying. If they finished NPTEL courses. We found differences when it comes to age and finishing NPTEL courses. Older students and those who finished NPTEL courses understand the colleges Outcome-Based Education strategy. They also know more about how the college calculates Program Educational Objectives and Program Outcomes. The Outcome-Based Education strategy and these factors are important, for helping students understand things better. The colleges Outcome-Based Education strategy is affected by these factors.

Table 2. Descriptive statistics of demographic data-based OBE based students' knowledge questionnaire

Characteristics	Frequency (Percentage)
Age group	
17-20	102 (40.2%)
20-23	132 (52%)
23-26	19 (7.5%)
26-29	1 (0.4%)
Gender	
Male	100 (39.4%)
Female	154 (60.6%)
Level of study	
UG	109 (42.9%)
PG	142 (55.9%)
Diploma	3 (1.2%)
Branch	
Arts	101 (39.8%)
Science	139 (54.7%)
Engineering & Technology	14 (5.5%)

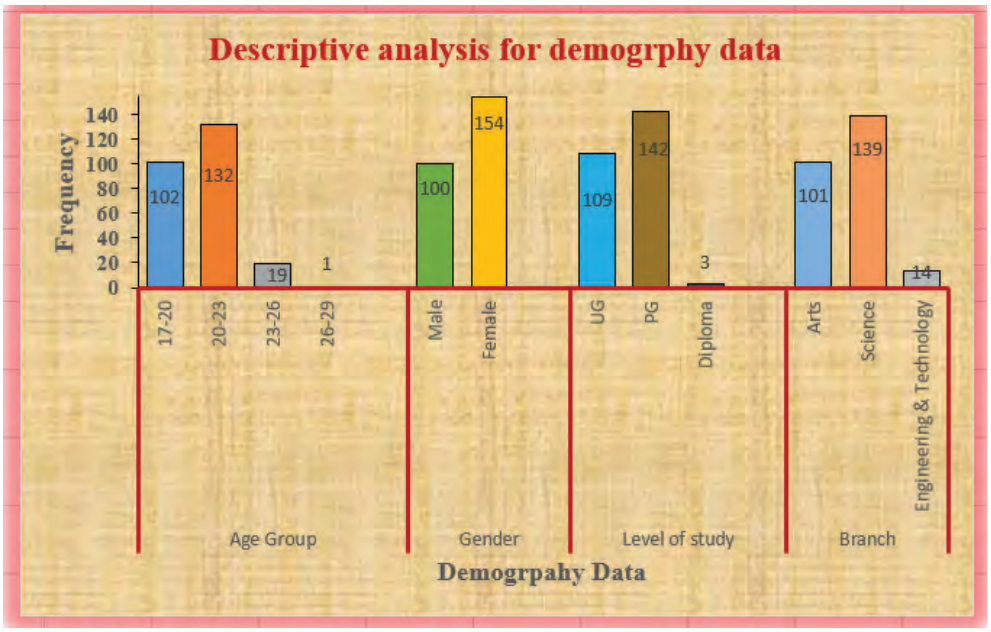


Figure 1. Descriptive analysis of survey respondents.

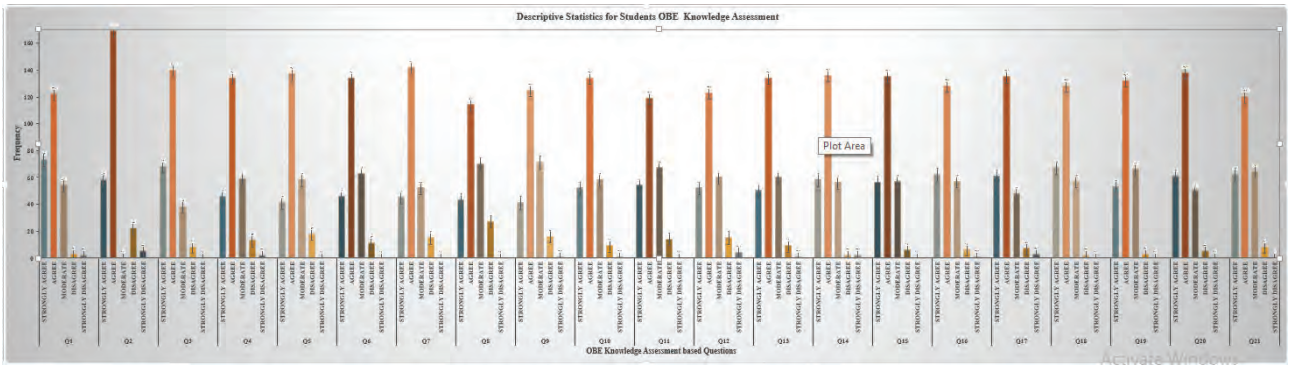


Figure 2. Descriptive analysis for Students awareness about OBE.

From the above reults, it is obseved that there is no statistical significant differene between male/female and academic degree branches. Furthermore, the completion of NPTEL courses significantly improved familiarity with Bloom's taxonomy and the various learning areas, demonstrating the importance of additional instructional materials. In most cases, the findings of the chi-square test indicate that the significant variables in improving students' understanding of educational frameworks are age and involvement, whereas gender and academic field seem to have minimal bearing on this knowledge.

The aim of the survey was to measure students' perceptions of the OBE approach and finding its association between various socio-economic and educational outcome variables. Chi-squre test was used to to find the association between these variables. For students' perceptions of having been briefed on OBE during Orientation Week, significant relationships were found for age ($p = 0.005$) and NPTEL course completion ($p = 0.009$), implying that students' perceptions differ based on age and NPTEL course completion status. However, there were no significant relationships found for students' perceptions of having been briefed on OBE during Orientation Week and other characteristics, namely gender, level of study, and branch of study.

Regarding 'Initial explanation of 'Course Outcomes (COs),' it was found that there were no significant associations with age, gender, level of study, branch of study, and NPTEL course completion. The results were shows that, the perceptions of the students about understanding the explanation of CO have similar trends based on different characteristics of the students. The perceptions regarding accessing 'OBE information' were found to have no

Table 3. Chi-square test results showing significant factors influencing OBE knowledge

Characteristics		Frequency (Percentage)	Age Vs χ^2 (p-value)	Gender Vs χ^2 (p-value)	Level of study Vs χ^2 (p-value)	Branch Vs χ^2 (p-value)	NPTEL Course Completion Vs χ^2 (p-value)
1. I am aware that my college is launching an OBE strategy that focuses on students' performance.	Strongly Agree	73 (28.7%)	21.077 (0.049*)	2.029 (0.731)	10.424 (0.237)	3.779 (0.876)	18.189 (0.001)
	Agree	122 (48%)					
	Moderate	54 (21.3%)					
	Disagree	3 (1.2%)					
	Strongly Disagree	2 (0.8%)					
2. I am aware of the educational program objectives (PEO's/PSO's) for my college	Strongly Agree	58 (22.8%)	17.815 (0.037*)	11.628 (0.009*)	16.417 (0.012*)	13.165 (0.040*)	5.643 (0.130)
	Agree	169 (66.5%)					
	Moderate	0 (0%)					
	Disagree	22 (8.7%)					
	Strongly Disagree	5 (2%)					
3. I am aware of the Program Outcomes (POs) of my college.	Strongly Agree	68 (26.8%)	11.441 (0.247)	3.128 (0.372)	4.565 (0.601)	5.408 (0.493)	17.262 (0.001*)
	Agree	140 (55.1%)					
	Moderate	38 (15%)					
	Disagree	8 (3.1%)					
	Strongly Disagree	0 (0%)					
4. I am familiar with the methodologies used to compute PEO's/PSO's.	Strongly Agree	46 (18.1%)	25.696 (0.012*)	5.571 (0.234)	8.162 (0.418)	11.823 (0.159)	10.655 (0.031)
	Agree	134 (52.8%)					
	Moderate	59 (23.2%)					
	Disagree	13 (5.1%)					
	Strongly Disagree	2 (0.8%)					
5. I am familiar with the methodologies used to calculate POs.	Strongly Agree	41 (16.1%)	12.858 (0.169)	1.894 (0.595)	6.349 (0.385)	2.957 (0.814)	5.563 (0.135)
	Agree	137 (53.9%)					
	Moderate	58 (22.8%)					
	Disagree	18 (7.1%)					
	Strongly Disagree	0 (0%)					
6. I am familiar with the procedures for measuring Course Outcomes (COs).	Strongly Agree	46 (18.1%)	15.104 (0.088)	0.860 (0.835)	5.272 (0.509)	2.257 (0.895)	12.181 (0.007*)
	Agree	134 (52.8%)					
	Moderate	63 (24.8%)					
	Disagree	11 (4.3%)					
	Strongly Disagree	0 (0%)					
7. I was aware of the learning domains (cognitive, emotive, and psychomotor) that were assigned to each CO.	Strongly Agree	45 (17.7%)	20.542 (0.015*)	0.727 (0.867)	6.686 (0.351)	7.167 (0.306)	8.701 (0.034*)
	Agree	142 (55.9%)					
	Moderate	52 (20.5%)					
	Disagree	15 (5.9%)					
	Strongly Disagree	0 (0%)					
8. I am familiar with Bloom's taxonomy.	Strongly Agree	43 (16.9%)	9.823 (0.365)	3.135 (0.371)	10.625 (0.101)	2.579 (0.860)	10.956 (0.012*)
	Agree	114 (44.9%)					
	Moderate	70 (27.6%)					
	Disagree	27 (10.6%)					
	Strongly Disagree	0 (0%)					
9. During Orientation Week, I was briefed on the OBE method.	Strongly Agree	41 (16.1%)	28.348 (0.005*)	1.160 (0.885)	7.546 (0.479)	11.961 (0.153)	13.602 (0.009*)
	Agree	125(49.2%)					
	Moderate	71 (28%)					
	Disagree	16 (6.3%)					
	Strongly Disagree	1 (0.4%)					

Table 3. Chi-square test results showing significant factors influencing OBE knowledge (continued)

Characteristics		Frequency (Percentage)	Age Vs χ^2(p-value)	Gender Vs χ^2(p-value)	Level of study Vs χ^2(p-value)	Branch Vs χ^2(p-value)	NPTEL Course Completion Vs χ^2(p-value)
10. The most of my course materials explained the COs of the courses initially in the semester.	Strongly Agree	52 (20.5%)	15.180 (0.232)	5.153 (0.272)	8.389 (0.396)	5.985 (0.649)	7.626 (0.106)
	Agree	134(52.8%)					
	Moderate	58 (22.8%)					
	Disagree	9 (3.5%)					
	Strongly Disagree	1 (0.4%)					
11. Everyone at my college has access to information about OBE.	Strongly Agree	54 (21.3%)	14.343 (0.111)	3.302 (0.347)	8.615 (0.196)	3.094 (0.797)	1.317 (0.725)
	Agree	119(46.9%)					
	Moderate	67 (26.4%)					
	Disagree	14 (5.5%)					
	Strongly Disagree	0 (0%)					
12. I feel that lecturer put more effort for teaching with OBE.	Strongly Agree	52 (20.5%)	11.825 (0.460)	2.000 (0.736)	9.420 (0.308)	6.309 (0.613)	12.725 (0.013)
	Agree	123(48.4%)					
	Moderate	60 (23.6%)					
	Disagree	15 (5.9%)					
	Strongly Disagree	4 (1.6%)					
13. I feel that lecturers work hard to ensure all students attain the learning outcome of their courses.	Strongly Agree	50 (19.7%)	12.151 (0.434)	7.364 (0.118)	12.443 (0.133)	19.772 (0.011*)	6.791 (0.147)
	Agree	134(52.8%)					
	Moderate	60 (23.6%)					
	Disagree	9 (3.5%)					
	Strongly Disagree	1 (0.4%)					
14. I feel that understanding the COs will help me to do better in my assessments.	Strongly Agree	58 (22.8%)	20.806 (0.053)	13.373 (0.010*)	8.433 (0.392)	12.434 (0.133)	10.449 (0.034*)
	Agree	136(53.5%)					
	Moderate	56 (22%)					
	Disagree	2 (0.8%)					
	Strongly Disagree	2 (0.8%)					
15. I am familiar of the evaluation standards used by my lecturers to grade students within classroom.	Strongly Agree	56 (22%)	17.004 (0.149)	7.027 (0.071)	8.163 (0.226)	16.841 (0.010)	7.518 (0.057)
	Agree	135(53.1%)					
	Moderate	57 (22.4%)					
	Disagree	6 (2.4%)					
	Strongly Disagree	0 (0%)					
16. I am completely aware of the association between assessments (quizzes, tests, assignments, practical tests, final exams, and so on) and the attainment of learning objectives.	Strongly Agree	62 (24.4%)	17.569 (0.041)	4.322 (0.364)	26.440 (0.001*)	13.465 (0.097)	6.524 (0.163)
	Agree	128(50.4%)					
	Moderate	57 (22.4%)					
	Disagree	6 (2.4%)					
	Strongly Disagree	1 (0.4%)					
17. The most of my lecturers encourage me to solve problems independently and to manage my independent learning.	Strongly Agree	61 (24%)	19.380 (0.080)	6.664 (0.155)	5.061 (0.751)	15.891 (0.044*)	4.933 (0.294)
	Agree	135 (53.1%)					
	Moderate	48 (18.9%)					
	Disagree	7 (2.8%)					
	Strongly Disagree	3 (1.2%)					
18. During class, I will get the opportunity to exhibit my communication abilities.	Strongly Agree	67 (26.4%)	15.940 (0.068)	3.868 (0.276)	1.475 (0.961)	1.516 (0.958)	1.363 (0.714)
	Agree	128 (50.4%)					
	Moderate	57 (22.4%)					
	Disagree	2 (0.8%)					
	Strongly Disagree	0 (0%)					

Table 3. Chi-square test results showing significant factors influencing OBE knowledge (continued)

Characteristics		Frequency (Percentage)	Age Vs χ^2 (p-value)	Gender Vs χ^2 (p-value)	Level of study Vs χ^2 (p-value)	Branch Vs χ^2 (p-value)	NPTEL Course Completion Vs χ^2 (p-value)
19. In class, students analyze, discuss, and reflect under our own opinion and perception.	Strongly Agree	53 (20.9%)	21.519 (0.011*)	6.948 (0.074)	37.361 (0.000*)	7.349 (0.290)	2.637 (0.421)
	Agree	132 (52%)					
	Moderate	66 (26%)					
	Disagree	3 (1.2%)					
	Strongly Disagree	0 (0%)					
20. My critical thinking abilities are improved by my professors' instructional techniques.	Strongly Agree	61 (24%)	16.735 (0.053*)	4.736 (0.192)	5.841 (0.441)	14.194 (0.028*)	0.706 (0.872)
	Agree	138 (54.3%)					
	Moderate	50 (19.7%)					
	Disagree	5 (2%)					
	Strongly Disagree	0 (0%)					
21. In class teacShing and learning activities, I am exposed to case studies or real-life industry experience.	Strongly Agree	62 (24.4%)	17.417 (0.043*)	2.979 (0.395)	9.710 (0.137)	10.419 (0.108)v	3.481 (0.323)
	Agree	120 (47.2%)					
	Moderate	64 (25.2%)					
	Disagree	8 (3.1%)					
	Strongly Disagree	0 (0%)					

significant differences based on age, gender, level of study, branch of study, and completion of the course from NPTEL. This shows that the perceptions of the students regarding accessing 'OBE information' have similar trends based on different characteristics of the students. Regarding 'lecturer's effort in teaching with OBE,' it was found that there is a significant association with NPTEL course completion ($p = 0.013$), indicating that students who completed NPTEL courses perceive more effort from the lecturer in teaching 'OBE.' However, there were no significant differences based on age, gender, level of study, and branch of study.

The perception of students on lecturers' hard work in ensuring that all students achieve learning outcomes varied significantly by branch of study ($p = 0.011$) but not by age, gender, level of study, and completion of NPTEL courses. This implies that different branches of study perceive lecturers' hard work differently. Understanding COs and its performance in assessments was significantly associated with age ($p = 0.053$), gender ($p = 0.010$), and completion of NPTEL courses ($p = 0.034$). This implies that age, gender, and completing additional courses could affect students' perception of understanding COs and its performance in assessments. Finally, understanding evaluation standards was significantly associated with branch of study ($p = 0.010$) but not with age, gender, level of study, and completion of NPTEL courses. This implies that students' understanding of evaluation standards is affected by the branch of study.

Chi square test were used to find the association between the variables. Form the results, it is observed that there is strong association between the variables. More specifically, the subject is closely linked with students' perceptions of encouragement for independent problem-solving, the development and acquisition of critical thinking skills, and the use of case studies or real-world industry experiences. Age significantly acts as a factor that affects classroom discussion and contemplation opportunities, as well as gains in critical thinking and real-world learning experiences. In contrast, variables such as gender and NPTEL course completion implies that little correlation with these perceptions, suggesting minimal influence on the student learning environment. As a whole, the data indicate that a student's age and field of study have a greater impact on their educational experience than their gender or involvement in supplementary courses, underlining the role of these demographic variables in shaping the learning environment.

Finally, the survey findings revealed that 65.86% of the students were knowledgeable about Outcome-Based Education (OBE). The chi-square test was used for additional analysis, which revealed that age and completion of NPTEL courses had a significant impact on students' knowledge of OBE. Older students showed a higher level of understanding about OBE. In addition, no significance was observed in the awareness of the students about gender and branches of academics. This indicated that these factors did not play a major role in the awareness of the students about OBE.

The findings of the present study indicate that a massive portion of students have a strong understanding of the basic principles of Outcome-Based Education (OBE), i.e., focus on learning outcomes, competency-based evaluation, and

student-centeredness. These findings are similar to those of other research works (e.g., Borsoto et al., [1]; Caguimbal et al., [8]), where it was highlighted that students' understanding of learning outcomes is greatly influenced by the transparent approach of Outcome-Based Education. However, what is new and different about the present research is its focus on exploring students' awareness of Outcome-Based Education based on other demographic characteristics, i.e., age and participation in supplementary courses (e.g., NPTEL). Although the majority of the existing literature (e.g., Harden [4]; Davis et al. [16]) has focused on the impact of OBE on the learning outcomes, this study is one of the first to consider the role of these demographic variables in the way that they impact the students' understanding of OBE. The data indicates that the older students and those involved in more courses have a better understanding of OBE, reinforcing the importance of these factors for the development of a more thorough understanding.

As far as the methods used in carrying out the OBE are concerned, the research reveals that these have had a positive effect on the learning processes and attitudes of students. Tan [6] and Rhaffor [15] has identified the benefits of these methods in ensuring greater student engagement. Some of the benefits of OBE, as identified in this research, include clearer learning objectives, greater alignment of curriculum and learning outcomes, and greater potential for self-directed learning, which are in conformity with the overall body of literature on the benefits of OBE. Some of the challenges affecting the implementation of OBE, as identified in this research, include unclear learning objectives, poor assessment methods, and faculty members. This is similar to the challenges identified in the studies done by Cook et al. [10] and Rhaffor et al. [15], where these challenges were also identified as impediments to the adoption of OBE. The challenges identified indicate the need to ensure that there are effective strategies to overcome these impediments for the effectiveness of OBE. Resistance from the faculty members is also a major challenge, which indicates the need to ensure that the faculty members are adequately prepared.

In response to these challenges, the students made various suggestions regarding the implementation of OBE. Among the suggestions made were better communication of the set learning goals, provision of more resources for the development of the faculty members, and more opportunities for the students to participate in the decision-making processes concerning the implementation of OBE. The suggestions made by the students were in compliance with the findings made by Shaheen [13] and Guico et al. [17] which highlighted the need for a supportive and participating environment for the successful implementation of OBE. The study also found that the students who undertook supplementary courses such as NPTEL had a better understanding of the principles of OBE. This indicates that the integration of extra-curricular learning activities can be an essential strategy in promoting students' awareness and understanding of OBE. From the findings, it also supports the suggestions made by Vennila et al. [14], which indicate that the integration of additional learning activities can play a significant role in promoting students' understanding of education frameworks like OBE.

CONCLUSION

This study aims to assess graduating students' awareness of Outcome-Based Education (OBE), and it was found that students have a good understanding of its principles, such as developing specific skills and competencies. Furthermore, it was found that students' age and participation in additional courses such as NPTEL have significant effects on students' awareness and attitudes towards Outcome-based education. Most students showed awareness and agreed on the significance of Outcome-based education, and it was found that students' educational background could affect students' awareness of Outcome-based education. This study adds new knowledge regarding students' awareness of Outcome-based education and its relationship with students' age and participation in additional courses, as it was found that these factors were not considered in previous literature. Future studies could be conducted to find how these factors can be utilized to better Outcome-based education practices, and the educational institutions could utilize the information to design effective strategies to improve student knowledge and engagement with Outcome-based education, leading to its successful integration into the educational system.

The simulations are comprehensive and offer important insights, but the publication does not cover the study's flaws or potential future research areas. If the publication addressed the possible drawbacks in the current study, such as limitations in simulation parameters, scope, or applicability, and suggested specific areas for future research, the authors think it might give a more thorough data analysis. By helping future research concentrate on unresolved issues and advance the field, this method will also increase the reliability of the study.

REFERENCES

[1] Borsoto LD, Lescano JD, Santorce MJN, Simbulan AF, Pagcaliwagan AM. Status of implementation and usefulness of outcomes-based education in the engineering department of an Asian university. Int J Multidiscip Acad Res 2014;2:14–25.

[2] Hansen JM. Outcome-based education: a smarter way to assess student learning. Clear House. 1989;63:172–1744. [CrossRef]

[3] Towers JM. Outcome-based education: another educational bandwagon. Educ Forum 1992;56:291–305. [CrossRef]

[4] Harden RM. Outcome-based education: the future is today. Med Teach 2007;29:625–629. [CrossRef]

[5] Mukhopadhyay S, Smith S. Outcome-based education: principles and practice. J Obstet Gynaecol 2010;30:790–794. [CrossRef]
[6] Tan HE, Oriah A, Senian M. Implementation of outcome-based education incorporating technology innovation. Proced Soc Behav Sci 2012;62:649–655. [CrossRef]
[7] De Asis GM. Yes, to OBE. Phoenix 2011;14:1–5.
[8] Caguimbal DA, Delacion DC, Medina A, Mendoza ML, Sanchez MJR. Level of awareness of maritime students on outcomes-based education. Educ Res Int 2013;2:7–12.
[9] Deneen CC, Brown GTL, Bond TG, Shroff RH. Understanding outcome-based education changes in teacher education: evaluation of a new instrument with preliminary findings. Asia Pac J Teach Educ 2013;41:441–456. [CrossRef]
[10] Cook DA, Hamstra SJ, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Comparative effectiveness of instructional design features in simulation-based education: systematic review and meta-analysis. Med Teach 2013;35:e867–e898. [CrossRef]
[11] Laguador JM, Pesigan M. Academic performance and measure of character and personality of engineering students with and without referral from counselling center. Asian Acad Res J Soc Sci Humanit 2013;1:281–293.
[12] Laguador JM, Dotong CI. Knowledge versus practice on outcomes-based education implementation of engineering faculty members in LPU. Int J Acad Res Prog Educ Dev 2014;3:63-74. [CrossRef]
[13] Shaheen S. Theoretical perspectives and current challenges of OBE framework. Net J Eng Educ 2019;1:122–129. [CrossRef]
[14] Vennila J, Basker P, Thenmozhi K, Rajeswari K, Nithyakala P. Analyzing the impact of inflammatory bowel disease (IBD) using R programming. JP J Biostat 2022;19:123–144. [CrossRef]
[15] Rhaffor K. Students' perception on outcome-based education (OBE) implementation: a preliminary study in UniKL MSI. In: Proceedings of the Colloquium on Manufacturing Engineering Technology; 2017; Lumut, Perak.
[16] Davis MH, Amin Z, Grande JP, O'Neill AE, Pawlina W, Viggiano TR, et al. Case studies in outcome-based education. Med Teach 2007;29:717–722. [CrossRef]
[17] Guico TM, Dolor G. Level of awareness and possible concerns of marine faculty members on outcomes-based education. J Int Acad Res Multidiscip 2013;1:159–167.

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1261–1282, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.1956

Research Article

Leveraging unexplored regional dynamics and temporal interdependencies in crop yield prediction: A graph theory approach

Mamta KUMARI[1,*], Suman[1], Devendra PRASAD[2]

[1]*Deenbandhu Chhotu Ram University of Science and Technology, Sonepat, 131039, India*
[2]*PIET Samalkha, 132102, India*

ARTICLE INFO

Article history
Received: 29 October 2024
Revised: 23 December 2024
Accepted: 06 March 2025

Keywords:
Crop Yield Prediction; Feature Engineering; Graph Convolutional Network (GCN); LSTM; MixHop; Spatial-Temporal Modelling

ABSTRACT

Agriculture is the pillar of India's economy and food security, yet accurate crop yield prediction remains a persistent challenge due to the complex interplay of environmental, agronomic, and socio-economic factors. In practice, districts often share crop-related characteristics with more than just their immediate neighbours and exhibit irregular temporal patterns in variables such as rainfall. The models that have been implemented frequently do not capture the higher-order spatial relationships between crop-producing areas and also do not take into account the non-uniform timing of the significant agricultural properties, which results in the decreased predictive capability. This paper describes a spatial graph hop with temporal enhancement (SGHTE) framework that utilizes multi-region spatial relationships through a longer MixHop GCN and irregular time dynamics through a time-conscious LSTM. SGHTE is able to capture cross-district interactions, long-range spatial-temporal correlations, and therefore is far more effective at predicting yields in the 32 districts of Rajasthan. The strategy is also enhanced by the use of a fully contained 15-attributes, among which are the underexplored aspects of saline and sodic soil composition, the multiple irrigation sources, the use of hybrid seeds, and the use of fertilizers, which adds depth to the feature space of making the predictions. With a dataset of 32 districts in Rajasthan with 13 years (20072019), SGHTE obtained a 0.1306 RMSE, 0.6775 R 2, and 0.8912 Pearson correlation coefficient to predict the yield of Bajra (Pearl Millet). These findings indicate apparent advances to the state of art approaches, showing that SGHTE has the ability to model intricate spatial-temporal interactions and produce high quality, practical predictions to aid decision-making in the policies of policymakers and farmers alike.

Cite this article as: Kumari M, Suman, Prasad D. Leveraging unexplored regional dynamics and temporal interdependencies in crop yield prediction: A graph theory approach. Sigma J Eng Nat Sci 2026;44(2):1261–1282.

***Corresponding author.**
*E-mail address: mamta1.tarar@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Globally, approximately 1.3 billion people are engaged in agricultural production, accounting for half of the world's labour force [1]. However, industrialized countries harbour only 9 percent of these agricultural workers and early 60 percent are in the developing countries [2,3]. The majority of agricultural workers are located in Asia with China having more than 40 per cent of all agricultural people in the world and India having more than 20 per cent [4]. Agriculture continues to be one of the main sources of the livelihood of millions. According to the world development Indicators database, India's population surpassed China's in 2023, reaching 1.417 billion compared to China's 1.412 billion [5]. This population growth raises concerns about future agricultural productivity due to global warming, which causes unpredictable and extreme climate conditions [6]. As indicated by Figure 1, India's population is expected to grow significantly in the coming years [7]. In an effort to address the socioeconomic problem of farmers and satisfy the needs of the fast-growing population, it is important to develop an efficient crop yield prediction model [8]. Nonlinear relationships among different attributes of agriculture reveal more accurate yield prediction with the use of the artificial intelligence (AI) systems to predict the yield [9]. Such systems may be excellent in the provision of proactive measures by policy makers to prevent feminine and mitigate crop failures to farmers [9]. This work is concerned with the forecast of the yield of Bajra (Pearl Millet), a food crop in the arid areas but particularly in Rajasthan, India. The factor that endears Bajra is its ability to withstand extreme climatic conditions in the northwestern and even central India [10]. It grows in semi-arid regions where it is spread over about 7.4 million hectares in India and yields about 9.2 million tonnes yearly [11]. Historically, the prediction of the crop harvest has needed complicated statistical algorithms and models. Nevertheless, the emergence of big data has led to the use of state-of-the-art techniques, such as machine learning and deep learning [12,13]. Such methods as convolutional neural networks (CNN) [14], graph convolutional networks (GCN) [15], recurrent neural networks (RNN) [14], Long short-term memory models [16] have become popular due to their prediction features. The literature review section presents a detailed analysis of past studies that used these techniques. Figure 1 shows the expected population changes of India and China, showing that in 2023, India will overtake China and have 1.4 billion inhabitants. This population change has direct effects on food security and agricultural planning and this has made it necessary to have the right yield prediction models like the one in the present study, hence the urgency to have the right models.

Literature Review

A detailed review of the attributes used for crop yield prediction in Rajasthan is presented, in Figure 2. Relevant Literature on CYP for Rajasthan was sourced from Dimension.ai [17], focusing on publications related to computer science and agricultural production. The identified attributes were categorized into six key groups: environmental, soil, crop, fertilizer, irrigation and seed attributes. According to the reviewed literature [12, 15, 18, 19] and Figure 2, environmental factors such as rainfall, minimum and maximum temperatures, sunlight duration, humidity, wind speed, and production area (in hectares) are the most

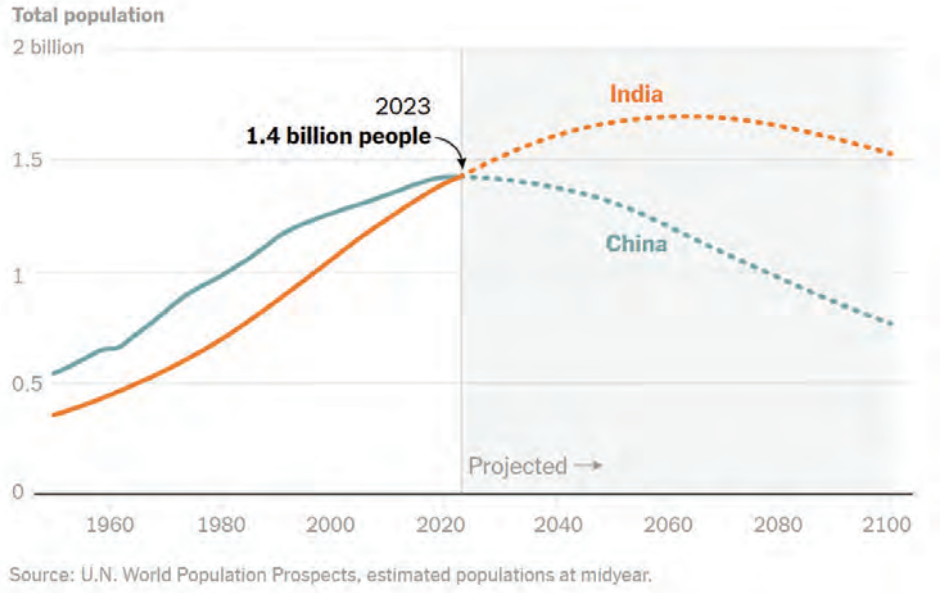


Figure 1. Projected population trends of India and China, highlighting India's expected overtaking of China and its implications for future agricultural demand.

frequently analysed attributes. These factors consistently appear in studies due to their significant impact on crop yield prediction. As Figure 2 also indicates, the agronomically critical characteristics like saline soil content, diversity of irrigation sources, type of hybrid seed and fertilizer are grossly underrepresented in earlier studies underpinning a significant research gap that our study directly fills.

In the research conducted in [12], the authors used the regression technique of Random Forest (RF) to model crop yields in the agroclimatic region of Rajasthan, where six major crops were used. Agrometeorological records of the past were used, such as the records of 1991 to 2020 in terms of rainfall, temperature, humidity, and crop yield. The RF algorithm obtained an accuracy of 92.3%. Likewise, the researchers in [20] used RF and decision trees and gradient-boosting methods to estimate yields of 10 districts and seven crops in Rajasthan. The variables included in the research were acreage, output and rainfall of 1987 to 2018 examined the variables of acreage, output and rainfall of 1987 to 2018. Further, [13] experimented with different machine learning techniques, such as K-Nearest Neighbours, Support Vector Machine, XGBoost regression, and lightGBM that combines the processed data with Smart Farm Ontology to predict yields better. Although they are successful, these machine learning models have a high level of hyperparameter tuning to reach the best performance. They also tend to overfit particularly where the dataset is small or there is a lot of complexity in the model. Moreover, they are vulnerable to noises because of their high-computational intensity, which may compromise the accuracy of predictions. In order to overcome the shortcomings of the classical machine learning methods, deep learning methods, including artificial Neural Network (ANN) were proposed in [21]. In [21] stepwise multiple linear regression (SMLR) was coupled with ANN, support vectors machines (SVM) and RF, and principal component analysis (PCA) coupled with these strategies were used in predicting mustard yield in Rajasthan. On the same note, in the study conducted in [22], ANNs were used together with machine learning algorithms such as RF and logistic regression to predict yield. Although ANNs have enhanced predictive ability, they have high demands on labelled data and extensive computing resources to be trained. This makes them less efficient and more resource-intensive compared to some machine learning techniques, particularly when working with limited datasets or constrained computing environments.

The study in [14] proposed a hybrid CNN-RNN model for predicting crop yields, combining CNNs to capture spatial dependencies in weather and soil data and RNNs to model temporal dependencies in crop yields. Unlike ANNs, CNNs excel at analysing spatial data, effectively detecting critical patterns such as variations in soil quality and weather conditions across different regions. The RNN component addresses temporal relationships by processing sequential data over time. The RNNs however have difficulties with the long term dependencies because of the vanishing and exploding gradient problems. Gradual gradient lead to slow learning of network weights and explosive gradient lead to very large changes in network weights, making the training process unstable. These issues may restrict the capability of the model to learn long time series data.

A CNN-GRU model is proposed in [23] to improve wheat yield estimates using three remotely sensed variables: vegetation temperature condition index (VTCI), leaf area index (LAI), and fraction of photosynthetically active radiation (FPAR). GRUs address the vanishing gradient problem that RNNs lack and facilitate them in capturing long-term dependencies more effectively. GRUs however, do not have a specific state of internal memory cells, which have the ability to store information across longer sequences. This hinders their ability to memorize information over prolonged periods of information. Conventional machine learning models have been used in the prediction of agricultural yield using tabular agronomic and environmental

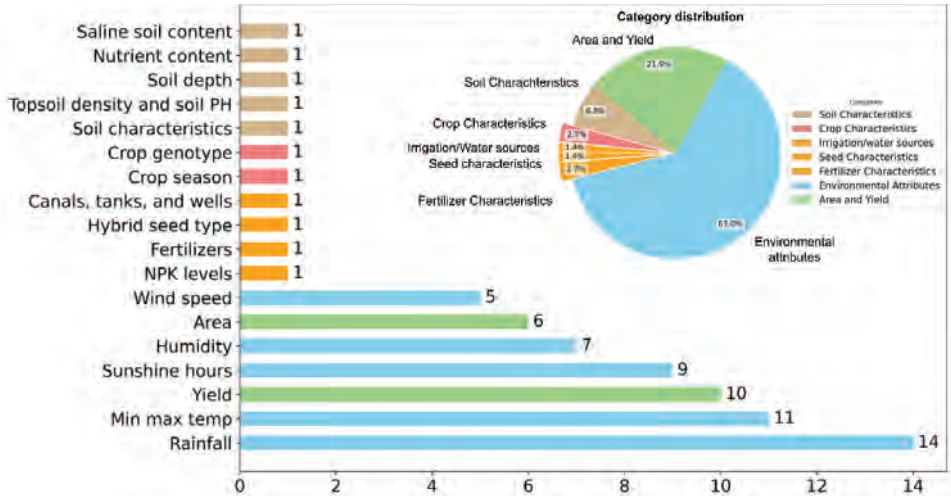


Figure 2. Frequency distribution of attributes for crop yield prediction in Rajasthan, showing environmental factors as most dominant.

data like Adaptive Boosted XGBoost (ABP-XGBoost) [24] and random forest [25]. ABP-XGBoost builds on existing boosting algorithms by dynamically attending to the hard-to-predict samples, whereas the Random Forest combines several decision trees in order to minimize overfitting and increase generalization. These models are also easy to compute, interpret, but do not always have a high-order spatial dependency and temporal anomaly, which can be used as a useful benchmark to measure the performance of more sophisticated spatio-temporal models such as SGHTE.

This study [16] presents a novel deep CNN-LSTM model that estimates soybean yield at the county level in the U.S. In line with previous research [14], the proposed approach employs CNNs for spatial feature extraction. However, for temporal feature learning, LSTM is used. LSTM addresses the constraints of RNNs by integrating memory cells capable of capturing long-term relationships within sequence data. LSTM models are designed to handle sequential data by systematically storing and updating a memory of previous inputs, hence acquiring temporal patterns via their recurrent architecture. However, vanilla LSTMs assume the time interval between events to be regular and uniform, which can lead to suboptimal performance in tasks where time gaps vary significantly [26]. Also, CNNs cannot effectively handle irregular, graph-structured data, and are not effective in capturing both local and global dependencies.

A SCM-GAT (Structural Causal Model Graph Attention Network) model is proposed in [27] that incorporates causal relationships between variables for better interpretability and robustness. GAT utilizes an attention-based mechanism to calculate the connections between nodes in a graph. These attention mechanisms assess the significance of surrounding nodes, enabling the model to concentrate on the most important relations of the nodes [28]. However, GATs can overfit the training data if not regularized, particularly in cases with limited labelled data.

The paper [15] introduces a novel model for accurate crop yield prediction by incorporating GCN to model dynamic spatial topology structures and a knowledge-guided Temporal Multi-head Attention Algorithm (KTMA) for temporal feature extraction. However, traditional GCN has a limited receptive area because, after each convolution layer, it only gathers information from its first-order neighbours.

Challenges to be addressed

Machine learning procedures [12,13] [20-22] and GAT [27] generally have problems with overfitting. CNNs [14,16] have challenges in learning spatial features, such as multifaceted interdependencies in the data. RNNs can experience the vanishing and exploding gradient issues during the processing of the temporal data. GRUs [20] have difficulties with long-term temporal dependencies, and LSTMs [16] fail to deal with irregular time series. Moreover, GCNs [15], GNN-RNN [29] only learn shallow spatial indicators, so they are likely to make inaccurate predictions of the yield, particularly when neighbouring regions follow the same trend of production. The main issues highlighted in the available literature are:

- Some important yield characteristics like saline and sodic soil [13], irrigation sources which are wells, canals and tanks [23], use of hybrid seeds, and use of fertilizers [30] are not discussed anywhere in the literature. These aspects have great impact on both accuracy and strength of prediction models hence poor yield prediction results.
- Environmental factors like soil quality, climatic conditions and irrigation methods bring together crop-producing regions in space [31]. This leads to the fact that such areas tend to exhibit spatial continuity, which yields similar yields [32]. However, most crop yield prediction models fail to capture these spatial connections, resulting in reduced prediction accuracy.
- Many models [14,20,32] assumes that CYP data attributes are uniformly distributed over time. However, several key attributes, such as rainfall and solar irradiation, exhibit temporal irregularities. The irregularity has largely been overlooked in previous search.

The gaps found in the literature, including failure to consider key crop prediction factors, including soil salinity, irrigation availability, and farm fertilizer application, inadequate extraction of spatial relations and time distortions have led to poor models of crop yield prediction. These issues highlight the necessity of more thorough methods. In an attempt to fill these gaps, this paper puts forward a new model that will utilize 15 attributes, including the infrequently explored ones, into the yield prediction model.

Table 2 reflects the gaps in the literature regarding the features of yield, spatial modelling, and the handling of temporal irregularity against the innovations suggested in the proposed SGHTE framework. It is a combination of untested agronomic characteristics, Multi-hop spatial Aggregation, Mixhop GCN, and irregular temporal dependency modelling with t-LSTM to give improved predictive power and robustness at forecasting of Bajra yield in Rajasthan.

Motivation and Contribution of the study

As illustrated in Figure 3, first order neighbors are districts that share a boundary with reference districts. For example, districts such as Sikar, Alwar, Dausa, Tonk, Ajmer and Nagaur are firs- order neighbors but not directly connected to the reference district. In the case of Jaipur, the second-order neighbors are as Jhunjhunu, Churu, Bharatpur, Sawai Madhopur, Bundi and Bhilwara. This rank of neighbors controls the circulation of information among the nodes (in this case, districts), according to their vicinity, and is of paramount importance in the extraction of spatial features. These spatial features that exist between neighboring nodes are extracted well by GCNs. This hierarchical data is important to spatial modelling because most

Table 1. Summary of Research Gaps and Proposed SGHTE Approach for Crop Yield Prediction

Aspect	What is known & research gaps	Proposed approach
Yield-related feature predictors	Previous studies on crop yield prediction in Rajasthan rely heavily on environmental factors—rainfall, min/max temperature, sunlight, humidity, wind speed, and cultivation area [12, 15, 18, 19]. Although these factors are important, there are other agronomic factors (e.g., soil salinity/sodicity, irrigation sources, use of hybrid seeds, use of fertilizers, etc.) that are not well represented, but have a major impact on yield [13, 23, 30]	To enhance predictive power, this research uses a more extensive list of 15 variables of the 32 districts of Rajasthan (e.g. underexplored yet powerful attributes like saline/sodic soil composition [13], accessibility of irrigation through wells, canals, and tanks [23], usage of hybrid seeds, and the application of fertilizers [30]) to enhance the predictive performance.
Spatial modelling (CNN, GCN, GAT, etc.)	CNN-based models [14,16] capture spatial patterns but assume independence between regions, overlooking spatial continuity—where neighboring regions often produce similar yields [32]. Spatial dependencies can be modeled by graph-based methods like GCN [15] and GAT [27]. GNN-RNN [32] captures the apatial connections between the districts but traditional algorithmsonly aggregate first-order neighbor information per layer, limiting their receptive field.	The SGHTE framework builds Mixhop GCN further to combine information about multiple hop neighbours so that wealthier spatial embeddings can be created, which consider local and global relationships between districts in Rajasthan.
Temporal irregularity	Other models such as CNN-LSTM [16], RNN [14]and BiLSTM, GNN-RNN [31] tend to assume time intervals which are uniform yet the data of rainfall and sunshine in the real world show uneven temporal representation. Such a discrepancy is a source of inefficient temporal modeling [32].	t-LSTM is another model in SGHTE that takes into account elapsed time between observations, modifying the memory cell states according to time gaps, and therefore, it exploits the irregular temporal dependencies better.

agricultural properties, including rainfall, soil, and irrigation tend to be spatially continuous beyond direct boundaries. The proposed Mix-hop GCN takes advantage of this Multi-hop connectivity to take into account both the local and the long-range spatial dependency, thus enhancing the accuracy of yield prediction.

As shown in Figure 4, traditional GCNs are limited by their small receptive field, which only allows them to gather information from first-order neighbours through each convolution layer. In this context of yield prediction, it is important to recognize that districts are often connected to others beyond their immediate neighbours [32, 33]. To address this, a Mix hop GCN has been proposed, which enables the model to capture spatial relationships from neighbours that are several hops away. By extracting features from k-different nodes at each convolution layer, as demonstrated in Figure 4[1], this approach improves spatial feature extraction and better captures long range dependencies among districts.

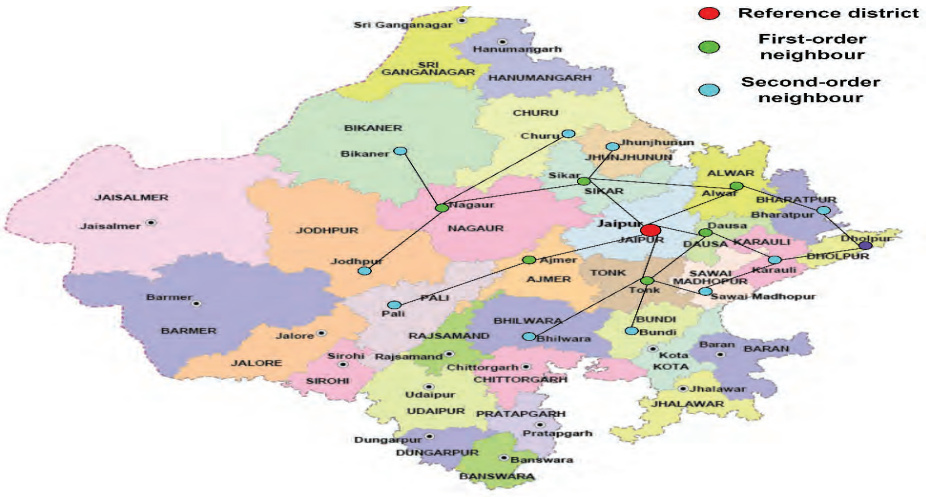


Figure 3. Spatial map of Rajasthan showing reference, first-order, and second-order neighboring districts for multi-hop dependency capture.

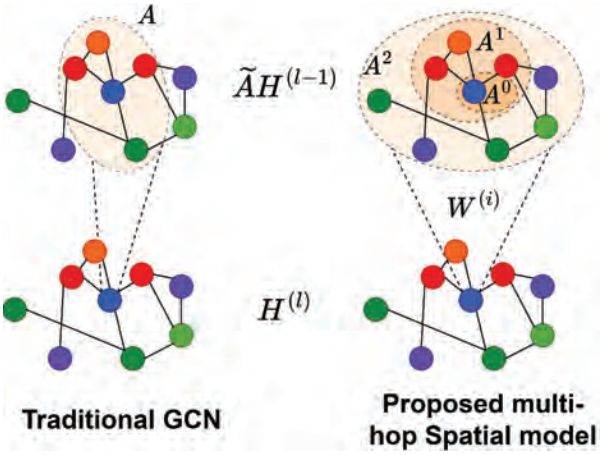


Figure 4. Comparison of feature extraction in traditional GCNs and the proposed Mix-hop GCN, showing how the latter captures higher-order spatial dependencies beyond immediate neighbours for improved yield prediction.

Additionally, it is necessary to note that the crop yield also depends on the spatial and temporal factors, such as precipitation, hours of the sunshine, and annual climatic variations. Figure 5 depicts a rainfall chart of the Ajmer district between the years 2010 and 2020 to emphasize the inconsistency of the data that is provided in the form of the rainfall. Vanilla LSTM Model presupposes that there can be regular time intervals between events [20]. Nonetheless, as it is shown in Figure 5, irregularity in the variability of the rainfall data, a rainfall data, is observed in 2012 and 2014, where uniformity is not observed. A time aware LSTM (t -LSTM) model is developed to overcome this problem of time irregularity in time-related data to extract more temporal features. T-LSTM model is used to provide better time-related feature extraction. T-LSTM integrates a memory, which will adapt the previous cell memory state in accordance with the irregular time intervals between events thus offering a better representation of the time data patterns [33, 34].

To summarise, the main contributions of this paper are:

- The present study presents a new regional data, which will be assembled based on different official sources that include attributes that have not been studied before but are also influential in predicting the yield of Bajra in Rajasthan, including saline and sodic soil composition [13], irrigation availability in the form of wells, canals, and tanks [23], hybrid seeding, and fertilizer application [30]. The fact that these other factors have been incorporated greatly increases the accuracy and strength of yield predictions.
- One of the main contributions of the study is the investigation of multihop neighbouring district relationships based on the graph theory where regional dynamics are used to examine the spatial relationships. To adequately capture the intricate interdependence of geographically related areas, the Mix-hop GCN method is used to make yield predictions more accurate at different locations.
- The given prediction framework can be used to solve time irregularities on the time series with the help of t-LSTM. This is a spatial depth and temporal variation methodology, which is known as Spatial Graph Hop with Temporal Enhancement (SGHTE). Taking into consideration unevenly distributed time intervals, the model is an apt measure of real world time dynamics, which results in higher performance of yield prediction and beats the state-of -the-art models.

Uniqueness of work: Compared to previous models [13,23,30], the paper introduces a regional dataset for Bajra with 15 attributes, including previously underused agronomic factors (saline/sodic soil, irrigation source mix, hybrid seeds, fertiliser usage). The paper focuses on generating multi-neighbour information of districts which contribute to yield prediction, compared to previous models like [15,27,32] which only considered immediate neighbour information of districts. The paper consider d the concept of temporal irregularity which is understudied (motivated

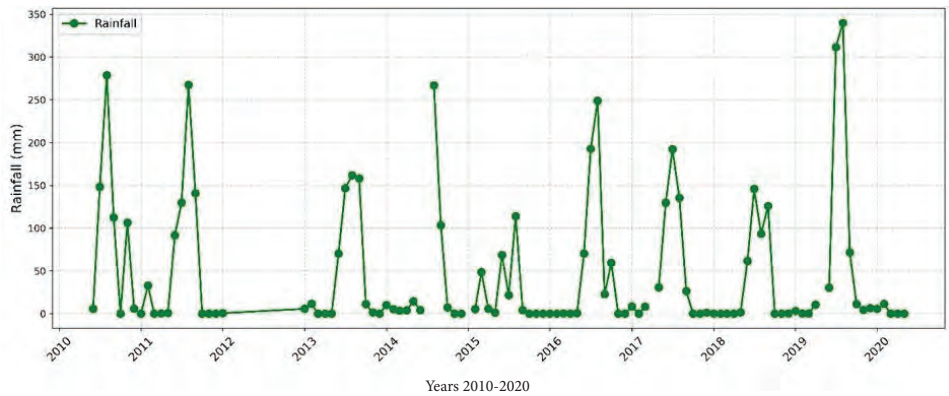


Figure 5. Rainfall variability in Ajmer district (2010–2020), demonstrating irregular temporal patterns that motivate the use of time-aware LSTM for handling non-uniform event intervals in climatic data.

by observed rainfall variability and limitations of vanilla LSTM/RNN) [14,16,21,35].

The main goal of the current research is the development and testing of a new Spatial Graph Hop with Temporal Enhancement (SGHTE) model that incorporates MixHop Graph Convolutional Networks (MixHop GCN) and Time-Aware Long Short-Term Memory (t-LSTM) to increase the prediction of yields in the Bajra crop in the district level in Rajasthan, India. The SGHTE model with its high-order spatial dependencies between districts and its capacity to take into account temporal anomalies in historical yield information should deliver a high level of predictive accuracy, in comparison with contemporary state-of-the-art solutions. The effectiveness of the model is proved by the use of the comprehensive experiments and comparative assessment involving the real-world agricultural data.

The research is an important contribution to the field of agriculture with a solution to a major problem in agriculture, which is the prediction of crop yield, and hence it will fill a gap between the engineering and natural sciences. SGHTE takes advantage of the irregular temporal patterns of the environment, agronomic, and socio-economic factors by integrating Mix-hop GCN to model the spatial dependency of multi-regions and time-sensitive LSTM to model the irregular temporal patterns. Although applied to Bajra yield prediction in Rajasthan, the approach can be applied to other crops, areas, and areas with key spatial-temporal interaction, including climate impact analysis and resource management [35, 36]. The enhanced predictive accuracy has a beneficial practical use by policy makers, farmers and planners in data-driven decision making to achieve food security and sustainable agriculture.

The remaining part of the paper is organized in the following way: In section 2, the proposed methodology is described, including the formulation of the problem, the model architecture, yearly process of the construction of a relational graph of district attributes, and the SGHTE model and the SGHTE model pseudocode. Section 3 explains the design and setup of the experiment, such as data set, preprocessing methods, training and testing procedures, and evaluation measures. Section 4 provides the outcomes of the SGHTE model, and the analysis of the results is done in comparison with the state-of-the-art approaches. Lastly, Section 5 will be the conclusion of the study which summarizes the important findings.

1. Proposed Methodology

This section describes the methodologies that the proposed study will use. The problem formulation section determines and identifies the major issues in crop yield prediction (CYP) and discusses the possible solutions. Section 2.2 presents SGHTE model, which is a combination of Mixhop GCN and t-LSTM to extract spatial feature and learn temporal features respectively. The following section describes how annual district attribute relational graphs are constructed, which is the input of Mixhop GCN model. Lastly, the operational structure of the SGHTE Model is studied further with a focus on how it addresses the issue of spatial temporal dependencies.

1.1. Problem Formulation

The two main unaddressed issues of CYP are the extraction of deeper and more profound spatial relations among the crop-producing regions and the incorporation of temporally uneven attributes. As seen from the literature, existing studies only utilize information from the first-order neighbouring nodes (nodes being districts, cities, counties, etc.) to construct the graphical embeddings ($H^{(l)}$) for yield prediction. Also, the uneven distribution of attributes over a time period has also been ignored in yield prediction. The proposed SGHTE model addresses both of these shortcomings.

Regions with similar latitudes and longitudes have similar climatic characteristics. These patterns show a substantial link between Bajra yields across districts. Indeed, it is seen that a district producing a substantial harvest of Bajra in a given year results in increased yields in the adjacent districts too. Therefore, to consider this spatial relationship, the connections based on regional dynamics by graph theory are incorporated. The data used and the collection procedure for the construction of this graph is explained in detail in section 4. The yearly district attribute relational graphs G_T, which are categorized by districts and provide comprehensive data on many attributes of crops like geographical data (x_{geo}): area, production, districts, year, meteorological data $x_{meteorlogical}$: (Annual rainfall, climate type), $x_{soil\ characteristics}$: (phosphorus, potassium, saline soil, alkaline soil), $x_{soiltype}$= (soil type 1 and soil type 2) $x_{yield\ boosters}$: hybrid seeds, and fertilizer use in Tonnes) and water resources $x_{waterres}$: (wells, canals and tanks) for each particular district on an annual basis. Thus, the set of crop features is represented as $x = \{x_{geo}, x_{meteorlogical}, x_{soil\ characteristics}, x_{soiltype}, x_{yield\ boosters}, x_{waterres}\}$. Thus, the spatial dependencies among crop-producing regions are captured as $f(x, A_t)$. The adjacency matrix A_t captures these attribute relations among x for each year with the help of cosine similarity.

The yearly district attribute relational Graphs are denoted as, $G_{t\in\{2007:2019\}}$ and the information about these graphs is stored as $A_{t\in\{2007:2019\}}$. Thus, to generate the year-wise spatial relations (E_t), the layer-wise (l) propagation is computed as,

$$f(x, A_t) = E_t^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}_t\widetilde{D}^{-\frac{1}{2}}E_t^{(l)}W^{(l)}\right) \quad (1)$$

Where, $\widetilde{D} = \sum \tilde{A}_t$ and $W^{(l)}$ is the layer-specific trainable weight matrix, $\sigma(.)$ is the activation function and $E_t^{(l)} \in \mathbb{R}^{N\times e}$ (e is the size of embeddings (year-wise spatial relations)) is denoted as the embeddings in l^{th} layer; $E_t^{(0)} = x$ However, the adjacency matrix is used in the term $\tilde{A}_t E_t^{(l-1)}$ in the vanilla GCN [36, 37], aggregates the features only from immediate neighbours as shown in Figure

3. Given that the convolution operation is carried out only once on the adjacency matrix , the GCN intrinsically prioritizes retrieving information from immediate neighbouring districts in each layer.

In this paper, this is addressed by aggregating information from both first-order, second-order neighbor districts and potentially higher-order spatial relations based on power P [1] as explained in Figure 3. This is enabled by using various powers of the adjacency matrix A in each layer. This facilitates the incorporation of information from k-hop neighbouring districts in a single layer. Thus, the equation (1) is modified as,

$$E_t^{(l+1)} = \|_{j \in P} \ \sigma\left(\tilde{A}^j E_t^{(l)} W_j^l\right) \tag{2}$$

where the hyper-parameter P is a set of integer adjacency powers, and $\tilde{A} = \{\tilde{A}_1, \tilde{A}_2, \cdots \tilde{A}_t\}$ denotes the adjacency matrices with self-connections, as shown in Figure 4 and $\|$ is the column-wise concatenation operator. Thus, the $H^{(l)}$ embedding is the concatenation of all the output year-wise embeddings $E_t^{(l)}$ and is given as

$$H^{(l)} = \odot_{t=1}^{t} E_t \tag{3}$$

To tackle the temporal irregularities as explained in Figure 5, the improved t-LSTM is employed. The input to the t-LSTM network consists of spatially aware node embeddings $H^{(l)}$ obtained from the output of mixhop GCN. Now the hidden state at time t is given by,

$$h^t = tLSTM(H^{(l)}, \Delta t) \tag{4}$$

Here Δt is the elapsed time between two irregular observations. Thus, the final predicted yield $\hat{y}_{d,t}$ is obtained by classification based on irregularity aware hidden state equation (4) and can be given as,

$$\hat{y}_{d,t} = FCN(h^t) \tag{5}$$

Proposed Model

The proposed SGHTE (spatial graph hop with temporal enhancements) model as seen in Figure 6 consists of three primary components: a thorough process of data collection and preprocessing, the Mixhop GCN for generating deep embeddings of the attribute relation of all the 32 districts, and t-LSTM to capture the temporal dependencies of the attributes and predict the yield. Data is parsed from various sources [37,38] The data exists in several modalities, complicating the extraction and collection process. Various pre-processing techniques are applied with the help of various Python libraries like Pandas and Regex to get the final data. This final data set contains 15 attributes regarding Rajasthan's 32 districts. This dataset comprises both temporal and spatial data of the districts. The pre-processed data is used by the SGHTE model. To construct the graphs for the spatial relations, the adjacency matrix is used, which is calculated by the cosine similarity between the edges of the nodes, as seen in step 2 of algorithm 1. The yearly district attribute relational Graphs (G_T) which contains the spatial information and relations are then

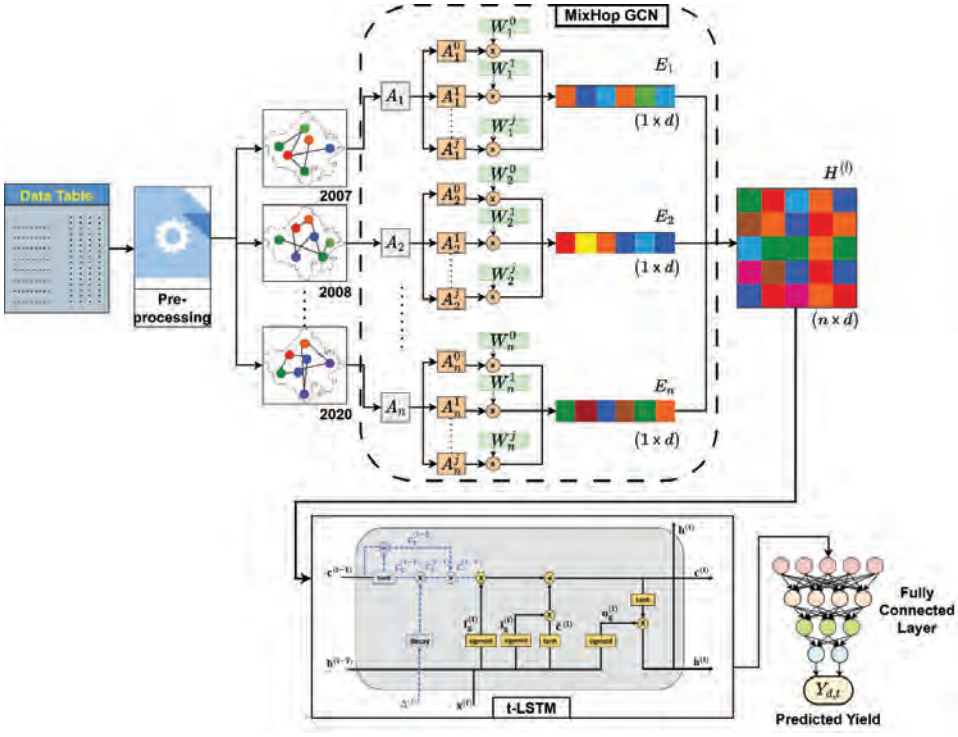


Figure 6. Process flow of the proposed SGHTE model, integrating Mix-hop GCN for spatial feature extraction with time-aware LSTM for modelling irregular temporal dependencies.

constructed using the adjacency matrices A_T (step 3). This G_T is propagated through various layers of Mixhop GCN. Mixhop GCN's convolutional layers gather information not just from neighboring nodes but also from the k distinct nodes around them (step 4). After the generation of these spatial embeddings for each year, these are aggregated in a final embedding $H^{(l)}$ (step 5). These embedding $H^{(l)}$ serve as the input for the model t-LSTM (step 6). The t-LSTM model now processes these embedding $H^{(l)}$ and addresses attributes like rainfall that are not linear over a given time. The SGHTE model is discussed in later sections in detail. The predicted yield is then enhanced by refining the model parameters using evaluation metrics (step 8 and 9).

Construction of Yearly District Attribute Relational Graphs (G_T)

The mixhop GCN model operates on graphs $G_T = (V, \mathcal{E})$ where V is the set of nodes and $\mathcal{E}$ is the set of edges of these nodes. Here Rajasthan's 32 districts serve the role of nodes, while edges represent the connections between them. The district nodes are $V = \{v_1, v_2, v_3 \cdots v_N\}$, whereas the connection among the pair-wise districts (v_i, v_k) is represented by the edge $\varepsilon_{ik} \in \mathcal{E}$. These connections are based on geographic proximity or environmental similarity. The G_T graphs are constructed with the help of the Adjacency matrices A_T. The adjacency matrices A_T represents the connections between the districts in the graph. With total N districts, the adjacency graph is a N x N matrix. A_{ij} contains the weight of the relation between districts i and k which is given by,

Algorithm 1. Overall algorithm

Input: ,
Output: Predicted yield

1. $H^{(1)} = 0$
   For $t = 1$ to t
   For each district $i = 1$ to N
2. Calculate $A_{i,j}$ using equation 6
   End for
3. Construct G_T
   For epoch in 1:100:
4. calculate yearly embeddings $E^{(t)}$ using equation 2
   End for
5. Final spatio-temporal embeddings $H^{(1)} = H^{(1)} \odot E_{(t)}$
   End for
   For epoch in 1:100:
6. Temporal relation mapping $h^t = tLSTM(H^{(l)}, \Delta t)$
   End for
7. Predicted Yield $\hat{y}_{d,t} = FCN(h^t)$
   For $[0,1, \cdots i]$ where $i = \{1, P\}$
8. Optimize *RMSE*, R^2, and r
   End for
   For learning rate= 0.0001 to 0.1
9. Optimize *RMSE*, R^2, and r
   End for
   Final predicted yield $\hat{y}_{d\,t}$

$$A_{i,k} = \frac{x_i^j . x_k^j}{|x_i^j| . |x_k^j|} \quad (6)$$

Where x_i^j, x_k^j represents the j^{th} feature matrix of the i^{th} and k^{th} districts in V respectively. Here $j \in [1, M]$ where M are the total number of features and $i, k \in [1, N]$. The following produced adjacency matrices A_T are used for the generation of the yearly district attribute relational graph, denoted as G_T. The aforementioned procedure is repeated annually, obtaining all required $G_T = \{G_1, G_2, \cdots G_t\}$ graphs.

Figure 7 represents one of the yearly district attribute relational graphs, G_9 showing the spatial relationships among various districts in Rajasthan for the year 2016 where each node represents a district, and the edges (lines) between them indicate the strength of the spatial correlation or dependency in various attributes between those districts. Here red and thick edges indicate strong spatial dependencies, suggesting that one district is highly correlated with that in the neighboring districts. While green and thin indicate weaker dependencies or correlations between districts. The graph displays clusters of districts characterized by significant interdependencies, which are accentuated by the red borders. For instance, districts like as Jaipur, Sikar, Ajmer, and Chittorgarh have significant geographical correlations, suggesting that the Bajra yield in these regions tends to follow similar patterns.

SGHTE (Spatial Graph Hop With Temporal Enhancements) Model

The SGHTE model consists of a combination of Mixhop GCN, and t-LSTM architectures as seen in Figure 6, which are used to capture spatial and temporal dimensions, respectively. CNN-based models [14,16] capture spatial patterns but assume independence between regions, overlooking spatial continuity—where neighboring regions often produce similar yields [32, 39]. Graph-based models such as GCN [10] [15] and GAT [27] can capture spatial dependencies. GNN-RNN [32] captures the apatial connections between the districts but traditional algorithmsonly aggregate first-order neighbor information per layer, limiting their receptive field. In order to achieve precise spatial attribute extraction, it is essential to use higher-order attribute capturing. This kind of capturing involves nodes receiving latent representations from their immediate (first-degree) neighbours as well as from subsequent k-degree neighbours at each convolution layer as seen in Figure 6. Considering this the authors use the mixhop GCN, a model that incorporates trainable aggregation parameters to determine the optimal combination of latent information from neighbouring nodes at different distances [40-42]. It is used as a spatial modelling technique to achieve larger and enriched embeddings, and its effectiveness has been shown. The model integrates data from successive K-order neighbours.

The graphs G_T are the input of the mixhop GCN model. They are trained over each year to obtain the spatial embeddings E_n for each year. The mixhop model aggregates

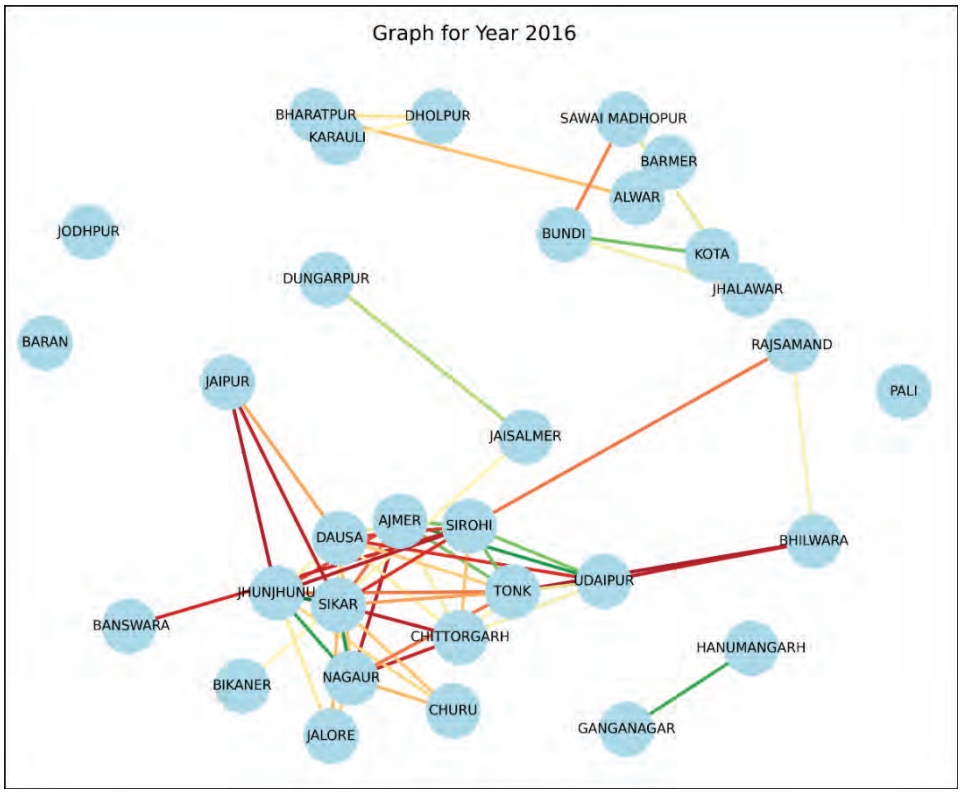


Figure 7. Attribute relation graph of Rajasthan's districts for the year 2016, visualizing spatial connections and feature linkages among neighboring regions.

information from the k-neighbouring nodes by taking the self-product of the adjacency matrix A_T. The number of times the self-product is performed is determined by the hyperparameter P. The adjacency matrix A_T is processed with the help of equation (2). Equation 2 is used to generalize the diffusion convolution layer. The model however is trained using the RMSE (Root Mean Square Error) loss function. The mathematical definition of RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_{d,t} - \mathrm{E_t})^2} \quad (7)$$

Here the variable m denotes the number of samples, $y_{d,t}$ represents the actual yield, and E_T represents the projected yield from yearly embeddings. A decreased root mean square error (RMSE) indicates a reduced number of prediction errors, indicating that the proposed model is exhibiting superior performance. The year-wise spatial embeddings derived from Equation (2), are trained using Equation (7). This results in highly feature-rich embedding E_1. Now the model is iterated for all 13 years resulting in $H^{(1)}$ which is the output of the mixhop GCN.

The $H^{(1)}$ embeddings are obtained as the output from the Mixhop GCN. The aforementioned embeddings serve as detailed representations of the spatial features that are acquired by the mixhop model for the input graphs D_{yg}.

Next, the latter component of the SGHTE model is introduced. To get precise forecasts of Bajra yield ($y_{d,t}$), it is essential to address temporal factors with the same level of consideration as spatial characteristics. Recurrent Neural Networks (RNNs) are a well-established choice for modeling sequential data due to their ability to store and update contextual information over time without relying on the restrictive Markov assumption [34]. The basic RNNs are enhanced by standard LSTMs that incorporate gated mechanisms to more effectively capture long term dependencies as well as address the problem of vanishing or exploding gradients. Yet, by their nature, they assume that time periods between observations are constant, which is hardly the case in agricultural data, where such variables as rainfall, irrigation, and fertilizer application tend to have uneven time periods. The first factor that needs to be taken into account is the fact that they should consider the temporal irregularities. In the present scenario, it is seen that the yearly rainfall data exhibits non-uniformity across a certain time frame, as detailed in Section 1.2 and shown in Figure 5. This problem is addressed by including a t-LSTM [2] to account for the potential impact of the uneven time between events. The main objective is to utilize the t-LSTM model illustrated in Figure 6 and produce the final yield $y_{d,t}$ from the embeddings $H^{(1)}$ obtained from equation (2), for the 13 years using the mixhop GCN model. The key difference in

t-LSTM compared to traditional LSTM involves modifying the memory of the prior cell state according to the elapsed time between the events.

The distinguishing change in t-LSTM is the modification of the previous cell memory $c^{(t-1)}$ state by factoring in the variable time difference. This modified part of the LSTM can be seen in Figure 6 by the blue lines of the temporal module of SGHTE. The t-LSTM first from the network calculates the short-term memory state $c_S^{(t-1)}$ which is given by,

$$c_S^{(t-1)} = \tanh\left(W_d c^{(t-1)} + b_d\right) \quad (8)$$

This short-term memory considers in the irregular time interval between the events like rainfall. Then memory discount is applied to this state, and the discounted short-term memory is given by,

$$\hat{c}_s^{(t-1)} = c_S^{(t-1)} \odot decay(\Delta^{(t)}) \quad (9)$$

The decay of this elapsed time is given by $decay(\Delta^{(t)}) = \frac{1}{\log(e + \Delta^{(t)})}$. Now finally the adjusted previous state memory $c_*^{(t-1)}$ is computed using the discounted short-term memory and the long-term memory $c_T^{(t-1)} = c^{(t-1)} - c_S^{(t-1)}$. The adjusted previous state memory is computed as,

$$c_*^{(t-1)} = c_T^{(t-1)} + \hat{c}_s^{(t-1)} \quad (10)$$

The t-LSTM only differs in this modification of the prior memory state, and the rest of the components of the t-LSTM are similar to vanilla LSTM. After this modification of the previous state, the t-LSTM follows the regular vanilla LSTM where firstly, the forget gate $f_g^{(t)}$ controls how much of the previous cell state $c^{(t-1)}$ should be retained. Then the input gate $i_g^{(t)}$ controls the amount of the $H^{(l)}$ matrix to be used as the new input to update the cell state. Candidate state $\tilde{c}^{(t)}$ refers to the possible new values that may be included in the cell state memory. Combining the prior cell state $c^{(t-1)}$ with the new candidate's memory $\tilde{c}^{(t)}$ the updated new cell state memory is $c^{(t)}$. Finally, the hidden state h^t is computed using the revised cell state. The output gate $o_g^{(t)}$ regulates the proportion of the cell state that should be sent as output. The below equations give how these gates are calculated.

$$f_g^{(t)} = \sigma(W_f H^{(l)} + U_f h^{(t-1)} + b_f) \quad \text{(forget gate) (12)}$$

$$i_g^{(t)} = \sigma(W_i H^{(l)} + U_i h^{(t-1)} + b_i) \quad \text{(input gate) (13)}$$

$$\tilde{c}^{(t)} = \tanh\left(W_c H^{(l)} + U_c h^{(t-1)} + b_c\right) \quad \text{(candidate memory gate) (14)}$$

$$c^{(t)} = f_g^{(t)} \odot c^{(t-1)} + i_g^{(t)} \odot \tilde{c}^{(t)} \quad \text{(current cell state) (15)}$$

$$o_g^{(t)} = \sigma(W_o H^{(l)} + U_o h^{(t-1)} + b_o) \quad \text{(output gate) (16)}$$

$$h^t = o^t \odot \tanh\left(c^{(t)}\right) \quad \text{(hidden state) (17)}$$

In these equations $\{W_f, W_i, W_c, W_o\}$, $\{U_f, U_i, U_c, U_o\}$, and $\{b_f, b_i, b_c, b_o\}$ are the network parameters of the forget, input, candidate, and output gates. $\{W_d, b_d\}$ are the parameters of the decomposition subspace. Using this t-LSTM model is explained in equations from (8-17).

Algorithm 2. Algorithm for spatial and temporal models of SGHTE

Input: G_T, l, P *and* Δt
Output: h^t **Predicted yield**

1. Initialize mixhop GCN with model parameters.
2. Let B = $E^{(l-1)}$) and $H^{(l)} = 0$
   For epoch 1 to 100
   For t = 1 to t
   For l = 1 to l
   For j = 1 to P
3. $B = \tilde{A}_{d,t} . B$
   If $j \in P$ **then**
4. $Q_j = B . W_j^{(l)}$
   End if
   End for
5. $E^{(l)} = \|_{j \in P} Q_j$
6. $H^{(l)} = H^{(l)} \odot E^{(t)}$
7. Calculate $f_g^{(t-1)}, i_g^{(t-1)}, \tilde{c}^{(t)}$, and $o_g^{(t)}$ using $H^{(l)}$ and equation 12,13,14, and 16
8. Modify current state memory $c^{(t)}$ using equations
9. Calculate h^t from equation 17

End for
End for

This model is trained over and the finally obtained predicted yield ($\hat{y}_{d,t}$) is,

$$\hat{y}_{d,t} = FCN(h^t) = FCN\left(o^t \odot \tanh\left(c^{(t)}\right)\right) \quad (18)$$

$\hat{y}_{d,t}$ is the predicted yield by the proposed SGHTE model. This predicted yield is now compared with the true yield for Bajra crop and the performance of the model is evaluated. In the next section design and settings for the experimental procedure of the SGHTE model is elaborated.

EXPERIMENTAL DESIGN AND SETTINGS

In this part, the authors present the experimental framework that they employed to measure the performance of the proposed model. It will contain the description of the dataset, the process in which the data collection was performed, the evaluation metrics which were used to evaluate the accuracy of the model, the training and testing process and the parameter settings. This holistic method provides the ability to replicate the results and to perform an accurate estimation of the performance of the model.

Dataset Description

The present study comprises data collected from many sources, including the DACNET website of the Ministry of Agriculture and Farmers' Welfare [38], as well as the agricultural site of Rajasthan [38]. The dataset covers 32 administrative districts (geographic resolution) of the state of Rajasthan, India, over a continuous time span from 2007 to 2020 (14 years), specifically focusing on the Kharif season crop known as Bajra (Pearl Millet). This results in a total of 416 district–year records, offering both spatial and temporal diversity in

the observations. Figure 8 shows an extract of Bajra dataset with such parameters as Area, Production, Yield, Canals, Tanks, Wells, Soil type, Phosphorus and Potassium content, Fertilizer, Hybrid seeds, Saline Soil, Alkaline Soil, Rainfall, Climate type, year, and District.

The selection of the aforementioned attributes is based on their significant contribution to the prediction of Bajra yield. As an example, the area allocated for cultivating a particular crop is denoted by the "Area." The attributes Cannals, Tanks, and Wells give details of the size of area that has been irrigated assisted by the various sources of water. The attribute Fertilizer refers to the amount of fertilizer applied expressed in tonnes. Rainfall is the average amount of meteorological precipitation taken and measured in millimetres (mm). The addition of soil nutrients including phosphorus and potassium are the main features of the soil. The salinity of soil is an important feature that can be used to refer to high levels of dissolved salts in the soil water. Considerable amounts of salt soil have been reported to have negative effects on the growth of plants which in turn results

District	Year	Area (Hectare)	Production (Tonnes)	Yield (Tonnes /Hectare)	Soil type 1	Soil type 2	Phosphorus	Pottasium	Saline Soil (Ha)	sodic or Alkaline Soil (Ha)	Annual Normal Rainfall (mm)	Climate Type	irrigated area by canals(hectare)	irrigated area by tanks(hectare)	irrigated area by wells(hectare)	hybrid seeds(quintal)	total fertilizer (tonnes)
1.AJMER	2011-12	74554	95330	1.28	lithosolsat foot	Sierozens	M	M	16712	19830	602	SEMI-ARID	4750	5256	0	2302	12649
2.ALWAR	2011-12	267792	552935	2.06	Alluvial prone t	Alluvial deposites ca	L	M	15976	97625	657	SEMI-ARID	488	0	0	10200	31421
3.BANSWARA	2011-12	101	127	1.26	Predominantly	welldrained calcare	M	M	2131	2130	950	SEMI-HUMID	59019	5005	0	0	20749
4.BARAN	2011-12	2849	3737	1.31	Clay Loam	Black of alluvial orig	M	H	1008	1584	874	HUMID	67280	5200	0	36	23249
5.BARMER	2011-12	878742	576877	0.66	Desert soils	sand dunes aeolian	M	M	1596	1989	266	ARID	1299	0	0	4430	6152
6.BHARATPUR	2011-12	118162	231978	1.96	Alluvial prone t	Alluvial deposites ca	L	M	32613	45217	664	SEMI-HUMID	3233	0	0	4580	29337
7.BHILWARA	2011-12	2882	2380	0.83	lithosolsat foot	alluvials in plains	M	H	27950	13470	683	SEMI-ARID	14624	5746	0	40	16522
8.BIKANER	2011-12	206510	200220	0.97	Desert soils	sand dunes aeolian	M	M	14134	14033	243	ARID	151568	0	0	437	9500
9.BUNDI	2011-12	2426	3182	1.31	Clay Loam	Black of alluvial orig	M	H	6009	9229	773	SEMI-ARID	113849	1203	0	137	19931
10.CHITTORGARH	2011-12	27	34	1.26	lithosolsat foot	alluvials in plains	L	M	17720	11397	842	MEDITERRANEAN	10445	936	0	0	28747
11.CHURU	2011-12	404192	293961	0.73	Desert soils	sand dunes aeolian	M	H	14134	250	355	SEMI-ARID	4265	0	0	11410	4416
12.DAUSA	2011-12	138706	228025	1.64	lithosolsat foot	Sierozens	L	H	4056	38437	561	MEDITERRANEAN	148	0	0	4112	17557
13.DHOLPUR	2011-12	82914	184950	2.23	Alluvial prone t	Alluvial deposites ca	M	L	5373	20121	745	SEMI-HUMID	8352	192	0	2686	12300
14.DUNGARPUR	2011-12	124	155	1.25	Predominantly	welldrained calcare	M	M	2819	3928	729	HUMID	9295	2489	0	0	6532
15.GANGANAGAR	2011-12	5858	10725	1.83	high soluble salts & exchangeable sodium	Alluvial deposites ca	M	H	14214	6517	274	ARID	595624	0	0	27	41416

Figure 8. Sample extract of the raw Bajra yield dataset before preprocessing, showing district-wise attributes across multiple years.

to a decline in agricultural productivity and worst, death of the plant in severe cases. The alkaline nature of the soil is a pointer of the levels of salts, which adversely affect the soil structure thereby decreasing crop production. Hybrid seeds: These are the genetically modified seeds which are utilized in the production of Bajra in quintals. The soil type and climate type are the particular composition of soil and the environment of various areas of existence.

Limitations on scope of data: The dataset is limited to Bajra crop in Rajasthan, future work will extend to multiple states and crops for broader applicability. District-level aggregation of some attributes may overlook within-district variability. Some district–year records were removed due to missing values, potentially biasing results, and label-encoding with scaling before the train–test split may cause data leakage. Rainfall parsing errors and similarity-based rather than geographic graph construction could also affect accuracy.

Data Collection and Preprocessing

The overall flow for data preprocessing is highlighted in Figure 9. The yield data is scraped from the DACNET website, which is maintained by the Ministry of Agriculture and Farmers' Welfare [38]. The provided dataset encompasses 32 districts in the state of Rajasthan over a period of 23 years. The gathered characteristics include area, production, year, and district name. The yield is computed by dividing the production data by the harvested area, as shown in columns 'Production' and 'Area'. The meteorological data, annual rainfall and the climatic type are undertaken. The dataset for annual rainfall is sourced from the official website of the Department of Water Resources, Rajasthan Government [40]. The values for the annual net irrigated area derived from water sources such as wells, canals, and tanks, as well as the quantities of hybrid seeds and fertilizers used were acquired from the Government of Rajasthan Directorate of Economics & Statistics [38]. The dataset was spread across many PDF files, whereby each file included the data corresponding to a certain year. Before analysis, each of these files underwent parsing in Python using packages such as PyPDF2 and regex. This study utilizes climatic data from reference [41] to include yield forecast for each district in Rajasthan. In accordance with [41], the geographical region of Rajasthan is classified into five distinct climatic categories, namely desert, semi-arid, semi-humid, humid, and Mediterranean. The climatic classification of each of the 32 districts was determined based on this categorization model. This has been done through regular expression functions whereby each district is assigned the respective climate. The following columns contain the categorical values: the Phosphorus, Pottasium, Soil type 1, Soil type 2, and Climate Type. The columns bearing the labels Phosphorus and Pottasium have the numerical data indicating the quantity of the elements in the soil of the sample like very low (VL), low (L), medium (M), and high (H). The columns that are marked as Soil type 1 and Soil type 2 contain the following classifications lithosols at the foothills, alluvial, prone to water logging, mostly of the reddish medium texture, clay loam, desert soils and so on. These categories are the pointers to the many different mixes of soils in each district. The climatic type includes numerous values, including dry, semi-arid, semi-humid, humid, and Mediterranean as the indicators of the climatic type of each region. Figure 10 is the sample of data that has been pre-processed.

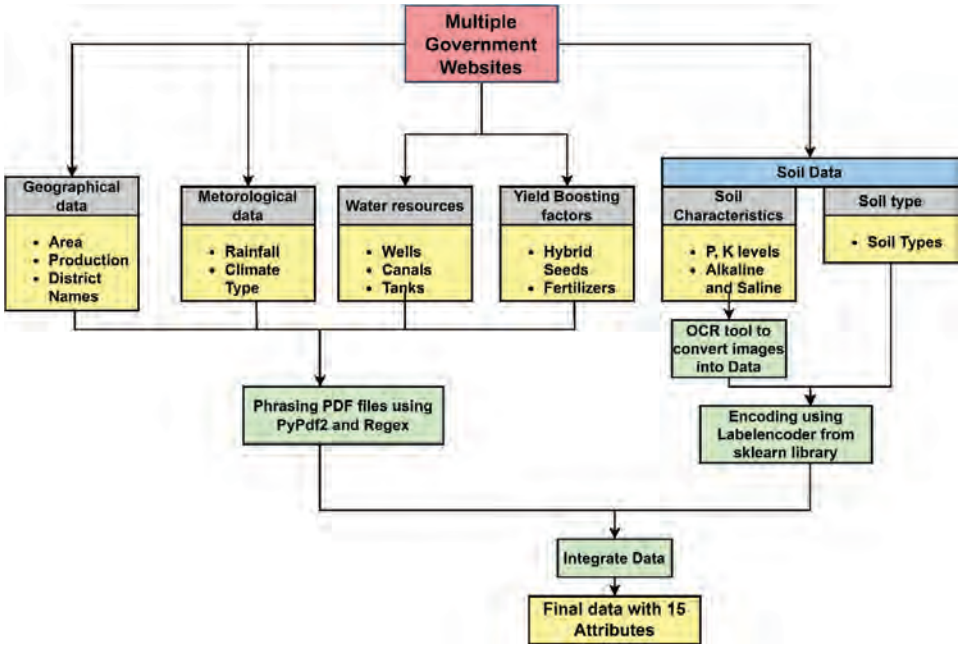


Figure 9. Flow diagram of the data preprocessing procedure, including handling of missing values, normalization, and encoding of categorical variables.

Districts	Year	Area (Hectare)	Production (Tonnes)	Yield (Tonnes/ Hectare)	Soil type 1	Soil type 2	Phosphorus	Potassium	Saline Soil Ha	Alkaline Soil
2.ALWAR	2011-12	0.264771	0.742558	0.872881	0	0	0.25	0.666667	0.000283	0.831
3.BANSWAR	2011-12	9.69E-05	0.000167	0.533898	0.428571	1	0.5	0.666667	3.74E-05	0.017926
4.BARAN	2011-12	0.002814	0.005015	0.555085	0.142857	0.142857	0.5	0	1.75E-05	0.013277
5.BARMER	2011-12	0.868836	0.774711	0.279661	0.285714	0.857143	0.5	0.666667	2.79E-05	0.016725
6.BHARATPU	2011-12	0.116827	0.31153	0.830508	0	0	0.25	0.666667	0.000578	0.384782
7.BHILWARA	2011-12	0.002847	0.003192	0.351695	1	0.571429	0.5	0	0.000495	0.114478
8.BIKANER	2011-12	0.20418	0.268881	0.411017	0.285714	0.857143	0.5	0.666667	0.00025	0.119272
9.BUNDI	2011-12	0.002396	0.004269	0.555085	0.142857	0.142857	0.5	0	0.000106	0.078369
10.CHITTOR	2011-12	2.37E-05	4.16E-05	0.533898	1	0.571429	0.25	0.666667	0.000314	0.096828
11.CHURU	2011-12	0.399634	0.39477	0.309322	0.285714	0.857143	0.5	0	0.00025	0.001919
12.DAUSA	2011-12	0.13714	0.306221	0.694915	1	0.428571	0.25	0	7.16E-05	0.327055
13.DHOLPUF	2011-12	0.081977	0.248374	0.944915	0	0	0.5	0.333333	9.49E-05	0.171107
14.DUNGARI	2011-12	0.00012	0.000204	0.529661	0.428571	1	0.5	0.666667	4.96E-05	0.033235
15.GANGAN	2011-12	0.005789	0.014399	0.775424	0.857143	0	0.5	0	0.000252	0.055278
16.HANUMA	2011-12	0.072056	0.135357	0.584746	0.857143	0	0.5	0.666667	0.001316	1
17.JAIPUR	2011-12	0.313714	0.663726	0.661017	1	0.428571	0.5	0.666667	0.001316	0.003707
18.JAISALME	2011-12	0.141881	0.098669	0.216102	0.285714	0.857143	0.5	0.666667	0.000221	0.1998

Alkaline Soil	Rainfall	Climate type	Cannals(Hectare)	Tanks(Hectare)	Wells(Hectare)	Hybrid seeds(Quintal)	Fertilizer(Tonnes)
0.831	0.075136	0.75	0.00076	0	0.885893	0.711744	0.423656
0.017926	0.108687	1	0.09186	0.174817	0.02685	0	0.267276
0.013277	0.099984	0.25	0.104717	0.181628	0.434257	0.002512	0.30391
0.016725	0.030365	0	0.002022	0	0.335507	0.30912	0.053382
0.384782	0.075938	1	0.005032	0	0.647739	0.319587	0.393119
0.114478	0.078114	0.75	0.022761	0.200699	0.309862	0.002791	0.205337
0.119272	0.027731	0	0.235906	0	0.273553	0.030493	0.102441
0.078369	0.088419	0.75	0.177199	0.042019	0.200551	0.00956	0.25529
0.096828	0.09632	0.5	0.016257	0.032693	0.35989	0	0.384473
0.001919	0.040556	0.75	0.006638	0	0.174166	0.796176	0.027944
0.327055	0.064144	0.5	0.00023	0	0.309451	0.28693	0.220503
0.171107	0.085213	1	0.012999	0.006706	0.206195	0.187426	0.14347
0.033235	0.083381	0.25	0.014467	0.086937	0.053274	0	0.05895
0.055278	0.031281	0	0.927053	0	0.001821	0.001884	0.570116
1	0.064487	0	0.593005	0	0.003272	0.179541	0.427422
0.003707	0.021204	0.75	0.01285	0.020293	0.58722	0.705882	0.387111
0.1998	0.042273	0	0.116294	0	0.087365	0.092108	0.046627

Figure 10. Example of encoded dataset after preprocessing, ready for input into the SGHTE model.

Data pre-processing like removing the rows with null values and integer-encoding the categorical values in order to transform the input data to a suitable format is necessary for machine learning models. The pre-processing steps involved are:

- Pandas' library is used to import the data shown in Figure 10.
- PyPDF2 is utilized to phrase rainfall, irrigated sources area, hybrid seeds, and fertilizer data
- The soil data was present in images. The data was extracted from the images using Microsoft's OCR tool, and stored in CSV files, which were added to the current dataset.
- Encoding of categorical values is performed using the LabelEncoder package from sklearn library.
- Scaling of the data was performed with StandardScaler and MinMaxScaler package from sklearn library
- Finally, the null values are removed with the help of Pandas library

After preprocessing, the yield data from 2010-2020 for 32 districts is considered for graph construction as shown in Figure 6. The yield prediction model proposed in Figure 6, is trained over this data and the settings for this are discussed in the next subsection.

Training and Testing Procedure With Hyperparameter Settings

The hyperparameter settings of the Mixhop GCN model and t-LSTM are discussed in this section. The

parameter setting for generating the year-wise embeddings ($H^{(1)}$) as shown in Figure 6 are: number of features are 15 (which represents the number of input features (x) for each district in the graph after preprocessing explained in above Section), 32 Hidden Channels, dropout Probability is 0.5, the learning rate for the Adam optimizer is 0.001. Based on the Figure 6, the t-LSTM model is trained on $H^{(1)}$ with the window size of 3, and the input, hidden, output sizes are 32,64, and 32, respectively. The training of both mixhop GCN and t-LSTM is conducted prior to testing. Initially, the Mixhop model undergoes training, which involves:

- Initializing the Mixhop GCN model with three GCN layers (conv1, conv2, conv3), where each layer will aggregate information from 0-hop (self), 1-hop, 2-hop, and 3-hop neighbours. A dropout layer for regularization is also included. This generates the year-wise embeddings as shown in Figure 6.
- The training process iterates over 100 epochs. The model processes the data of every year (spatial data of districts) one at a time in each epoch.
- A mixhop GCN model is applied to the data of each year, which will produce node embeddings (spatial features) of the districts. These embeddings are gathered and also to determine the RMSE loss between the embeddings and the actual yield values.
- The annual loss is calculated with the help of RMSE and the mean loss on the years is calculated. This is followed by the loss being backpropagated to adjust the model parameters using the optimizer.

The steps involved in the training of the t-LSTM model are:

- The t-LSTM model takes the embeddings created by mixhop GCN as its input and the time series data is divided into training and testing sets. This study uses a 70:30 data division i.e. 70 percent of the data is taken as a training data and 30 percent is taken as a testing data.
- The t-LSTM model is trained with a fully connected layer with learning rate 0.001. The model is defined by processing the data in 3 sliding window, with each window being a sequence of embeddings across a series of time steps.
- For each test year, the TLSTM model will use input sequences generated from prior years. After training is completed the SGHTE model is evaluated on the test data. Using the same window sliding approach the model is iterated on the testing data. Finally, evaluation metrics are calculated.

Details on Evaluation Parameters

Three evaluation measures, namely *RMSE*, R^2, and Pearson correlation coefficients, are used to assess the performance of the SGHTE model. The quantifies the mean of the deviations between the projected and actual yield. The metric provides an indication of the degree to which the model's predictions align with the observed data, with lower values indicating superior performance. Equation (6) can be used to calculate RMSE. coefficient of determination is an important parameter in regression analysis, which is used to estimate the level of goodness of fit to a model. R^2 is defined to be the portion of the variation in the dependent variable which can be predicted by the independent variables. It has a value between 0 and 1 with a larger value having a better fit. Equation (13) can be used to estimate R^2. Here $y_{d,t}$ is an average of the actual yield.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_{d,t} - \hat{y}_{d,t})^2}{\sum_{i=1}^{m}(y_{d,t} - \bar{y}_{d,t})^2} \tag{13}$$

The Pearson correlation coefficient is used to determine the direction and strength of the linear relationship between two variables. It has the value of -1 to 1, a value of one to -1 indicates how strong the linear relationship is, and a value of zero to -1 indicates how weak the linear relationship is. The value of positive correlation is also high and this implies that the model has a high adherence to the actual yield values implying good model performance. According to equation (14), correlation coefficient (r) is provided. A correlation analysis can be used to determine which features (rainfall, temperature, soil type, etc.) are most related to crop yield to help select features to include in the model $y_{d,t}$ is the average of predicted yields.

$$r = \frac{\sum_{i=1}^{m}(y_{d,t} - \bar{y})(\hat{y}_{d,t} - \bar{\hat{y}}_{d,t})}{\sqrt{\sum_{i=1}^{m}(y_{d,t} - \bar{y}_{d,t})^2 \sum_{i=1}^{m}(\hat{y}_{d,t} - \bar{\hat{y}}_{d,t})^2}} \tag{14}$$

Table 3 summarizes the experimental setup of the proposed SGHTE model, which makes the methodology completely transparent as to be reproducible. The algorithmic process included these stages: data cleansing, categorical encoding, scaling, and consolidation of heterogeneous data into a single data set of 15 attributes of 32 districts in 13 years. The spatial input of the Mixhop GCN module was constructed by creating the graphs every year with cosine similarity between the district-level feature vectors. This spatial module used three convolution layers with multi-hop neighbourhood aggregation and then a t-LSTM time based module that was used to forecast irregularities in time dependence in attributes like rainfall. Each module was trained in 100 epochs with a sliding window training method with a 70:30 split between training and testing. RMSE was used as the primary loss function, while RMSE, R^2, and Pearson's correlation coefficient served as evaluation metrics. These details collectively provide a comprehensive blueprint for reproducing the results.

After setting up the model the experimentation is performed and predicted yield results are obtained. These results are discussed in the following section.

RESULTS AND DISCUSSION

The results of the experimentation for the proposed SGHTE model are divided into three parts: 1) Sensitivity analysis, where the optimal hyperparameter setting of the

Table 3. Model architecture, hyperparameters, preprocessing steps, and training/testing settings for the SGHTE model

Aspect	Details
Data Preprocessing	Removal of rows with null values (Pandas), integer encoding of categorical variables (LabelEncoder, sklearn), scaling using StandardScaler and MinMaxScaler (sklearn), OCR extraction for soil data (Microsoft OCR), PDF parsing for rainfall, irrigation, hybrid seeds, and fertilizer data (PyPDF2, regex).
Dataset	DAWNET-based data were used as the source of 20072012-2017-2020 Bajra crop statistics of 32 districts of Rajasthan sourced through DACNET, Department of water resources, Rajasthan, Directorate of Economics and statistics and literature-based climate classification. Finally data set: 15 variables, 416 entries, time-varying (temporal) and spatial (spatial) data.
Graph Construction	Nodes: 32 districts; Edges: cosine similarity-based weights between district feature vectors; Graphs constructed yearly from adjacency matrices A^t.
Model Architecture	**Spatial module:** Mixhop GCN with 3 layers (conv1, conv2, conv3), aggregating 0-hop, 1-hop, 2-hop, and 3-hop neighbours; dropout layer (0.5). **Temporal module:** t-LSTM with input size 32, hidden size 64, output size 32, incorporating time-aware memory decay. Fully Connected Layer (FCN) for final yield prediction.
Hyperparameters – Mixhop GCN	Input features: 15; Hidden channels: 32; Dropout: 0.5; Learning rate: 0.001 (Adam optimizer); Epochs: 100.
Hyperparameters – t-LSTM	Window size: 3; Input size: 32; Hidden size: 64; Output size: 32; Learning rate: 0.001; Epochs: 100.
Training/Testing Split	70% training, 30% testing (sliding window approach).
Loss Function	RMSE (Root Mean Square Error).
Evaluation Metrics	RMSE, R^2 (Coefficient of Determination), Pearson's correlation coefficient (r).

model is selected. 2) Predicted results, where the predicted output of the model is compared to true yield, and 3) State-of-the-art comparison, where the proposed model and its variants are compared with other effective crop yield models.

Sensitivity Analysis

The proposed model's performance is sensitive to mixhop GCN parameter and learning rate =[0.0001,0.001,0.01,0.1] The optimal tuning of these sensitive parameters is required for optimal performance metrics.

In the map given for illustration purposes in Figure 3, for Jaipur, beyond three hops , the relevance of neighboring districts typically diminishes. The districts that would fall into fourth-order neighbours are often too far geographically and may not have significant socio-economic or cultural ties to the reference district. Thus, here testing up to three hops is evaluated, focusing on districts where influences are most likely to be felt. The learning rate is also evaluated for varying range from 0.0001 to 0.1.

Figure 11 presents a comparative bar plot illustrating the performance metrics of the SGHTE model by varying the MixHop GCN hyperparameter which is a set of integer adjacency powers that refer to the hops over which the model aggregates information. The learning rate is kept fixed at a value of 0.001. A value of equal to 0 implies that the model simply combines information from its immediate neighbors, which is analogous to a conventional GCN. From Figure 11, it is evident that as the hyperparameter set includes more hops (e.g., [0,1,2,3]), the performance improves. The RMSE values remain low, while the R^2 and correlation values are relatively high, especially in the hop set [0,1,2,3], which seems to give the best overall results with an R^2 of 0.6775 and a of 0.8912. This implies that adding a larger range of hops, i.e. considering districts linked by distant districts rather than simply their immediate neighbours, leads to more accurate and trustworthy predictions by the Mixhop GCN module in SGHTE model.

Figure 12 consists of three column graphs that depict the model assessing performance metrics: Test shown in blue, Test shown in red, and the Test is shown in green. The x-axis indicates various learning rates. The y-axis corresponds to the numerical values of the performance metrics. The optimal hop value from the above experiment in Figure 12 is used for learning rate sensitivity analysis.

The best learning rate of 0.001 would give the minimum RMSE of 0.1306, and it means that the model would be able to train without overfitting or underfitting data sets. The reduced values of RMSE show improved model accuracy. R 2 value of 0.6775, which is the largest at the learning rate of 0.001, implies that this learning rate allows the model to best represent the variation in the Bajra yield data. The highest correlation coefficient of 0.8912 is achieved when learning rate is 0.001 indicating that learning rate provides the best compromise of accurate predictions with strong correlation. The larger values of the correlation coefficient and R 2 denote the superiority of the model. This indicates that with a learning rate of 0.001, the model is successfully extracting knowledge of the relational patterns in the

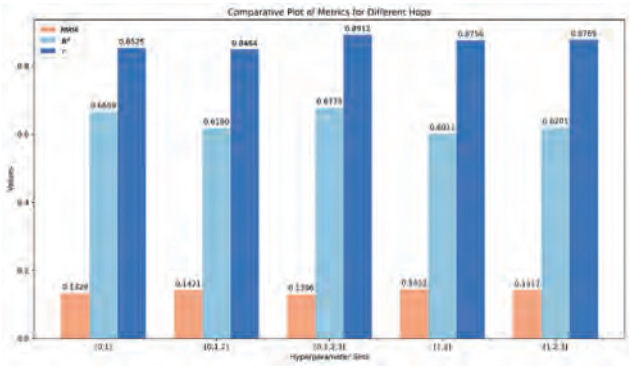


Figure 11. Effect of varying the Mix-hop hyperparameter on model performance.

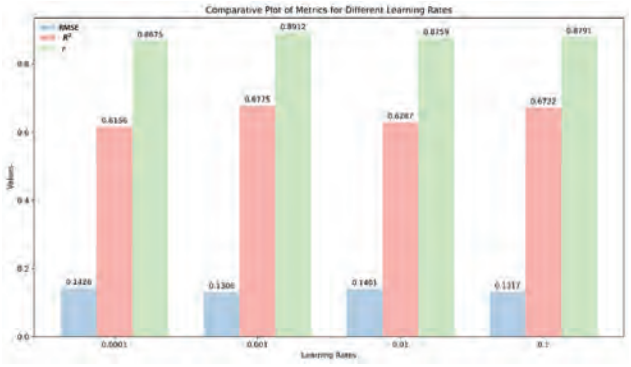


Figure 12. Impact of different learning rates on proposed model.

district-wise data without being too aggressive in weight adjustments (which may result in overfitting) or being too overly conservative in weight adjustments (which may result in underfitting).

Comparative Analysis

This section evaluates the performance of the proposed model by analysing true versus predicted yield, conducting a comparative analysis with state-of-the-art models, and examining model variations through feature reduction studies.

Figure 13 shows the true yield values vs the predicted yield values of all the 32 districts for the final year of data 2019-2020. The model was trained on the yields for all the previous years and the targeted year predicted yields for the year 2019-2020 are given in the Figure 14. The model has a high predictive level of the yields in all 32 districts, and the margin of error is rather low. In training, the RMSE is decreased to 0.1306 which is a better predictive result. The blue dashed line represents the real yield values of the districts and the red solid line is the projections of the model. The fact that in the majority of the districts the RMSE of the actual and predicted lines is smaller proves that the model is a good reflection of the spatial and temporal patterns. This illustrates the ability of SGHTE model to represent Mix-hop spatial interactions and temporal patterns that result to accurate yield predictions.

The proposed model is compared with its variants alongside with other state-of-the-art models to show the improvement in yield prediction. The modified versions of the proposed model for component wise analysis are:

- The proposed model (SGHTE) is trained without the features and denoted as LASGHTE (Limited Attribute SGHTE)
- In the proposed model seen in Figure 6, the irregularity aware temporal part is replaced with traditional LSTM (MGCN+LSTM).
- Mutli-hop spatial part is replaced with traditional GCN (GCN+t-LSTM).
- Irregularity aware temporal part is replaced with GRU (MGCN+GRU).
- Both spatial and temporal parts are replaced with traditional GCN and LSTM (GCN+LSTM).

All comparisons are performed after optimizing the model parameters, ensuring consistency by utilizing the same datasets.

A comparison of the predicted vs. actual yield values for the proposed SGHTE model and component wise analysis is provided in Figure 14. This figure illustrates the regression analysis of predicted vs. actual yield values along with their regression coefficients (R values). The red line represents the regression line, showing the model's overall trend in predicting yield values. It is a line of best fit for the predicted yield (y-axis) against the true yield (x-axis) in Figure 14. As compared to the shaded red region around the regression line, which denotes the uncertainty or variability of the predictions of the model. When the spread is narrow it means that there is more confidence in the predictions whereas when the spread is wide the prediction is less certain. The blue dots, that are presented in Figure 14, point to each single data point, and they plot the actual yield (x-axis) against the yield predicted by the model (y-axis).

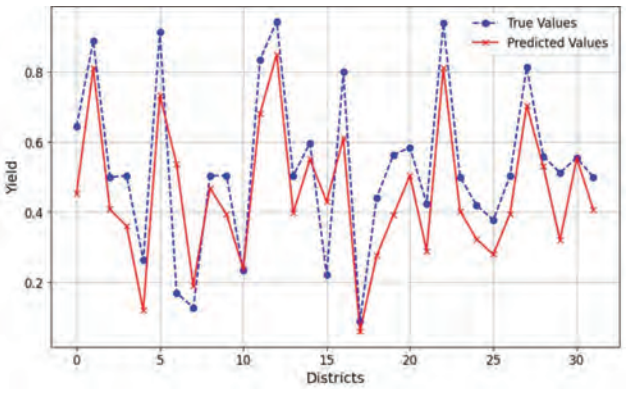


Figure 13. Comparative plot of actual versus predicted Bajra yield using the proposed model, showing high alignment and predictive accuracy.

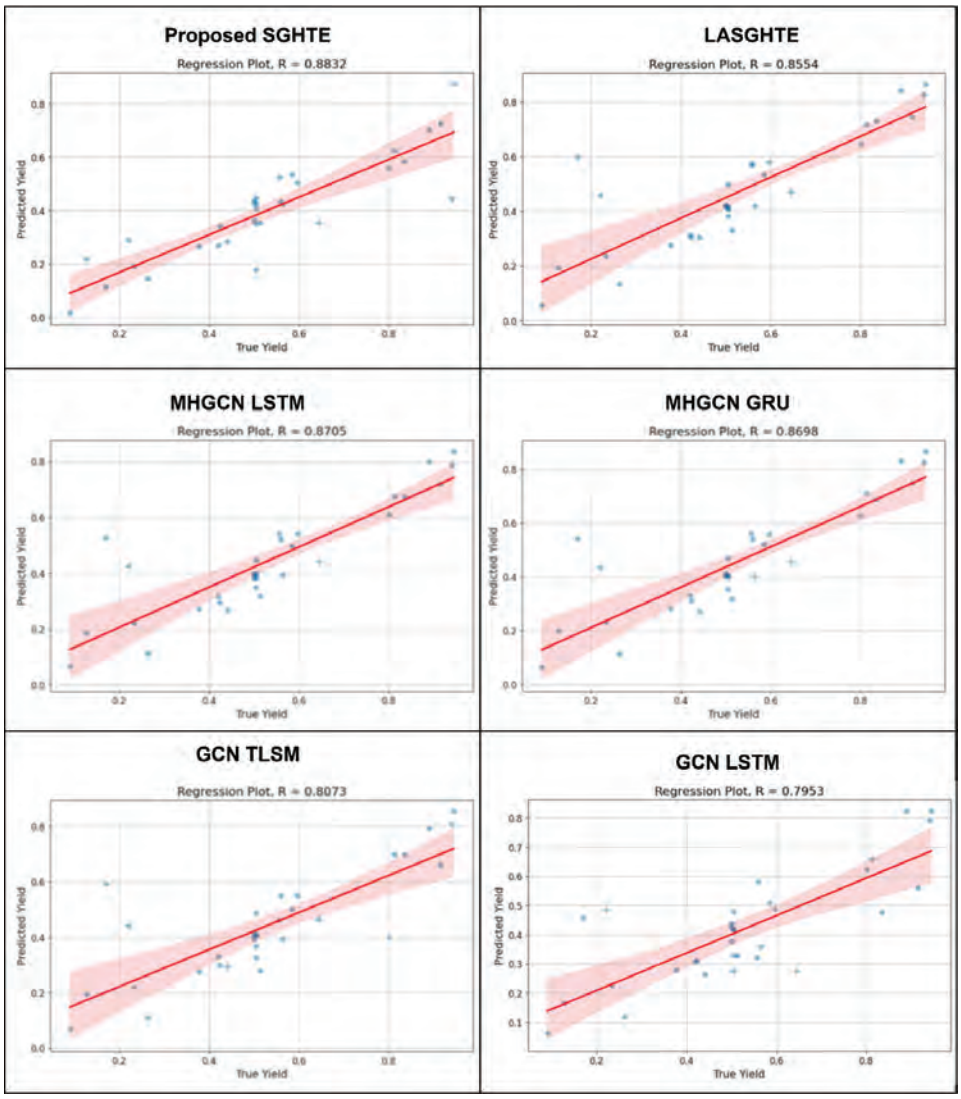


Figure 14. Regression plot of the proposed SGHTE model along with component-wise analysis, demonstrating the contributions of spatial and temporal modules to prediction performance.

The closeness of the blue dots to the red line is the extent to which the model predictions match the actual ones.

The proposed model shows highest regression coefficient $R = 0.8832$, indicating strong predictive accuracy for crop yield. The red regression line closely corresponds with the actual yield values, but the blue dots are densely grouped around it, indicating minimal variation and enhanced yield prediction. This is because of the depth wise district connections based on regional dynamics and handling temporal irregularities with the help of t-LSTM. However, when LASGHTE is evaluated, it shows a slightly lower $R = 0.8554$ compared to the proposed SGHTE, indicating a small decrease in predictive accuracy. This variant excludes certain features (like yield boosters and water resources), which leads to a reduction in the predictive power. This indicates that the inclusion of fertilizers and seed information helped with the yield prediction improvement of 3.2% in the proposed model compared to LASGHTE.

The variant MHGCN LSTM shows R-value of 0.8705, close to that of the proposed SGHTE, indicating a strong predictive alignment, though slightly less than the proposed model. Here, the t-LSTM component is replaced with a traditional LSTM and MixHop is retained, indicating that model still captures complex spatial dependencies well keeping the R high, but do not account for irregular temporal variations as explained in Figure 7, this is why it shows 0.19% decrease in R value compared to proposed model. Extension of LSTM model, GRU is also applied for ablation study, it shows R-value of 0.8698, similar to MHGCN LSTM, indicating a strong predictive performance. Replacing t-LSTM with GRU still yields good results, but GRU might not handle irregular temporal data as effectively as t-LSTM. While GRU is generally capable in time-series modelling, it

may introduce a small bias, showing 0.015% of decrease in R compared to proposed model.

GCN t-LSTM shows lower regression coefficient of R = 0.8073, indicating a weaker alignment between predicted and actual yields. By replacing MixHop GCN with a traditional GCN, this model likely captures fewer complex spatial relationships, which reduces its predictive capability. However, the t-LSTM component still helps in managing irregular temporal patterns, which keeps the regression coefficient from dropping further. Whereas the variant GCN-LSTM shows R = 0.7953 suggesting weaker alignment, with more bias and error in predictions.

State-of-art models comparison (SOTA)

The performance of proposed model is compared with the state-of-the-art models used for yield prediction like LSTM [20], CNN-RNN [21], and GNN-RNN [32]. Commonly used for time-series forecasting models in crop yield prediction is LSTM [20], which captures long-term temporal relationships. CNN-RNN [21] and GNN-RNN [32] models are the ensemble models that capture both spatial and temporal features and these architectures provide a strong comparison point for the SGHTE model with GNN-RNN [32] model showing improvements in using graph based spatial feature extraction [43]. These state-of-the-art models are evaluated using the pre-processed Bajra crop data, maintaining all the parameters and methodology approaches described in the study identically.

Table 4 provides a comparison of the performance of various state-of-the-art models, including the proposed MixHop GCN with Time-aware LSTM (denoted as MGCN t-LSTM), across three metrics: ,, and (pearson correlation). The proposed model outperforms the recent models with the best ,, and CORR scores of 0.1306, 0.6775, and 0.8912, respectively. Table 4 shows that the SGHTE achieves high performance compared to the LASGHTE variation as it includes irrigation accessibility, fertilizer, hybrid seeds, and soil types, that impacts significantly on the crop yield. The model where LSTM is used for temporal variations MGCN LSTM and LSTM [20] shows high RMSE, which is 11.83% and 23.7% compared to proposed model. Whereas MGCN LSTM and LSTM [20] shows low R^2 and r which is 1.23%,1.64% and 5.54%, 1.26% compared to proposed model respectively. This is because the time irregularity employed in proposed model helps capture temporal patterns more effectively than the standard LSTM.

Table 4. Comparison of proposed model with SOTA models

Model	Improved Attribute Dataset		
	RMSE	R^2	*r*
SGHTE (Proposed)	**0.1306**	**0.6775**	**0.8912**
LASGHTE	0.1549	0.5459	0.8662
MGCN LSTM	0.1482	0.6692	0.8766
MGCN GRU	0.1546	0.6062	0.8686
GCN t-LSTM	0.1420	0.6187	0.8075
GCN LSTM	0.1492	0.5790	0.7841
LSTM [4]	0.1712	0.640	0.880
CNN-RNN [8]	0.21	0.59	0.50
GNN-RNN [32]	0.4914	0.6236	0.8549
ABP-XGBoost [44]	0.3224	0.5223	0.6905
Random Forest [45]	0.582	0.4995	0.6244

The SGHTE model incorporates a high-order spatial representation, which allows it to model crop yield predictions across districts that are not only geographically proximate but also considers shared agricultural characteristics, such as soil types, irrigation facilities, and fertilizer usage. This comprehensive spatial modelling helps SGHTE capture variations in yield due to regional factors, providing a robust context for districts that might share similar environmental and resource conditions even if they are not immediate neighbours.

The proposed SGHTE model shows the lowest RMSE, indicating superior accuracy in yield prediction. The model GNN-RNN [32] shows high RMSE value compared to the proposed model, reflecting its limitations in capturing spatial relations that extend beyond immediate geographical proximities. This leads to low shows R^2 and r which is 7.96% and 4.07% compared to proposed model. Similarly, CNN-RNN model which captures the spatial relations directly from the crop data instead of incorporating the graph structure shows high RMSE leading to 12.92% and 43.90% of decrease in and compared to proposed model.

The proposed model shows 14.54% and 12% of improvement in and compared to GCN-LSTM which captures the immediate neighbours ignoring the second-order and third order neighbours leading to inferior yield prediction. When the proposed model is compared with MGCN GRU where instead of t-LSTM, GRU is employed, it shows 10.52% and 2.54% decrease in R^2 and r. This shows that the proposed model is not only able to capture the spatial relations of district with its neighbour districts at multiple hop-distances (P = [0,1,2,3]) but also account for temporal irregularities through decay time which is not addressed through GRU. Thus, the SGHTE model achieves robust and reliable yield predictions by effectively combining high-order spatially-informed graph structures with advanced temporal processing mechanisms, significantly outperforming SOTA models.

The suggested SGHTE framework is superior to previous models that have been mentioned in the literature. An example is that older machine learning methods (Random Forest and Gradient Boosting) [12, 20] could be highly accurate in some cases, but fail to represent higher-order spatial patterns and non-periodic time scales, which tend to overfit local region-specific patterns. Deep learning methods such as ANN [21,43] and hybrid CNN–RNN architectures [5, 19, 44] addressed spatial or temporal dependencies individually, but CNNs were constrained by regular grid

structures and RNN-based models struggled with long-term dependencies or irregular event timing. Similarly, GCN-based methods [45,46] effectively modelled spatial relationships but were restricted to first-order neighbours, thereby missing broader regional influences. Conversely, Mix-hop GCN incorporated in SGHTE allows capturing multi-hop spatial dependencies when a single layer is used, whereas the t-LSTM extension has been demonstrated to be well motivated to handle non-uniform time intervals as demonstrated by an RMSE of 0.1306, a R2 of 0.6775, and a Pearson correlation of 0.8912 on the Rajasthan dataset. Compared to CNN to RNN hybrids [14,16] and GCN to LSTM models [46], SGHTE lowers the RMSE by the magnitude of 23.7% and it can obtain higher correlation values, which confirm the increased ability of this model to reflect the complex patterns of space-temporal variations. These enhancements confirm that SGHTE is not only filling the gaps that were found in the previous research but also offers a better generalizable and solid method of crop yield prediction.

CONCLUSION

The Spatial Graph Hop with Temporal Enhancement (SGHTE) model integrates Mix Hop Graph Convolutional Networks (Mix Hop GCN) with Time-Aware Long Short-Term Memory (t-LSTM) to enhance district-level yield predictions for the Bajra crop across 32 districts in Rajasthan, India. Leveraging a comprehensive dataset span from 2010 to 2020, the model incorporates key agronomic features such as soil classification, hybrid seed quantity, and irrigation sources, enhancing predictive accuracy. We propose SGHTE, a hybrid model that leverages multi-region spatial dependencies via an extended MixHop GCN and models irregular temporal patterns with a time-aware LSTM. By capturing cross-district interactions and long-range spatial–temporal correlations, SGHTE significantly improves crop yield prediction accuracy across Rajasthan's 32 districts. Spatial Graph Hop with Temporal Enhancement (SGHTE) is a model that is based on Mix Hop Graph Convolutional Network (Mix Hop GCN) and the Time-Aware Long Short-Term Memory (t-LSTM) on top of the former to predict on a district-level yield of the Bajra crop in 32 district of Rajasthan, India. The model uses several important agronomic characteristics, including soil types, quantity of hybrid seeds and source of irrigation which were obtained through a period of time of 10 years (2010-2020) which increases its predictive power. SGHTE model has been shown to be better at forecasting than the state-of-the-art models with a RMSE of 0.1306, coefficient of determination(R2) of 0.6775 and Pearson correlation coefficient of 0.8912. SGHTE outperforms a vanilla LSTM variant by an 11.84% decrease in RMSE and 1.24% increase in R 2, which reflects its efficiency in describing spatial interaction, dealing with irregularities, and including various agricultural variables, which are important in predicting crop yields in geographically and agriculturally diverse areas. The comparative analysis proves its capability of covering the complexity of agricultural data and it is therefore quite useful in data driven decision making in agriculture.

Limitations and Future Work: Although SGHTE has high performance on predicting yields of Bajra in the state of Rajasthan, its extrapolation to other crops or regions is not tested and will have to be re-trained with region-specific data. The model can be affected by absent or noisy inputs (e.g., rainfall, fertilizer use), and the MixHop GCN and t-LSTM components will cost more computationally at scale. The existing constraints are that the aggregation is at the district level, removals of missing values, and that preprocessing might lead to data leakage, and the construction of a graph is not based on strict geography, but similarity. The work will be extended into a variety of states and crops, tested in out-of-distribution regimes, use higher-resolution data, and run case studies in which feature-importance analysis is used to enhance robustness, interpretability, and policy actions. In addition to Bajra forecasting, the spatial-temporal framework of SGHTE provides flexible addressing of different problems in engineering and natural sciences.

REFERENCES

[1] Pathare AA, Sethi D. Review and Study of Smart Net Meter Unit for Solar Roof Top Photo-Voltaic (SRTPV) Systems Enabled with Internet of Things (IoT) for EV Infrastructure. Modern Computing Technologies for EV Efficiency and Sustainable Energy Integration 2025;317–340.

[2] Pathare AA, Sethi D. Development of IoT-enabled solutions for renewable energy generation and net-metering control for efficient smart home. Discov Internet Things 2024;4:11. [CrossRef]

[3] Pathare AA, Singh RP, Sethi D. An IoT-Enabled Smart Net-Metering System for Real-Time Analysis of Renewable Energy Generation in MATLAB/Simulink. J Ins Eng India Ser B 2024;105:1583–1598. [CrossRef]

[4] Forastieri V. The ILO Programme on Safety and Health in Agriculture: The challenge for the new century–providing occupational safety and health services to workers in agriculture1. Health Safety Agric 2000;1:1.

[5] Yunker JA. Economic growth in China and India: The potential role of population. World Dev Sustain 2024;4:100130. [CrossRef]

[6] Malhi GS, Kaur M, Kaushik P. Impact of climate change on agriculture and its mitigation strategies: A review. Sustain 2021;13:1318. [CrossRef]

[7] United Nations. World Population Prospects 2022: Summary of Results. Available at: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf. Acessed on 4 Apr 2026.

[8] Sisman Z, Tekiner-Mogulkoc H. Using malmquist TFP index for evaluating agricultural productivity: Agriculture of Türkiye NUTS2 regions. Sigma J Eng Nat Sci 2022;40:513–528. [CrossRef]
[9] Ansarifar J, Wang L, Archontoulis SV. An interaction regression model for crop yield prediction. Sci Rep 2021;11:1–14. [CrossRef]
[10] Satyavathi CT, Sankar SM, Singh SP, Kapoor C, Soumya SL, Ambawat S, et al. Breeding Climate Resilient Pearl Millet Cultivars for India. Springer 2024;31–55. [CrossRef]
[11] Pearl Millet (Bajra) Millet Statistics | Millet Statistics by ICAR (Indian Council of Agricultural Research) & IIMR (Indian Institute of Millet Research). (n.d.). https://www.milletstats.com/pearl-millet-bajra/ Accessed on Apr 09, 2026.
[12] Sharma SK, Sharma DP, Gaur K. crop yield predictions and recommendations using random forest regression in 3a agroclimatic zone, rajasthan. J Data Acquisition Proc 2023;38:1635.
[13] Choudhary NK, Chukkapalli SSL, Mittal S, Gupta M, Abdelsalam M, Joshi A. Yieldpredict: A crop yield prediction framework for smart farms. IEEE Xplore 2020;2340–2349. [CrossRef]
[14] Khaki S, Wang L, Archontoulis SV. A CNN-RNN framework for crop yield prediction. Front Plant Sci 2020;10:1750. [CrossRef]
[15] Qiao M, He X, Cheng X, Li P, Zhao Q, Zhao C, et al. KSTAGE: A knowledge-guided spatial-temporal attention graph learning network for crop yield prediction. Inf Sci 2023;619:19–37. [CrossRef]
[16] Sun J, Di L, Sun Z, Shen Y, Lai Z. County-level soybean yield prediction using deep CNN-LSTM model. Sensors 2019;19:4363. [CrossRef]
[17] Dimensions AI: The Most Advanced Scientific Research Database. Available at: https://www.dimensions.ai/. Accessed on 4 Apr 2026.
[18] Jhajharia K, Mathur P, Jain S, Nijhawan S. Crop yield prediction using machine learning and deep learning techniques. Proc Comput Sci 2023;218:406–417. [CrossRef]
[19] Morales A, Villalobos FJ. Using machine learning for crop yield prediction in the past or the future. Front Plant Sci 2023;14:1128388. [CrossRef]
[20] Jhajharia K, Mathur P. Machine Learning Based Crop Yield Prediction Model in Rajasthan Region of India. Iraqi J Sci 2024. [CrossRef]
[21] Vashisth A, Goyal A. Prediction of mustard yield using different machine learning techniques: a case study of Rajasthan, India. Int J Biometeorol 2023;67:539–551. [CrossRef]
[22] Sharma SK, Sharma DP, Gaur K. Machine Learning Techniques for Crop Yield Forecasting in Semi-Arid (3A) Zone, Rajasthan (India). Curr Agric Res J 2023;11. [CrossRef]
[23] Wang J, Wang P, Tian H, Tansey K, Liu J, Quan W. A deep learning framework combining CNN and GRU for improving wheat yield estimates using time series remotely sensed multi-variables. Comput Electron Agric 2023;206:107705. [CrossRef]
[24] Radhika A, Masood MS. Crop Yield Prediction by Integrating Et-DP Dimensionality Reduction and ABP-XGBOOST Technique. J Internet Serv Inf Secur 2022;12:177–196. [CrossRef]
[25] Asamoah E, Heuvelink GB, Chairi I, Bindraban PS, Logah V. Random Forest machine learning for maize yield and agronomic efficiency prediction in Ghana. Heliyon 2024;10. [CrossRef]
[26] Wang F, Li J, Peng D, Yi Q, Zhang X, Zheng J, et al. Estimating Soybean Yields using Causal Inference and Deep Learning Approaches with Satellite Remote Sensing Data. IEEE Xplore 2024. [CrossRef]
[27] Wang H, Zhang L, Zhao J. Application of a Fusion Attention Mechanism-Based Model Combining Bidirectional Gated Recurrent Units and Recurrent Neural Networks in Soil Nutrient Content Estimation. Agronomy 2023;13:2724. [CrossRef]
[28] Khaki S, Wang L. Crop yield prediction using deep neural networks. Front Plant Sci 2019;10:621. [CrossRef]
[29] Lamba V, Hooda S, Ahuja R, Kaur A. Wheat yield prediction using feedforward neural networks. IEEE Xplore 2021;1–6. [CrossRef]
[30] Dharmaraja S, Jain V, Anjoy P, Chandra H. Empirical analysis for crop yield forecasting in India. Agri Res 2020;9:132–138. [CrossRef]
[31] Fan J, Bai J, Li Z, Ortiz-Bobea A, Gomes CP. A GNN-RNN approach for harnessing geospatial and temporal information: application to crop yield prediction. Proc AAAI Conf Artif 2022;36:11873–11881. [CrossRef]
[32] Balasubramanian A, Elangeswaran SVJ. A novel power aware smart agriculture management system based on rnn-lstm. Electr Eng 2025;107:2347–2368. [CrossRef]
[33] Baytas IM, Xiao C, Zhang X, Wang F, Jain AK, Zhou J. Patient subtyping via time-aware LSTM networks. 23rd ACM SIGKDD international conference on knowledge discovery and data mining. Halifax, Canada; 2017. p. 65–74. [CrossRef]
[34] Nathawat R, Singh SK, Sajan B, Pareek M, Kanga S, Đurin B, et al. Integrating Cloud-Based Geospatial Analysis for Understanding Spatio-Temporal Drought Dynamics and Microclimate Variability in Rajasthan: Implications for Urban Development Planning. J Indian Soc Remote Sens 2025. [CrossRef]
[35] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv 2016.
[36] Area under cultivation of Bajra.Area and Production Statistics, Ministry of Agriculture and Farmers Welfare. Available online: https://aps.dac.gov.in/Home.aspx?ReturnUrl=%2f

[37] Government of Rajasthan Directorate of Economics & Statistics. Rajasthan Agriculture Statistics Agriculture Statistics-Government of Rajasthan. Available at: https://rajas.rajasthan.gov.in/Index.aspx Accessed on Apr 09, 2026.
[38] Rajasthan agriculture department https://agriculture.rajasthan.gov.in/home Accessed on Apr 09, 2026.
[39] Annual rainfall data date. Department of water resources, Government of Rajasthan. Available online: https://water.rajasthan.gov.in/wr/#/department-order/142/23/2776/30900 Accessed on Apr 09, 2026.
[40] Gunawat A, Dubey SK, Sharma D. Development of indices for aridity and temperature changes pattern through GIS mapping for Rajasthan, India. Clim Chang Environ Sustain 2016;4:178–189. [CrossRef]
[41] Abu-El-Haija S, Perozzi B, Kapoor A, Alipourfard N, Lerman K, Harutyunyan H, et al. Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. ArXiv 2019;21–29.
[42] Mishra A, Rawat S, Gautam S, Mishra EP. Comparison between Different Mustard Yield Prediction Models Developed using Various Techniques for Udaipur Region of Rajasthan. Int J Environ Clim Chang 2022;12:475–485. [CrossRef]
[43] Cedric LS, Adoni WYH, Aworka R, Zoueu JT, Mutombo FK, Krichen M, et al. Crops yield prediction based on machine learning models: Case of West African countries. Smart Agric Technol, 2022;2:100049. [CrossRef]
[44] Meena RP, Dhama A. Relevance of thermal units in deciding sowing time and yield prediction of groundnut (Arachis hypogaea L.) under irrigated condition of western Rajasthan. J Agrometeorol 2004;6:62–69. [CrossRef]
[45] Sharma V, Singh PK. Performance of AquaCrop model for predicting yield and biomass of okra (Abelmoschus esculentus) crop. Indian J Agric Sci 2023;93:899–905. [CrossRef]
[46] Vats S, Shankar VG, Mohapatra S, Singh H. Enhancing Crop yield through recommendation system using Fuzzy Inference Model. IEEE Xplore 2023;1–5. [CrossRef]

Symbol	Meaning	Symbol	Meaning
G_T	The yearly district attribute relational	$\|\|$	Column-wise concatenation operator
$H^{(l)}$	Output of Mixhop GCN	t	Time
k	Number of nearby nodes	Δt	Elapsed time
x	Feature/attribute matrix	FCN	Fully connected layer
A_t	Adjacency matrix each year	V	Graph nodes
D	Degree matrix of the adjacency matrix	$\mathcal{E}$	Node edges
I	Identity matrix	N	Number of districts
$\tilde{A}_t$	Normalized adjacency matrix with self-connections	$c_S^{(t-1)}$	Short-term memory gate
W_n^j	Trainable weight matrix	$\hat{c}_s^{(t-1)}$	Discounted short-term gate
l	Number of layers of mixhop GCN	$c_T^{(t-1)}$	Long-term memory gate
$y_{d,t}$	Actual yield	$c_*^{(t-1)}$	Adjusted previous state memory gate
I	Identity matrix	$f_g^{(t)}$	Forget gate
$A_{i,j}$	Cosine similarity of the and districts	$i_g^{(t)}$	Input Gate
m	Number of samples of data	$o_g^{(t)}$	
E_t	Yearly embedding output of mixhop GCN	$\tilde{c}^{(t)}$	
d	District name	c^t	
σ	Activation function	h^t	
		r	

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2036

Research Article

A novel generalized lifetime *q*-distribution

Nurgül OKUR[1], Kaoubara DJONG-MON[2,*]

[1]*Department of Data Science and Analytics, Giresun University, Giresun, 28200, Türkiye*
[2]*Department of Mathematics, Gazi University, Ankara, 06560, Türkiye*

ARTICLE INFO

Article history
Received: 23 October 2024
Revised: 13 January 2025
Accepted: 05 February 2025

Keywords:
Lifetime *q*-Distribution; *q*-Calculus; *q*-Parameter Estimation

ABSTRACT

The *q*-analogue of probability distributions provides a framework that offers a broader range of definitions for them and also includes their classical forms. This paper introduces the *q*-analogue of a new one-parameter generalized lifetime distribution and thoroughly examines its distributional and statistical characteristics. The study includes modeling the *q*-analogues of the probability density function and cumulative distribution function, along with an exploration of their shapes through rigorous mathematical analysis. Moreover, the *q*-analogues of moments and related measures are derived for the proposed *q*-distribution. Furthermore, the *q*-analogues of the reliability functions, central moment, and moment generating function for non-negative *q*-continuous random variables are defined and presented for the proposed *q*-distribution. In addition, this paper presents the Lindley *q*-distribution derived from the proposed *q*-distribution, comparing its *q*-distributional properties with those of the classical form. Finally, this paper focuses on estimating the parameters of the proposed *q*-distribution. While the method of moments is commonly used for continuous *q*-distributions due to their complexity, discrepancies occur between empirical and theoretical *q*-moments as *q* deviates from 1. To address this, we propose a modified method for calculating empirical *q*-moments, ensuring consistency and reliability even for small *q* values.

Cite this article as: Okur N, Djong-Mon K. A novel generalized lifetime *q*-distribution. Sigma J Eng Nat Sci 2026;44(2):1283–1298.

INTRODUCTION

The fundamental concepts of *q*-calculus have been emerging in the literature since the early 20th century, with Jackson's work [1] serving as a key milestone. In fact, the origins of this concept can be traced back to Euler [2], who first introduced *q* in his exploration of Newton's infinite series in the Introduction. The "Quantum Calculus" published by Kac and Cheung [3] covers most of the key aspects of *q*-calculus. Researchers are continually making numerous contributions to the advancement of *q*-calculus as given in [4-6].

Probability *q*-distributions offer a versatile framework that extends classical probability distributions by introducing the *q*-parameter, providing a broader spectrum of definitions. Notable contributions in this area include studies on discrete *q*-distributions by various scientists, such

***Corresponding author.**
*E-mail address: kaoubaradjongmon@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

as [7-10], among others. Recently, Djongmon and Okur [11] introduced a generalized q-binomial distribution and a new q-multinomial distribution, inspired by the work of Charalambides [12].

Significant research on continuous q-distributions includes:

The Erlang q-distribution in two forms is obtained by Charalambides [12]. The Gaussian and k-generalized gamma q-distributions are presented introduced D'iaz, Ortiz, and Pariguan [13,14]. The multivariate q-continuous random variables and related probabilistic measures, as well as q-ordered random variables are given by Vamvakari [15,16]. The q-analogues of the gamma and beta, the characterization of the exponential q-distribution through the q-memorylessness property, the extension of k-gamma q-distribution and the q-inversion method to simulate data from a q-distribution are presented by Boutouria, Bouzida, and Masmoudi [17,18]. The Lindley q-distribution in two forms, and the parameters estimation and simulation studies of Lindley, gamma and exponential q-distributions are elaborated by Bouzida and Zitouni [19,20]. This work is significant as the first to introduce q-simulation using the inverse method in q-probability.

Moreover, the lifetime distributions are powerful tools for modeling a wide range of data sets in a more flexible and accurate manner. They are particularly useful in fields where extreme values and long lifetimes are important. These types of distributions allow for more accurate predictions by overcoming the limitations of traditional distributions (Table 1). Some key works in this area are as follows:

Drawing on recent advancements, this paper introduces a new generalized oneparameter lifetime distribution and its q-analogues are proposed, and their special forms consist of the aforementioned distributions in Table 2 in the current paper.

Furthermore, this paper presents the distributional and statistical properties of the proposed q-distribution. For this purpose, the q-analogues of the reliability functions, central moment, and moment generating function for non-negative q-continuous random variables are defined and obtained the proposed q-distribution. Additionally, this article introduces the Lindley q-distribution, derived from the proposed q-distribution, by thoroughly comparing its q-distributional properties with those of the classical form.

In addition, this paper aims to estimate the parameters of the proposed q-distribution. While various parameter estimation methods are used in classical distributions, the method of moments is primarily employed for continuous q-distributions due to their complexity. To apply the method of moments, both theoretical and empirical q-moments are required. However, discrepancies arise between empirical and theoretical q-moments as q-deviates from 1. To address this, we propose a modified method for calculating empirical q-moments, ensuring consistency and reliability even for small values of q.

PRELIMINARIES

This section outlines the fundamental principles of q-calculus and q-probability theory. In this entire study, unless otherwise stated, it is assumed that $0 < q < 1$.

Definition 1 ([3]). Let x, q be real numbers. The q-number $[x]_q$ is defined

$$[x]_q = \frac{1-q^x}{1-q}, (q \neq 1).$$

Table 1. Overview of some lifetime distributions

Lifetime distributions	Probability density functions	Introducer (Year)
Exponential	$f(\tau;\alpha) = \alpha e^{-\alpha\tau}$	---
Lindley [21]	$f(\tau;\alpha) = \frac{\alpha^2}{\alpha+1}(1+\tau)e^{-\alpha\tau}$	Lindley (1958)
Sujatha [22]	$f(\tau;\alpha) = \frac{\alpha^3}{\alpha^2+\alpha+2!}(1+\tau+\tau^2)e^{-\alpha\tau}$	Shanker (2016a)
Amarendra [23]	$f(\tau;\alpha) = \frac{1}{\sum_{k=0}^{3} k!\,\alpha^{-(k+1)}}\sum_{k=0}^{3}\tau^k e^{-\alpha\tau}$	Shanker (2016b)
Devya [24]	$f(\tau;\alpha) = \frac{1}{\sum_{k=0}^{4} k!\,\alpha^{-(k+1)}}\sum_{k=0}^{4}\tau^k e^{-\alpha\tau}$	Shanker (2016c)
Shambu [25]	$f(\tau;\alpha) = \frac{1}{\sum_{k=0}^{5} k!\,\alpha^{-(k+1)}}\sum_{k=0}^{5}\tau^k e^{-\alpha\tau}$	Shanker (2016d)

For $n \in \mathbb{N}$, a natural q-number, it reduces to $[n]_q = \sum_{k=0}^{n-1} q^x$.

Definition 2 ([3]). The q-Gauss binomial formula is given by

$$(x+y)_q^n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q q^{\binom{k}{2}} y^k x^{n-k}, \ (-\infty < x, y < \infty) \tag{1}$$

The q-binomial coefficients are provided by

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{[n]_q!}{[n-k]_q!\,[k]_q!} = \frac{[n]_{k,q}}{[k]_q!}, (k = 0,1,\dots,n)$$

Where $[n]_q! = [n]_q[n-1]_q \dots [2]_q[1]_q$ and $[n]_{k,q} = [n]_q[n-1]_q \dots [n-k+1]_q$.

Definition 3 ([3]). The q-exponential function are presented:

$$E_q^{\tau} = \sum_{k=0}^{\infty} q^{\binom{k}{2}} \frac{\tau^k}{[k]_q!}, \ (\tau \in \mathbb{R}), \tag{2}$$

$$e_q^{\tau} = \sum_{k=0}^{\infty} \frac{\tau^k}{[k]_q!}, \ (|\tau| < 1/(1-q)), \tag{3}$$

Definition 4 ([3]). The q-derivative of f is defined as

$$D_q f(\tau) = \frac{d_q\big(f(\tau)\big)}{d_q(\tau)} = \frac{f(q\tau) - f(\tau)}{q\tau - \tau}, (q \neq 1)$$

where $d_q(.)$ is the q-differential operator. For $n \in \mathbb{N}$ and $x \neq a$, $a \in \mathbb{R}$,

$$D_q e_q^{ax} = a e_q^{ax}, \qquad D_q E_q^{ax} = a E_q^{qax},$$
$$D_q (x-a)_q^n = [n]_q (x-a)_q^{n-1}, \quad D_q (x-a)_q^{-n} = [-n]_q (x - q^n a)_q^{-(n+1)},$$
$$D_q (a-x)_q^n = -[n]_q (a - qx)_q^{n-1}, \quad D_q (a-x)_q^{-n} = [n]_q (a-x)_q^{-(n+1)}.$$

Definition 5 ([3]). The well-known Jackson q-integral of f is given by

$$\int_a^b f(\tau) d_q\tau = (1-q) \sum_{n=0}^{\infty} q^n \big(bf(q^n b) - af(q^n a)\big).$$

which is the Riemann integral when q approaches 1, inherently linear, and

$$\int_{u(x)}^{u(y)} f(u(\tau)) d_{q^{1/\beta}} u(\tau); \ u(\tau) = \delta\tau^{\beta}, (\delta, \beta \in \mathbb{R}).$$

Also, the q-integral by parts is presented as follows:

$$\int_x^y g(\tau) D_q f(\tau) d_q\tau = f(\tau) g(\tau)\big|_x^y - \int_x^y f(q\tau) D_q g(\tau) d_q\tau. \tag{4}$$

Definition 6 ([3]). The generalized q-integral is given by

$$\int_{-\infty}^{\infty} f(\tau) d_q\tau = (1-q) \sum_{n=0}^{\infty} q^n f(q^n).$$

Definition 7 ([3,4]). The q-gamma functions are given for $\alpha > 0$

$$\Gamma_q(\alpha) = \int_0^{[\infty]_q} \tau^{\alpha-1} E_q^{-q\tau} d_q\tau, \tag{5}$$

$$\gamma_q(\alpha) = \int_0^{\infty} \tau^{\alpha-1} e_q^{-\tau} d_q\tau. \tag{6}$$

Identities derived from the q-gamma functions can be given for $n \in \mathbb{N}$

$$\Gamma_q(\alpha + n) = [\alpha]_{n,q} \Gamma_q(\alpha), \quad \Gamma_q(n+1) = [n]_q!,$$

$$\gamma_q(\alpha + n) = q^{-\binom{n}{2} - \alpha n} [\alpha]_{n,q} \gamma_q(\alpha), \quad \gamma_q(n+1) = q^{-\binom{n+1}{2}} [n]_q!.$$

Definition 8 ([5]). The q-Laplace transforms are defined for $\tau \geq 0$ and $s > 0$

$$\mathbb{L}_q^I\big(\mathbb{f}_q(\tau), s\big) = \int_0^{[\infty]_q} f(\tau) E_q^{-qs\tau} d_q\tau, \ \ \mathbb{L}_q^{II}\big(\mathbb{f}_q(\tau), s\big) = \int_0^{\infty} f(\tau) e_q^{-s\tau} d_q\tau.$$

Identities derived from the q-Laplace transforms can be given for $\alpha > 0$

$$\mathbb{L}_q^I(\tau^{\alpha}, s) = \frac{\Gamma_q(\alpha+1)}{s^{\alpha+1}}, \ \ \mathbb{L}_q^{II}(\tau^{\alpha}, s) = \frac{\gamma_q(\alpha+1)}{s^{\alpha+1}}, \ \ (s > 0)$$

And for $n \in \mathbb{N}$

$$\mathbb{L}_q^I(\tau^n; s) = \frac{[n]_q!}{s^{n+1}}, \tag{7}$$

$$\mathbb{L}_q^{II}(\tau^n; s) = \frac{[n]_{1/q}!}{q^n s^{n+1}} = \frac{q^{-\binom{n+1}{2}} [n]_q!}{s^{n+1}}, \tag{8}$$

$$\mathbb{L}_q^I\big(\tau^n E_q^{a\tau}; s\big) = \sum_{r=0}^{\infty} \frac{q^{\binom{r}{2}} a^r}{[r]_q!} \frac{[r+n]_q!}{s^{r+n+1}}, \tag{9}$$

$$\mathbb{L}_q^{II}\big(\tau^n E_q^{a\tau}; s\big) = \frac{[n]_{\frac{1}{q}}!}{(s-a)_{\frac{1}{q}}^{n+1}} = \frac{q^{-\binom{n}{2}} [n]_q!}{(s - q^{-n}a)_q^{n+1}}, s > a, \tag{10}$$

$$\mathbb{L}_q^I\big(\tau^n e_q^{a\tau}; s\big) = \frac{[n]_q!}{(s-a)_q^{n+1}}, s > a, \tag{11}$$

$$\mathbb{L}_q^{II}\big(\tau^n e_q^{a\tau}; s\big) = \sum_{r=0}^{\infty} \frac{a^r}{[r]_q!} \frac{q^{-\binom{r+n+1}{2}} [r+n]_q!}{s^{r+n+1}} \tag{12}$$

Definition 9 ([15]). A random variable (RV) ξ is considered q-continuous if there exists a non-negative function $f_q(\tau)$ for $\tau \geq 0$ such that

$$P(a < \xi < b) = \int_a^b f_q(\tau) d_q\tau.$$

Definition 10 ([15]). The q-cumulative distribution function (q-CDF) of the non-negative q-continuous random random variable ξ is defined as

$$F_q(\tau) = P(\xi \le \tau) = \int_0^\tau f_q(u) d_q u\,, (\tau > 0),$$

satisfying the relation $P(\alpha < \xi \le \beta) = F_q(\beta) - F_q(\alpha)$. Then, we derive the q-probability density function (q-PDF) of ξ as follows:

$$f_q(\tau) = D_q F_q(\tau) = \frac{F_q(\tau) - F_q(q\tau)}{(1-q)\tau} = \frac{P(q\tau \le \xi \le \tau)}{(1-q)\tau}, (q \ne 1), \quad (13)$$

Definition 11 ([16]). Consider $\zeta = (\xi, \eta)$ as a bivariate q-continuous RV with $f_q^{\zeta}(\tau, \nu)$. The joint q-CDF of ζ can be given as

$$F_q^{\zeta}(\tau, \nu) = P(\xi \le \tau, \eta \le \nu) = \int_0^\tau \int_0^\nu f_q^{\zeta}(u, v)\, d_q u d_q v.$$

Thus, the joint and the marginal q-PDFs can be obtained, respectively, as follows:

$$f_q^{\zeta}(\tau, \nu) = \frac{\partial_q F_q^{\zeta}(\tau, \nu)}{\partial_q \tau \partial_q \nu}, \quad f_q^{\xi}(\tau) = \int_0^\infty F_q^{\zeta}(\tau, \nu) d_q \nu\,,$$

$$f_q^{\eta}(\nu) = \int_0^\infty F_q^{\zeta}(\tau, \nu) d_q \tau.$$

If $f_q^{\eta}(\nu) > 0$ for $\nu > 0$, the conditional q-PDF of ξ given η with $f_q^{\eta}(\nu)$ can be determined as

$$f_q^{\xi|\eta}(\tau|\,\nu) = P(q\tau \le \xi \le \tau | q\nu \le \eta \le \nu) = \frac{f_q^{\zeta}(\tau, \nu)}{f_q^{\eta}(\nu)}. \quad (14)$$

MODELING THE GENERALIZED LIFETIME DISTRIBUTION AND ITS *q*-ANALOGUES

Modeling the Generalized Lifetime Distribution

The ordinary generalized lifetime distribution ($GLD(n, \alpha)$; $\alpha > 0$, $n \in \mathbb{N}$) can be defined as a mixture of the gamma distributions with parameters (k, α) as follows:

$$f(\tau; \alpha) = \sum_{k=0}^{n} p_k g(\tau; k, \alpha) = a \sum_{k=0}^{n} \tau^k e^{-\alpha\tau}\,;$$
$$a = \frac{1}{\sum_{k=0}^{n} k!\, \alpha^{-(k+1)}}, \quad (15)$$

where the corresponding PDF of the gamma distribution as

$$g(\tau; k, \alpha) = \frac{\alpha^{k+1}}{\Gamma(k+1)} \tau^k e^{-\alpha\tau}, (\tau > 0, k \ge 0),$$

and $\sum_{k=0}^{n} p_k = 1$ such that the corresponding k-th mixing proportion is represented as follows

$$p_k = \frac{G(k, \alpha)}{T(\alpha)}, \quad G(k, \alpha) = \frac{\Gamma(k+1)}{\alpha^k}, \quad T(\alpha) = \sum_{k=0}^{n} G(k, \alpha).$$

Remark 1. For n = 0,1,2,3,4,5, $GLD(n, \alpha)$ corresponds to the exponential, Lindley, Sujatha, Amarendra, Devya and Shambu distributions in Table 1.

Modeling the *q*-Analogues of the Generalized Lifetime Distribution

This subsection focuses on constructing the q-analogues of the generalized lifetime distribution GLD. The q-PDF of the generalized lifetime q-distribution GLD_q can be given for $i = I, II$ byy

$$f_q^{i(m)}(\tau; [\alpha]_q) = \sum_{k=0}^{n} p_k^{i(k)} g^{C_m(k)}(\tau; k, [\alpha]_q)\,,$$
$$m = 1,2, \ldots, 2^{n+1}, n \in \mathbb{N}\backslash\{0\}, \quad (16)$$

and $f_q^{i(m)}(\tau;[\alpha]_q) = g^i(\tau; 0, [\alpha]_q)$, ($m = 1$, $i = I, II$) for $n = 0$, where $p_q^{i(k)}(i = I, II; k = 0,1,\ldots, n, n \in \mathbb{N})$ represents the i-th type of q-analogue of p_k for $[\alpha]_q > 0$

$$p_q^{i(k)} = \frac{G_q^i(k, [\alpha]_q)}{T_q^i(n, [\alpha]_q)}, \quad (17)$$

With $\sum_{k=0}^{n} p_q^{i(k)} = 1, T_q^i(n, [\alpha]_q) = \sum_{k=0}^{n} G_q^i(k, [\alpha]_q)$ and

$$G_q^i(k, [\alpha]_q) = \begin{cases} \dfrac{\Gamma_q(k+1)}{[\alpha]_q^k}, & i = I \\ \dfrac{\gamma_q(k+1)}{[\alpha]_q^k}, & i = II \end{cases} \quad (18)$$

Also, $g_q^{\,i}(i = I, II)$ represents the two types of q-PDFs of the gamma distribution, and $C_m(k)$ determines which type of q-PDF, $g_q^{\,I}$ or $g_q^{\,II}$, is used for each combination of m and k. Therefore, the q-analogues of the gamma distribution are needed. From [20], the first type of q-PDF of the gamma q-distribution can be obtained:

$$g_q^I(\tau; k, [\alpha]_q) = \frac{[\alpha]_q^{k+1}}{\Gamma_q(k+1)} \tau^k E_q^{-q[\alpha]_q\tau} \quad (19)$$

for $k \ge 0$ and $[\alpha]_q > 0$ with $E_q^{\,\tau}$ and Γ_q (.) in Eqs.(2) and (5), respectively. The first type of gamma q-distribution is insufficient to model all instances of the generalized lifetime q-distribution by itself. To address this, a new q-distribution with the following q-PDF is introduced:

$$g_q^{II}(\tau; k, [\alpha]_q) = \frac{[\alpha]_q^{k+1}}{\gamma_q(k+1)} \tau^k e_q^{-[\alpha]_q\tau}. \quad (20)$$

for $k \ge 0$ and $[\alpha]_q > 0$ with $e_q^{\,\tau}$ and γ_q(.) in Eqs.(3) and (6), respectively. It can be easily verified that the function given

in Eq.(20) satisfies the necessary conditions to qualify as a q-PDF and converges to the PDF of the classical gamma distribution as $q \to 1$. This property is the basis for naming it the second type of gamma q-distribution.

Each function $f_q^{i(m)}$ given in Eq.(16) satisfies the necessary conditions to qualify as a q-PDF, and converges to the generalized lifetime distribution GLD as $q \to 1$. To be able to analyze the distributional and statistical properties of the generalized lifetime q-distribution GLD_q, let $f_q^{i(m)} = f_q^{i}$ and $C_m(k) = i$ for some values of m (m = 1, 2,…, 2^{n+1}, $n \in \mathbb{N}$) given in Eq.(16). Thus, this study focuses exclusively on the following q-PDFs:

$$f_q^i(\tau; [\alpha]_q) = \sum_{k=0}^{n} p_k^{i(k)} g^i(\tau; k, [\alpha]_q), \tag{21}$$

where $p_q^{i(k)}$ and $g^i(\tau; k, [\alpha]_q)$ are as given in Eqs.(16)-(20), respectively.

Definition 12. If a random variable ξ is characterized by the following q-PDFs, it is said to follow the generalized lifetime q-distribution GLD_q

$$f_q^I(\tau; [\alpha]_q) = a_q^I \sum_{k=0}^{n} \tau^k E_q^{-q[\alpha]_q \tau} 1_{[0,[\infty]_q]}(\tau), \tag{22}$$

$$f_q^{II}(\tau; [\alpha]_q) = a_q^{II} \sum_{k=0}^{n} \tau^k e^{-[\alpha]_q \tau} 1_{[0,\infty)}(\tau), \tag{23}$$

where $1_A(\tau)$ represents the characteristic function for $A \subseteq \mathbb{R}$, and

$$a_q^I = \frac{1}{\sum_{k=0}^{n} [k]_q! \, [\alpha]_q^{-(k+1)}}, \quad a_q^{II} = \frac{1}{\sum_{k=0}^{n} q^{-\binom{k+1}{2}} [k]_q! \, [\alpha]_q^{-(k+1)}}.$$

As q tends 1, the q-PDF $f_q^i(i = I, II)$ converges to the PDF f as given in Eq.(15). The first and second type of the generalized lifetime q-distribution GLD_q are denoted as GLD_q^I and GLD_q^{II}, respectively.

Proposition 1. The q-PDF $f_q^i(i = I, II)$ satisfies non-negativity and normality conditions.

Proof. Clearly, the q-PDF $f_q^i(i = I, II) \geq 0$ for all $\tau \geq 0$. Using Eqs.(7) and (8), respectively

$$\int_0^{[\infty]_q} f_q^I(\tau; [\alpha]_q) d_q\tau = a_q^I \int_0^{1/(1-q)} \sum_{k=0}^{n} \tau^k E_q^{-q[\alpha]_q \tau} d_q\tau$$

$$= a_q^I \sum_{k=0}^{n} \mathbb{L}_q^I(\tau^k; [\alpha]_q) = \frac{1}{\sum_{k=0}^{n} [k]_q! \, [\alpha]_q^{-(k+1)}} \sum_{k=0}^{n} \frac{[k]_q!}{[\alpha]^{k+1}} = 1,$$

$$\int_0^{[\infty]_q} f_q^{II}(\tau; [\alpha]_q) d_q\tau = a_q^{II} \int_0^{\infty} \sum_{k=0}^{n} \tau^k e_q^{-q[\alpha]_q \tau} d_q\tau$$

$$= a_q^{II} \sum_{k=0}^{n} \mathbb{L}_q^{II}(\tau^k; [\alpha]_q) = \frac{1}{\sum_{k=0}^{n} q^{-\binom{k+1}{2}} [k]_q! \, [\alpha]_q^{-(k+1)}} \sum_{k=0}^{n} \frac{q^{-\binom{k+1}{2}} [k]_q!}{[\alpha]^{k+1}} = 1.$$

Theorem 1. Let $\xi \backsim GLD_q(n, [\alpha]_q)$ be for $[\alpha]_q > 0$. The q-CDFs of ξ are given as two kinds by

$$F_q^I(\tau; [\alpha]_q) = \begin{cases} 0, & \tau < 0 \\ 1 - a_q^I \sum_{k=0}^{n} \sum_{j=0}^{k} [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j} E_q^{-[\alpha]_q \tau}, & \tau \in [0, [\infty]_q] \\ 1, & \tau > [\infty]_q \end{cases} \tag{24}$$

$$F_q^{II}(\tau; [\alpha]_q) = \begin{cases} 0, & \tau < 0 \\ 1 - a_q^{II} \sum_{k=0}^{n} \sum_{j=0}^{k} q^{-\binom{k+1}{2}} [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j} e_q^{-q^j [\alpha]_q \tau}, & \tau \in [0, \infty) \\ 1, & \tau > \infty \end{cases} \tag{25}$$

where $a_q^I (i = I, II)$ is as given in Definition 12, and

$$F_q(\tau; \alpha) = \lim_{q \to 1} F_q^i(\tau; [\alpha]_q) = \begin{cases} 0, & \tau < 0 \\ 1 - a \sum_{k=0}^{n} \sum_{j=0}^{k} (k)_j \alpha^{-(j+1)} \tau^{k-j} e_q^{-\alpha\tau}, & \tau \in [0, \infty) \\ 1, & \tau > \infty \end{cases}$$

Proof. Applying the q-integration by parts as in Eq.(4) and using the rules of the q-derivation as in Definition 4, respectively, we have

$$\begin{aligned} F_q^I(\tau; [\alpha]_q) &= 1 - \int_{\tau}^{[\infty]_q} f_q^I(u; [\alpha]_q) d_q u \\ &= 1 - a_q^I \int_{\tau}^{[\alpha]_q/(1-q)} \sum_{k=0}^{n} u^k E_q^{-q[\alpha]_q u} d_q u \\ &= 1 - a_q^I \sum_{0}^{n} \frac{1}{[\alpha]_q^{k+1}} \int_{[\alpha]_q \tau}^{[\alpha]_q/(1-q)} z^k E_q^{-qz} d_q z \\ &= 1 + a_q^I \sum_{0}^{n} \frac{1}{[\alpha]_q^{k+1}} \sum_{j=0}^{k} [k]_{j,q} z^{k-j} E_q^{-z} \Big|_{[\alpha]_q \tau}^{[\alpha]_q/(1-q)} \\ &= 1 - a_q^I \sum_{k=0}^{n} \sum_{j=0}^{k} [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j} E_q^{-[\alpha]_q \tau}, \end{aligned}$$

$$\begin{aligned} F_q^{II}(\tau; [\alpha]_q) &= 1 - \int_{\tau}^{\infty} f_q^{II}(u; [\alpha]_q) d_q u \\ &= 1 - a_q^{II} \int_{\tau}^{\infty} \sum_{k=0}^{n} u^k e_q^{-[\alpha]_q u} d_q u \\ &= 1 - a_q^{II} \sum_{0}^{n} \frac{1}{[\alpha]_q^{k+1}} \int_{[\alpha]_q \tau}^{\infty} z^k e_q^{-z} d_q z \\ &= 1 + a_q^{II} \sum_{0}^{n} \frac{1}{[\alpha]_q^{k+1}} \sum_{j=0}^{k} q^{-\binom{j+1}{2}} [k]_{j,q} z^{k-j} e_q^{-q^j z} \Big|_{[\alpha]_q \tau}^{\infty} \\ &= 1 - a_q^{II} \sum_{k=0}^{n} \sum_{j=0}^{k} q^{-\binom{j+1}{2}} [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j} e_q^{-q^j [\alpha]_q \tau}, \end{aligned}$$

Remark 2. For n = 0,1,2,3,4,5, the generalized lifetime q-distribution GLD_q^i (n, α) ($i = I, II$) corresponds to the q-analogue of the exponential, Lindley, Sujatha, Amarendra, Devya, and Shambu distributions, and converges to their classical forms as $q \to 1$ in Table 2.

To explore the behavior of the generalized lifetime q-distribution GLD_q^i (n, α) ($i = I, II$) under varying parameter conditions, we present the q-PDFs and q-CDFs for a range of randomly chosen values of α, n and q. Figures 1–8 illustrate how different parameter combinations influence the shape and variability of the distribution. Notably, as $q \to 1$, these curves converge to the PDF and CDF of the ordinary

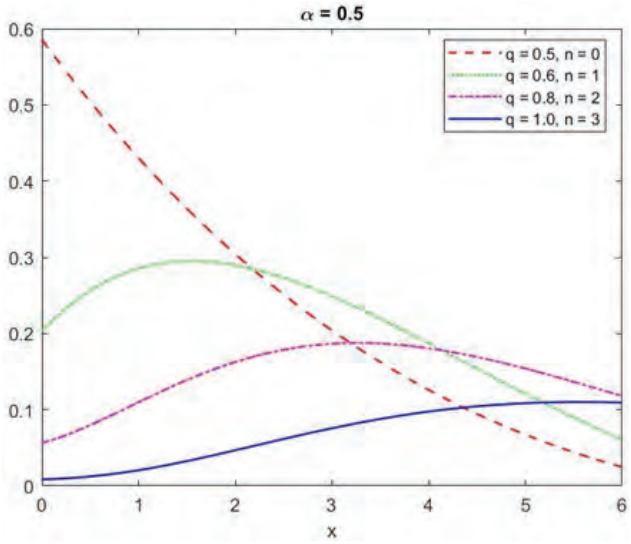


Figure 1. The q-PDF curves of $GLD_q^{\ I}$ for $\alpha = 0.5$.

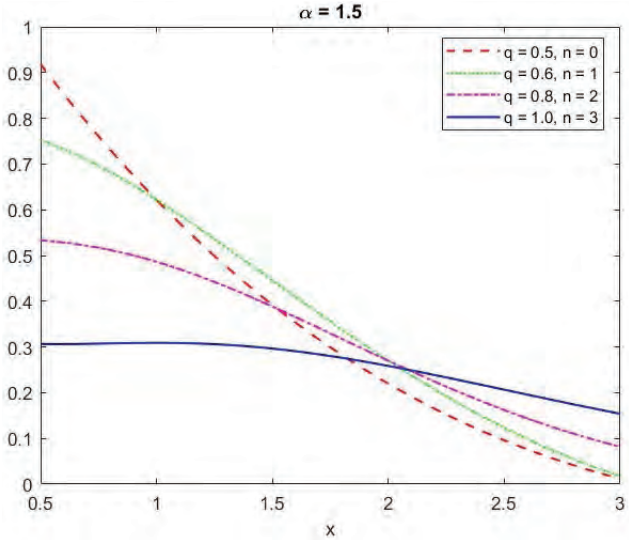


Figure 2. The q-PDF curves of $GLD_q^{\ I}$ for $\alpha = 1.5$.

generalized lifetime distribution GLD, demonstrating the limiting behavior of the q-generalization.

q-RELIABILITY FUNCTIONS

This section delves into the q-reliability functions for a q-continuous non-negative random variable and provides them for $\xi \backsim GLD_q$.

Lemma 1. Let ξ be a q-continuous non-negative random variable. The following q-equalities hold for $\tau \geq 0$

$$\frac{P(q\tau \leq \xi \leq \tau | \xi \geq q\tau)}{(1-q)\tau} = \frac{f_q(\tau)}{1-F_q(q\tau)};$$

$$\mathbb{E}_q(\xi - \tau | \xi > \tau) = \frac{1}{1-F_q(\tau)} \int_\tau^\infty \Big(1 - F_q(qu)\Big)\, d_q u$$

where f_q and F_q represent the q-PDF and the q-CDF of ξ, respectively.

Proof. Let $F_q(\tau) = P(\xi \leq \tau)$ be the q-CDF of ξ for $\tau \geq 0$ Then, we have

$$P(\xi > \tau) = 1 - F_q(\tau).$$

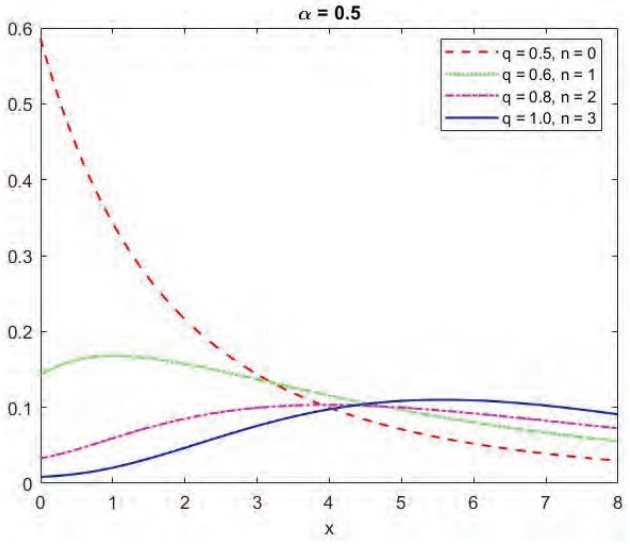


Figure 3. The q-PDF curves of $GLD_q^{\ II}$ for $\alpha = 0.5$.

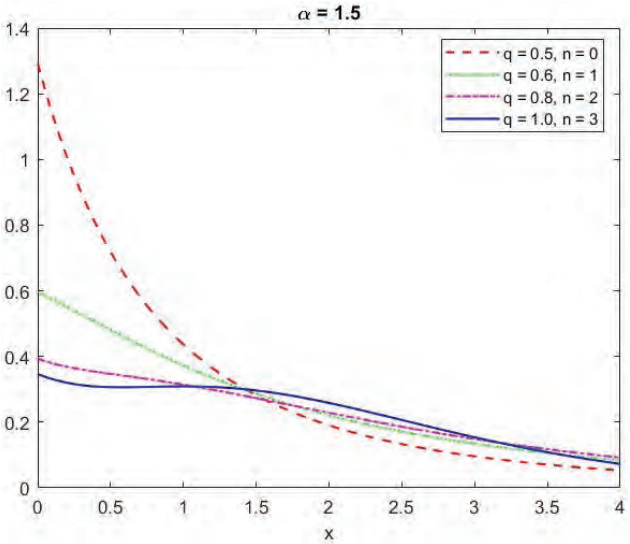


Figure 4. The q-PDF curves of $GLD_q^{\ II}$ for $\alpha = 1.5$.

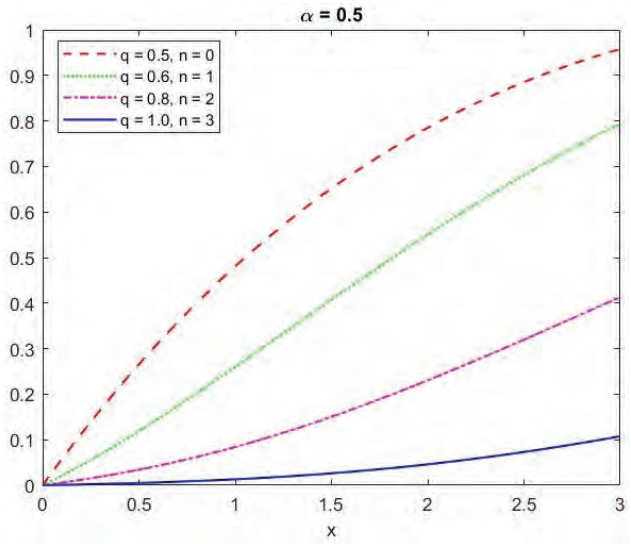


Figure 5. The q-CDF curves of $GLD_q^{\ I}$ for $\alpha = 0.5$.

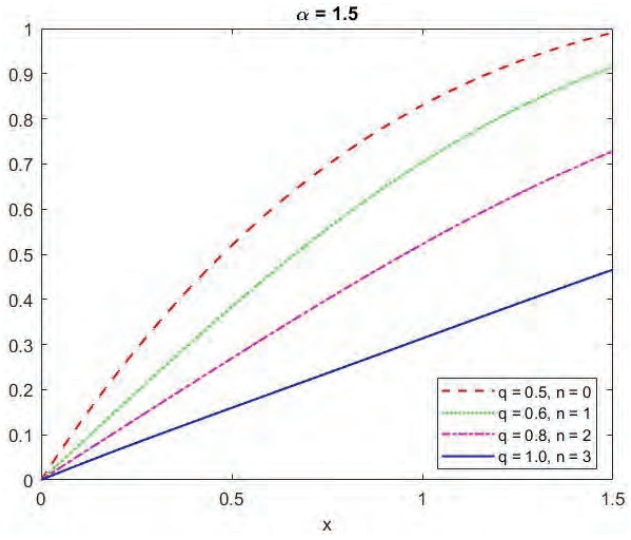


Figure 6. The q-CDF curves of $GLD_q^{\ I}$ for $\alpha = 1.5$.

Taking into consideration Eqs.(13) and (14), we have

$$\frac{P(q\tau \le \xi \le \tau | \xi \ge q\tau)}{(1-q)\tau} = \frac{P(q\tau \le \xi \le \tau, \xi \ge q\tau)}{(1-q)\tau P(\xi \ge q\tau)}$$
$$= \frac{P(q\tau \le \xi \le \tau)}{(1-q)\tau P(\xi \ge q\tau)} = \frac{f_q(\tau)}{1-F_q(q\tau)}$$

Futhermore, by using the q-integration by parts given in Eq.(4), we get

$$\mathbb{E}_q(\xi - \tau | \xi > \tau) = \int_\tau^\infty \frac{(u-\tau) f_q(u) d_q u}{P(\xi > \tau)} = \frac{1}{1-F_q(\tau)} \Bigg((u-\tau)\left(1-F_q(u)\right)|_\tau^\infty + \int_\tau^\infty \left(1-F_q(qu)\right) d_q u \Bigg) = \frac{1}{1-F_q(\tau)} \int_\tau^\infty \left(1-F_q(qu)\right) d_q u.$$

Based on Lemma 1, the following definition is proposed:

Definition 13. Let ξ be a q-continuous non-negative random variable. The q-RFs of ξ are defined for $\tau \ge 0$ as follows

(1) the q-survival function (q-CCDF)

$$\mathbb{S}_q(\tau) = P(\xi > \tau) = 1 - F_q(\tau),$$

(2) the q-hazard rate function (q-HRF)

$$h_q(\tau) = \frac{P(q\tau \le \xi \le \tau | \xi \ge q\tau)}{(1-q)\tau} = \frac{f_q(\tau)}{\mathbb{S}_q(q\tau)}, \qquad \left(\mathbb{S}_q(q\tau) \ne 0\right),$$

(3) the q-mean residual life function (q-MRLF)

$$mrl_q(\tau) = \mathbb{E}_q(\xi - \tau | \xi > \tau) = \frac{1}{\mathbb{S}_q(\tau)} \int_\tau^\infty \mathbb{S}_q(qu) d_q u, \ \left(\mathbb{S}_q(q\tau) \ne 0,\right)$$

where f_q and F_q represent the q-PDF and the q-CDF of ξ, respectively. As q tends 1, the q-RFs converge to the classical forms of RFs, represented by $\mathbb{S}(\tau\,;\alpha)$, $h(\tau\,;\alpha)$, and $mrl(\tau\,;\alpha)$, respectively.

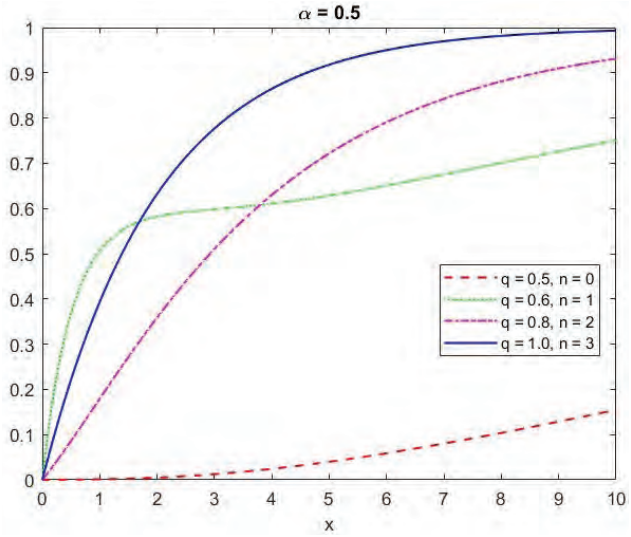


Figure 7. The q-CDF curves of $GLD_q^{\ II}$ for $\alpha = 0.5$.

Theorem 2. Let $\xi \backsim GLD_q\ (n, [\alpha]_q)$ where $[\alpha]_q > 0$. The q-RFs of ξ are given as follows

$$\mathbb{S}_q^I(\tau; [\alpha]_q) = a_q^I \sum_{k=0}^n \sum_{j=0}^k [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j} E_q^{-[\alpha]_q \tau};$$

$$h_q^I(\tau; [\alpha]_q) = \frac{\sum_{k=0}^n \tau^k}{\sum_{k=0}^n \sum_{j=0}^k [k]_{j,q} [\alpha]_q^{-(j+1)} (q\tau)^{k-j}};$$

$$mrl_q^I(\tau; [\alpha]_q) = \frac{1}{\sum_{k=0}^n \sum_{j=0}^k [k]_{j,q} [\alpha]_q^{-(j+1)} \tau^{k-j}} \times \sum_{k=0}^n \sum_{j=0}^k \sum_{m=0}^{k-j} [k]_{j,q} [k-j]_{m,q} q^{k-j} [\alpha]_q^{-(j+m+2)} \tau^{k-j-m};$$

$$\mathbb{S}_q^{II}(\tau;[\alpha]_q) = a_q^{II}\sum_{k=0}^{n}\sum_{j=0}^{k} q^{-\binom{j+1}{2}}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}e_q^{-q^j[\alpha]_q\tau};$$

$$h_q^{II}(\tau;[\alpha]_q) = \frac{\sum_{k=0}^{n}\tau^k e_q^{-[\alpha]_q\tau}}{\sum_{k=0}^{n}\sum_{j=0}^{k} q^{-\binom{j+1}{2}}[k]_{j,q}[\alpha]_q^{-(j+1)}(q\tau)^{k-j}e_q^{-q^{j+1}[\alpha]_q\tau}};$$

$$mrl_q^{I}(\tau;[\alpha]_q) = \frac{1}{\sum_{k=0}^{n}\sum_{j=0}^{k} q^{k-\binom{j+1}{2}}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}\,e_q^{-q^j[\alpha]_q\tau}} \times\sum_{k=0}^{n}\sum_{j=0}^{k}\sum_{m=0}^{k-j} q^{k-j-\binom{j+1}{2}-\binom{m+1}{2}}[k]_{j,q}[k-j]_{m,q} [\alpha]_q^{-(j+m+2)}\tau^{k-j-m}e_q^{-q^{j+m}[\alpha]_q\tau};$$

where $a_q^{I}(i = I, II)$ is as given in Definition 12

Proof. From Definition 13, we can simply obtain the q-CCDFs and the q-HRFs of ξ. The first type of the q-MRLF of ξ is obtained as follows:

$$mrl_q^{I}(\tau;[\alpha]_q) = \frac{1}{\mathbb{S}_q(\tau;[\alpha]_q)}\int_{\tau}^{[\infty]_q}\mathbb{S}_q(qu;[\alpha]_q)d_qu$$

$$= \frac{\int_{\tau}^{[\infty]_q}\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+1)}(qu)^{k-j}E_q^{-q[\alpha]_qu}\,d_qu}{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}E_q^{-[\alpha]_q\tau}}$$

$$= \frac{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+2)}q^{k-j}\int_{[\alpha]_q\tau}^{\frac{[\alpha]_q}{1-q}}z^{k-j}E_q^{-qz}d_qz}{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}E_q^{-[\alpha]_q\tau}}$$

$$= \frac{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+2)}q^{k-j}\sum_{m=0}^{k-j}[k-j]_{m,q}z^{k-j-m}E_q^{-z}\big|_{[\alpha]_q\tau}^{[\alpha]_q/(1-q)}}{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}E_q^{-[\alpha]_q\tau}}$$

$$= \frac{1}{\sum_{k=0}^{n}\sum_{j=0}^{k}[k]_{j,q}[\alpha]_q^{-(j+1)}\tau^{k-j}} \times\sum_{k=0}^{n}\sum_{j=0}^{k}\sum_{m=0}^{k-j}[k]_{j,q}[k-j]_{m,q}q^{k-j}[\alpha]_q^{-(j+m+2)}\tau^{k-j-m}$$

Likewise, the second type of the q-MRLF of ξ can be easily derived.

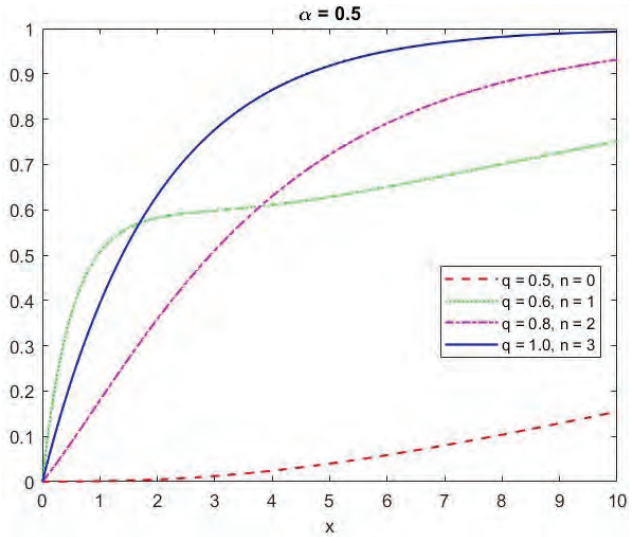


Figure 8. The q-CDF curves of GLD_q^{II} for $\alpha = 1.5$.

Corollary 1. As q tends 1, the q-RFs given in Theorem 2 converge to

$$\mathbb{S}(\tau;\alpha) = \frac{1}{\sum_{k=0}^{n}k!\,\alpha^{-(k+1)}}\sum_{k=0}^{n}\sum_{j=0}^{k}(k)_j\alpha^{-(j+1)}\tau^{k-j}e^{-\alpha\tau};$$

$$h(\tau;\alpha) = \frac{\sum_{k=0}^{n}\tau^k\,e^{-\alpha\tau}}{\sum_{k=0}^{n}\sum_{j=0}^{k}(k)_j\alpha^{-(j+1)}e^{-\alpha\tau}};$$

$$mrl(\tau;\alpha) = \frac{1}{\sum_{k=0}^{n}\sum_{j=0}^{k}(k)_j\alpha^{-(j+1)}e^{-\alpha\tau}} \times\sum_{k=0}^{n}\sum_{j=0}^{k}\sum_{m=0}^{k-j}(k)_j(k-j)_mq^{k-j}\alpha^{-(j+m+2)}\tau^{k-j-m}.$$

Corollary 2. Let the conditions of Theorem 2 be met. Then, the equalities $h_q^{I} = f_q^{I}$, $h_q^{II} = f_q^{II}$, $mrl_q^{I} = \mu_q^{I}$, $mrl_q^{II} = \mu_q^{II}$ hold for $\tau = 0$

q-MOMENTS AND *q*-MOMENT GENERATING FUNCTIONS

This section consists of two subsections.

q-Moments and Related Measures.

This subsection explores the q-moments, q-mean, q-variance and the q-central moment of $\xi \backsim GLD_q$.

Definition 14 ([15]). If the absolute q-moment of ξ^r exists and is finite, then the r-th q-moment of ξ is defined as

$$\mu_q^{(r)} = \mathbb{E}_q(\xi^r) = \int_0^{\infty}\tau^r f_q(\tau)d_q\tau \tag{26}$$

for $\tau \geq 0$, and $\lim_{q\to 1}\mu_q^{(r)} = \mu^{(r)}$ for $r = 1, 2, \ldots$ where $\mu^{(r)}$ represents the r-tth moment of ξ.

In particular, as q tends to 1, the q-mean $E_q(\xi) = \mu_q$ and the q-variance $\mathbb{V}_q(\xi) = \mu_q^{(2)} - \mu_q^{2}$ converge e to the mean μ and the variance $V(\xi)$, respectively.

Theorem 3. The r-th q-moment $\mu_q^{(r)}$ $(r = 1, 2, \ldots)$ of $\xi \backsim GLD_q$ is given in two types as follows:

$$\mu_q^{I(r)} = a_q^{I}\sum_{k=0}^{n}\frac{[k+r]_q!}{[\alpha]_q^{k+r+1}}; \tag{27}$$

$$\mu_q^{II(r)} = a_q^{II}\sum_{k=0}^{n}\frac{q^{-\binom{k+r+1}{2}}[k+r]_q!}{[\alpha]_q^{k+r+1}}; \tag{28}$$

where $a_q^{I}(i = I, II)$ is as given in Definition 12, and

$$\mu^{(r)} = \lim_{q\to 1}\mu_q^{i(r)} = \frac{1}{\sum_{k=0}^{n}k!\,\alpha^{-(k+1)}}\sum_{k=0}^{n}\frac{(k+r)!}{\alpha^{k+r+1}}.$$

Proof. Initially, assume that $\xi \backsim GLD_q$ Through the substitution of Eq.(22) into Eq.(26), we obtain:

$$\mu_q^{I(r)} = \int_0^{[\infty]_q} \tau^r f_q^I(\tau;[\alpha]_q) d_q\tau = \int_0^{\frac{1}{1-q}} a_q^I \sum_{k=0}^{n} \tau^{k+r} E_q^{-q[\alpha]_q\tau}\, d_q\tau$$
$$= a_q^I \sum_{k=0}^{n} \mathbb{L}_q^I(\tau^{k+r};[\alpha]_q).$$

Then, $\mu_q^{I(r)}$ is verified using Eq.(7). Similarly, $\mu_q^{II(r)}$ is readily deduced by applying Eq.(23) into Eq.(26), and then using Eq.(8).

Corollary 3. Given the conditions of Theorem 3, one has

$$\mu_q^I = a_q^I \sum_{k=0}^{n} \frac{[k+1]_q!}{[\alpha]_q^{k+2}};\ \ \mu_q^{II(r)} = a_q^{II} \sum_{k=0}^{n} \frac{q^{-\binom{k+2}{2}}[k+1]_q!}{[\alpha]_q^{k+2}},$$

$$\mathbb{V}_q^I(\xi) = a_q^I \sum_{k=0}^{n} \frac{[k+2]_q!}{[\alpha]_q^{k+3}} - \left(a_q^I \sum_{k=0}^{n} \frac{[k+1]_q!}{[\alpha]_q^{k+2}}\right)^2,$$

$$\mathbb{V}_q^{II}(\xi) = a_q^{II} \sum_{k=0}^{n} \frac{q^{-\binom{k+3}{2}}[k+2]_q!}{[\alpha]_q^{k+3}} - \left(a_q^{II} \sum_{k=0}^{n} \frac{q^{-\binom{k+2}{2}}[k+1]_q!}{[\alpha]_q^{k+2}}\right)^2,$$

where $a_q^I (i = I, II)$ is as given in Definition 12, and as q tends 1,

$$\mu = a \sum_{k=0}^{n} \frac{(k+1)!}{\alpha^{k+2}};\ \ V(\xi) = a \sum_{k=0}^{n} \frac{(k+2)!}{\alpha^{k+3}} - \left(a \sum_{k=0}^{n} \frac{(k+1)!}{\alpha^{k+2}}\right)^2,$$

where a is as given in Eq.(15).

Lemma 2. The r-th q-central moment of a q-continuous random variable is given as follows:

$$m_q^{(r)} = \sum_{k=0}^{r} \begin{bmatrix} r \\ s \end{bmatrix}_q q^{\binom{s}{2}} (-1)^s \mu_q^s \mu_q^{r-s}, \quad s \le r \tag{29}$$

for $r = 1, 2, \ldots$ and $\lim_{q\to 1} m_q^{(r)} = m^{(r)}$ for $r = 1, 2, \ldots$, where $m^{(r)}$ represents the r-th central moment of ξ.

Proof. Using the q-Gauss binomial formula given in Eq.(1) and Eq.(26), we get

$$m_q^{(r)} = \mathbb{E}_q(\xi - \mu_q)_q^r = \int_0^{\infty} (\tau - \mu_q)_q^r f_q(\tau) d_q(\tau)$$
$$= \int_0^{\infty} \sum_{k=0}^{r} \begin{bmatrix} r \\ s \end{bmatrix}_q q^{\binom{s}{2}} (-1)^s \mu_q^s \tau^{r-s}\ f_q(\tau) d_q(\tau)$$
$$= \sum_{k=0}^{r} \begin{bmatrix} r \\ s \end{bmatrix}_q q^{\binom{s}{2}} (-1)^s \mu_q^s \mu_q^{r-s}.$$

Theorem 4. The r-th q-central moment $m^{(r)}$ $(r = 1, 2, \ldots)$ of a q-continuous random variable $\xi \backsim GLD_q$ is derived in two types as follows:

$$m_q^{I(r)} = \sum_{k=0}^{r} \begin{bmatrix} r \\ s \end{bmatrix}_q q^{\binom{s}{2}} (-1)^s \left(a_q^I\right)^{s+1} \left(\sum_{k=0}^{n} \frac{[k+1]_q!}{[\alpha]_q^{k+2}}\right)^s \sum_{k=0}^{n} \frac{[k+r-s]_q!}{[\alpha]_q^{k+r-s+1}};$$

$$m_q^{II(r)} = \sum_{k=0}^{r} \begin{bmatrix} r \\ s \end{bmatrix}_q q^{\binom{s}{2}} (-1)^s \left(a_q^{II}\right)^{s+1} \left(\sum_{k=0}^{n} \frac{q^{-\binom{k+2}{2}}[k+1]_q!}{[\alpha]_q^{k+2}}\right)^s \sum_{k=0}^{n} \frac{q^{\binom{k+r-s+1}{2}}[k+r-s]_q!}{[\alpha]_q^{k+r-s+1}};$$

where $a_q^I (i = I, II)$ is as given in Definition 12, and

$$m^{(r)} = \lim_{q\to 1} m_q^{i(r)} = \sum_{k=0}^{r} \binom{r}{s} (-1)^s a^{s+1} \left(\sum_{k=0}^{n} \frac{(k+1)!}{\alpha^{k+2}}\right)^s \sum_{k=0}^{n} \frac{(k+r-s)!}{\alpha^{k+r-s+1}},$$

where a is as given in Eq.(15).

Proof. The proof is completed by inserting Eqs.(27) and (28) into Eq.(29).

***q*-Moment Generating Functions.**

In this subsection, the q-moment generating function (q-MGF) is introduced for any q-continuous non-negative random variable ξ, and it is derived for $\xi \backsim GLD_q$.

Definition 15. The q-MGF of any q-continuous non-negative random variable ξ is expressed in two distinct forms as follows:

$$\mathbb{M}_q^I(t) = \mathbb{E}_q\left(E_q^{qt\xi}\right) = \int_0^{\infty} E_q^{qt\tau} f_q(\tau) d_q\tau, \tag{30}$$

$$\mathbb{M}_q^{II}(t) = \mathbb{E}_q\left(e_q^{t\xi}\right) = \int_0^{\infty} e_q^{t\tau} f_q(\tau) d_q\tau \tag{31}$$

for those values of t for which the q-expectations $E_q(E_q^{qt\xi})$ and $E_q(e_q^{t\xi})$ are well defined. If these q-expectations do not converge for values of t close to 0, the q-MGF is considered undefined.

Also, $M(t) = \lim_{q\to 1} \mathbb{M}_q^i(t)$ $(i = I, II)$, where $M(t)$ represents the MGF of ξ.

Proposition 2. Let the q-MGF of the q-continuous non-negative random variable ξ exist. Some of its key properties are given as follows:

(1) Relationship between the q-Laplace transform and the q-MGF:

$$\mathbb{M}_q^I(t) = \mathbb{L}_q^I\left(f_q(\tau;-t)\right);\ \ \mathbb{M}_q^{II}(t) = \mathbb{L}_q^{II}\left(f_q(\tau;-t)\right),$$

(2) Generating q-moments using the q-MGF

$$D_q^r\left(\mathbb{M}_q^I(t)\right)\Big|_{t=0} = q^{\binom{r+1}{2}} \mu_q^{(r)},\ D_q^r\left(\mathbb{M}_q^{II}(t)\right)\Big|_{t=0} = \mu_q^{(r)}, (r = 1,2,\ldots).$$

Proof. Considering Definitions 2.8 and 5.1, the relationship between the q-Laplace transform and the q-MGF can be easily derived. Utilizing Definition 4, we have

$$D_q^r\left(\mathbb{M}_q^I(t)\right) = \mathbb{E}_q\left(q^{\binom{r+1}{2}} \xi^r E_q^{qt\xi}\right) \Rightarrow D_q^r\left(\mathbb{M}_q^I(t)\right)\Big|_{t=0} = q^{\binom{r+1}{2}} \mu_q^{(r)},$$

$$D_q^r\left(\mathbb{M}_q^{II}(t)\right) = \mathbb{E}_q\left(\xi^r e_q^{t\xi}\right) \Rightarrow D_q^r\left(\mathbb{M}_q^{II}(t)\right)\Big|_{t=0} = \mu_q^{(r)}.$$

Thus, the proof is complete.

Theorem 5. Let $\xi \backsim GLD_q$ $(n, [\alpha]_q)$ with $[\alpha]_q > 0$. The q-MGFs of ξ are given as follows:

$$\mathbb{M}_q^{I(I)}(t) = a_q^I \sum_{r=0}^{\infty} \frac{q^{\binom{r+1}{2}} t^r}{[r]_q!} \sum_{k=0}^{n} \frac{[k+r]_q!}{[\alpha]_q^{k+r+1}} \tag{32}$$

$$\mathbb{M}_q^{I(II)}(t) = a_q^{II} \sum_{r=0}^{\infty} \frac{q^{\binom{r+1}{2}} t^r}{[r]_q!} \sum_{k=0}^{n} \frac{q^{-\binom{r+k+1}{2}} [k+r]_q!}{[\alpha]_q^{k+r+1}} \quad (33)$$

$$\mathbb{M}_q^{II(I)}(t) = a_q^{I} \sum_{r=0}^{\infty} \frac{t^r}{[r]_q!} \sum_{k=0}^{n} \frac{[k+r]_q!}{[\alpha]_q^{k+r+1}} \quad (34)$$

$$\mathbb{M}_q^{II(II)}(t) = a_q^{II} \sum_{r=0}^{\infty} \frac{t^r}{[r]_q!} \sum_{k=0}^{n} \frac{q^{-\binom{r+k+1}{2}} [k+r]_q!}{[\alpha]_q^{k+r+1}} \quad (35)$$

where $\mathbb{M}_q^{i(j)}(t)$ ($i, j = I, II$) denotes the i-th tyoe q-MGF of the j-th kind GLD_q, and $a_q^{I}(i = I, II)$ is as given in Definition 12, and

$$M(t) = \lim_{q \to 1} \mathbb{M}_q^{i(j)}(t) = \frac{1}{\sum_{k=0}^{n} k!\, \alpha^{-(k+1)}} \sum_{k=0}^{n} \frac{k!}{\alpha^{k+1}}, \quad t < \alpha.$$

Proof. By inserting Eq.(22) into Eq.(30), we have

$$\mathbb{M}_q^{I(I)}(t) = \mathbb{E}_q\left(E_q^{qt\xi}\right) = \int_0^{[\infty]_q} E_q^{qt\tau} f_q^{I}(\tau; [\alpha]_q) d_q\tau$$

$$= a_q^{I} \sum_{k=0}^{n} \mathbb{L}_q^{I}(\tau^k E_q^{qt\tau}; [\alpha]_q).$$

Thus, by using Eq.(9), we obtain Eq.(32). By substituting Eq.(23) into Eq.(30), and then using the Taylor expansion of $E_q^{qt\tau}$ as given in Eq.(2), we have

$$\mathbb{M}_q^{I(II)}(t) = \mathbb{E}_q\left(E_q^{qt\xi}\right) = \int_0^{\infty} E_q^{qt\tau} f_q^{II}(\tau; [\alpha]_q) d_q\tau$$

$$= a_q^{II} \int_0^{\infty} \sum_{r=0}^{\infty} \frac{q^{\binom{r}{2}} (t\tau)^r}{[r]_q!} \sum_{k=0}^{n} \tau^k e_q^{-[\alpha]_q\tau}\, d_q\tau$$

$$= a_q^{II} \sum_{r=0}^{\infty} \frac{q^{\binom{r}{2}} t^r}{[r]_q!} \mathbb{L}_q^{II}(\tau^{k+r}; [\alpha]_q)\, d_q\tau.$$

Thus, by using Eq.(8), we obtain Eq.(33). By applying Eq.(22) into Eq.(31), then using the Taylor expansion of $e_q^{t\tau}$ as given in Eq.(3) and using Eq.(7), we get Eq.(34). Finally, by inserting Eq.(23) into Eq.(31), and then using Eq.(12), we have Eq.(35).

Remark 3. As per Theorem 5, the q-moment $\mu_q^{i(r)}$ ($i = I, II$; $r = 1, 2, \ldots$) in Eqs.(27) and (28) can be calculated by extracting the coefficient of $q^{\binom{r}{2}}\, t^r/[r]_q!$ from $\mathbb{M}_q^{I(i)}$, and $t^r/[r]_q!$ from $\mathbb{M}_q^{II(i)}$.

Proposition 3. Assuming that the conditions outlined in Theorem 5 are met, alternative forms of q-MGFs are expressed for $t \neq [\alpha]_q$ as follows:

$$\mathbb{M}_q^{I(II)}(t) = a_q^{II} \sum_{r=0}^{n} \frac{[k]_{1/q}!}{\left([\alpha]_q - t\right)_{1/q}^{k+1}}, \quad (36)$$

$$\mathbb{M}_q^{II(I)}(t) = a_q^{I} \sum_{r=0}^{n} \frac{[k]_q!}{\left([\alpha]_q - t\right)_q^{k+1}} \quad (37)$$

Proof. By substituting Eq.(22) into Eq.(30), and then using Eq.(10), we have Eq.(36). Also, by applying Eq.(23) into Eq.(31), and then using Eq.(12), and then using Eq.(11), we obtain Eq.(37).

Corollary 4. Under the conditions of Proposition 3, one has for $r = 1, 2, \ldots$

$$D_q^r\left(\mathbb{M}_q^{II(I)}(t)\right)\Big|_{t=0} = \mu_q^{I(r)}; \quad D_q^r\left(\mathbb{M}_q^{I(II)}(t)\right)\Big|_{t=0} = \mu_q^{II(r)}.$$

Proof. The r-th derivatives of Eqs.(36) and (37) with respect to the variable t are obtained for $t \neq [\alpha]_q$, respectively

$$D_q^r\left(\mathbb{M}_q^{II(I)}(t)\right) = a_q^{I} \sum_{r=0}^{n} \frac{[k+r]_q!}{\left([\alpha]_q - t\right)_q^{k+1}};$$

$$D_q^r\left(\mathbb{M}_q^{I(II)}(t)\right) = a_q^{II} \sum_{r=0}^{n} \frac{q^{-\binom{r+k+1}{2}} [k+r]_q!}{\left([\alpha]_q - q^{-k}t\right)_q^{k+1}}$$

Thus,

$$D_q^r\left(\mathbb{M}_q^{II(I)}(t)\right)\Big|_{t=0} = \mu_q^{I(r)}; \quad D_q^r\left(\mathbb{M}_q^{I(II)}(t)\right)\Big|_{t=0} = \mu_q^{II(r)}.$$

AN ILLUSTRATIVE EXAMPLE

The main purpose of this section is to introduce an important q-distribution, whose classical form is frequently used in applied fields, derived from the generalized lifetime q-distribution $GLD_q^{\,i}(n, \alpha)(i = I, II)$. It corresponds to the Lindley q-distribution for $n = 1$, as derived from Eq.(16) such that

$$f_q^{i(m)}(\tau; [\alpha]_q) = \sum_{k=0}^{n} p_k^{i(k)} g^{C_m(k)}(\tau; k, [\alpha]_q), \quad m = 1,2,3,4, \quad (38)$$

where

$$p_q^{I(0)} = \frac{[\alpha]_q}{[\alpha]_q + 1}, \quad p_q^{I(1)} = 1 - p_q^{I(0)}, \quad p_q^{II(0)} = \frac{q[\alpha]_q}{q[\alpha]_q + 1},$$

$$p_q^{I(1)} = 1 - p_q^{II(0)}$$

and $C_m(k)$ determines which of the following gamma q-distributions is used for each combination m and k:

$$g_q^{I}(\tau; k, [\alpha]_q) = \frac{[\alpha]_q^{k+1}}{\Gamma_q(k+1)} \tau^k E_q^{-q[\alpha]_q\tau};$$

$$g_q^{II}(\tau; k, [\alpha]_q) = \frac{[\alpha]_q^{k+1}}{\gamma_q(k+1)} \tau^k e_q^{-[\alpha]_q\tau}.$$

Thus, $f_q^{i(m)}$ ($i = I, II$; $m = 1,2,3,4$) is given by

$$f_q^I(\tau;[\alpha]_q)=\frac{[\alpha]_q^2}{[\alpha]_q+1}\begin{cases}(1+\tau)E_q^{-q[\alpha]_q\tau}, & \tau\in[0,[\infty]_q]\\ (1+\tau)e_q^{-[\alpha]_q\tau}, & \tau\in[0,\infty)\\ \left(E_q^{-q[\alpha]_q\tau}+\tau e_q^{-[\alpha]_q\tau}\right), & \tau\in[0,\infty)\\ \left(e_q^{-[\alpha]_q\tau}+\tau E_q^{-q[\alpha]_q\tau}\right), & \tau\in[0,\infty)\end{cases}$$

$$f_q^{II}(\tau;[\alpha]_q)=\frac{q[\alpha]_q^2}{q[\alpha]_q+1}\begin{cases}(1+\tau)E_q^{-q[\alpha]_q\tau}, & \tau\in[0,[\infty]_q]\\ (1+\tau)e_q^{-[\alpha]_q\tau}, & \tau\in[0,\infty)\\ \left(E_q^{-q[\alpha]_q\tau}+\tau e_q^{-[\alpha]_q\tau}\right), & \tau\in[0,\infty)\\ \left(e_q^{-[\alpha]_q\tau}+\tau E_q^{-q[\alpha]_q\tau}\right), & \tau\in[0,\infty)\end{cases}$$

Each function $f_q^{i(m)}$ in Eq.(38) the conditions of a q-PDF for the Lindley q-distribution, and converges to the PDF as $q\to 1$. As throughout this paper, to analyze its some properties, let $f_q^{i(m)}=f_q^i$ and $C_m(k)=i$ for some m. Hence,

$$f_q^I(\tau;[\alpha]_q)=\frac{[\alpha]_q^2}{[\alpha]_q+1}(1+\tau)E_q^{-q[\alpha]_q\tau},\ \tau\in[0,[\infty]_q] \quad (39)$$

$$f_q^I(\tau;[\alpha]_q)=\frac{q[\alpha]_q^2}{q[\alpha]_q+1}(1+\tau)e_q^{-[\alpha]_q\tau},\quad \tau\in[0,\infty) \quad (40)$$

Thus, its corresponding q-CDFs are obtained as follows:

$$F_q^I(\tau;[\alpha]_q)=\begin{cases}0, & \tau<0\\ 1-\dfrac{([\alpha]_q(1+\tau)+1)E_q^{-[\alpha]_q\tau}}{[\alpha]_q+1} & \tau\in[0,[\infty]_q]\\ 1, & \tau>[\infty]_q\end{cases} \quad (41)$$

$$F_q^{II}(\tau;[\alpha]_q)=\begin{cases}0, & \tau<0\\ 1-\dfrac{q[\alpha]_q(1+\tau)e_q^{-[\alpha]_q\tau}+e_q^{-q[\alpha]_q\tau}}{q[\alpha]_q+1} & \tau\in[0,\infty)\\ 1, & \tau>\infty\end{cases} \quad (42)$$

Furthermore, as q tends 1, we have

$$f(\tau;\alpha)=\frac{\alpha^2}{\alpha+1}(1+\tau)e^{-\alpha\tau};$$

$$F(\tau;\alpha)=\begin{cases}0, & \tau<0\\ 1-\dfrac{(\alpha(1+\tau)+1)e^{-\alpha\tau}}{\alpha+1} & \tau\in[0,\infty).\\ 1, & \tau>\infty\end{cases}$$

Figures 1–8 for $n=1$ illustrate the effect of different parameter combinations on the shape and variability of the Lindley q-distribution's q-PDFs and q-CDFs. Moreover, the q-RFs of the Lindley q-distribution are obtained from Theorem 2 for $n=1$:

$$\mathbb{S}_q^I(\tau;[\alpha]_q)=\frac{([\alpha]_q(1+\tau)+1)E_q^{-[\alpha]_q\tau}}{[\alpha]_q+1};$$

$$\mathbb{S}_q^{II}(\tau;[\alpha]_q)=\frac{q[\alpha]_q(1+\tau)e_q^{-[\alpha]_q\tau}+e_q^{-q[\alpha]_q\tau}}{q[\alpha]_q+1};$$

$$h_q^I(\tau;[\alpha]_q)=\frac{[\alpha]_q^2(1+\tau)}{[\alpha]_q(1+q\tau)+1}$$

$$h_q^{II}(\tau;[\alpha]_q)=\frac{[\alpha]_q^2(1+\tau)e_q^{-[\alpha]_q\tau}}{q[\alpha]_q(1+\tau)e_q^{-q[\alpha]_q\tau}+e_q^{-q^2[\alpha]_q\tau}}$$

$$mrl_q^I(\tau;[\alpha]_q)=\frac{[\alpha]_q(1+q\tau)+[2]_q}{[\alpha]_q([\alpha]_q(1+\tau)+1)}$$

$$mrl_q^{II}(\tau;[\alpha]_q)=\frac{[\alpha]_q(1+q\tau)e_q^{-[\alpha]_q\tau}+[2]_qe_q^{-q[\alpha]_q\tau}}{[\alpha]_q\left([\alpha]_q(1+q\tau)e_q^{-[\alpha]_q\tau}+e_q^{-q[\alpha]_q\tau}\right)}$$

and $h_q{}^I=f_q{}^I$, $h_q{}^{II}=f_q{}^{II}$, $mrl_q{}^I=\mu_q{}^I$, $mrl_q{}^{II}=\mu_q{}^{II}$ hold for $\tau=0$, and as Type equation here. tends 1:

$$S(\tau;\alpha)=\frac{(\alpha(1+\tau)+1)e^{-\alpha\tau}}{\alpha+1};\ h(\tau;\alpha)=\frac{\alpha^2(1+\tau)}{\alpha(1+\tau)+1};$$

$$mrl(\tau;\alpha)=\frac{\alpha(1+q\tau)+2}{\alpha(\alpha(1+\tau)+1)}.$$

In addition, the r-th q-moment and the r-th q-central moment of the Lindley q-distribution are derived from Theorem 3 and Theorem 4 for $n=1$, respectively

$$\mu_q^{I(r)}=\frac{[r]_q!\,([\alpha]_q+[r+1]_q)}{[\alpha]_q^r([\alpha]_q+1)};$$

$$\mu_q^{II(r)}=\frac{q^{-\binom{r+1}{2}}[r]_q!\,(q[\alpha]_q+q^{-r}[r+1]_q)}{[\alpha]_q^r([\alpha]_q+1)};$$

$$m_q^{I(r)}=\sum_{k=0}^{r}\begin{bmatrix}r\\ s\end{bmatrix}_q q^{\binom{s}{2}}(-1)^s\left(\frac{[\alpha]_q+[2]_q}{[\alpha]_q+1}\right)^s\frac{([\alpha]_q+[r-s+1]_q)}{[\alpha]_q+1}\frac{[r-s]_q!}{[\alpha]_q^r};$$

$$m_q^{II(r)}=\sum_{k=0}^{r}\begin{bmatrix}r\\ s\end{bmatrix}_q q^{\binom{s}{2}}(-1)^s\left(\frac{[\alpha]_q+q^{-2}[2]_q}{q[\alpha]_q+1}\right)^s\frac{q([\alpha]_q+q^{-(r-s+1)}[r-s+1]_q)}{q[\alpha]_q+1}\frac{q^{\binom{r-s+1}{2}}[r-s]_q!}{[\alpha]_q^r}$$

for $r=1,2,\ldots$, and their classical form are derived as q tends 1:

$$\mu=\frac{r!\,(\alpha+r+1)}{\alpha^r(\alpha+1)};$$

$$m^{(r)}=\sum_{k=0}^{r}\binom{r}{s}(-1)^s\left(\frac{\alpha+2}{\alpha+1}\right)^s\frac{(\alpha+r-s+1)}{\alpha+1}\frac{(r-s)!}{\alpha^r};$$

Thus, the q-mean and q-variance of the Lindley q-distribution are obtained:

$$\mu_q^I = \frac{([\alpha]_q + [2]_q)}{[\alpha]_q([\alpha]_q + 1)};$$

$$\mathbb{V}_q^I(\xi) = \frac{[2]_q([\alpha]_q + [3]_q)}{[\alpha]_q^2([\alpha]_q + 1)} - \left(\frac{([\alpha]_q + [2]_q)}{[\alpha]_q([\alpha]_q + 1)}\right)^2;$$

$$\mu_q^{II} = \frac{([\alpha]_q + q^{-2}[2]_q)}{[\alpha]_q(q[\alpha]_q + 1)};$$

$$\mathbb{V}_q^{II}(\xi) = \frac{q^{-2}[2]_q([\alpha]_q + q^{-3}[3]_q)}{[\alpha]_q^2(q[\alpha]_q + 1)} - \left(\frac{([\alpha]_q + q^{-2}[2]_q)}{[\alpha]_q(q[\alpha]_q + 1)}\right)^2;$$

$$\mu = \lim_{q\to 1} \mu_q^{i(r)} = \frac{(\alpha + 2)}{\alpha(\alpha + 1)};$$

$$V(\xi) = \lim_{q\to 1} \mathbb{V}_q^i(\xi) = \frac{\alpha^2 + 4\alpha + 2}{\alpha^2(\alpha + 1)^2}; i = I, II.$$

The q-MGFs of the Lindley q-distribution are proved from Theorem 5 for $n = 1$:

$$\mathbb{M}_q^{I(I)}(t) = \sum_{r=0}^{\infty} q^{\binom{r+1}{2}} \left(\frac{[\alpha]_q + [r+1]_q}{[\alpha]_q + 1}\right)\left(\frac{t}{[\alpha]_q}\right)^r;$$

$$\mathbb{M}_q^{I(II)}(t) = \sum_{r=0}^{\infty} q^{\binom{r+1}{2}} \left(\frac{q([\alpha]_q + q^{-(r+1)}[r+1]_q)}{q[\alpha]_q + 1}\right)\left(\frac{t}{[\alpha]_q}\right)^r$$

$$\mathbb{M}_q^{II(I)}(t) = \sum_{r=0}^{\infty} \left(\frac{[\alpha]_q + [r+1]_q}{[\alpha]_q + 1}\right)\left(\frac{t}{[\alpha]_q}\right)^r$$

$$\mathbb{M}_q^{II(II)}(t) = \sum_{r=0}^{\infty} q^{-\binom{r+1}{2}} \left(\frac{q([\alpha]_q + q^{-(r+1)}[r+1]_q)}{q[\alpha]_q + 1}\right)\left(\frac{t}{[\alpha]_q}\right)^r$$

$$M(t) = \lim_{q\to 1} \mathbb{M}_q^{i(j)}(t) = \frac{\alpha^2}{\alpha + 1}\left(\frac{t}{\alpha - t} + \frac{1}{(\alpha - t)^2}\right), i = I, II, t < \alpha.$$

q-PARAMETER ESTIMATION

This section aims to present a modified method of moments (MoM) approach for parameter estimation, specifically tailored to align with the framework of continuous q-distributions. While the MoM has been utilized in earlier studies, such as Bouzida and Zitouni [20], the empirical q-moment calculations in those applications relied on the classical moment calculation approach. However, this traditional approach does not fully capture the structure and properties of continuous q-distributions.

To demonstrate this limitation, a comparison is made between the calculations of the first few q-moments using both the theoretical q-moments and the classical moment calculation approach. This comparison underscores the inadequacy of the classical approach within the q-distribution framework and motivates the development of a modified method. The new technique offers a refined empirical q-moment calculation process that is specifically designed for continuous q-distributions, ensuring both accuracy and consistency in parameter estimation.

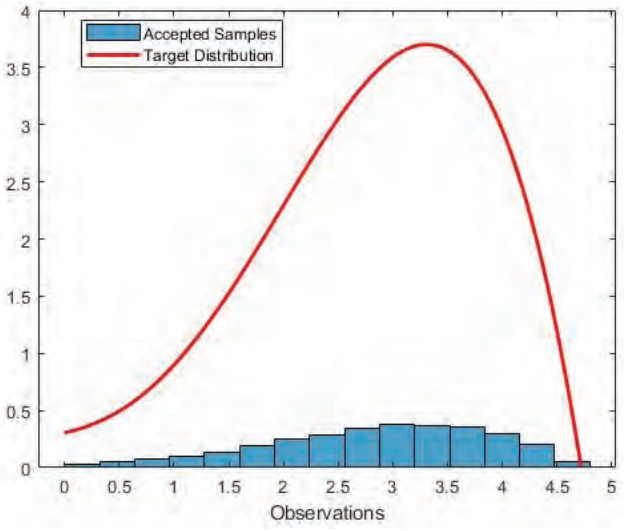


Figure 9. Sample Histogram vs q-PDF curve $\alpha = 1.5$. $q = 0.24$.

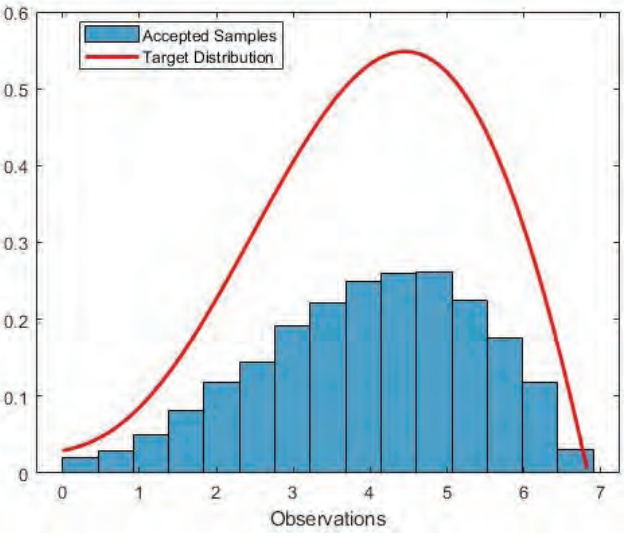


Figure 10. Sample Histogram vs q-PDF curve $\alpha = 0.5$. $q = 0.5$.

Data Generation via Monte Carlo Simulation

We begin by generating data from GLD_q^I using a Monte Carlo simulation approach. Specifically, we utilize the accept-reject method, chosen for its straightforwardness compared to the inverse transform method employed in prior studies, such as that of Bouzida and Zitouni [20]. The q-uniform continuous distribution serves as the prior, while GLD_q^I (case $n = 3$) is the target distribution. Notably, discrepancies between the sample histogram and the q-PDF curve are observed when plotting both, raising questions about the fidelity of the sample to the target distribution (see Figures 9-11). These discrepancies are particularly pronounced for q values near 1, where the curve aligns better with the histogram, illustrating how q-distributions differ fundamentally from classical distributions.

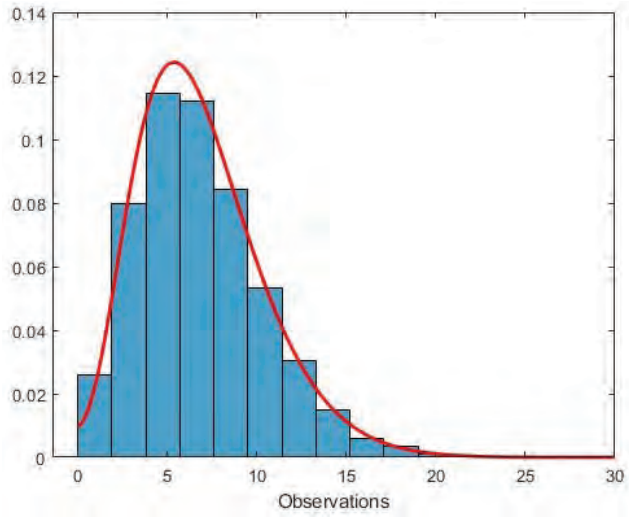


Figure 11. Sample Histogram vs q-PDF curve $\alpha = 0.5$. $q = 0.95$.

Limitations of Classical Empirical q-Moments

The classical method for calculating q-empirical moments $\hat{\mu}_q^{(r)}$ results in significant divergence from theoretical q-moments $\hat{\mu}_q^{(r)}$, regardless of sample size (N). This divergence is especially pronounced for smaller q values away from 1, rendering the classical approach impractical for parameter estimation using the MoM. As summarized in Table 2, these deviations underscore the necessity of transitioning to a modified method for reliable computations. For example, the deviations between the two approaches remain substantial even with large sample sizes, ighlighting the inadequacy of classical techniques for q-distributions.

Modified Empirical q-Moment

Our methodology leverages kernel density estimation (KDE) to refine the calculation of empirical q-moments. Given normalized histogram data, a KDE curve is constructed to approximate the q-PDF curve. For larger sample sizes, this KDE-based approach closely mimics the true q-PDF. The modified empirical q-moment is defined as follows:

$$\hat{\mu}_q^{(r)} = cb^{r+1}(1-q)\sum_{i=1}^{M} q^{(r+1)(i-1)}\hat{f}(bq^{i-1}) = cU_r,$$

where:

- U_r represents the weighted sum of KDE values, and aligns with the Jackson integral definition,
- c is the scaling factor, representing the area under the q-PDF curve,
- b is the largest observation in the sample,
- M is the threshold lower bound of bins with geometrically decreasing widths $(1 - q)bq^{i-1}$, where $i = 1, 2,..., \infty$, and that assumed final bin contains negligible or no observations,
- $\hat{f}$ is the KDE over the sample histogram, and $c\hat{f}$ approximates the true q-PDF.

While determining c analytically is infeasible due to the complexity of q-involved expressions, numerical approximations provide a practical solution. Importantly, the modified method does not rely heavily on exact knowledge of c, as demonstrated in our numerical results.

Parameter Estimation via Method of Moments

The modified MoM involves matching theoretical q-moments to their empirical counterparts up to a certain order. In our case, given one parameters and one unknown c, we use the first two moments:

$$\mu_q^{(r)} = \hat{\mu}_q^{(r)} = cU_r, \quad r = 1, 2.$$

$$\mu_q^{(2)} - \mu_q U_2/U_1 = 0.$$

To solve this system for the parameter α, we utilize MATLAB's built-in lsqnonlin function. This specialized optimization tool is designed to minimize the sum of squared residuals in nonlinear least-squares problems, making it an efficient and robust choice for addressing nonlinear systems [26-30].

Table 2. Empirical vs theoretical moments for varying N, q, α

N	q	α	$\hat{\mu}_q$	μ_q	$\hat{\mu}_q^{(2)}$	$\mu_q^{(2)}$
2000	0.13	0.2	15.12	2.933	245.85	8.738
	0.35	1	2.558	1.294	7.486	1.895
	0.67	3	0.777	0.659	0.860	0.633
7500	0.13	0.2	15.092	2.933	245.22	8.738
	0.35	1	2.540	1.294	7.394	1.895
	0.67	3	0.784	0.659	0.870	0.633
17000	0.13	0.2	15.049	2.933	2.44.19	8.738
	0.35	1	2.562	1.294	7.496	1.895
	0.67	3	0.775	0.659	0.854	0.633

Steps For Implementation

To prepare the technique for implementation, follow these steps:

1. **Estimate the Kernel Density**: Begin by estimating the KDE for the given sample. This involves setting up $\hat{f}$ using the KDE formula. Alternatively, MATLAB's built-in ksdensity function can be used to perform this step efficiently.
2. **Calculate Required Values**: Determine the threshold M, representing the maximum number of bins, and b, the largest observation in the sample. Use these to compute U_r up to the desired number of equations.
3. **Formulate the System of Equations**: Construct the system of equations based on the computed values, ensuring it is properly defined and ready to be solved using a numerical equation solver.

With the parameter estimation methodology firmly established, we are now ready to proceed to the simulation studies, which will validate the proposed approach.

Numerical Insights

Table 2 highlights the ratio between theoretical qmoments and the weighted sum U_r, providing insights into the estimation process. The ratio grows larger for smaller q values and approaches 1 as q nears 1. For a sample size of 10,000, for each pair (q, α) the ratios derived from the first three moments exhibit negligible variation, indicating robustness. These findings confirm that the area under the q-PDF curve deviates from 1 except for $q = 1$, further supporting the need for a modified empirical q-moment formulation (Table 3).

Parameter Estimation Results

Table 4 presents parameter estimation results across varying α, sample sizes, and q values, focusing on scenarios with small q, which are underrepresented in the literature. Our methodology demonstrates significant improvements in precision and reliability, particularly for small q values. Key findings include:

1. **Small Variance and MSE:** The variance of parameter estimates remains consistently low across 10 replications, indicating high precision. The computed MSE values further reinforce the accuracy of the method, with minimal deviations from true parameter values.
2. **Impact of Sample Size:** Increasing sample size substantially reduces MSE and absolute bias (AB). Figure 12 illustrates this trend, showing the MSE of parameter estimates for $\alpha = 0.45$. across varying sample sizes.

Table 3. Theoretical moments, weighted sums, and ratios for $N = 10000$

(q, α)	r-th Order	$\mu_q^{(r)}$	U_r	$\mu_q^{(r)}/U_r$
(0.25, 0.5)	$r = 1$	1.924	0.1522	12.647
	$r = 2$	3.817	0.300	12.692
	$r = 3$	7.620	0.598	12.735
(0.55, 1)	$r = 1$	1.747	1.021	1.710
	$r = 2$	3.496	2.031	1.721
	$r = 3$	7.371	4.258	1.731
(0.95, 2.6)	$r = 1$	0.773	0.755	1.024
	$r = 2$	1.044	1.022	1.021
	$r = 3$	1.878	1.843	1.019

Table 4. Simulation results

q	α	0.5	1	2	2.5
0.3	Mean	0.503	1.010	2.038	2.578
	MSE	1.80E-05	1.63E-04	0.002	0.007
	Var	3.80E-06	7.17E-05	7.07E-04	0.001
0.5	Mean	0.503	1.005	2.015	2.524
	MSE	1.32E-05	4.19E-05	3.704E-04	9.89E-04
	Var	2.30E-06	9.60E-06	1.619E-04	4.28E-04
0.9	Mean	0.498	0.995	1.988	2.492
	MSE	3.53E-05	2.98E-05	1.75E-04	1.34E-04
	Var	2.62E-06	1.45E-05	4.79E-05	8.71E-05

The mathematical expressions for MSE and variance (Var) are provided below:

$$MSE = \frac{1}{R}\sum_{i=1}^{R}(\hat{\alpha}_i - \alpha)^2, \quad Var = \frac{1}{R-1}\sum_{i=1}^{R}(\hat{\alpha}_i - \bar{\alpha})^2.$$

Overall, the results validate the effectiveness of our methodology for parameter estimation in q-distributions. The robustness to variations in q, combined with high precision and accuracy, underscores the method's applicability to diverse scenarios, particularly those involving small q values. These advancements address critical gaps in the literature, confirming the reliability and extendability of the modified MoM approach.

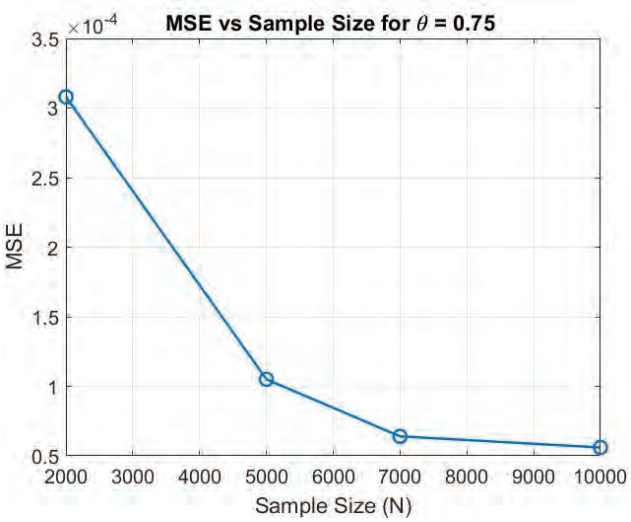


Figure 12. Variation of MSE of α with N for $GLD_q^I(3, [0.75]_q)$.

CONCLUSION

In this paper, we introduce a novel generalized lifetime q-distribution and analyze its distributional and statistical properties in detail. We provide a comprehensive mathematical analysis of this new q-distribution using a method that is unique in the literature, including its q-modeling, q-reliability functions, q-moments, q-moment generating functions. Moreover, we examine the distributional properties of the Lindley q-distribution derived from the proposed q-distribution. To address parameter estimation challenges associated with this distribution, we employ the method of moments. As the parameter q approaches 1, the proposed q-distribution converges to its classical form. To further evaluate the proposed q-distribution, we conduct simulations using data for different values of q and sample sizes (n).

In conclusion, the evolution of q-distributions represents a natural progression in the development of q-calculus. q-calculus serves as a parametric generalization of classical calculus, with the classical framework being recovered in the limit as $q \to 1$. This property makes q-calculus a more expansive and inclusive framework, extending beyond the confines of traditional analysis. Our findings suggest that the proposed q-distribution holds significant promise and may have widespread applications across various fields. In future research, we plan to explore a new two-parameter generalized lifetime q-distribution.

ACKNOWLEDGEMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Jackson FH. On q-functions and a certain difference operator. Earth Environ Sci Trans R Soc Edinb 1909;46:253-281. [CrossRef]

[2] Euler L. Introductio in Analysin Infinitorum. Vol. 2. Apud Bernuset, Delamolliere, Falque & soc.; 1797.

[3] Kac VG, Cheung P. Quantum Calculus. Vol. 113. New York: Springer; 2002. [CrossRef]

[4] De Sole A, Kac V. On integral representations of q-gamma and q-beta functions. arXiv [Preprint] 2003:math/0302032.

[5] Chung WS, Kim T, Kwon HI. On the q-analog of the Laplace transform. Russ J Math Phys 2014;21:156-168. [CrossRef]
[6] Uçar F, Albayrak D. On q-Laplace type integral operators and their applications. J Differ Equ Appl 2012;18:1001-1014. [CrossRef]
[7] Dunkl CF. The absorption distribution and the q-binomial theorem. Commun Stat Theory Methods 1981;10:1915-1920. [CrossRef]
[8] Crippa D, Simon K. q-distributions and Markov processes. Discrete Math 1997;170(1-3):81-98. [CrossRef]
[9] Kupershmidt BA. q-probability: I. Basic discrete distributions. J Nonlinear Math Phys 2000;7:73-93. [CrossRef]
[10] Kemp AW. Certain q-analogues of the binomial distribution. Sankhyā Indian J Stat Ser A 2002:293-305.
[11] Djongmon K, Okur N. On a generalized q-binomial distribution and new q-multinomial distribution. Commun Stat Theory Methods 2025;54:6277-6294. [CrossRef]
[12] Charalambides CA. Discrete q-Distributions. New York: John Wiley & Sons; 2016. [CrossRef]
[13] Díaz R, Pariguan E. On the Gaussian q-distribution. J Math Anal Appl 2009;358:1-9. [CrossRef]
[14] Díaz R, Ortiz C, Pariguan E. On the k-gamma q-distribution. Cent Eur J Math 2010;8:448-458. [CrossRef]
[15] Kyriakoussis A, Vamvakari M. Heine process as aq-analog of the Poisson process—waiting and interarrival times. Commun Stat Theory Methods 2017;46:4088-4102. [CrossRef]
[16] Vamvakari M. On q-order statistics. arXiv [Preprint] 2023:arXiv:2311.12634.
[17] Boutouria I, Bouzida I, Masmoudi A. On characterizing the gamma and the beta q-distributions. Bull Korean Math Soc 2018;55:1563-1575.
[18] Boutouria I, Bouzida I, Masmoudi A. On characterizing the exponential q-distribution. Bull Malays Math Sci Soc 2019;42:3303-3322. [CrossRef]
[19] Bouzida I, Zitouni M. The Lindley q-distribution: Properties and simulations. Ric Mat 2023:1-16.
[20] Bouzida I, Zitouni M. Estimation parameters for the continuous q-distributions. Sankhya A 2023;85:948-979. [CrossRef]
[21] Lindley DV. Fiducial distributions and Bayes' theorem. J R Stat Soc Series B Methodol 1958:102-107. [CrossRef]
[22] Shanker R. Sujatha distribution and its applications. Stat Transit New Ser 2016;17:391-410. [CrossRef]
[23] Shanker R. Amarendra distribution and its applications. Am J Math Stat 2016;6:44-56.
[24] Shanker R. Devya distribution and its applications. Int J Stat Appl 2016;6:189-202.
[25] Shanker R. Shambhu distribution and its applications. Int J Probab Stat 2016;5:48-63.
[26] Turhan S, Okur N, Maden S. Hermite-Hadamard type inequality for strongly convex functions via sugeno integrals. Sigma J Eng Nat Sci 2017;8:1-10. [CrossRef]
[27] Djongmon K, Okur N. An analysis of the q-analogues of the Aradhana lifetime distribution. Bull Korean Math Soc 2025;62:1695-1720.
[28] Okur N, Djong-Mon K. Generalized mixture q-distribution: Modelling, properties and parameter estimation. Commun Fac Sci Univ Ank Ser A1 Math Stat 2025;74:645-669. [CrossRef]
[29] Sabir PO. Some remarks for subclasses of bi-univalent functions defined by Ruscheweyh derivative operator. Filomat 2024;38:1255-1264. [CrossRef]
[30] Sabir PO, Lupas AA, Khalil SS, Mohammed PO, Abdelwahed M. Some classes of Bazilevič-type close-to-convex functions involving a new derivative operator. Symmetry 2024;16:836. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2037

Research Article

Tribological analysis of textured ring-cylinder liner pair under varying operating conditions

Atulkumar PATIL[1], Pankaj BELDAR[1,*], Vishnu WAKCHAURE[2], Prashant KUSHARE[1]

[1]*Department of Mechanical Engineering, K.K.Wagh Institute of Engineering Education and Research, Savitribai Phule Pune University, 422003, India*

[2]*Department of Mechanical Engineering, Amrutvahini College of Engineering, Savitribai Phule Pune University, 422608, India*

ARTICLE INFO

Article history
Received: 14 November 2024
Revised: 01 January 2025
Accepted: 07 March 2025

Keywords:
Coefficient of Friction; Cylinder Liner; Laser Surface Texturing; Piston Ring; Temperature

ABSTRACT

Friction reduction at the ring-liner interface significantly enhances internal combustion engine performance. This study investigates the frictional behavior of laser-textured piston ring-liner pairs under varying operating conditions using a friction testing machine. Parameters such as dimple shape (circular, elliptical, rectangular, trapezoidal, non-textured), dimple density (7% to 35%), load (10N to 90N), speed (60 rpm to 1500 rpm), and temperature (30°C to 150°C) were analyzed using a Taguchi L25 array. Results indicate that a circular dimpled ring with 35% density, 90N load, 1500 rpm speed, and 150°C temperature yielded the lowest coefficient of friction, reducing it by 40.81%. Rectangular dimples reduced friction by 60.12% under specific conditions. Elliptical, trapezoidal, and non-textured rings also showed significant friction reduction. Overall, circular dimples at high density, speed, load, and temperature offered the best frictional resistance.

Cite this article as: Patil A, Beldar P, Wakchaure V, Kushare P. Tribological analysis of textured ring-cylinderlinerpairundervaryingoperatingconditions.SigmaJEngNatSci2026;44(2):1313–1327.

INTRODUCTION

In an internal combustion engine low thermal and mechanical efficiency are the major concerns along with net energy availability as fuel energy is majorly dissipated in the form of heat as well as friction [1]. Many of the scientists are working on wear and friction analysis of compression piston rings and wall of cylinder liner pair to enhance the performance of the engine [2]. It can be accomplished through proper use of lubricants with additives, surface modification of ring or liner, use of proper coating on it [3]. The engines are working at variable speed, load and temperature conditions throughout all four strokes leads to rapid change in lubricants film thickness [4]. The piston ring experiences a very high normal load during compression and power stroke results into increase in wear and friction [5]. During the initial part of compression as well as combustion stroke ring experiences boundary lubrication, from the mid stoke to top dead center as well as between mid-stroke to bottom dead center at larger load conditions circumstances mixed lubrication is observed [6]. Since from past few years, friction can be reduced by modifying

***Corresponding author.**
*E-mail address: pankajrbell@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

the surface with micro textures which plays an important role in friction reduction. These micro textures are generally created intentionally on the surface in the form of cavities or undulations. The lubricant is generally entrapped in these cavities and by providing the hydrodynamic effect it helps in the reduction of the friction. The various studies has been conducted for the evaluation of performance of the textured surfaces but still the behaviour of such surfaces is not very clear [7]. Hence frictional and wear minimization related to dimpled ring and liner needs to be evaluated.

Parul Mishra et al. studied the tribological behaviour of textured piston ring liner system under mixed and boundary regimes and concluded that the average friction coefficient under boundary lubrication was reduced up to 12%. They found that un-textured rings behaves better than the textured ones under hydrodynamic lubrication regime due to extra fluid entrained in dimples. Also concluded that friction reduction of 12% and reduction of wear up to 31% was observed with base oil without additives as the dimples on ring obstructs the appropriate tribochemical film formation from additives in lubricant on the textured surfaces. Hence, additives abolishes the beneficial effect of texturing [7,8]. Yali Zhang et al. studied frictional performance of textured piston rings under different sliding speeds and loads and concluded that rectangular shaped textured rings has shown reduction of average friction coefficient of 36.6% as compared to un-textured rings [9]. V. Ezhilmaran et al. investigated the frictional behaviour of textured ring liner interface and concluded that the optimal value for aspect ratio is 0.2 and for area density is 16% which had given low friction coefficient almost at all loading conditions [10]. Textured ring samples had shown a significant decrease in wear rate of liner at about 72% comparing with non-textured sample. They had also carried out friction trials with micro-channeled piston rings textured from femtosecond and nanosecond laser pulses at 10N to 130 N loads [11]. They observed that coefficient of friction for textured samples was lowered in comparison with non-textured samples [12]. Cong Shen et al. shown that laser pockets led to a reduction of friction up to about 15 % between cylinder liner and piston assembly over a wide range of operating speeds [13]. Atulkumar Patil et al. explored tribological nature of the dimpled ring and cylinder liner pair for a medium passenger vehicle with different dimple shapes like spherical, square by varying loads at 80 N, 150 N and at a constant temperature of 150°C with SAE 20W40 lubricant. They concluded that ring with dimples of spherical shape of 150μm diameter has shown enhanced frictional resistance as compared to ring with dimples of square shape of 150μm size under large loads [14]. Babu et al. explored the effect of protruding texture and reported that texture height of 25μm and area density of 10% has proven better for good wear and friction resistance of positive textured ring-liner interface [15]. Ankit Tyagi et al. used a Taguchi approach for analyzing wear of carbon coating applied on piston ring and concluded that the important factors for wear was sliding speed contributing 59.89%, temperature contributing 28.71% and load contributing 9.18% [16]. Shuwen Wang et al. analyzed the tribological effects of the dimple shape and texture area density on the sliding contacts with dimple shapes like circular, elliptical and grooves of 10% and 20% area density. They concluded that the elliptical dimples found to be suitable for wear reduction, circular dimples found to be suitable for friction reduction and groves proven to be good for friction stabilization [17]. Haytam Kasem et al. studied the frictional responses of the textured stainless steel surfaces with the cast iron counter surface in oil for a sliding contact and concluded that smaller diameter dimples led to a decrease in the coefficient of friction in comparison with larger diameter dimples as well as surface pattern with grids [18]. Amirabbas Akbarzadeh et al. evaluated the running in behaviour and frictional nature of a piston ting with various surface treatments on a specially made engine test rig and found that for a number of steady state and running in experiments combinations of surfaces of piston rings used were coated, textured, coated textured, textured coated and plain out of which the combination of texturing followed by coating had showed enhancement in frictional resistance by 12.5 % and enhancement in break in time up to 50% in comparison with other cases when a single surface modification treatment was used [19]. K. Tripathi et al. investigated the wear and friction nature of grey cast iron textured cylinder liner in lubricated conditions. The friction coefficient of the textured specimen was reduced approximately with 32% in comparison to the specimen with polish. As well as 15% dimple density specimen has shown lowest friction coefficient as compared to all the dimpled specimens working with low as well as high viscosity oils. The oil with high viscosity enhances the frictional resistance and reduces the wear resistance in comparison with the oil with low viscosity [20].

Atulkar et al. compiled the data related to ring liner contact used in an internal combustion engine for studying tribological performances of it in order to save fuel and reduce emissions with the modification of ring surface and studied the combined effect of nano lubricant and partially textured ring [21-26]. They found that with combined effect fuel consumption lowered up to 4.58%, and with only lubricant added with nano particles fuel consumption reduced up to 3.23% [27]. Xiang Rao et al. studied the interactions of various types of surface textures and its effect on friction of ring and liner used in diesel engine at hot conditions. They concluded that proper use of surface textures like 3 mm wide grooves on liner proven to be effective as compared to non-textured surface in order to enhance the performance of engine [28]. R. N. Bathe et al. reviewed the application of texturing with the help of laser to be used for ring and liner contact situations along with its future prospects in vehicle frictional applications in order to get rid of the problems of fuel consumption, emissions and friction [29]. Chenwei Miao et al. investigated the running in nature of textured liner-ring pair with grooves on liner and dimples on ring at

300 N load and speeds of 50 rpm to 100 rpm. They found that dimples on rings acts as mobile oil storage while grooves on liner acts as stationary oil storage which helps in reducing the friction between the pair [30]. Baby et al. (2024) conducted an uncertainty analysis on the friction and wear-rate behavior of cylinder liner-piston ring tribo pairs under boundary lubrication conditions. By considering both measurement and experimental uncertainties, they provided a systematic approach to evaluate tribological performance in engines operating under low lubrication conditions. Their findings underscore the importance of accounting for uncertainties in assessing wear rates and friction, which can lead to more reliable predictions of engine component longevity and performance [31,32]. Wang et al. [33] explored the tribological properties of surface-modified piston rings under extreme conditions. They examined different surface treatments to identify enhancements in friction reduction and wear resistance. Their study demonstrated that specific surface modifications significantly improve performance under high load and extreme environmental conditions. This work is crucial for applications where piston rings are subjected to heavy-duty cycles, as it highlights the potential of surface engineering in enhancing durability and minimizing frictional losses. In another study, Cesur et al. [34] investigated the tribological characteristics of ring-cylinder couplings coated with various materials. By applying different coatings to the piston ring and cylinder liner surfaces, they optimized frictional performance and wear resistance. Their research employed an optimization approach to identify the most effective coating materials, revealing that certain coatings can significantly reduce wear, thereby extending the functional life of the components. This study adds to the body of knowledge by providing a comparative analysis of materials that can mitigate wear and enhance tribological performance in internal combustion engines. Qiu et al. [35] studied the influence of ring materials on the thermodynamic performance of self-lubricating spherical plain bearings. They highlighted how material selection impacts heat generation and distribution, affecting overall tribological performance. Although the focus was on spherical plain bearings, their insights into material thermodynamics and self-lubrication mechanisms are relevant to the development of piston ring-cylinder liner systems where temperature management is critical to maintaining low friction and wear.

With an emphasis on oil feeding conditions, Liu et al. [36] examined the lubricating properties of oil-control rings in connection to cylinder liners. Their results show that friction levels are directly impacted by the oil supply, which also has a significant impact on the lubrication process. According to the study, better oil distribution across the liner-piston interface can lower wear rates and friction, increasing engine efficiency. In high-performance engines, where ideal lubrication is crucial for lowering frictional losses and averting premature wear, our study emphasizes the necessity of accurate oil control systems. Yadvendrakumar Mishra et al. proposed some optimization methods which are beneficial for achieving the responses. They also found that Nd: YAG laser is possibly used to drill geometrically accurate holes in fibre materials [37]. Rita Ferreira et al. [38] used a homemade tribometer to test utilized dimples with varying texture diameters and densities. They had tested the friction coefficient for all the samples under consideration and evaluated the best for wear resistance. The best tribological results were obtained for the texture with a density area of 15% and an aspect ratio of 0.25. Jang S. has found that majority of wear damage and friction loss occur at the top dead center location due to the thin-film thickness caused by slow-down and reversal contact velocities as well as high applied load, hydrodynamic lubrication film formation and pressure between the piston ring and cylinder liner are thoroughly examined. Consideration is given to the cylinder liner's surface roughness from the honing process, which is comparable in size to the thickness of the lubricant coating at the TDC point. They had compared the cylinder surfaces of the honed roughness with those of patterned designs in terms of frictional power loss and minimal film thickness [39-41].

From the above stated comprehensive literature review it has been found that the frictional analysis of laser fabricated textured ring and liner pair operating at varying engine operating conditions like load, speed and temperature along with the varying surface modification parameters like texture shape and texture density was rarely observed. It has also been found that many of the researchers have worked on the textured liners but very rare work has been found on textured ring interacting with liners. Also the various dimple shapes along with varying engine parameters were rarely experimented by the earlier researchers and the said acquaintance has been considered during this research work. Therefore tribological trials of textured ring and liner interface were conducted on reciprocating friction testing machine. The purpose of the study is to study the frictional behaviour of dimpled ring-liner interface at different engine operating conditions in order to find the best possible combination at which the frictional performance will be enhances.

MATERIALS AND METHODS

Specimen Preparation

The Electro Discharge Machining (EDM) process was used to cut the specimens in the form of segments from the commercially available cylinder liner and piston ring pieces used in the gasoline/diesel engines of medium duty passenger vehicles. The Electro Discharge Machining equipment is used for the preparation of specimens is shown in Figure 1. The CAD drawing was used in order to cut both the specimens in an accurate manner as per requirement of the friction testing machine (linear reciprocating parameter).

Figure 1. Electric discharge machining equipment.

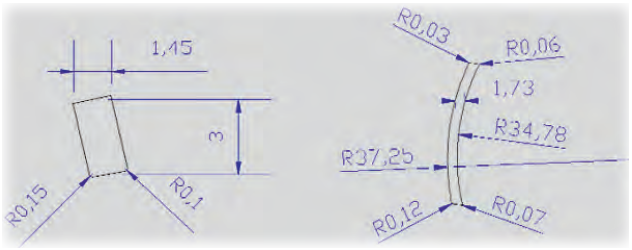


Figure 2a. Profile of piston ring and profile of cylinder liner (dimensions).

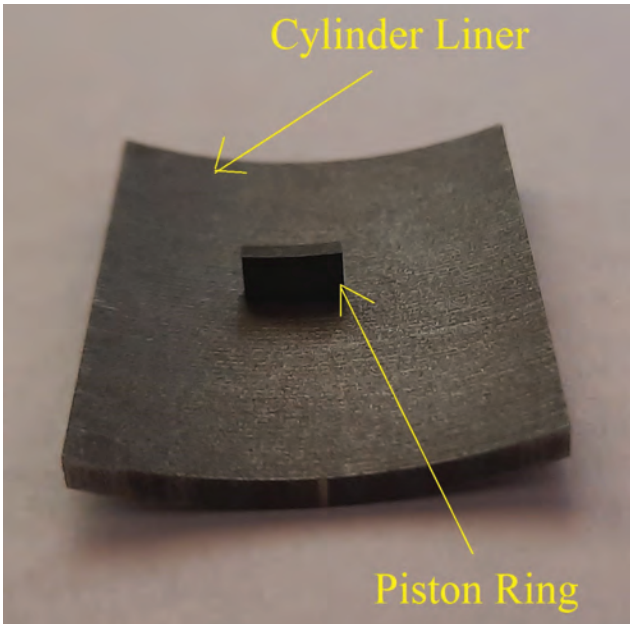


Figure 2b. Piston ring- cylinder liner contact pair.

The piston ring and cylinder liner is made up of high carbon steel and cast iron respectively. During experimentation piston ring segments of dimensions 5.75mm curved length, 1.5mm thickness and 2.5mm width while cylinder liner of dimensions 25mm curved length, axis length of 35mm were used. The details of profile dimensions and images of cylinder liner and piston ring pair are shown in Figure 1, Figure 2a and 2b.

The chemical composition of various constituents of piston ring as well as cylinder liner material is as shown in Table 1.

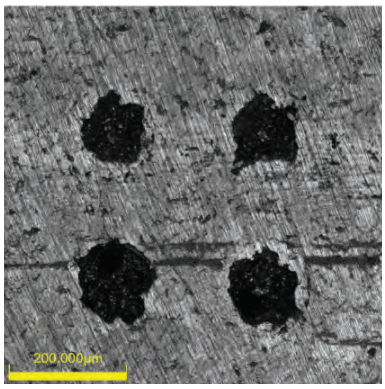
Laser Surface Texturing and Characterization of Textures

The outer surface of piston ring was textured with a nanosecond pulsed commercial laser from Marco Laser Machine, Germany having 20W power, 1064nm wavelength, 20Hz to 200 KHz frequency, maximum laser speed of 15000 mm/s, pulse duration of 26ns to 250 ns. The confocal microscope of Olympus OLS-5000 model was used to characterize the laser texturing details of the ring specimen along with the liner. The details of laser textured piston ring geometric parameters like dimple densities, dimple shape, dimple dimensions etc. are as shown in figure. The various dimple shapes of circular, rectangular, elliptical and trapezoidal were used along with 7%, 14%, 21%, 28% and 35% dimple densities respectively.

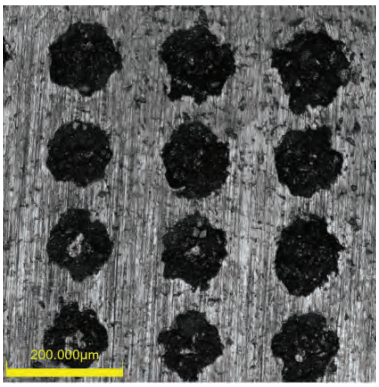
Figure 3 shows the confocal microscope images of some of the textured rings with different dimple shapes along with ring without texture and liner used for the experimentation.

Table 1. Chemical composition of ring and liner

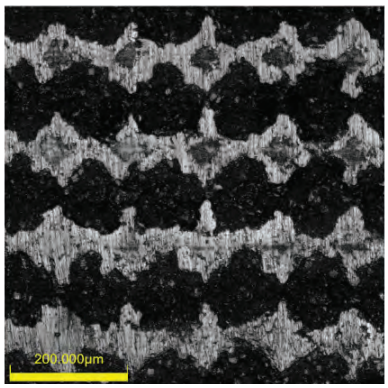
Specimen	Constituents in %					
	C	Mn	Si	P	S	Fe
Piston Ring	3.6	0.80	2.80	0.085	0.043	92.672
Cylinder Liner	2.85	0.58	3.09	0.18	0.072	93.228



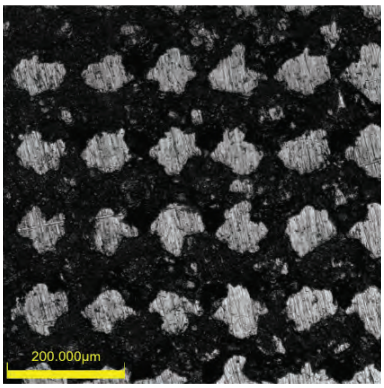
Cir. 7%



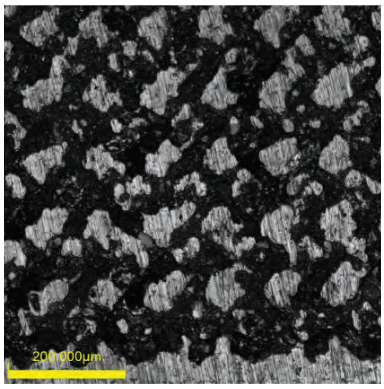
Cir. 14%



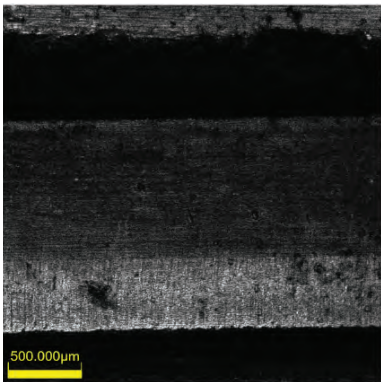
Cir. 21%



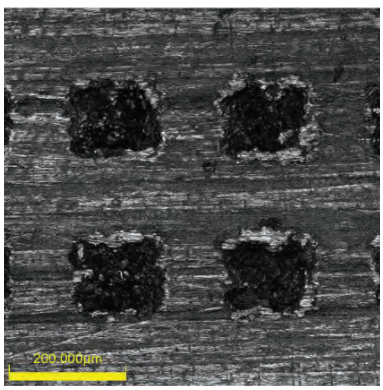
Cir. 28%



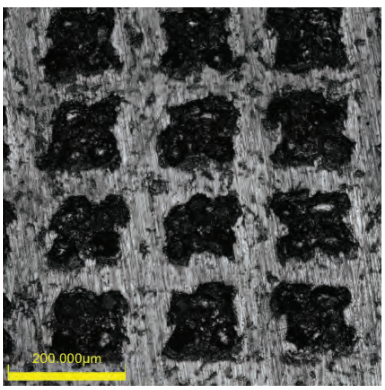
Cir.35%



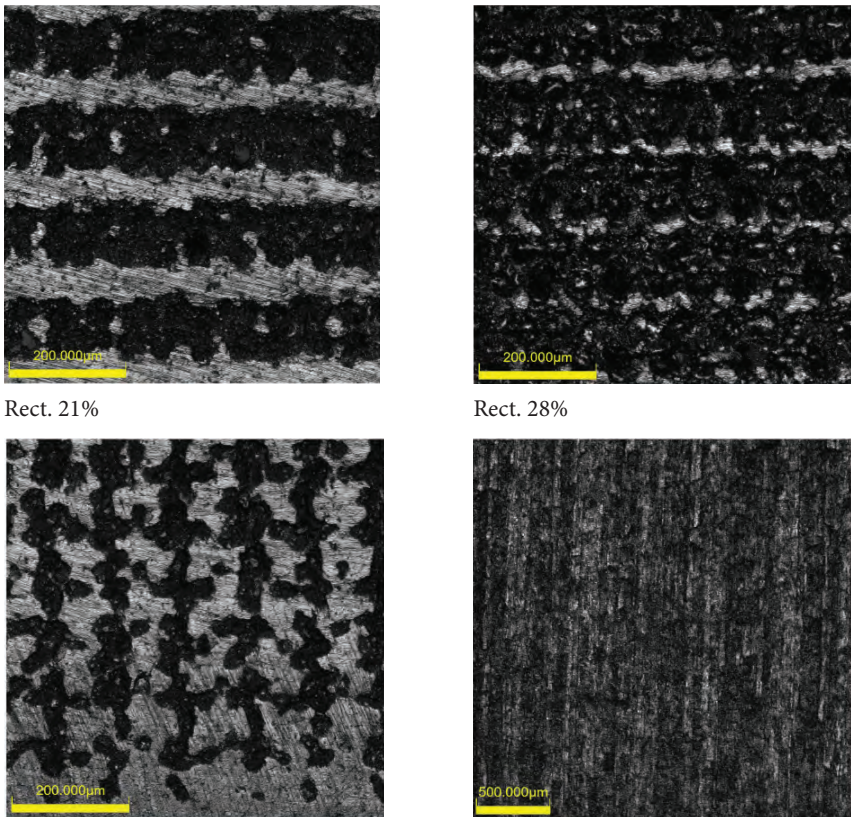
Non Textured Ring



Rect. 7%



Rect. 14%



Rect. 21% Rect. 28%

Rect. 35% Liner before test

Figure 3. Confocal images of laser textured rings and liner.

Experimental Setup

A reciprocating friction testing machine (from Ducom Instruments, Bangalore) which mimics the reciprocating motion of a ring and liner interface, was utilized to analyze the frictional nature of piston rings textured with the help of laser process interacting with liner. The segmented specimens of textured piston ring and regular cylinder liner were fitted in an upper ring and lower liner holders respectively for the testing purpose. During the experimentation textured ring was loaded by the use of a holder exclusively customized for the ring specimens through a circuit operated by air. Normal and lateral force sensors having a very good precision were used to measure the normal and friction forces. The Win-Ducom software was used to collect the test data through a data acquisition system connected to a computer. Figure 4 shows a schematic diagram of an experimental set up with specimen of liner and ring along with the holders.

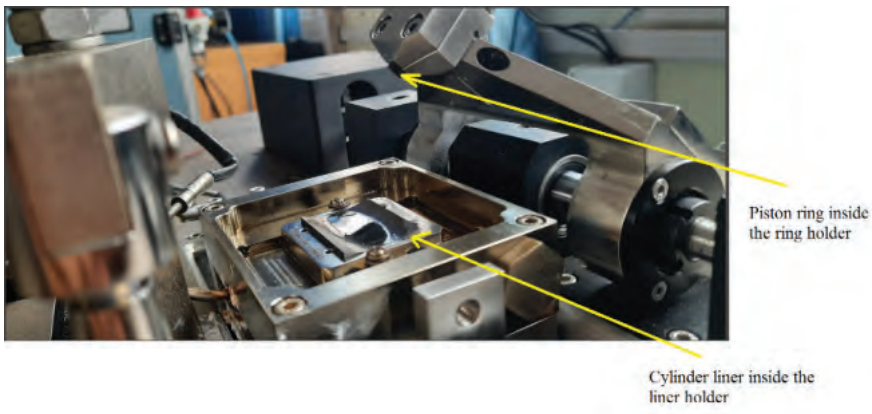


Figure 4. Experimental set up showing ring and liner samples in a linear reciprocating tribometer.

The tribological trials were conducted for 40 minutes of duration by varying various tribological parameters like dimple shape on the textured ring, load, speed and temperature. All the tests were conducted twice and the results were analyzed by averaging the values of frictional force and coefficient of friction. All the three modes of lubrication like boundary, mixed and hydrodynamic were observed during motion of ring and liner interface so as to evaluate frictional behaviour of liner and textured ring at all the modes of lubrication. The tribological test conditions were varied like dimple shape from non-textured, circular, rectangular, elliptical, trapezoidal, dimple density from 7% to 35%, load from 10N to 90N, speed from 60rpm to 1500rpm, temperature from 30°C to 150 °C, oil of SAE5W30, stroke length from 5mm to 25mm, frequency of 5Hz to 25Hz. All the tests were conducted for 40 minutes of duration. ASTM G181-11 standard was used for frictional study of ring and liner interface to conduct the trails. Initially both the test specimens were neatly cleaned with acetone after which dried with the help of air before conducting the trails. The

Table 2. Details of levels and associated control factors

Control Factor	Unit	Level				
		1	2	3	4	5
A- Dimple Shape	Type	Circular	Rectangular	Elliptical	Trapezoidal	Non Textured
B-Density	[%]	7	14	21	28	35
C-Load	[N]	10	30	50	70	90
D-Speed	[rpm]	60	480	720	1020	1500
E-Temperature	[°C]	30	60	90	120	150

Table 3. L_{25} orthogonal array of experimental work (5 factors 5 levels)

Trial no.	Dimple shape	Dimple density [%]	Load [N]	Speed [rpm]	Temperature [°C]
1	Circular	7	10	60	30
2	Circular	14	30	480	60
3	Circular	21	50	720	90
4	Circular	28	70	1020	120
5	Circular	35	90	1500	150
6	Rectangular	7	30	720	120
7	Rectangular	14	50	1020	150
8	Rectangular	21	70	1500	30
9	Rectangular	28	90	60	60
10	Rectangular	35	10	480	90
11	Elliptical	7	50	1500	60
12	Elliptical	14	70	60	90
13	Elliptical	21	90	480	120
14	Elliptical	28	10	720	150
15	Elliptical	35	30	1020	30
16	Trapezoidal	7	70	480	150
17	Trapezoidal	14	90	720	30
18	Trapezoidal	21	10	1020	60
19	Trapezoidal	28	30	1500	90
20	Trapezoidal	35	50	60	120
21	Non-textured	N-T	90	1020	90
22	Non-textured	N-T	10	1500	120
23	Non-textured	N-T	30	60	150
24	Non-textured	N-T	50	480	30
25	Non-textured	N-T	70	720	60

proper fixation of both the specimens was done inside the specially made specimen holders of linear reciprocating friction testing machine. The output parameters like frictional force and frictional coefficient of the liner ring interface were noted through a software by Win-Ducom. The characterization of the frictional nature of liner and textured ring interacting surfaces was done with the help of averaging the output parameters of the trails.

Design of Experiments

Taguchi design approach is used for planning the experiments during the study so as to assess the frictional nature of the liner and ring pair [21]. The 5 factors under consideration for the experimental design are dimple shape, dimple density, load, speed and temperature at 5 levels of each [22]. Factors and their respective levels are summarized in Table 2.

The selection of various parameters of dimple shape, dimple density, load, speed and temperature were done by considering the operation of a medium duty passenger car engine as explained in earlier section [23]. As the 5 control factors and 5 levels of each are considered then according to Taguchi's orthogonal array design L_{25} [24-25] orthogonal array was used for the experimentation purpose and the general details of it along with the actual values are as per Table 3. As all of the above experimental trails were carried out at different load, sliding speed and temperature conditions by varying dimple shape and its density to analyze the frictional nature of interface.

RESULTS AND DISCUSSION

All the tribological tests were conducted at various operating conditions like varying dimple shapes, dimple densities, loads, speeds and temperatures therefore the frictional behaviour was analyzed dimple shape wise for all types of dimples with varying operating conditions as specified earlier. The variation of coefficient of friction was plotted against the test duration in seconds for each type of dimple shape with the varying operating conditions. The data of coefficient of friction obtained from the data acquisition system varying as per the time of test was noted for a specific interval of time and used for plotting the graphs as shown in following figures. Following section give the details of variation of frictional coefficient verses time for all types of dimpled ring along with the non-textured one.

Variation of Coefficient of Friction with Circular Dimple Ring

Figure 5 shows the variation of coefficient of friction (COF) of circular dimple ring and liner pair. It can be seen that circular dimple ring with 35% dimple density, at 90N load, 1500 rpm and at 150°C has shown good frictional resistance as compared to all other rings while sliding against the cylinder liner. Ezhilmaran et al. [10] proved that irrespective of varying texture parameters at dead centers boundary lubrication was observed. As ring moves from dead centers to mid stroke positions it travels with slightly increase in velocity which shifts the lubrication regime from boundary to mixed one. Furthermore when ring attains maximum velocity the lubrication regime shifts to hydrodynamic where a thin film of lubricant exists to lubricate the interacting surfaces [11]. It can be observed that the circular dimple ring-liner interface is operating at 1500rpm which is at a higher speed than the other rings resulted in a hydrodynamic lubrication regime led to minimization of the coefficient of friction due to availability of a reasonable amount of fluid film thickness between the interacting surfaces [10]. Figure 5 also shows that circular dimple ring with 14% dimple density at 30N load, running at 480rpm and 60°C shows higher coefficient of friction as compared to other rings due to low velocity of sliding resulted into

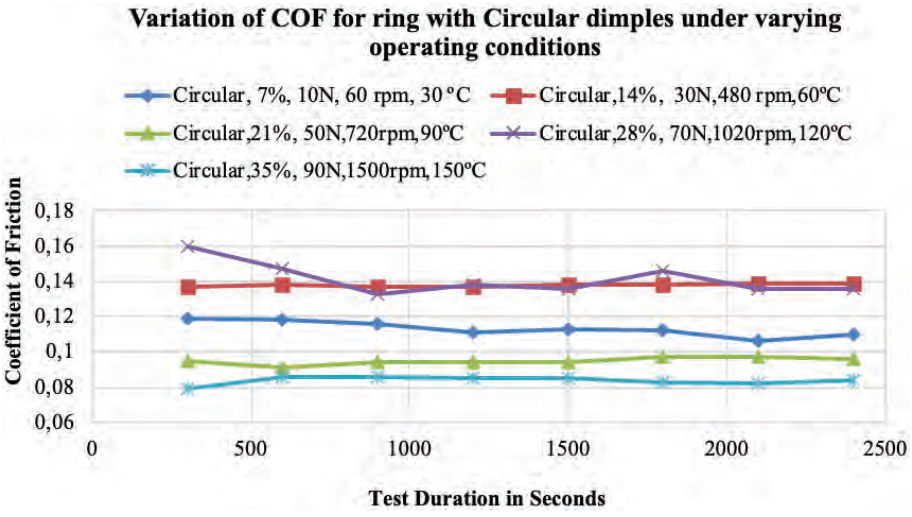


Figure 5. Coefficient of friction versus time for circular shape dimple ring.

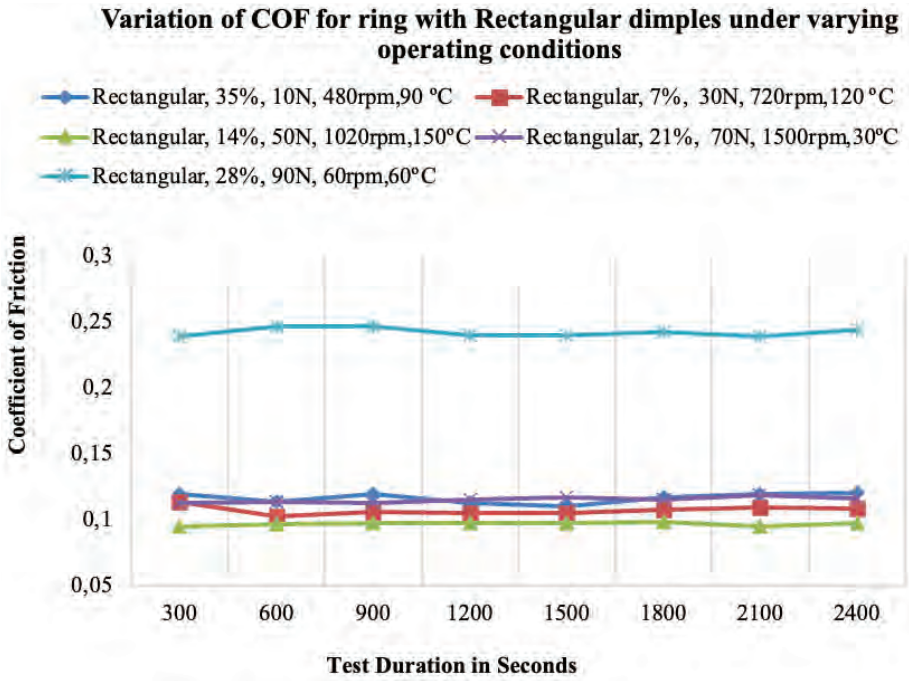


Figure 6. Coefficient of friction versus time for rectangular shape dimple ring.

boundary lubrication regime responsible for raising the coefficient friction [11]. It has been found that nearly 40.81% decrease in coefficient of friction was observed with a circular dimple ring with 35% dimple density, at 90N load, 1500 rpm and at 150°C.

Variation of Coefficient of Friction with Rectangular Dimple Ring

Figure 6 depicts the variation of coefficient of friction of rectangular dimple ring liner pair. It can be seen that rectangular dimple ring with 14% dimple density, at 50N load, 1020rpm and at 150°C has shown good frictional resistance as compared to all other rings. Ezhilmaran et al. [10] concluded that boundary lubrication was observed at dead centers irrespective of varying texture parameters. As ring moves from dead centers to mid stroke positions it travels with slightly increase in velocity which shifts the lubrication regime from boundary to mixed one. Furthermore when ring attains maximum velocity the lubrication regime shifts to hydrodynamic where a thin film of lubricant exists to lubricate the interacting surfaces [11]. It can be observed that the rectangular dimple ring-liner interface is also operating at 1020 rpm which is nearly at a higher speed than the other rings resulted in a hydrodynamic lubrication regime led to minimization of the coefficient of friction due to availability of a sufficient amount of fluid film thickness between the interacting surfaces [10]. Figure 6 also shows that ring with dimple density of 28% at a load of 90N running at 60rpm and 60°C shows enhanced coefficient of friction because a pure boundary lubrication regime as it is rotating at a very low speed of 60 rpm which had lowered the lubricant film thickness tends to have increase in coefficient friction [11]. It has been found that nearly 60.12% decrease in coefficient of friction for rectangular dimpled ring with a dimple density of 14%, at a load of 50N, running at a speed of 1020rpm and at 150°C.

Variation of Coefficient of Friction with Elliptical Dimple Ring

Figure 7 shows the variation of coefficient of friction of elliptical dimpled ring-liner pair. The elliptical dimple ring with 21% dimple density, loaded with 90N, running at 480rpm and operating at 120°C has shown good frictional resistance as compared to all other rings sliding against the cylinder liner. Ezhilmaran et al. [10] concluded that boundary lubrication was observed at dead centers irrespective of varying texture parameters. As ring moves from dead centers to mid stroke positions it travels with slightly increase in velocity which shifts the lubrication regime from boundary to mixed one. Furthermore when ring attains maximum velocity the lubrication regime shifts to hydrodynamic where a thin film of lubricant exists to lubricate the interacting surfaces [11]. This case implies that running speed of 480rpm which is very closer to boundary lubrication regime but still shows a good frictional resistance because here the dimple density of used dimple is of 21% which gives sufficient dimpled area to store the lubricant in it which may be available during the sliding of the surfaces by forming a sufficient lubricant film thickness between the interacting pair led to reduction in friction [12]. It can be observed that ring with 14% dimple density at 70N load running at 60rpm and 90°C shows very poor frictional

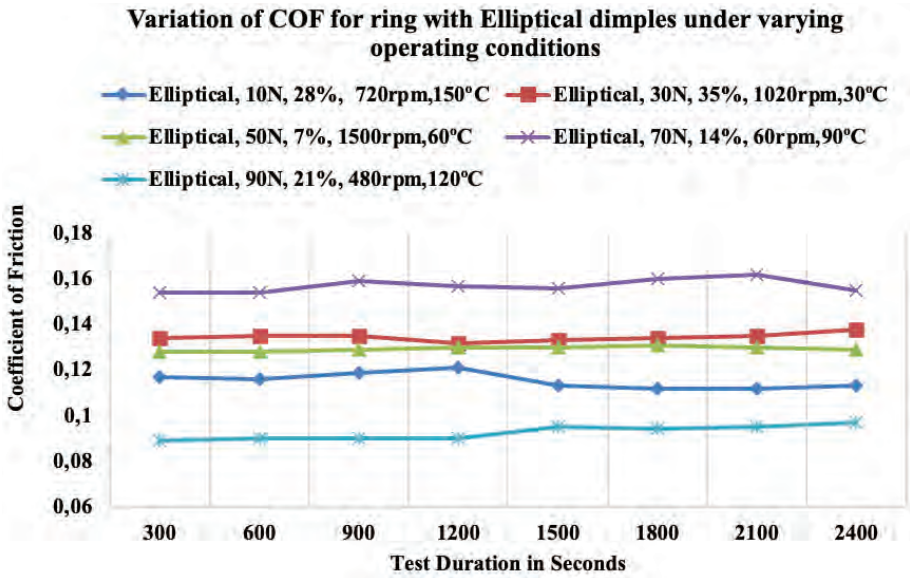


Figure 7. Coefficient of friction versus time for elliptical shape dimple ring.

resistance due to pure boundary lubrication regime which had decreased the lubricant film thickness which had led to enhancement of frictional coefficient [11]. It has been found that nearly 41.12% decrease in coefficient of friction for elliptical dimple ring with 21% dimple density, loaded with 90N, running at 480rpm and operating at 120°C.

Variation of Coefficient of Friction with Trapezoidal Dimple Ring

Figure 8 shows the variation of coefficient of friction of trapezoidal dimple ring-liner. The trapezoidal dimple ring with 28% dimple density loaded at 30N, running at 1500rpm and operating at 90°C has shown good frictional resistance as compared to all other rings sliding against the cylinder liner. Ezhilmaran et al. [10] concluded that boundary lubrication was observed at dead centers irrespective of varying texture parameters. As ring moves from dead centers to mid stroke positions it travels with slightly increase in velocity which shifts the lubrication regime from boundary to mixed one. Furthermore when ring attains maximum velocity the lubrication regime shifts to hydrodynamic where a thin film of lubricant exists to

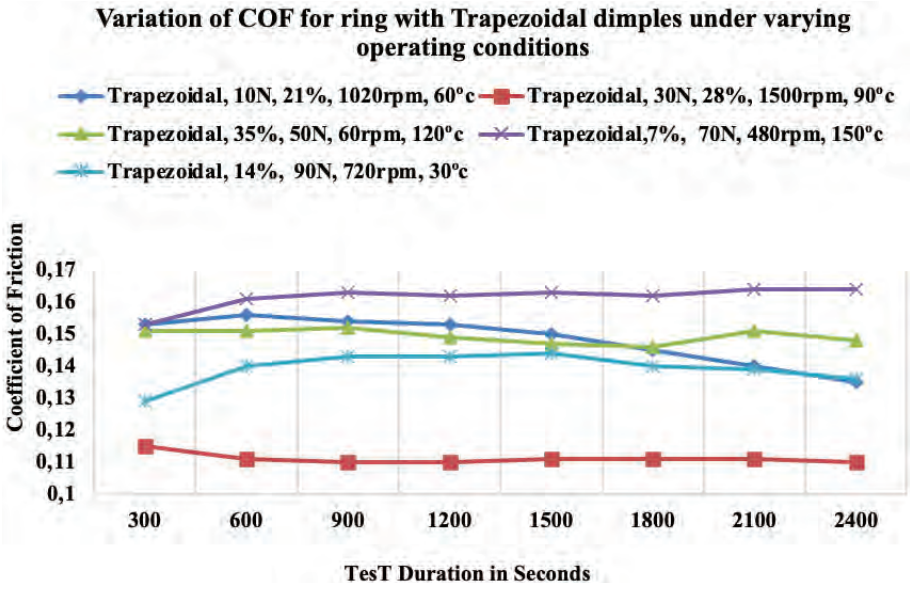


Figure 8. Coefficient of friction versus time for trapezoidal shape dimple ring.

lubricate the interacting surfaces [11]. In this case also it may be observed that the trapezoidal dimple ring-liner interface which is operating at 1500 rpm resembling the hydrodynamic lubrication regime led to minimization of the coefficient of friction of the pair due to availability of sufficient amount of lubricant at the interface [12]. It can be observed from the trapezoidal ring with 7% dimple density at 70N load, running at 480rpm and 150°C shows poor frictional resistance as the pair observed boundary lubrication regime due to low speed of rotation which had reduced the film thickness of lubricant and in turn responsible for the enhancement of the frictional coefficient [11]. It has been found that nearly 31.19% decrease in coefficient of friction for trapezoidal dimple ring with 28% dimple density loaded at 30N, running at 1500rpm and operating at 90°C.

Variation of Coefficient of Friction with Non-Textured Ring

Figure 9 shows the variation of coefficient of friction for non-textured ring-liner pair. The non-textured ring loaded with 90N, running at 1020rpm and at 90°C has shown good frictional performance as compared to all other non-textured rings while sliding against the cylinder liner. Ezhilmaran et al. [10] concluded that boundary lubrication was observed at dead centers irrespective of varying texture parameters. As ring moves from dead centers to mid stroke positions it travels with slightly increase in velocity which shifts the lubrication regime from boundary to mixed one. Furthermore when ring attains maximum velocity the lubrication regime shifts to hydrodynamic where a thin film of lubricant exists to lubricate the interacting surfaces [11]. It can be observed that the ring liner interface is operating in a hydrodynamic lubrication regime as it is operating at 1020rpm at higher loading conditions which had led to minimization of the coefficient of friction [10]. Also the non-textured ring with loaded by 30N, running at 60rpm and 150°C shows higher coefficient of friction because as the interacting pair slides in a pure boundary lubrication mode due to very low running speed of 60 rpm which had lowered the lubricant film thickness and enhanced the frictional coefficient [11]. It has been found that nearly 82.21%decresae in coefficient of friction for non-textured ring loaded with 90N, running at 1020rpm and at 90°C.

Figure 10 shows theComparative analysis of different dimple shapes (circular, rectangular, and elliptical) with varying dimple densities. The optimal configurations, as indicated, include Circular Dimples with 35% dimple density, Rectangular Dimples with 14% dimple density, and Elliptical Dimples with 21% dimple density, demonstrating improved surface characteristics. The study reveals that dimple shape and density play a crucial role in determining the surface characteristics and tribological performance of the material. Among circular dimples, a higher density of 35% exhibits superior performance compared to 14%, indicating enhanced lubricant retention and reduced friction. For rectangular dimples, the 14% dimple density performs better than the 28% density, suggesting that an optimal balance between surface contact and lubricant trapping is essential. In the case of elliptical dimples, a moderate increase in density to 21% results in improved performance compared to 14%, demonstrating that a well-calibrated density enhances lubrication efficiency without excessive surface modification. Overall, the optimal dimple configurations identified for enhanced tribological behavior are 35% density for circular dimples, 14% for rectangular dimples, and 21% for elliptical dimples. These findings highlight the importance of selecting appropriate dimple shapes and densities to optimize wear resistance and lubrication efficiency.

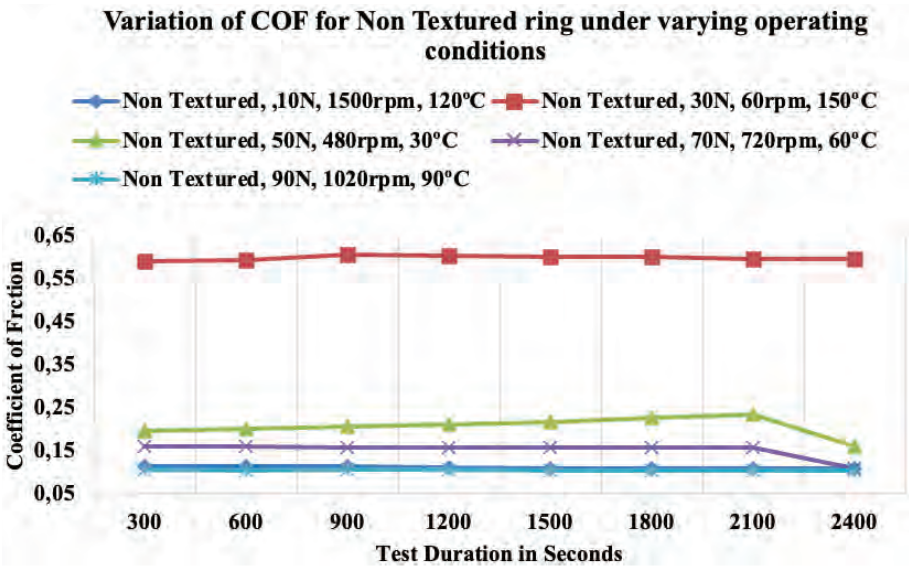
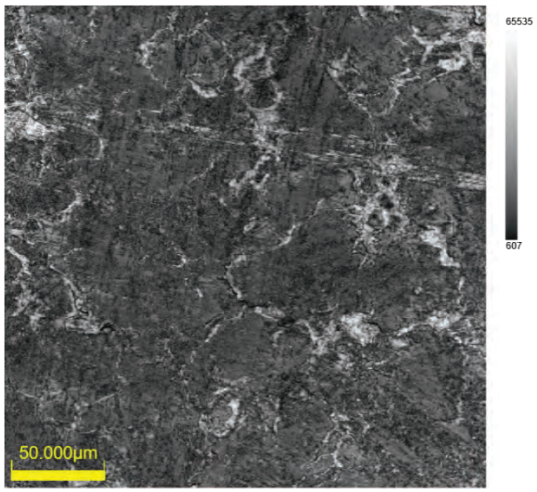
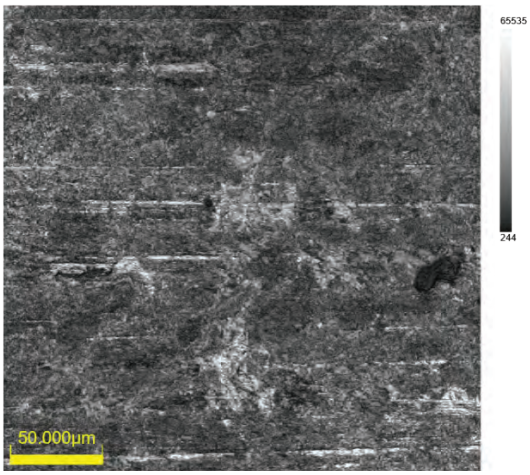


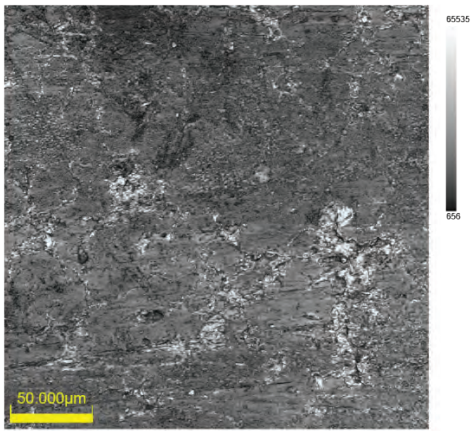
Figure 9. Coefficient of friction versus time for non-textured ring.



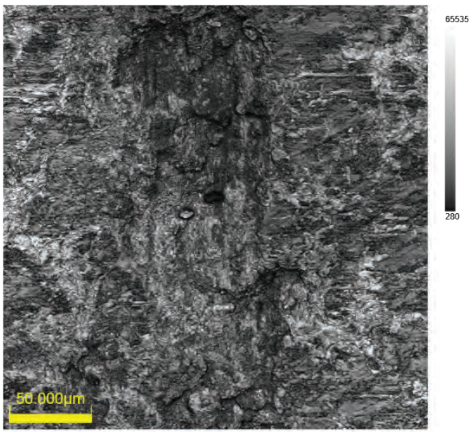
a)Circular dimple with 14% dimple density



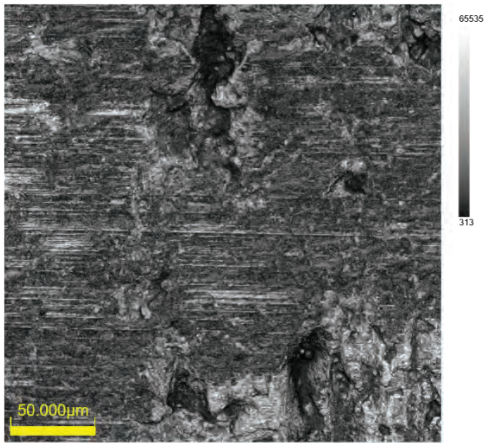
b) Circular dimple with 35% dimple density (good)



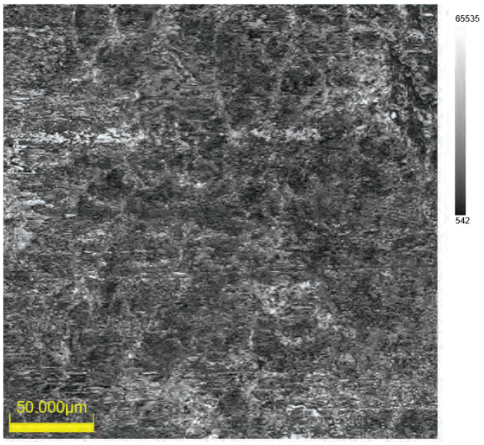
c) Rectangular dimple with 14% dimple density (good)



d) Rectangular dimple with 28% dimple density



e) Elliptical dimple with 14% dimple density



f) Elliptical dimple with 21% dimple density (good)

Figure 10. Optical microscopy images of elliptical dimple with different dimple density.

CONCLUSION

The frictional analysis of laser surface textured ring-liner pair was carried out at varied operating conditions of dimple shape, dimple density, load, speed and temperature by considering the reference of medium duty passenger car engine. The following concluding remarks can be summarized for the work under consideration.

- Dimple shape and dimple density are playing a vital role in analyzing the frictional performance of textured ring and liner pair and which must be selected properly as per the application of engine.
- The circular dimple shape and the dimple densities of 28% and 35% are proving better for reducing the friction coefficient of the textured ring and liner interface.
- The sliding velocity/speed of rotation of the interacting pair and applied load on the ring are also very significant factors while analyzing the performance of the pair.
- It can be observed that except trapezoidal ring-liner pair all other pairs have given a good frictional resistance nearly at higher loading conditions like 50N and 90N.
- It can also be observed that except elliptical ring-liner pair all other pairs have given a good frictional resistance nearly at higher operating speeds like 1020rpm, 1500 rpm etc. which corresponds to hydrodynamic lubrication mode.
- It can be observed that dimple density for a textured ring-liner pair with 21% and 35% found better performer for circular and elliptical dimple ring respectively while textured ring-liner pair with 14% and 28% found better performer for rectangular and trapezoidal dimple ring respectively.
- It has been found that nearly 40.81% decrease in coefficient of friction was observed with a circular dimple ring with 35% dimple density, at 90N load, 1500 rpm and at 150°C.
- It has been found that nearly 60.12% decrease in coefficient of friction for rectangular dimpled ring with a dimple density of 14%, at a load of 50N, running at a speed of 1020rpm and at 150°C.
- It has been found that nearly 41.12% decrease in coefficient of friction for elliptical dimple ring with 21% dimple density, loaded with 90N, running at 480rpm and operating at 120°C.
- It has been found that nearly 31.19% decrease in coefficient of friction for trapezoidal dimple ring with 28% dimple density loaded at 30N, running at 1500rpm and operating at 90°C.
- It may be concluded in general that considering the individual effect of various dimples fabricated on the ring outer surface, the frictional nature of the liner and textured ring pair has improved at nearly medium to higher dimple density, higher speed of rotation, higher loading conditions and at higher operating temperatures.

Limitations and Future Scope

This study has been carried out at simulated operating conditions therefore it is having some limitations like the availability of combustion and allied factors but the load applied at the back of the ring, speed of rotation and the temperature has been simulated as per the engine under consideration. Also due to experimental set up limitation the speed has been considered up to 1500 rpm which is somewhat smaller than the real engine conditions but still it is sufficient one for analyzing the performance of the pair. Though there are certain limitations for the study under consideration still the results will be surely helpful to the engine component designers, researchers in the concern domain. The future work like the testing of laser textured ring in an actual engine and evaluating its tribological performance will be of great help but which may require more amount of funding. Also the future work like tribological evaluation of texture ring and liner pair with combination of circular, elliptical, rectangular and trapezoidal textures on a single ring and liner as its counterpart may be the another area to be considered.

ABBREVIATIONS

EDM	Electric Discharge Machining
ASTM	American Society for Testing of Materials
SAE	Society of Automotive Engineers
COF	Coefficient of Friction

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Ryk G, Kligerman Y, Etsion I. Experimental investigation of laser surface texturing for reciprocating automotive components. Tribol Trans 2002;45:444-449. [CrossRef]

[2] Ryk G, Kligerman Y, Etsion I, Shinkarenko A. Experimental investigation of partial laser surface texturing for piston ring friction reduction. Tribol Trans 2005;48:583-588. [CrossRef]

[3] Ryk G, Etsion I. Testing piston rings with partial laser surface texturing for friction reduction. Wear 2006;261:792-796. [CrossRef]
[4] Bolander N, Sadeghi F. Surface modification for piston ring and liner. In: IUTAM Symposium on Elastohydrodynamics and Microelastohydrodynamics; 2006. p. 271-283. [CrossRef]
[5] Etsion I, Sher E. Improving fuel efficiency with laser surface textured piston rings. Tribol Int 2009;42:542-547. [CrossRef]
[6] Shen C, Khonsari M. The effect of laser machined pockets on the lubrication of piston ring prototypes. Tribol Int 2016;101:273-278. [CrossRef]
[7] Mishra P, Ramkumar P. Effect of micro texture on tribological performance of piston ring-cylinder liner system under different lubrication regimes. SAE Tech Pap 2018;28-0052. [CrossRef]
[8] Mishra P, Ramkumar P. Effect of additives on a surface textured piston ring–cylinder liner system. Tribol Mater Surf Interfaces 2019;13:67-75. [CrossRef]
[9] Zhang Y, Zhang X, Wu T, Xie Y. Effects of surface texturing on the tribological behavior of piston rings under lubricated conditions. Ind Lubr Tribol 2016;68:158-169. [CrossRef]
[10] Ezhilmaran V, Vasa N, Vijayaraghavan L. Investigation of Nd3+ YAG laser aided surface texturing to improve tribological characteristics of piston ring. J Micro Nano Eng 2017;12:195-202. [CrossRef]
[11] Ezhilmaran V, Vasa N, Vijayaraghavan L. Investigation on generation of laser assisted dimples on piston ring surface and influence of dimple parameters on friction. Surf Coat Technol 2018;335:314-326. [CrossRef]
[12] Ezhilmaran V, Vasa N, Vijayaraghavan L, Krishnan S. Influence of pulse width in laser assisted texturing on molychrome films. Appl Phys A 2018;124:167. [CrossRef]
[13] Shen C, Khonsari M. Tribological and sealing performance of laser pocked piston rings in a diesel engines. Tribol Lett 2016;64:26. [CrossRef]
[14] Patil A, Shirsat U. Effect of laser textured dimples on tribological behaviour of piston ring and cylinder liner contact at varying load. Mater Today Proc 2021;44:1005-1020. [CrossRef]
[15] Babu V, Syed I, Benbeera S. Experimental investigation on effects of positive texturing on friction and wear reduction of piston ring and cylinder liner system. Mater Today Proc 2020;24:1112-1121. [CrossRef]
[16] Tyagi A, Pandey S, Murtuza Q, Walia R, Tyagi M. Tribological behaviour of carbon coating for piston ring applications. Mater Today Proc 2020;25:759-764. [CrossRef]
[17] Wang S, Yan F, Chen A. Tribological effects of laser surface texturing and residual stresses. Ind Lubr Tribol 2018;70. [CrossRef]
[18] Kasem H, Stav O, Grutzmacher P, Gachot C. Effect of low depth surface texturing on friction reduction in lubricated sliding contact. Lubricants 2018;6:62. [CrossRef]
[19] Akbarzadeh A, Khonsari M. Effect of untampered plasma coating and surface texturing on friction and running in behaviour of piston rings. Coatings 2018;8:110. [CrossRef]
[20] Tripathi K, Joshi B, Gyawali G, Amanov A, Lee S. A study on the effect of laser surface texturing on friction and wear behavior of graphite cast iron. J Tribol 2016;138:011601. [CrossRef]
[21] Islam M, Pramanik A. Comparison of design of experiments via traditional and Taguchi method. J Adv Manuf Syst 2016;15:151-160. [CrossRef]
[22] Park S. Robust Design and Analysis for Quality Engineering. London: Chapman & Hall; 1996.
[23] Montgomery D. Design and Analysis of Experiments. John Wiley & Sons; 2008.
[24] Taguchi S. Roust Engineering (Taguchi Methods). Workshop Materials. Kuala Lumpur: UTM; 2013.
[25] Ross P. Taguchi Techniques for Quality Engineering. New York: McGraw-Hill; 1988.
[26] Atulkar A, Pandey R, Subbarao P. Role of textured piston rings/liners in improving the performance behaviour of IC engines: A review with vital findings. Surf Topogr Metrol Prop 2021;9:023002. [CrossRef]
[27] Atulkar A, Pandey R, Subbarao P. Synergistic effect of textured piston ring and nano-lubricant on performance parameters and emissions of IC engine. Surf Topogr Metrol Prop 2021;9:035009. [CrossRef]
[28] Rao X, Sheng C, Guo Z, Zhang X, Yin H, Xu C, et al. Effects of textured cylinder liner piston ring on performances of diesel engine under hot engine tests. Renew Sustain Energy Rev 2021;146:111193. [CrossRef]
[29] Bathe R, Padmanabham G, Thirumalini S, Vignesh R. Impact of laser surface texturing (LST) on the tribological characteristics of piston rings and cylinder liners- a review. Part 2: application of the process. Trans IMF 2022;100:119-127. [CrossRef]
[30] Miao C, Guo Z, Yuan C. Tribological behavior of co-textured cylinder liner-piston ring during running-in. Friction 2022;10:878-890. [CrossRef]
[31] Patil A, Wakchaure V, Shirsat U. Experimental investigation of micro-textured piston ring and cylinder liner pair at mid stroke operating conditions. Int J Eng Trends Technol 2022;70:381-392. [CrossRef]
[32] Baby AK, Rajendrakumar PK, Lawrence DK. Uncertainty analysis of friction and wear-rate of cylinder liner–piston ring tribo pair under boundary lubrication conditions. J Tribol 2024;146. [CrossRef]
[33] Wang Y, Sun Z, Huang R, Zhao Z, Zhang W. Tribological properties of several surface-modified piston rings under extreme conditions. J Tribol 2024;146. [CrossRef]

[34] Cesur I, Akgündüz M, Çelik HA, Çay Y, Ergen G. Tribological analysis and optimization of ring-cylinder couple coated with different materials. Arab J Sci Eng 2024;49. [CrossRef]

[35] Qiu M, Zhang Y, Lu T, Pang X. Influence of ring material on thermodynamic performance of self-lubricating spherical plain bearing. J Aerosp Power 2020;35.

[36] Liu G, Sun J, Li B, Zhu S. Lubrication analysis of oil-control-ring and cylinder liner frictional pair considering oil feeding condition. J Mech Eng 2019;55. [CrossRef]

[37] Mishra YK, Gupta SK, Mishra S, Singh DP. Laser beam drilling of fiber reinforced composites using Nd: YAG and CO2 Laser: A review. Mater Today Proc 2023.

[38] Ferreira R, Carvalho O, Sobral L, Carvalho S, Silva F. Laser texturing of piston ring for tribological performance improvement. Friction 2023;11. [CrossRef]

[39] Jang S. Computational study on the frictional power loss reduction of piston ring with laser surface texturing on the cylinder liner. Int J Automot Technol 2022;23. [CrossRef]

[40] Morris N, Leighton M, Rahmani R, King P. Friction reduction through surface texturing at the piston ring-liner interface. In: Proceedings of the LUBMAT 2014 Conference; 2014.

[41] Li X, Zhang Y, Wu X, Wang H. Influence of surface texturing on tribological performance of piston ring-cylinder liner pair. Tribol Lett 2023.

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1328–1343, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00055

Sigma Journal of Engineering and Natural Sciences
Mühendislik ve Fen Bilimleri Dergisi

Research Article

Solving multi-depot vehicle routing problem with incoterms: A case study for export logistics

Gökçe ÖZDEN-GÜRCAN[1,*], Suna KÖSE[2]

[1]*Department of Machinery and Metal Technologies, Eskisehir Osmangazi University, Eskisehir, 26040, Türkiye*
[2]*Department of Business Administration, Pamukkale University, Denizli, 20160, Türkiye*

ARTICLE INFO

Article history
Received: 04 December 2024
Revised: 15 January 2025
Accepted: 12 March 2025

Keywords:
Export Operations; Heterogeneous Fleet; Incoterms; Logistics; Multi-depot Vehicle Routing; Multi Product

ABSTRACT

In today's global market, the international trade methods used in exports have great importance in terms of transportation, including pre-shipment, and post-shipment services due to their impact on buyers and senders. To carry out export operations at low costs and on time, models, that address the multi-depot, multi-product, and heterogeneous fleet vehicle routing problems, together with delivery methods and costs based on the responsibilities of the seller and buyer, are needed. In this study, a mixed integer programming model is presented for companies with production facilities and warehouses in different regions to carry out export operations using Incoterms delivery methods. The objective is to minimize total costs, including holding costs, transportation costs, and the costs associated with delivery terms (incoterms). The customer's distance from each depot, the various vehicle capacities, the number of vehicles in use, the holding costs at each depot, and the delivery terms (incoterms) are all taken into account during the modeling and solving process to find the best solutions. The results show that the model is applicable to exporting firms with a distribution network characterized by a multi-depot, multi-product, heterogeneous fleet.

Cite this article as: Özden-Gürcan G, Köse S. Solving multi-depot vehicle routing problem with incoterms: A case study for export logistics. Sigma J Eng Nat Sci 2026;44(2):1328–1343.

INTRODUCTION

Sociocultural, political, and technological developments in the globalizing world, as well as rapid economic development, have led to the emergence of intense competition. In particular, increasing customer expectations have forced manufacturers to prioritize logistics services and invest in establishing the necessary systems. This situation has led to the continuous development of vehicle routing for reasons such as short life cycles, the emergence of e-commerce, an increase in the number of products transported in densely populated areas, and traffic problems.

Dantzig and Ramser [1] first proposed the "truck dispatching problem" and modeled a homogeneous fleet that could serve many gas stations from a central location to minimize travel distance. Clarke and Wright [2] generalized this problem as the "linear optimization problem," which is frequently encountered in logistics. One of the most widely studied topics in this field is known as the

***Corresponding author.**
*E-mail address: gozden@ogu.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

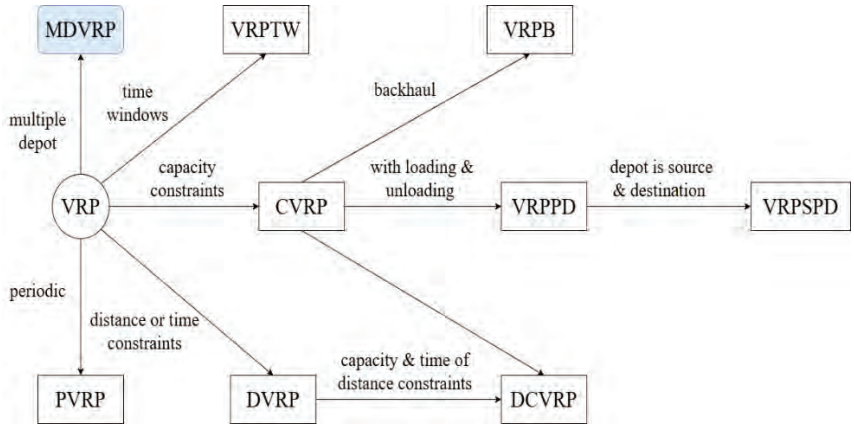


Figure 1. VRP variant hierarchy.

"vehicle routing problem (VRP)"[3]. VRP is an important topic widely researched in logistics science due to its potential cost savings and service improvement opportunities for organizations dealing with the physical distribution of products [4]. When only customers are clearly grouped around each depot, a distribution network problem can be solved as multiple individual single-depot VRPs; otherwise, when customers are to be served from any of the depots using the available fleet of vehicles, a multi-depot-based solution must be utilized [5].

The VRP variant hierarchy is illustrated in Figure 1. One of these variants considers a common, more realistic case in which goods are distributed from several depots to final customers [6,7]. The distribution network problem can be resolved as multiple distinct single-depot VRPs if customers are clearly grouped around each depot. If not, a multi-depot-based strategy must be employed, allowing customers to be served from any depot using the available vehicle fleet. In this study, we address the Multiple Depot Vehicle Routing Problem (MDVRP), a version of the vehicle routing problem involving multiple depots.

The solution of the multi-depot vehicle routing problem with multi-product, heterogeneous fleet, and incoterms conditions is of great importance for exporting companies. The study focuses on the Multi-Depot Vehicle Routing Problem (MDVRP) and develops a mathematical model that incorporates Incoterms delivery conditions in the context of export logistics. The model aims to minimize total costs, including transportation, storage, and delivery expenses, by optimizing the transportation of various products from different depots to customers. According to the reviewed literature, there have been no problem structures and mathematical models with the indicated conditions (incoterms delivery terms) and features (multi-depot, multi-product, heterogeneous fleet) presented in this paper. Integrating all of these features allows the solution to be more valuable and realistic for some industries. Through the use of a multi-product structure and a heterogeneous fleet of vehicles, the model offers solutions suitable for real-life scenarios. It has been tested in several conditions and shows effectiveness in route and cost optimization.

Literature Review

Studies in the literature, categorized according to objective function, fleet type, product type, and solution method, are depicted in Table 1.

Table 1. Studies in the literature according to objective function, fleet type, product type, and solution method

Authors	Fleet type	Product type	Objective function	Solution method
Liu and Lee 2003 [8]	Homogeneous	Single-product	Minimizing the sum of depot-establishing, transportation, and inventory costs.	Heuristic method
Currie et al. 2003 [9]	Heterogeneous	Multi-product	Minimizing total costs (fixed vehicle and operating costs)	Exact and heuristic methods
Kang et al. 2005 [10]	Homogeneous	Single-product	Minimizing transportation cost	Exact algorithms-MIP
Ho et al. 2008 [11]	Homogenous	Single-product	Minimizing the total delivery time	Hybrid algorithms (GA, SA,TS)

Table 1. Studies in the literature according to objective function, fleet type, product type, and solution method (continued)

Authors	Fleet type	Product type	Objective function	Solution method
Moin et al, 2011 [12]	Homogeneous	Multi-product	Minimizing inventory and transportation costs, variable travel costs, and fixed vehicle costs	MILP, hybrid genetic algorithm
Aras et al, 2011 [13]	Homogeneous	Single-product	Maximizing total profit	MILP and TS-based heuristic
Benslimane et al, 2013 [14]	Heterogeneous	Single-product	Minimizing total distribution cost	MILP, Ant Colony Algorithm
Ramos et al, 2014 [15]	Heterogeneous	Multi-product	Minimizing distance and CO^2 emissions	MILP
Zhang et al, 2014 [16]	Homogeneous	Multi-product	Minimizing delivery cost	MILP, GA
Rahmani et al. 2016 [17]	Homogeneous	Multi-product	Minimizing total travel cost, opening cost of selected processing centers, and fixed vehicle costs	MILP, Nearest Neighbor Heuristic, Best Sequential Insertion, Hybrid Clustering Algorithm
Tirkolaee et al, 2017 [18]	Heterogeneous	Single-product	Minimizing total costs, including vehicle usage costs, total traveled distance, and earliness and tardiness costs of services	MILP
Saragih, et al. 2019 [19]	Heterogeneous	Single product	Minimizing the sum of the fixed costs of locating opened depots, routing costs, expected inventory costs, and truck transportation cost of retailers, depots, and suppliers.	MINLP and SA-based Heuristic
Yao et al. 2019 [20]	Homogenous	Single-product	Minimizing the total cost of the delivery process	Ant Colony Optimization (ACO)
Wang et al. 2020 [21]	Heterogeneous	Multi-product	Minimizing the makespan and the total travel time.	MILP, Adaptive Large Neighborhood Search (ALNS)
Moonsri et al. 2022 [22]	Heterogeneous	Multi-product	Minimizing transportation, labor, and holding cost.	MILP, HSADE
Kaveh et al. 2022 [23]	Heterogeneous	Multi-product	Minimizing travel cost	Exact Method
Erdem, 2023 [24]	heterogeneous	Multi-product	Minimizing total energy costs	MILP and AGVNS heuristic
Nguyen, et al. 2022 [25]	heterogeneous	Multi-product	Minimizing the number of unserved customers, number of vehicles needed, and total travel distance	MILP, A-ALNS
Lin et al. (2023) [26]	heterogeneous	Multi-product	Minimizing the sum of the total fixed vehicle cost and the total travel distance	MILP and ALNS algorithms
Çil et al, 2023 [27]	homogeneous	Multi-product	Minimizing the total cost of the workstation and transportation cost.	CP, MILP, MILNP, SA, and MS-SA
Suarez et al., 2024 [28]	homogeneous	Multi-product	Maximizing dispatch, Minimizing wastage	MILP, DP
Arenas-Vasco et al., 2024 [29]	heterogeneous	Multi-product	Minimize the total time elapsed in the operation.	MILP
Maleki et al., 2024 [30]	heterogeneous	Multi-product	Minimize the total cost of the network consisting of transportation, establishment, and waste processing costs	Augmented ε-constraint method, LP-Metric method, Fuzzy method
Li, N., 2025 [31]	homogeneous	Multi-product	Minimize supplier's total cost	ACO and Fuzzy Adaptive GA
Li and Wang, 2025 [32]	homogeneous	Multi-product	Minimize the total transportation cost, the penalty cost violating online time windows and the split delivery cost.	ALNS

MIP: Mixed integer programing; MILP: Mixed-integer linear programing; MINLP: Mixed-integer nonlinear programing; HSADE: Hybrid self-adaptive differential evolution algorithms; AGVNS: Adaptive General-Variable Neighborhood Search; A-ALNS: Applying an adaptive large neighborhood search; CP: Constraint programing; SA: Simulated annealing; MS-SA: Multi-start simulated annealing; DP: Dynamic programing; ACO: Ant colony optimization FAGA: Fuzzy adaptive genetic algorithm; ALNS: Adaptive large neighborhood search; GA: Genetic algorithm

Liu and Lee [8] designed a mathematical model for a single-product multi -depot location and routing problem with inventory control decisions. The aim is to minimize the sum of the warehouse-opening, shipping, and inventory costs. The fleet type is homogenous (vehicle capacity is the same). A two-stage heuristic method was proposed to solve this problem. The proposed heuristic method is more effective than existing methods that do not consider inventory control decisions. Ramos et al. [15] discussed the planning of recyclable wastewater systems due to economic and environmental concerns. They define service areas and vehicle routes within logistics networks with more than one warehouse where different products are collected. This problem was modeled as a multiple product, multi-depot vehicle routing problem, and the aim was to minimize distance and CO_2 emissions. The MILP method for solving this problem was developed and applied to a real case study. Six different scenarios were examined within the scope of service areas and objective functions. As a result, economic and environmental targets were achieved by saving up to 22% and decreasing CO_2 emissions by 27%. Yao et al. [20] discussed the problem of fresh seafood delivery and aimed to find the most appropriate routes to minimize the total cost during the entire delivery process. For this reason, the fresh seafood distribution problem was modeled as a multi-depot vehicle routing problem with a homogeneous fleet. The Ant Colony Optimization (ACO) algorithm is used to solve this problem. In light of the findings, the proposed model is found to be applicable. Moonsri et al. [22] studied multi-product (different sizes of eggs), multi-depot vehicle routing for a heterogeneous fleet, inventory, and time window restrictions to solve an egg distribution problem in Thailand. To solve this problem, a Mixed Integer Linear Programming (MILP) model is designed. The aim is to minimize the total cost. Researchers suggested Hybrid Self-Adaptation Differential Evolution Algorithms (HSADE) for this study. The results demonstrate that HSADE's performance is better than the current method and that HSADE can provide an average improvement of 14.13% in total cost. Erdem [24] aimed to minimize total energy costs by studying milk collection problems with heterogeneous fleets. At the same time, sustainable transport targets were taken into consideration through an environmentally friendly approach. The mathematical model was formulated, and AGVNS was proposed to solve this problem. The findings obtained during the analysis indicate that the problem has been successfully solved in terms of solution quality and time. Lin et al. [26] focused on the post-harvest operations of fruits and vegetables. Mobile pre-processing units are relatively new agricultural applications. Mobile pre-processing services reduce total operating costs compared to traditional center pre-processing. In this context, the current problem of pre-processing operations in short food supply chains was formulated using a heterogeneous fleet and a mixed integer programming model, and an ALNS was proposed. These findings demonstrate the potential of mobile pre-processing units in small-scale agricultural applications.

In the literature, studies show that objective functions containing transportation costs based on road lengths are considered. In addition, it is observed that multi-product and heterogeneous fleet models have been discussed more recently. In this study, the problem structure of heterogeneous fleets and multiple products is considered, and, unlike in other studies, a model structure, including delivery terms and costs, is used for companies that export abroad. The multiple problem structure enabled the calculation of total inventory costs in each warehouse which varied according to product type. On the other hand, the heterogeneous fleet structure allows the use of vehicles of varying capacities on the basis of product types.

Our contributions: The structure of the problem and the mathematical model offers a novel contribution to the literature by integrating the vehicle routing problem with Incoterms (international trade delivery conditions). While there are many studies on the multi-depot vehicle routing problem in the literature, this study stands out by introducing the integration of delivery terms and costs within the context of export logistics. The study has a unique place within the fields of international trade and logistics due to the new and innovative ideas.

- A distribution model is presented for companies with production facilities and warehouses in different regions to conduct export operations using Incoterms delivery methods.
- This study focuses on the export logistics activities of the case study. The problem is formulated as a multi-depot, multi-product, heterogeneous fleet vehicle routing problem that can be used in various real-world export operations for factories. The objective is to minimize total costs, including transportation costs, holding costs, and delivery terms (incoterms).
- To solve this problem, the customer's distance from each depot, the different capacities of each vehicle, the numbers of vehicles used, the holding cost at each depot, and delivery terms (incoterms) are considered in the modeling and solving processes to obtain optimal solutions. Mixed-Integer Linear Programming (MILP) model is used to solve the indicated problem.

Export Operations in Logistics

The concept of Incoterms, which stands for "International Commercial Terms", explains the supply chain requirements for the sale and purchase of goods. These international trade methods express the practical ways in which the obligations, risks, and costs for the seller and buyer are accepted during the exchange of goods [33]. The International Chamber of Commerce developed Incoterms regulations, which are especially important for maintaining stability in international trade and preventing expressions in contracts about international trade from being interpreted incorrectly in different countries. The responsibilities of buyers and sellers are shown in Table 2.

International sales transactions carry significant risks for both buyers and sellers. These risks include the following: the possibility of the goods being delivered without problems if a product is sent from one country to another, the complete cost of the goods, the protection of the goods, transportation risks, customs issues and misunderstandings about commercial terms due to unfamiliar commercial legislation in the buyer's and seller's countries. In Table 2, some of the incoterms: delivery methods are explained, such as FAS, FOB, CFR, and CIF [34].

Free alongside ship (FAS) is the delivery term of the seller (exporter) and refers to the placement of the goods alongside the ship at the loading port. From that point onward, the buyer (importer) firm, takes over all costs for the goods and risks of damage to the goods.

Free on board (FOB) is the delivery term of the seller who delivers the goods up to the point at which they pass the ship's rail in the loading port. The seller (exporter) is obliged to load the goods on the ship. From that point onward, the buyer, importer, firm assumes all costs, losses, and damage risks related to the goods.

Cost and freight (CFR) is the delivery terms of the seller (exporter) company, which indicate the costs required for the transportation of the goods to the specified arrival (port), and freight. "The cost of goods and freight" means that the seller delivers the goods to the loading port after the goods have passed the ship's rail. Additional costs arising from various situations that result in damage from the moment of delivery are the responsibility of the buyer company.

Cost insurance and freight (CIF) means that the seller delivered the goods to a loading port where the goods crossed the ship's rail. The seller (exporter) is obligated to pay the costs of transporting the goods to the specified destination and the price of the freight. However, the additional expenses caused by damage that occurs after the goods pass the handrail are the responsibility of the seller [35].

Multi- Depot Vehicle Routing Problem for Export Logistics

The MDVRP for export logistics was designed for companies that produce tile ceramics and export them to other countries. The shipping transactions of the company can be carried out from either the production factory or the distribution warehouses throughout the country. The products exported by the company are shipped on standard size palettes. The product types differ. In other words, although the products have different colors and features, the standard packaging dimensions are the same. In this study, the warehouses and customers discussed are shown on a map of Türkiye (Fig. 2).

The nodes shown in Figure 2 indicate Çanakkale Çan Depot, Yozgat Depot, İzmir Depot, Gemlik-Borusan Port, Haydarpaşa Port, İskenderun Port, İstanbul Ambarlı Port, İzmir Port, İzmit- Derince Port, Mersin Port, Halkalı Train Station, and Ankara, respectively.

The solution aims to develop a distribution plan that minimizes the total cost of the company according to different scenarios. In each scenario, the delivery method used for customers' orders exported by the company was changed,

Table 2. Buyer (B) / seller (S) cost distribution according to the incoterms chosen.

Operations	Incoterms			
	FAS	FOB	CFR	CIF
Packing	S	S	S	S
Loading containerization	S	S	S	S
Pre_carriage	S	S	S	S
Customs export formalities	S	S	S	S
Consolidation platform-departure	B	S/B	S	S
Principal transportation	B	B	S	S
Transportation insurance	B	B	B	S
Consolidation platform-arrival	B	B	B	B
Customs export formalities duties and taxes	B	B	B	B
Post carriage	B	B	B	B
Unloading	B	B	B	B

Figure 2. Depots and customer locations in Türkiye.

and the mathematical model was solved. For example, in the first scenario, products are distributed to the 5th node according to the FOB or FAS delivery method, (incoterms), whereas the second scenario is realized according to CIF delivery to the 5th node. In different scenarios, the company's delivery and road costs results were obtained by changing the methods of delivery to customers.

Proposed Mathematical Model

In this study, a mixed integer programming model was proposed to solve multi-depot, multi-product and heterogeneous fleet vehicle routing problem. The following assumptions were considered in this model.

- ✓ Each customer can be assigned to a single route (vehicle).
- ✓ The sum of customer demands on a route cannot exceed the capacity of the vehicle on that route.
- ✓ Sub tour elimination: All routes must include customer service to be considered complete.
- ✓ A vehicle can leave from a certain point.
- ✓ Each customer can receive services from a single route.
- ✓ Transportation costs are calculated according to the travelled distance. Each distance unit is also equal to a unit of cost.
- ✓ The sum of customer demands assigned to a warehouse should be less than the warehouse's capacity for each product.
- ✓ The location and number of warehouses are predetermined.
- ✓ A customer can be assigned to a warehouse only if there is a route from that warehouse to that customer.

Notations

Sets

j: customers
i: depots
k: vehicles
g: products
y: delivery terms

Parameters

c_{ij}: transportation cost from i to j,
m_{ig}: capacity for depot i and product g,
h_{ig}: holding cost for depot i and product g
q_{jg}: demand of customer j for product g,
v_k: vehicle capacity of k
s_y: incoterms cost for delivery terms y

Decision Variables

$$x_{ijk} = \begin{cases} 1, \text{if } i \text{ immediately precedes } j \text{ on route } k \\ 0, \text{otherwise} \end{cases}$$

$$z_{ij} = \begin{cases} 1, \text{if customer } j \text{ is alloted to the depot } i \\ 0, \text{otherwise} \end{cases}$$

U_{lk}: auxiliary variable for the sub-tour elimination constraint along route k

$$e_{yj} = \begin{cases} 1, \text{if customer } j \text{ is delivered by delivery terms y} \\ 0, \text{otherwise} \end{cases}$$

Mathematical Formulation

$$min \sum_{i\in I\cup J}\sum_{j\in I\cup J}\sum_{k\in K} c_{ij}\, x_{ijk} + \sum_{i\in I\cup J}\sum_{j\in J}\sum_{g\in G} h_{ig}\, q_{jg} z_{ij} + \sum_{g\in G}\sum_{y\in Y}\sum_{j\in J} q_{jg}\, e_{yj} s_y \tag{1}$$

$$\sum_{k\in K}\sum_{i\in I\cup J} x_{ijk} = 1 \qquad j \in J \tag{2}$$

$$\sum_{y\in Y} e_{yj} = 1 \qquad j \in J \tag{3}$$

$$\sum_{j\in I\cup J}\sum_{g\in G} q_{jg}\, x_{ijk} \le v_k \qquad k \in K \tag{4}$$

$$U_{lk} - U_{jk} + N x_{ijk} \le N - 1 \qquad l, j \in J, k \in K \tag{5}$$

$$\sum_{j\in I\cup J} x_{ijk} - \sum_{j\in I\cup J} x_{jik} = 0 \qquad k \in K, i \in I \cup J \tag{6}$$

$$\sum_{i\in I}\sum_{j\in J} x_{ijk} \le 1 \qquad k \in K \tag{7}$$

$$\sum_{j\in J} q_{jg} z_{ij} \le m_{ig} \qquad i \in I, g \in G \tag{8}$$

$$-z_{ij} + \sum (x_{ilk} + x_{ljk}) \leq 1 \qquad i \in I, j \in J, k \in K \qquad (9)$$

$$x_{ijk} \in \{0,1\} \qquad i \in I, j \in J, k \in K \qquad (10)$$

$$z_{ij} \in \{0,1\} \qquad i \in I, j \in J \qquad (11)$$

$$e_{yj} \in \{0,1\} \qquad y \in Y, j \in J \qquad (12)$$

$$U_{lk} \geq 0 \qquad l \in J, k \in K \qquad (13)$$

The objective function is to minimize total costs, including transportation, holding and delivery terms (incoterms) costs (1). The transportation cost is obtained by multiplying the travel distance by the unit travel cost. The holding cost is obtained by multiplying the holding cost for each depot i, for product g, by the demand of customer j for product g to be sent. The Incoterms costs are calculated by multiplying the costs associated with the delivery terms by the number of deliveries to which these terms are applied, if chosen.

The cost of delivery terms is proportional to incoterms. For example, if FOB is selected, the incoterms cost of the FOB includes packing, loading, containerizing, pre-carriage, and customs-export formalities. If CIF is selected, the incoterms cost for CIF includes packing, loading, containerization, pre-carriage, customs-export formalities, consolidation platform departure, and principal transportation.

Table 3. Incoterms for customers in case 1

2 depot- max 10 customers	**Node label**	**Delivery method**
node 1 and node 2	depot	
node 3	customer	internal
node 4	customer	export (FOB/ CFR)
node 5	customer	export (FAS/ FOB)
node 6	customer	export (CIF)
node 7	customer	export (FAS/ FOB/ CFR)
node 8	customer	export (FOB)
node 9	customer	export (FOB/ CFR)
node 10	customer	export (FAS)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal
3 depot- max 9 customers		
node 1 and node 2 node 3	depot	
node 4	customer	export (FOB/ CFR)
node 5	customer	export (FAS/ FOB)
node 6	customer	export (CIF)
node 7	customer	export (FAS/ FOB/ CFR)
node 8	customer	export (FOB)
node 9	customer	export (FOB/ CFR)
node 10	customer	export (FAS)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal
4 depot- max 8 customers		
node 1 and node 2 node 3 node 4	depot	
node 5	customer	export (FOB/ CFR)
node 6	customer	export (CIF)
node 7	customer	export (FAS/ FOB/ CFR)
node 8	customer	export (FOB)
node 9	customer	export (FOB/ CFR)
node 10	customer	export (FAS)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal

Thus, unit incoterms costs must exceed for CFR compared to FOB. Constraint set (2) guarantees that each customer is assigned a single route. Constraint set (3) ensures that one delivery term must be selected for each customer. For example, for node 4 (customer), there are two options, FOB and CFR, while for node 5 (customer), there are also two options, FAS and FOB. For this constraint set, FAS, FOB, CFR, CIF, and internal customers are numbered 1, 2, 3, 4, 5 for the set of delivery terms "y", respectively. The capacity constraint for a set of vehicles is given by Constraint set (4). Sub-tour elimination constraint (5). The constraint set (6) shows the flow conservation. Constraint set (7) ensures that each route is served at most once. The capacity constraints for the depots are given in Constraint set (8). Constraint set (9) ensures that a customer can be assigned to a depot only if a route from that depot passes through that customer. The binary requirement on the decision variables is indicated in the constraint sets (10-12). Positive values of auxiliary variables are indicated in the constraint set (13).

COMPUTATIONAL EXPERIMENTS

Test Instances

Two different cases were presented for the derived test problems. These cases are handled as 2 depots, 3 depots, and 4 depots. The location and number of warehouses are predetermined. The number of warehouses and customers

Table 4. Incoterms for customers in case 2

2 depot- max 10 customers	**Node label**	**Delivery method**
node 1 and node 2	depot	
node 3	customer	internal
node 4	customer	export (CFR/ CIF)
node 5	customer	export (CIF)
node 6	customer	export (FAS/ FOB/ CFR)
node 7	customer	export (FOB/ CFR)
node 8	customer	export (CFR/ CIF)
node 9	customer	export (CFR)
node 10	customer	export (FOB/ CFR)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal
3 depot- max 9 customers		
node 1 and node 2 node 3	depot	
node 4	customer	export (CFR/ CIF)
node 5	customer	export (CIF)
node 6	customer	export (FAS/ FOB/ CFR)
node 7	customer	export (FOB/ CFR)
node 8	customer	export (CFR/ CIF)
node 9	customer	export (CFR)
node 10	customer	export (FOB/ CFR)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal
4 depot- max 8 customers		
node 1 and node 2 node 3 node 4	depot	
node 5	customer	export (CFR/ CIF)
node 6	customer	export (FAS/ FOB/ CFR)
node 7	customer	export (FOB/ CFR)
node 8	customer	export (CFR/ CIF)
node 9	customer	export (CFR)
node 10	customer	export (FOB/ CFR)
node 11	customer	export (CFR/ CIF)
node 12	customer	internal

may change according to different cases. In the first case, the delivery terms (incoterms) to be applied to the customer nodes and the related information about the internal and external customers are given in Table 3.

In the second case, the delivery terms (incoterms) to be applied to the customer nodes and the related information about the internal and external customers are given in Table 4.

In Tables 3 and 4, various delivery methods (Incoterms) were used according to the number of different warehouses and customers for testing. Generally, external customers are obtained from the 4th node (customer), and export delivery methods are defined for each of these customers. The internal customers are defined after the 11th node and before the 4th node. External customers can choose delivery methods that are preferred for each customer in cases 1 and 2. For example, when node 4 is defined as an external customer in Table 4, only FOB or CFR delivery methods can be used for this customer. In Table 5, when node 8 is defined as an external customer, only CFR or CIF delivery methods can be used with this customer. The case studies are handled using hypothetical data of parameters (such as demands, depot and vehicle capacities) and real road distances related transportation costs (https://doi.org/10.7910/DVN/AFKIGA). 7 different product types and the demand for these products are considered between [500,5000] units, the demand of each demand point for each product varies between [0,500] units. Vehicle capacities are determined between [800,1500] units. A similar number of depots were used in some studies to derive case studies for MDVRP scenarios [36].

RESULTS AND DISCUSSION

The mathematical model was solved using the Gams Cplex solver. A Casper Excalibur computer with Intel(R) Core (TM) i7 processor, 32 GB RAM, and 4.50 GHz speed was used. Optimal solutions are obtained for each test problem. The optimality gap is zero with the proposed mathematical model.

Case 1: The solution results are shown in Table 5, including delivery terms used by the exporter company, the routes taken to send the products to the ports, the objective function value, and the solution time.

Table 5. Results of the test problems for 2 depots in case 1

Test problem	Depot	Node	Solution	Time (s)	Route	Incoterms
1	2	7	100,152	0.45	1-3-1 (vehicle 2) 1-7-5-1 (vehicle 3) 2-4-6-2 (vehicle 1)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS
2	2	8	116,906	0.54	1-4-1 (vehicle 4) 1-5-7-1 (vehicle 3) 2-3-8-2 (vehicle 1) 2-6-2 (vehicle 2)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS, Node 8=FOB
3	2	9	126,166	1.38	1-5-7-1 (vehicle 3) 1-9-4-1 (vehicle 4) 2-3-8-2 (vehicle 1) 2-6-2 (vehicle 2)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS, Node 8=FOB, Node 9=FOB
4	2	10	140,654	20.30	1-7-5-1 (vehicle 3) 1-9-4-1 (vehicle 4) 2-8-3-2 (vehicle 1) 2-10-6-2 (vehicle 5)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS, Node 8=FOB, Node 9=FOB, Node 10=FAS
5	2	11	158,951	790.32	1-4-1 (vehicle 2) 1-7-1 (vehicle 4) 1-11-5-9-1 (vehicle 3) 2-3-8-2 (vehicle 1) 2-6-10-2 (vehicle 5)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS, Node 8=FOB, Node 9=FOB, Node 10=FAS, Node 11=CFR
6	2	12	163,088	120.59	1-3-1 (vehicle 2) 1-5-9-1 (vehicle 4) 1-7-11-1 (vehicle 1) 2-4-8-12-2 (vehicle 3) 2-10-6-2 (vehicle 5)	Node 3=internal, Node 4= F0B, Node 5=FAS, Node 6= CIF, Node 7= FAS, Node 8=FOB, Node 9=FOB, Node 10=FAS, Node 11=CFR, Node 12= internal

In Table 5, test problems that include 2 depots and up to 13 customers were solved. As the number of customers increased from 7 to 12, the total costs also increased. For example, the test problem, which includes two depots and 10 nodes (8 customers), was solved, and it was found that all two depots were used for the distribution. The 7th and 5th nodes were delivered using the 1st and 3rd vehicles, respectively. The 9th and 4th nodes were distributed from the 1st warehouse via vehicle 4. In addition, delivery was made to the 3rd and 8th nodes with the 1st vehicle and to the 10th and 6th customers with the 5th vehicle from the 2nd depot. Node 3 was delivered to an internal customer. According to the solution result of the model, the allocation or determination is FOB for node 4, FAS for node 5, CIF for node 6, FAS for node 7, FOB for node 8, FOB for node 9, and FAS for node 10. The objective function value of this test problem was found to be a cost of 140,654. The currency unit is abbreviated by 'cu'. In solving test problems with 2 depots, in case 1, the solution time increased from 7 nodes to 12 nodes. Although the software was run for about 2 hours to solve test problem 13 (node 13), no results were obtained.

In the test problems with 3 depots (Table 6), goods were distributed from depots 2 and 3. This may be because the costs of the 1st depot are higher than those of the 3rd depot, and the 3rd depot is also more advantageous because it is closer to customers. When the solutions were examined, it was found that the total cost was 140,654 cu when the 1st and 2nd depots were used in a test problem with 2 depots and 10 customers (Table 5), while the total cost was 131,110 cu for the use of 2nd and 3rd depots (Table 6).

The 3rd and 4th depots were generally used for distribution in test problems with four depots, while the 2nd, 3rd, and 4th depots were used in test problems with twelve customers and four depots. For example, in the test problem with 4 depots and 12 nodes, nodes 10 and 6 are served from depot 2 and vehicle 1, node 8 is served from depot 3 and vehicle 4,

Table 6. Results of the test problems for 3 depots in case 1

Test problem	Depot	Node	Solution	Time (s)	Route	Incoterms
1	3	7	91,145	0.30	2-6-2 (vehicle 2) 3-4-3 (vehicle 1) 3-5-7-3 (vehicle 3)	Node 4=FOB, Node 5= FAS, Node 6=CIF, Node 7= FAS
2	3	8	107,469	0.29	2-6-2 (vehicle) 3-4-8-3 (vehicle 1) 3-7-5-3 (vehicle)	Node 4=FOB, Node 5= FAS, Node 6=CIF, Node 7= FAS, Node 8=FOB
3	3	9	116,622	7.11	2-6-2 (vehicle 1) 3-5-7-3 (vehicle 3) 3-8-3 (vehicle 2) 3-9-4-3 (vehicle 4)	Node 4=FOB, Node 5=FAS, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB
4	3	10	131,110	7.77	2-10-6-2 (vehicle 1) 3-7-5-3 (vehicle 3) 3-8-3 (vehicle 4) 3-9-4-3(vehicle 2)	Node 4=FOB, Node 5=FAS, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS
5	3	11	149,429	9.62	2-6-10-2 (vehicle 1) 3-4-9-3 (vehicle 2) 3-7-5-3 (vehicle 3) 3-11-8-3 (vehicle 4)	Node 4=FOB, Node 5=FAS, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS, Node 11=CFR
6	3	12	153,280	4711.80	2-10-6-2 (vehicle 1) 2-12-2 (vehicle 2) 3-8-3 (vehicle 4) 3-9-5-4-3 (vehicle 3) 3-11-7-3 (vehicle 5)	Node 4=FOB, Node 5=FAS, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS, Node 11=CFR, Node 12= internal
7	3	13	156,947	16137.03	2-10-6-2 (vehicle 5) 2-12-2 (vehicle 2) 3-8-13-3 (vehicle 4) 3-9-5-4-3 (vehicle 3) 3-11-7-3 (vehicle 1)	Node 4=FOB, Node 5=FAS, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS, Node 11=CFR, Node 12= internal, Node 13=internal

Table 7. Results of the test problems for 4 depots in Case 1

test problem	depot	node	solution	time (s)	route	incoterms
1	4	7	76,934	0.27	4-5-7-4 (vehicle 3) 4-6-4 (vehicle 2)	Node 5=FOB, Node 6=CIF, Node 7= FAS
2	4	8	93,2246	0.25	3-8-3 (vehicle 2) 4-6-4 (vehicle 1) 4-7-5-4 (vehicle 3)	Node 5= FOB, Node 6=CIF, Node 7= FAS, Node 8=FOB
3	4	9	10,2066	3.18	3-8-3 (vehicle 2) 4-5-7-4 (vehicle 3) 4-6-9-4 (vehicle 1)	Node 5=FOB, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB
4	4	10	11,5907	3.04	3-8-3 (vehicle 2) 4-5-7-4 (vehicle 3) 4-6-10-4 (vehicle 1) 4-9-4 (vehicle 4)	Node 5=FOB, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS
5	4	11	13,4204	48.93	3-8-9-3 (vehicle 2) 4-5-4 (vehicle 4) 4-6-10-4 (vehicle 3) 4-11-7-4 (vehicle 1)	Node 5=FOB, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS, Node 11=CFR
6	4	12	13,8262	48.25	2-10-6-2 (vehicle 1) 3-8-3 (vehicle 4) 4-7-11-4 (vehicle 5) 4-9-12-5-4 (vehicle 3)	Node 5=FOB, Node 6=CIF, Node 7=FAS, Node 8= FOB, Node 9= FOB Node 10=FAS, Node 11=CFR, Node 12= internal

Table 8. Results of the test problems for 2 depots in case 2

Test problem	Depot	Node	Solution	Time (s)	Route	Incoterms
1	2	7	116,652	0.25	1-3-1 (vehicle 2) 1-7-5-1 (vehicle 3) 2-4-6-2 (vehicle 1)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB
2	2	8	138,406	0.40	1-4-1 (vehicle 4) 1-5-7-1 (vehicle 3) 2-3-8-2 (vehicle 1) 2-6-2 (vehicle 2)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8= CFR
3	2	9	150,166	2.37	1-4-9-1 (vehicle 4) 1-5-7-1 (vehicle 3) 2-6-2 (vehicle 2) 2-8-3-2 (vehicle 1)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR
4	2	10	170,654	27.09	1-7-5-1 (vehicle 3) 1-9-4-1 (vehicle 4) 2-8-3-2 (vehicle 1) 2-10-6-2 (vehicle 5)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB
5	2	11	188,951	160.55	1-4-1 (vehicle 4) 1-7-1 (vehicle 2) 1-9-5-11-1 (vehicle 3) 2-8-3-2 (vehicle 5) 2-10-6-2 (vehicle 1)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB, Node 11= CFR
6	2	12	193,088	133.29	1-3-1 (vehicle 2) 1-5-9-1 (vehicle 4) 1-11-7-1 (vehicle 1) 2-6-10-2 (vehicle 5) 2-12-8-4-2 (vehicle 3)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB, Node 11= CFR, Node 12= internal

nodes 7 and 11 are served from depot 4 and vehicle 5, and nodes 5, 9, and 12 are served from depot 4 and vehicle 3.

Case 2: According to the second case, the solution results are shown, including delivery terms are used by the exporter company to send products to the ports, the routes are used, the objective function value, and the solution time.

In Table 8, test problems that include 2 depots and up to 13 customers were solved. As the number of customers increases from 7 customers to 12, the total costs also increased. For example, test problem which includes 2 depots and 10 nodes (8 customers) was solved, and it was found that all 2 depots were used for the distribution. The 7th and 5th nodes were served by the 3rd vehicle and the 1st depot. The 9th and 4th nodes were routed from the 1st depot via the 4th vehicle. In addition, delivery to the 3rd and 8th nodes were completed using the 1st vehicle from the 2nd depot, and the 10th and 6th customers were served using the 5th vehicle from the 2nd depot. Node 3 was delivered as an internal customer. According to the solution result of the model, CFR for node 4, CIF for node 5, FAS for node 6, FOB for node 7, CFR for node 8, CFR for node 9, and FOB for node 10 are obtained. The objective function value of this test problem was found to be 170,654 cu. In the solution of 2 depots test problems, according to case 2, the solution time increased as the number of nodes increased from 7 to 12. Although the software was run for about 2 hours for the solution of the test problem 13 (node 13), no results were obtained.

In Table 9, test problems that include 3 depots and up to 12 customers were solved. For example, the test problem, which includes 3 depots and 11 nodes (8 customers), was solved, and it was obtained that 2 depots (2nd and 3rd) were used for the distribution. The 10th and 6th nodes were delivered using the 2nd depot and vehicle 1; the 4th and 9th nodes were delivered using the 3rd depot and vehicle 2; the 5th and 7th nodes were delivered using the 3rd depot and vehicle 3; and the 11th and 8th nodes were delivered using the 3rd depot and vehicle 4. No solution was reached within a reasonable time for the test problem with 3 depots and 13 nodes after 2 hours. The Multi-Depot Vehicle Routing Problem (MDVRP) and the Vehicle Routing Problem with a Heterogeneous Fleet (HFVRP) are two well-known NP-hard problems that are generalized in the MDHFVRP; therefore, the MDHFVRP is NP-hard [37]. Consequently, even with the most effective MIP solvers like CPLEX, it is hard, if not impossible, to reach an optimal solution for large-scale problems in a reasonable amount of computation time [38]. Therefore, heuristic and meta-heuristic

Table 9. Results of the test problems for 3 depots in case 2

Test problem	Depot	Node	Solution	Time (s)	Route	Incoterms
1	3	7	107,625	0.26	2-6-2 (vehicle 2) 3-4-3 (vehicle1) 3-5-7-3 (vehicle 3)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB
2	3	8	128,969	0.30	2-6-2 (vehicle 2) 3-4-8-3 (vehicle 1) 3-7-5-3 (vehicle 3)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8= CFR
3	3	9	140,622	7.46	2-6-2 (vehicle 4) 3-5-7-3 (vehicle 3) 3-8-3 (vehicle 1) 3-9-4-3 (vehicle 2)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR
4	3	10	161,110	6.09	2-10-6-2 (vehicle 1) 3-7-5-3 (vehicle 3) 3-8-3 (vehicle 4) 3-9-4-3 (vehicle 2)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB
5	3	11	179,429	6.76	2-10-6-2 (vehicle 1) 3-4-9-3 (vehicle 2) 3-5-7-3 (vehicle 3) 3-11-8-3 (vehicle 4)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB, Node 11= CFR
6	3	12	183,280	3093.63	2-6-10-2 (vehicle 1) 2-12-2 (vehicle 4) 3-4-5-9-3 (vehicle 3) 3-7-11-3 (vehicle5) 3-8-3 (vehicle 2)	Node 3= internal, Node 4=CFR, Node 5= CIF, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10= FOB, Node 11= CFR, Node 12= internal

Table 10. Results of the test problems for 4 depots in case 2

Test problem	Depot	Node	Solution	Time (s)	Route	Incoterms
1	4	7	73,934	0.15	4-5-7-4 (vehicle 3) 4-6-4 (vehicle 2)	Node 5= CFR, Node 6=FAS, Node 7= FOB
2	4	8	95,246	0.29	3-8-3 (vehicle 2) 4-6-4 (vehicle 1) 4-7-5-4 (vehicle 3)	Node 5= CFR, Node 6=FAS, Node 7= FOB, Node 8=CFR
3	4	9	106,566	2.75	3-8-3 (vehicle 2) 4-5-7-4 (vehicle 3) 4-9-6-4 (vehicle 4)	Node 5= CFR, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR
4	4	10	126,407	4.33	3-8-3 (vehicle 2) 4-5-7-4 (vehicle 3) 4-6-10-4 (vehicle 4) 4-9-4 (vehicle 2)	Node 5= CFR, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10=FOB
5	4	11	144,704	45.21	3-8-9-3 (vehicle 4) 4-5-4 (vehicle 2) 4-7-11-4 (vehicle 1) 4-10-6-4 (vehicle 3)	Node 5= CFR, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10=FOB, Node 11=CFR
6	4	12	148,762	1216.25	2-6-10-2 (vehicle 5) 3-8-3 (vehicle 4) 4-9-12-5-4 (vehicle 3) 4-11-7-4 (vehicle 1)	Node 5= CFR, Node 6=FAS, Node 7= FOB, Node 8=CFR, Node 9=CFR, Node 10=FOB, Node 11=CFR, Node 12=internal

Figure 3. Solution to test problem 2 depot 10 nodes for case 2 on the Türkiye map.

algorithms are proposed to address this group of large-scale problems [39].

In Table 10, test problems featuring 4 depots and up to 12 customers were solved. For example, a test problem that includes 4 depots and 10 nodes (6 customers) was solved, and it was found that both depots (2nd and 3rd) were used for the distribution. The 3rd node, was delivered using the 3rd depot and vehicle 4, the 5th and 7th nodes, were delivered using the 4th depot and vehicle 3, the 6th and 10th nodes, were delivered using the 4th depot and vehicle 4, and the 9th node was delivered using the 4th depot and vehicle 2.

In Figure 3, the test problem which includes 2 depots and 10 nodes (8 customers) was solved. The 7th and 5th nodes were serviced by the 1st depot and 3rd vehicle (orange lines). The 9th and 4th nodes were serviced by the 4th vehicle from the 1st depot (green lines). In addition, delivery to the 3rd and 8th nodes were distributed using the 1st vehicle (purple lines) and the 10th and 6th customers using the 5th vehicle (yellow lines) from the 2nd depot.

The specified problem structure and mathematical model can be integrated into different industries, besides ceramic industry products. For example, its use in different sectors such as distribution of agricultural products [40] in exports (considering cold chain costs), distribution of automobile service parts [41], and distribution of consumer durable goods will provide convenience and ideas to logistics service providers or companies. Cold chain transportation equipment requires different vehicle types suitable for the conditions of the route. Therefore, the costs of cold chain transportation are significant for industrial products. Transportation practices and transportation costs of different industries will be useful to consider when determining new distribution models.

CONCLUSION

In this study, a multi-depot, multi-product and heterogeneous fleet vehicle routing problem with delivery methods and costs based on the responsibilities of the seller and buyer is considered. A distribution model is presented for companies with production facilities and warehouses in different regions to conduct export operations using incoterms delivery methods.

- Investigations were conducted in terms of distribution costs under different scenarios (in case of different delivery terms to different customers).
- Vehicle routes are determined using different delivery terms according to the depot's and customer locations of the manufacturers who deliver to international companies.
- It is observed how the distribution system works in the case of using vehicles of different capacities and product types.

In future research, heuristic and meta-heuristic algorithms can be used to solve large-scale test problems. In the multi-product problem structure, different palletization systems of different sizes can be considered. New applications can be realized for various sectors, such as automotive, agricultural products, and consumer durable products, by considering due dates and time windows.

AUTHORS' CONTRIBUTIONS

All authors contributed to the study. Problem description G.O.G and S.K. Design and implementation of the model G.O.G. Analyses of the results and discussion G.O.G and S.K. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Dantzig GB, Ramser JH. The Truck Dispatching Problem. Manag Sci 1959;6:80–91. [CrossRef]

[2] Clarke G, Wright JW. Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. Oper Res 1964;12:568–581. [CrossRef]

[3] Braekers K, Ramaekers K, Nieuwenhuyse IV. The vehicle routing problem: State of the art classification and review. Comput Ind Eng 2016;99:300–313. [CrossRef]

[4] Li Y, Soleimani H, Zohal M. An improved ant colony optimization algorithm for the multi-depot green vehicle routing problem with multiple objectives. J Clean Prod 2019;227:1161–1172. [CrossRef]

[5] Torres JRM, Franco JL, Isaza SN, Jiménez HF, Padilla NH. A literature review on the vehicle routing problem with multiple depots. Comput Ind Eng 2015;79:115–129. [CrossRef]

[6] Weise T, Podlich A, Gorldt C. Solving Real-World Vehicle Routing Problems with Evolutionary Algorithms. New York: Springer; 2009. p. 29–53. [CrossRef]
[7] Crevier B, Cordeau JF, Laporte G. The multi-depot vehicle routing problem with inter-depot routes. Eur J Oper Res 2007;176:756–773. [CrossRef]
[8] Liu SC, Lee SB. A two-phase heuristic method for the multi-depot location routing problem taking inventory control decisions into consideration. Int J Adv Manuf Technol 2003;22:941–950. [CrossRef]
[9] Currie RH, Salhi S. Exact and heuristic methods for a full-load, multi-terminal, vehicle scheduling problem with backhauling and time windows. J Oper Res Soc 2003;54:390–400. [CrossRef]
[10] Kang KH, Lee YH, Lee BK. An Exact Algorithm for Multi Depot and Multi Period Vehicle Scheduling Problem. Springer 2005;350–359. [CrossRef]
[11] Ho W, Ho GTS, Ji P, Lau HCW. A hybrid genetic algorithm for the multi-depot vehicle routing problem. Eng Appl Artif Intell 2008;21:548–557. [CrossRef]
[12] Moin NH, Salhi S, Aziz NAB. An efficient hybrid genetic algorithm for the multi-product multi-period inventory routing problem. Int J Prod Econ 2011;133:334–343. [CrossRef]
[13] Aras N, Aksen D, Tekin MT. Selective multi-depot vehicle routing problem with pricing. Transp Res Part C Emerg Technol 2011;19:866–884. [CrossRef]
[14] Benslimane MT, Benadada Y. Ant colony algorithm for the multi-depot vehicle routing problem in large quantities by a heterogeneous fleet of vehicles. Inf Syst Oper Res 2013;51:31–40. [CrossRef]
[15] Ramos TRP, Gomes MI, Barbosa-Póvoa AP. Economic and environmental concerns in planning recyclable waste collection systems. Transp Res Part E Logist Transp Rev 2014;62:34–54. [CrossRef]
[16] Zhang Y, Chen XD. An Optimization Model for the Vehicle Routing Problem in Multi-product Frozen Food Delivery. J Appl Res Technol 2014;12:239–250. [CrossRef]
[17] Rahmani Y, Ramdane Cherif-Khettaf W, Oulamara A. The two-echelon multi-products location-routing problem with pickup and delivery: formulation and heuristic approaches. Int J Prod Res 2016;54:999–1019. [CrossRef]
[18] Tirkolaee EB, Goli A, Bakhsi M, Mahdavi I. A robust multi-trip vehicle routing problem of perishable products with intermediate depots and time windows. Numer Algebra Control Optim 2017;7:417–433. [CrossRef]
[19] Saragih NI, Bahagia SN, Suprayogi, Syabri I. A heuristic method for location-inventory-routing problem in a three-echelon supply chain system. Comput Ind Eng 2019;127:875–886. [CrossRef]
[20] Yao B, Chen C, Song X, Yang X. Fresh seafood delivery routing problem using an improved ant colony optimization. Ann Oper Res 2019;273:163–186. [CrossRef]
[21] Wang L, Kinable J, Van Woensel T. The fuel replenishment problem: A split-delivery multi-compartment vehicle routing problem with multiple trips. Comput Oper Res 2020;118:104904. [CrossRef]
[22] Moonsri K, Sethanan K, Worasan K, Nitisiri K. A Hybrid and Self-Adaptive Differential Evolution Algorithm for the Multi-Depot Vehicle Routing Problem in Egg Distribution. Appl Sci 2021;12:35. [CrossRef]
[23] Kaveh F, Shirouyehzad H, Zolfani SH, Arabzad SM. Proposing A Mathematical Model of Balancing the Inventory of Multi-Zone Bycle Sharing Systems with Mobile Stations and Applying Maintenance Constraints. Transport 2022;37:145–160. [CrossRef]
[24] Erdem M. Optimisation of the electric truck route for milk collection problem: a real case study. Transp Lett 2023;15:193–210. [CrossRef]
[25] Nguyen VS, Pham QD, Nguyen TH, Bui QT. Modeling and solving a multi-trip multi-distribution center vehicle routing problem with lower-bound capacity constraints. Comput Ind Eng 2022;172:108597. [CrossRef]
[26] Lin N, Akkerman R, Kanellopoulos A, Hu X, Wang X, Ruan J. Vehicle routing with heterogeneous service types: Optimizing post-harvest preprocessing operations for fruits and vegetables in short food supply chains. Transp Res Part E Logist Transp Rev 2023;172:103084. [CrossRef]
[27] Cil ZA, Oztop H, Diri Kenger Z, Kizilay D. Integrating distributed disassembly line balancing and vehicle routing problem in supply chain: Integer programming, constraint programming, and heuristic algorithms. Int J Prod Econ 2023;265:109014. [CrossRef]
[28] Suárez CA, Guaño WA, Pérez CC, López HR. Multi-objective optimization for perishable product dispatch in a FEFO system for a food bank single warehouse. Oper Res Persp 2024;12:100304. [CrossRef]
[29] Vasco AA, Rivera JC, Baldoquín MG. Effect of formulations over a Periodic Capacitated Vehicle Routing Problem with multiple depots, heterogeneous fleet, and hard time-windows. PLOS One 2024;19:e0311303. [CrossRef]
[30] Maleki A, Hemmati V, Reza Abazari S, Aghsami A, Rabbani M. Optimal distribution and waste management of Covid-19 vaccines from vaccination centers' satisfaction perspective – A fuzzy time window-based VRP. Transp Res Part E Logist Transp Rev 2024;183:103454. [CrossRef]
[31] Li N. A Two-stage Algorithm for Production Distribution Optimization of Fresh Products. Int J Ind Eng 2025;32.
[32] Li N, Wang Z. Vehicle routing problem for omnichannel retailing including multiple types of time windows and products. Comput Oper Res 2025;173:106828. [CrossRef]

[33] Davis J, Vogt J. Incoterms® 2020 and the missed opportunities for the next version. Int J Logist Res Appl 2022;25:1263–1286. [CrossRef]

[34] Hien N, Laporte G, Roy J. Business Environment Factors, Incoterms Selection and Export Performance. OSCM An Int J 2014;63–78. [CrossRef]

[35] Gursoy Y. Dış Ticaret İşlemleri. Bursa: Ekin Kitabevi; 2005. p. 136-137.

[36] Uslu A, Çetinkaya C, Isleyen SK. Vehicle Routing Problem in Post-Disaster Humanitarian Relief Logistics: A Case Study in Ankara. Sigma J Eng Nat Sci 2017;35:481–499.

[37] Bolaños RI, Escobar JW, Echeverri MG. A meta-heuristic algorithm for the multi-depot vehicle routing problem with heterogeneous fleet. Int J Ind Eng Comput 2018;9:461–478. [CrossRef]

[38] Salhi S, Imran A, Wassan NA. The multi-depot vehicle routing problem with heterogeneous vehicle fleet: Formulation and a variable neighborhood search implementation. Comput Oper Res 2014;52:315–325. [CrossRef]

[39] Mirabi M, Fatemi Ghomi SMT, Jolai F. Efficient stochastic hybrid heuristics for the multi-depot vehicle routing problem. Robot Com Int Manuf 2010;26:564–569. [CrossRef]

[40] Wang EJ, Tsai DM, Su TS, Lin KY. Simulated Annealing for Cost-Effective Transport of Live Aquaculture Products. Aquacult Econ Manag 2012;16:68–95. [CrossRef]

[41] Gao J, Gu F, Hu P, Xie Y, Yao B. Automobile chain maintenance parts delivery problem using an improved ant colony algorithm. Adv Mech Eng 2016;8:1687814016665297. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1344–1368, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2041

Research Article

Current research aspects in minimum quantity lubrication for turning operations: Cutting fluid as mono and hybrid nanofluid

Altaf NALBANDH[1,2], Neeraj CHAVDA[1,3,*], Rakesh BUMATARIA[2]

[1]Gujarat Technological University, Ahmedabad, Gujarat,382424, India
[2]Department of Mechanical Engineering, Government Polytechnic, Gujarat, 360577, India
[3]Department of Mechanical Engineering, G. H. Patel College of Engineering and Technology, The Charutar Vidya Mandal (CVM) University, Vallabh Vidya Nagar, Anand, Gujarat, 388120, India

ARTICLE INFO

Article history
Received: 19 October 2024
Revised: 07 December 2024
Accepted: 06 March 2025

Keywords:
Cutting Fluid; Hybrid Nanofluids; Minimum Quantity Lubrication; Nanofluids; Turning

ABSTRACT

The simplest machining method for eliminating undesired material from a workpiece using a single point cutting tool is turning. Appropriate cooling and lubrication save production lead time, machining costs, and environmental impacts controlled by cutting fluids. The impact of cooling cum lubrication techniques, such as minimum quantity lubrication with typical cutting fluids, minimum quantity lubrication with nanofluids, and minimum quantity lubrication with hybrid nanofluids, on turning performance metrics were thoroughly reviewed. The effect of these cooling solutions on results, such as cutting temperature, cutting force, surface roughness, tool wear, and chip morphology, has been extensively reviewed and addressed in the literature. Moreover, the nanoparticle types, size, concentration, base fluid types, and lubrication for nanofluid and hybrid nanofluids with minimum quantity lubrication have been considered during the study. The review of the relevant literature indicates that outcomes are significant. The impact of these cooling solutions on cutting temperature, cutting force, surface roughness, tool wear, and chip morphology has been thoroughly examined and addressed in the literature. The review of relevant literature indicates that outcomes are significantly improved when minimum quantity lubrication with nanofluids and hybrid nanofluids is contrasted with alternative cooling techniques. It is possible to reduce substantially pressures, cutting temperatures, surface roughness, and tool wear. Different kinds of nanoparticles, particle size, concentration, and nozzle inclination angle are the most critical factors in optimizing heat transfer rate and minimizing tool wear. Aluminium oxide (Al_2O_3) is the most commonly used nanoparticle in minimum quantity lubrication applications involving nanofluids and hybrid applications because it forms a coating between the workpiece and cutting tool contact, reducing friction. Additionally, the coefficient of friction significantly slowed. In addition, the minimum quantity lubrication with nanofluids also improves the colour and shape of the chip's morphology.

Cite this article as: Nalbandh A, Chavda N, Bumataria R. Current research aspects in minimum quantity lubrication for turning operations: Cutting fluid as mono and hybrid nanofluid. Sigma J Eng Nat Sci 2026;44(2):1344–1368.

***Corresponding author.**
*E-mail address: neeraj_chavda@yahoo.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

During each machining operation, heat is generated at the point of contact between the cutting tool and the workpiece, drastically lowering the tool›s service life and production rate. It increases cutting temperature and surface quality (roughness), and these are both undesirable in a production machining environment. The most straightforward and most popular cutting technique for removing unwanted material from a workpiece is turning, which is done with a conventional single-point cutting tool. It involves several machining processes, such as milling, grinding, and drilling. Cutting fluids are crucial for their ability to cool the interface between the cutting tool and the workpiece, remove chips, and lubricate. Dry or wet/flooded cooling methods are used, although dry cooling creates more heat than flooded cooling due to more friction. Using a blend of compressed air and standard cutting fluids, minimum quantity lubricant reduces the amount of coolant used compared to flood cooling [1].

Minimum Quantity Lubrication

The traditional coolant application system for turning operations is a flood-type system. The workpiece-cutting tool interface zone involves a substantial amount of cutting fluid jet; therefore, it is not economical, safe, or ecologically friendly. In addition, the accessibility of cutting fluid within the cutting zone is a concern. The flood cooling constraints could be circumvented by replacing the cutting fluid with wet machining or dry machining [2]. A novel, economical, and practical approach to minimum quantity lubrication machining technology is created to eliminate the issues above, produced using conventional cutting fluid in the metal industry [3]. The alternative strategy of mist lubrication, where an air and cutting fluid mixture is supplied to the cutting zone and has greater accessibility due to the high pressure, has proven to be a superior option based on the availability of coolant system applications [4]. The nozzle was employed to mix the lubricant/coolant with oil, generally known as an aerosol. To achieve better cooling results, spray the aerosol in the cutting zone at high pressure [5].

Minimum quantity lubrication combines compressed air and the minor oil in thin drops, generating a spray on the workpiece-tool cutting zone [6]. The oil flow rate in these systems is in the lower range, which is in contrast to the type of lubrication and cooling systems that are typically used. A typical lubricating system with a minimum quantity, as shown in Figure 1, consists of a pump, a fluid reservoir, an air-oil mixing unit, a compressor, and a nozzle, in addition to pipelines that connect all of the components. [7].

Minimum quantity lubrication cooling/lubrication systems are classified into two types: external and internal [1]. In Figure 2, It is illustrated here that a schematic diagram of the two ways that were discussed earlier is shown. The external minimum quantity lubrication method involves spraying an aerosol through an external nozzle. This method is being investigated by a number of researchers for its potential to improve turning operation. The atomization process generates aerosol, transforming the air and liquid mixture into mist or spray.

Several researchers have reported on the internal minimum quantity lubrication system, which uses aerosol supplied through an internal passage [9], [10]. There are two techniques to prepare aerosol: (i) external and (ii) internal. The aerosol was prepared externally and fed through internal channels in the cutting tool, referred to as external. The internal atomizer in a spindle mixes a mixture first, and secondly, it is supplied via an internal channel in the cutting tool, referred to as internal. The common machining processes mainly used by minimum quantity lubrication are turning, drilling, milling, grinding, and electro-discharge machining (EDM).

Nano Cutting Fluids

The heat is generated during the machining process due to friction in between cutting tool and workpiece. This heat

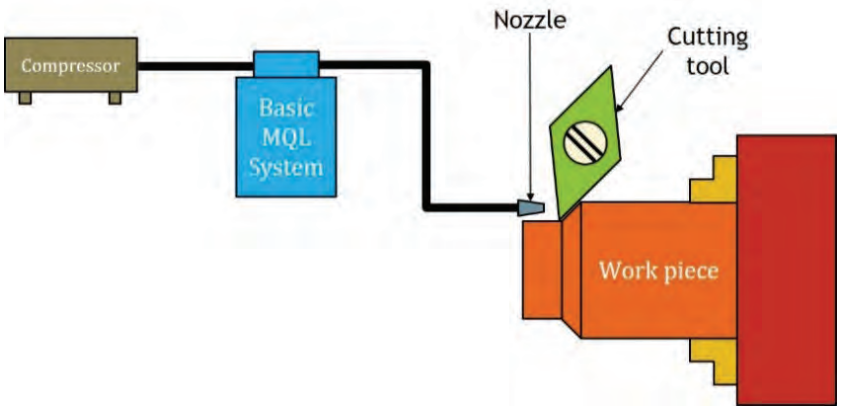


Figure 1. Typical arrangement of turning operation with a minimum quantity lubrication system.

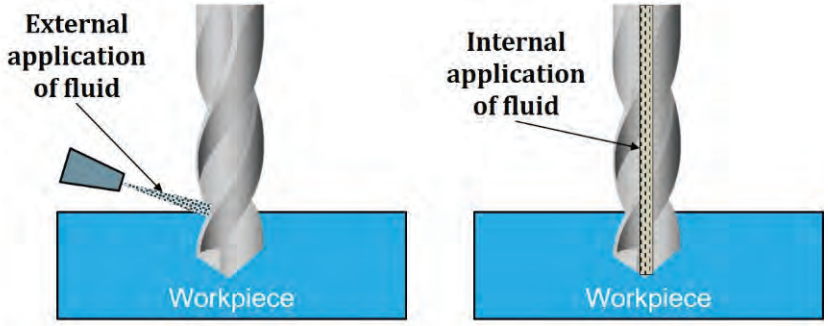


Figure 2. (a) External minimum quantity lubrication and (b) Internal minimum quantity lubrication cooling/lubrication technique [From Okokpujie et al. [8], with permission from from Elsevier.]

is undesirable as it retards cutting tool life and workpiece surface quality. Conventional cutting fluids with cooling and lubrication properties can reduce both heat and friction. Liquids separate surfaces by generating a layer, which reduces frictional resistance, and cooling capability assits to control the heat generation of the cutting tool, chip, and workpiece [11]. However, the application and disposal of these materials have detrimental consequences on personal health and the environment [12]. Nanofluid is a mixture of metal and non-metal nanoparticles (< 100 nm) in usually used cutting fluids [13]. The various shapes, such as cylindrical and spherical nanoparticles are produced from the pure metals (Au, Ag, Fe, Cu), metal oxides (Al_2O_3, CuO, Fe_2O_3, Fe_3O_4, MgO, SiO_2, TiO_2, ZnO), carbides (SiC, TiC), nitrides (AlN, SiN), and various kinds of carbon (graphite, diamond, single or multi-wall carbon nanotubes (MWCNT)) are combined with varying concentrations of base fluids: water, cutting oil, and vegetable oil [14].

Hybrid Nano-Cutting Fluids

Heat transfer between the workpiece-cutting tool interface affects cutting temperature, tool wear, and, consequently, tool life. Mono nanofluids are single-component nanofluids that have been examined for their enormous potential to retard machining output parameters. In addition, growth in nanofluid-based cutting fluids has been hampered by the enhanced demand for more excellent heat transfer capabilities of nanofluids. The creation of hybrid nanolfluids are achieved by combining several nanoparticles in varying proportions with a base fluid. Recent breakthroughs in nanotechnology employ hybrid nanofluids to resolve nanoparticle thermal conductivity and stability concerns [15].

The arrangement of information in the article is sequenced by (i) begins with an introduction, then moves on (ii) to separate headings discussing minimum quantity lubrication plus conventional cutting fluids (CFsMQL), minimum quantity lubrication plus nano cutting fluids (NFsMQL), and (iii) minimum quantity lubrication plus hybrid nano cutting fluids, and (iv) the impacts of various machining input parameters on cutting forces, cutting temperature, chip morphology, and surface roughness. It concludes by discussing the impacts of main input turning parameters on cutting temperatures, cutting forces, and surface roughness.

Nanofluids are a relatively recent practice that enhances machining efficiency and productivity. Varied types of nanofluids have a broad scope of applications in different kinds of machining operations to cool down and lubricate the heat-affected zone in machining. Numerous researchers have carried out experimental work & analysis on the application of different kinds of nano-fluids as a main cutting fluid in lathe, milling, and grinding processes using dry, wet, or minimum quantity lubrication methods and with different tool material, workpiece material, and machining parameters.

There are only a handful of studies that have been published in the academic literature that discuss the impact of various hybrid nanofluids, particularly the combination of Al_2O_3, Cu, CuO, ZnO, TiO_2, CNT, and MWCNT as a cutting fluid application in turning operations. For the purpose of utilizing the various cutting fluid combinations, it is necessary to make use of effective parameters such as surface roughness, cutting temperature, and cutting force as minimum quantity lubrication. An investigation into the similarities and differences between a number of cutting fluids, including conventional fluid, nanofluid, and hybrid nanofluid, has been carried out. Workpiece material, cutting tool material, and cutting fluids are all variables that are taken into consideration as effective parameters in this review process. Tabular representations have been used to summarize the most important findings. This article presents a statistical analysis and critique of the work, as well as future scope, economy, sensitivity analysis, and constraints. Additionally, the paper describes their shortcomings.

LITERATURE RELATED TO WORKING MEDIUM

MQL with Conventional Cutting Fluids

Dhar et al. [16] experimentally investigated that minimum quantity lubrication was turning retard cutting

temperature, surface roughness and tool wear in contrast to the dry & wet cooling environment in turning AISI-4340 steel. Scanning Electron Microscopy) SEM views of the cutting tool showed retardation in the notch growth and wore on the cutting edge. Dhar et al. [17] experimentally studied the impact of minimum quantity lubrication in machining AISI-1040 steel. The retardation in cutting temperature, good form, the colour of steel chips, and improvement in dimensional accuracy was obtained.

Bruni et al. [18] studied the impacts of wet, minimum quantity lubrication, and dry environments in turning AISI 420B stainless steel with both usual and inserts by wiper technology in turning machines equipped with a bed made of cast iron and polymer concrete. They conclude that the cooling environment does not impact tool wear. Dhar et al. [19,20] studied the operational performance of turning AISI-1040 steel using minimum quantity lubrication to dry lubricant. With the application of minimum quantity lubrication, results revealed that retardation of 5-15% in cutting force, enhancement in surface finish, and accuracy of dimensional parameter due to less flank wear and sizeable retardation in cutting temperature.

Davim et al. [20] experimentally turned commercial brass on CNC to compare the minimum quantity lubrication and wet lubricant system. They concluded that both lubricant systems had less variation in cutting power. The lowest cutting force at a rate of 200 m/min, surface roughness enhanced with an enhancement in feed rate and retardation with lubricant. Khan et al. [21] experimentally analyzed the effects of minimum quantity lubrication by vegetable oil on AISI 9310 alloy steel turning. In contrast to wet machining, they revealed a 10% variation in the temperature of the chip-tool interface, and the rear surfaces of chips generated with minimum quantity lubrication machining were better and smoother. Additionally, minimum quantity lubrication machining has been proven to prolong tool life and defer tool wear significantly.

Hwang et al.[22] studied minimum quantity lubrication and wet machining of AISI 1045 steel with an experimental plan designed by fractional factorial design, i.e., selection of significant factor with MINITAB, and central composite structure, i.e., finding of optimal level with Response Surface Methodology (RSM). It was found that experimental equations suggested by RSM validated cutting force in both lubricant machining but not for surface roughness. Leppert [23] studied the impacts of dry and minimum quantity lubrication cutting over conventional emulsion cooling in turning, while AISI 1045 steel was utilized as workpiece material. They concluded that a rise in cutting force in the absence or retardation of cutting fluid happened at a low feed rate. Significant results were obtained regarding surface roughness, surface topography, profile bearing ratio, and waviness with minimum quantity lubrication machining. Also, feed rate has a considerable influence on output machining parameters.

Che Haron et al. [24] studied the impacts of machining parameters on the surface entirety of Inconel 718. The RSM design of the experiment for a combination of surface roughness, microstructure, surface topography, and microhardness utilized the box-Behnken method. The surface roughness value found ranged from 0.4 to 2.9 meters. According to the ANOVA analysis, feed rate significantly impacted surface roughness. Leppert [25] experimentally looked into the impact of lubrication and cooling modes on the tool›s surface texture, cutting force, and topography of AISI 316L steel. Turning tests were performed in a variety of conditions, including dry (D), minimum quantity lubrication, and emulsion (E). The overall maximum (force F = 458 N) when the cutting interface zone was cooled in (E) lubrication scenario and the least (force F = 312 N) with (D), and minimum quantity lubrication had no significant effect on cutting force. In addition, it had been shown that minimum quantity lubrication at low value of machine feed rates and the employed cutting speeds led to retardation in surface roughness range from 1.34 to 1.50 μm in contrast to dry machining range from 1.54 to 1.82 μm and emulsion range from 1.68 to 2.26 μm. Analysis of surface topography using SEM images has also demonstrated a positive impact of the minimum quantity lubrication approach on conditions of surface quality.

Lohar [26] investigated the effects of wet and dry turning machining over minimum quantity lubrication when using the CBL tool to turn hardened alloy steel AISI 4340. When employing minimum quantity lubrication instead of dry and wet machining, Taguchi›s design and analysis of response factors led to a reduction of 40% in cutting force, a reduction of 36% in cutting temperature, and an improvement of 30% in surface finish. Hadad et al. [27] examined experimentally how machining forces, temperature, and surface roughness are affected by minimum quantity lubrication, nozzle position, and cutting parameters when turning AISI 4140 steel alloy. It is founded on results that nozzle position significantly affects low cutting temperature and retardation of cutting force in minimum quantity lubrication. Also, a temperature model was developed with a heat source and hind sink mechanism for machining plus minimum quantity lubrication, and this model confirmed the validity with experimental results.

Saini et al. [28] investigated how machining parameters affected the cutting forces and tool tip temperature when turning AISI 4340 steel with carbide inserts coated with PVD and CVD. The study was carried out to examine the impact of machine cutting speed, feed rate, and approach angle on temperature changes and the three cutting force components; experiments were devised so that machine cutting speed, feed rate, and approach angle were adjusted to four levels. In contrast, the depth of the cut remains constant. In terms of cutting force and temperature, PVD-coated carbide inserts outperformed CVD-coated carbide inserts. Agrawal et al. [29] experimentally evaluated the effects of aloe vera oil and servo cut s oil in machining M2

steel. Results were obtained after using Taguchi's design method as a design of experiments strategy, and it was found that surface roughness was retard by a value of 6.7%, and tool wear was retard by a value of 0.14% with aloe vera oil compared to conventional oil. Gordana et al. [30] compared previous literature on the impact of minimum quantity lubrication in turning carbon steel C45E with higher feed and cutting speed. Comparing minimum quantity lubrication to standard lubrication, the cutting force, tool wear, and surface roughness were typically found to be more favourable with minimum quantity lubrication lubrication, which was also cost-effective and environmentally benign.

Liu et al. [31] presented a more flexible Coupling Response Surface Methodology (CRSM) under minimum quantity lubrication conditions in machining very hard material titanium alloy with input factors like cutting speed, feed rate, and depth of cut. The Taguchi experimental design was employed with output parameters like machine processing efficiency, machine cutting force, and surface quality. Using ANOVA, the analysis found that feed rate was the primary factor affecting cutting force and surface quality. Additionally, it showed that these parameters had the minimum values at the least depth of cut, the least feed rate, and the highest cutting speed. New developed CRSM method found close agreement with the performed experimental value.

Masoudi et al. [32], experimentally in the machining of AISI 1050 steel, studied the impact of nozzle position (at a flank face, rake face, and both faces) with minimum quantity lubrication and studied hardness of workpiece, machining conditions, and types of tool each with four levels. It is predicated on improved cutting force reduction, rake face surface roughness, and optimal nozzle and tool wear rate reductions that are markedly accelerated by increased workpiece hardness. Sivaiah et al. [33] performed turning of Inconel 718 material, a hybrid textured tool (circular pitting holes) with minimum quantity lubrication declined cutting temperature, tool flank wear, and surface finish roughness by 36%, 59%, and 46%, respectively, contrasted to an untextured tool with minimum quantity lubrication environment.

Sivaiah et al. [34] investigated the cooling strategies for minimum quantity lubrication + textured tool and minimum quantity lubrication + untextured tool when turning AISI 52100 steel. They found the retardation in surface roughness, flank wear, and cutting junction temperature to be 42%, 40%, and 25%, respectively.

Singh et al. [35] investigated the effect of cutting parameters on the performance of turning Hastelloy C-276 in a vegetable oil minimum quantity lubrication environment using a CNMG120408 insert. With the developed RSM-based Central Composite Design (CCD) model, the optimal cutting temperature, surface roughness, and chip retardation coefficient were discovered in this setting as the optimal parameters.

Bhadoria et al. [36] studied the impact of the minimum quantity lubrication environment with ceramic insert in turning EN31 steel. The minimum quantity lubrication + ceramic insert condition helped produce the most petite surface roughness and lower flank wear. In addition, they demonstrated that high-speed cutting and feeding are feasible, enhancing production without sacrificing surface quality.

Sivaiah et al. [37] studied improved performance in machining AISI 304 steel with a circular micro pit textured tungsten carbide tool with minimum quantity lubrication. A new weighted gray relation analysis model was proposed for best-fit parameters. This cooling environment led to a 54 % decrease in surface roughness, a 7% decrease in tool wear, and a 27% enhancement in material removal rate.

It has been stated that a comparative analysis of the research works linked with minimum quantity lubrication using conventional cutting fluid has been carried out. This study was based on the material of the workpiece, the material of the cutting tool, the different types of cutting fluids, the mode of lubrication, and critical observations in Table 1.

Table 1. Comparative analysis of studies related to minimum quantity lubrication that use traditional cutting fluid

Researchers	Workpiece material	Cutting tool material	Types of cutting Fluids	Lubrication method	Findings	Year
Dhar et al. [16]	AISI-4340 steel	Carbide SNMM 120408	NM	minimum quantity lubrication, dry, wet	Retard CT, TW, and SR. Shown retardation in dimensional deviation.	2006
Dhar et al. [17]	AISI-4340 steel	Carbide SNMM 120408	NM	minimum quantity lubrication, dry, wet	Retard SR and TW. Shown retardation in dimensional deviation.	2006

Table 1. Comparative analysis of studies related to minimum quantity lubrication that use traditional cutting fluid (continued)

Researchers	Workpiece material	Cutting tool material	Types of cutting Fluids	Lubrication method	Findings	Year
Bruni et al. [18]	AISI 420B steel	CVD-coated carbide	Vegetable oil	minimum quantity lubrication, dry, wet	minimum quantity lubrication could not have a significant advantage over dry machining for TW and SR.	2006
Dhar et al. [19]	AISI-1040 steel	Carbide SNMM 120408	Mobil Cut-102	minimum quantity lubrication, dry	Retard CF by 5-15%, and CT. Retardation in SR, TW, and dimensional deviation was obtained.	2007
Davim et al. [20]	Commercial brass (CuZn39Pb3)	K10 inserts CC650 WG	Emulsion oil Microtrend 231L	minimum quantity lubrication, wet	Results showed a slight variation in SR and CP for a given lubrication mode.	2007
Khan et al. [21]	AISI 9310 alloy steel	Uncoated carbide SNMG 120408	Vegetable oil	minimum quantity lubrication, dry, wet	Retard CT by 10%, and CT. Retarded the SR and TW and eliminated built-up edge.	2009
Hwang, and Lee [22]	AISI 1045 steel	CNMG 120404 FG	NM	minimum quantity lubrication, wet	Regarding SR and CF, minimum quantity lubrication has more advantages than wet turning.	2010
Leppert [23]	AISI 1045 steel	MSS 2525–12-EB	Biodegradable vegetable oil	minimum quantity lubrication, dry, emulsion	Improved outcomes for SR and CF. Satisfactory results concerning ST.	2011
Che Haron et al. [24]	Inconel 718	CNMG 12 04 08-QM	NM	minimum quantity lubrication	FR showed a significant effect on SR, as supported by ANOVA.	2011
Leppert [25]	AISI. 316L stainless steel	Carbide SNMG 120408-TF	Accu-Lube LB 8000 oil	minimum quantity lubrication, dry, emulsion	minimum quantity lubrication won't impact CF as much as dry and wet machining. Reduce the SR and build-up edge.	2012
Lohar, and Nanavaty [26]	AISI 4340 steel	CBN TNMA 160404 PB250	Cutting oil	minimum quantity lubrication, dry, wet	Retard CF by 40%, CT by 36%, and SR by 30%.	2013
Hadad, and Sadeghi [27]	AISI 4140 steel	HSS tool steel	Ester oil (RS 1642)	minimum quantity lubrication, dry, wet	Retard CT, CF, and SR. It showed that the minimum quantity lubrication nozzle position has influenced CT and SR.	2013
Gordana et al. [30]	Carbon steel C45E	Carbide SNMG 1204 08 NMX	NM	minimum quantity lubrication, wet	minimum quantity lubrication is an excellent alternative to wet lubrication in CF, TW, and SR. Better results were obtained in machining cost and ecology when minimum quantity lubrication was used.	2013
Liu et al. [31]	Ti-6Al-4V alloy	Coated tungsten carbide	Vegetable oil	minimum quantity lubrication, dry, wet	Shown enhancement in CS and DC leads to the rise in CF. SR was affected mainly due to CS.	2013
Saini et al. [28]	AISI 4340 steel	Coated carbide (PVD and CVD coated)	Soluble cutting oil	minimum quantity lubrication, dry	Compared to CVD-coated inserts, better results were shown in PVD-coated inserts in CF and CT. minimum quantity lubrication improved performance in CVD-coated inserts for CF and PVD-coated inserts for CT.	2014

Table 1. Comparative analysis of studies related to minimum quantity lubrication that use traditional cutting fluid (continued)

Researchers	Workpiece material	Cutting tool material	Types of cutting Fluids	Lubrication method	Findings	Year
Masoudi et al. [32]	AISI 1050 steel	HSS and Tungsten carbide	Ester oil	minimum quantity lubrication, dry, wet	minimum quantity lubrication had significantly retard CF, SR, and TW. minimum quantity lubrication›s nozzle position produced superior outcomes at the rake face as opposed to the flank face.	2017
Agrawal, and Patil [29]	M2 steel	Carbide TNMG	Aloe vera oil, and Servo cut S oil	minimum quantity lubrication.	Lowered SR by 6.7% compared to servo cut S oil cutting fluid, and SR decreased as CS enhanced. TW lowered by 0.14% in aloe vera oil cutting fluid.	2018
Sivaiah et al. [33]	Inconel 718	Tungsten carbide (PVD coated)	Emulsion based flood	minimum quantity lubrication	minimum quantity lubrication with the hybrid textured tool decreased CT, TW, and SR by 36%, 59%, and 46% compared to the untextured device.	2019
Sivaiah et al. [34]	AISI 52100 steel	Tungsten carbide	Emulsion based flood	minimum quantity lubrication, wet	minimum quantity lubrication + textured tool cooling strategy showed retardation in CT, TW, and SR compared to minimum quantity lubrication + untextured tool. minimum quantity lubrication technique uses less fluid consumption than conventional cooling and is more environmentally friendly.	2020
Singh et al. [35]	Hastelloy C-276	CNMG uncoated	Vegetable oil	minimum quantity lubrication, wet, dry	Less CT, SR, and lower chip retardation coefficient were achieved with minimum quantity lubrication. An optimum parameter value was also found from the developed RSM-based CCD model.	2021
Bhadoria, and Bartarya [36]	EN31 steel	Uncoated ceramic insert	Servocut s	minimum quantity lubrication, dry	The most negligible value of SR and lower TW was found in the minimum quantity lubrication environment. minimum quantity lubrication revealed that high speed and feed working without compromising surface quality are indicators of better productivity than dry machining.	2021
Sivaiah et al. [37]	AISI 304	Tungsten carbide coated	Emulsion based flood	minimum quantity lubrication.	A new weighted gray rational model was developed for parameter fitment. Retardation in TW, SR and significant improvement in material removal rate was obtained.	2021

The concise published review articles relate to minimum-quantity lubrication with conventional cutting fluids [4], [5], [38]. The effects of minimum-quantity lubrication, in contrast to dry and flooded cooling methods, in the primary machining processes of turning, drilling, grinding, and milling are the subject of important reviews.

MQL with Nanofluids (MQLNFs)

Krishna et al. [39] investigated experimentally turning AISI 1040 steel using nano boric acid (50 nm) lubricant suspensions in base oils such as SAE-40 and coconut oil. Nanofluid was prepared with different weight propositions at room temperature and mixed with a sonicator for one hour. The impacts of variations in machine

cutting temperature, surface roughness, and average tool flank were studied. Rao et al. [40] investigated how cutting tool-workpiece cutting pressures, cutting temperatures, sample surface quality, and tool wear were affected by Carbon Nanotube (CNT) nanofluid when turning AISI 1040 steel. It was found in the results that the above parameters had a good effect, up to 2% CNT inclusion. Vasu et al. [41] investigated experimentally how various conditions—dry, minimum quantity lubrication, and minimum quantity lubrication + Al_2O_3 nanoparticles—affect temperature dissipation, cutting force, tool wear, and surface roughness when turning Inconel 600 alloy. Taguchi method was used for experiments design with input machining parameters and S/N ratios for different environmental conditions, i.e., dry, minimum quantity lubrication, minimum quantity lubrication with 4v% of Al_2O_3 nanoparticles, and minimum quantity lubrication with 6v% of Al_2O_3 nanoparticles.

Khandekar et al. [42] studied the impact of nanofluid as a cutting fluid, prepared by adding 1wt% Al_2O_3 nanoparticles fluid on wear, cutting force, workpiece surface roughness, and chip thickness in turning of AISI 4340. Comparative studies of tools amid dry cutting, conventional cutting fluid, and nanofluid as a cutting fluid were studied. From experimental data, it was found that retardation of 50% and 30% in cutting force retardation in surface roughness by 54.5% and 28.5%, respectively. Amrita et al. [43] studied mist using nanographite (80 nm) when turning AISI 1040 steel with cemented carbide and HSS; soluble oil was tested in lubricating circumstances such as dry, flood, and mist without nanographite. The results showed significant retardation in tool wear and cutting forces with both cutting tools compared to dry and flood lubrication conditions. Furthermore, the findings of the experiment showed that using 0.5 weight per cent nano graphite with soluble oil and both cemented carbide and HSS cutting tools resulted in the least amount of cutting forces and tool wear at a flow rate value of 15 ml/min. Amrita et al. [44] studied the usage of Graphite (80 nm), boric acid (100 nm), and MoS_2 (100 nm) nanoparticles in base fluid as emulsifier oil with minimum quantity lubrication in turning of AISI 1040 steel. The stability of nanofluids containing 0.3 wt % nanoparticles was evaluated, and these cutting fluids were administered at 10 ml/min.

Sayuti et al. [45], examined experimentally how SiO_2 nano lubricants with minimum quantity lubrication affected the AISI 4140 workpiece›s machining. To make sure the SiO_2 particles were uniformly suspended, the mixture was sonified for 48 hours at 240 W, 40 kHz, and 500 W after adding an average of 5–10 nm-sized particles to mineral oil. The experimental design was based on the Taguchi design (L16 orthogonal array). The fuzzy model was trained to determine the association between the output parameters (surface roughness and tool wear) and the input parameters (nano lubrication fraction, air pressure, and nozzle angle). A 60° nozzle tilt, a 2 bar air stream pressure, and a 0.5% weight concentration of nanoparticles in the mineral oil produced the least degree of tool wear. On the other hand, surface quality was improved by a 30° nozzle orientation angle, low air stream pressure, and 0.5 weight per cent nanoparticles in the mineral oil. Shabgard et al. [46] studied the impact of base fluid, a soluble cutting fluid, and CuO nanoparticles on cutting force and surface roughness when turning AISI 4340 steel. For contrasting nanofluid with soluble and dry cutting fluids, tests were conducted. The findings show that adding a 1 wt% fraction of CuO nanoparticle as soluble oil to the base fluid lessens machining cutting force and surface roughness. Furthermore, CuO nanofluid drastically reduces surface quality and machining cutting force by 49% and 24%, respectively. Khalil et al. [47] investigated the impact of Al_2O_3 on tool wear using AISI 10150 mild steel as the workpiece and SDBS + minimum quantity lubrication as the cutting fluid. It is resulted that tool wear got retarded, and enhancement of tool life.

Sharma et al. [48] performed experimentation with the MWCNT's effect on AISI D2 steel spinning at different speeds. MINITAB tests were designed using Taguchi's method. NFminimum quantity lubrication is superior to traditional minimum quantity lubrication in surface quality and cutting junction temperature. Sharma et al. [49] and [50] investigated how minimum quantity lubrication affected the machining of AISI 1040 steel using TiO_2 and Al_2O_3 nanoparticles distributed in a base fluid as a vegetable oil + water emulsion. It resulted that with comparison to dry, mist, and wet machining, nanofluid machining resulted in considerable retardation in surface roughness and tool wear, as well as betterment in chip quality (silver-coloured chips with a curl and segmented morphology). Gupta et al. [51] investigated experimentally the effects of cutting fluid with minimum quantity lubrication on Al_2O_3 (40 nm), MoS_2 (40 nm), and graphite (40 nm) nanoparticles. For Titanium (grade-2) alloy turning, the following factors were selected: machining cutting force, cutting temperature, tool wear, and surface finish. RSM and combined goal functions were utilized to create experiments. Particle Swarm Optimization (PSO) and Bacterial Forging Optimization (BFO) approaches produced superior experimental outcomes, while graphite-based nanofluids were adjusted to decrease output responses.

Paturi et al. [52] demonstrated experimentally the impact of Tungsten disulfide WS_2 as a solid lubricant in turning Inconel 718. Nanofluid for minimum quantity lubrication applications was created by adding nano additives in emulsifier oil at a fraction of 0.05wt%. Multi-liner models were developed and validated using a correlation between cutting parameters. Compared to conventional minimum quantity lubrication machining, nanofluid-based minimum quantity lubrication machining retard surface roughness by an average of 35%. Rapeti et al. [53] applied based fluids such as coconut oil, canola oil, and sesame oil with nano-cutting fluids containing MoS_2 nanoparticles to AISI 1040 steel turning. Experiment design, lubricant, nanoparticle additions, and cutting velocity were optimized

using Taguchi's L27 orthogonal array. For statistical analysis, they considered feed rate to be four separate variables, each with three distinct values. In addition, Grey Relational Analysis, which has its roots in the work of Taguchi, was applied to achieve multi-objective optimization. Coconut oil + 0.5% nano MoS_2 improved machining performance at the value of 40 m/min machine cutting speed and a 0.14 mm/rev machine feed rate. A cost estimate of its application further assessed nano-cutting fluids› industrial viability.

Patole et al. [54] experimentally studied the machining parameters optimization in minimum quantity lubrication using nanofluids suspended with MWCNT (15 nm) for turning of AISI 4340 steel workpiece. Experiments involving cutting technique and machining surface roughness are examined using Grey Relational Grade as a response variable. The optimal conditions for the response variable›s value were ascertained by the application of Grey Rational Analysis (GRA). Also, the S/N ratio of GRA yielded excellent results; therefore, the ANOVA findings nearly matched GRA. In addition, a comparison investigation revealed a superior surface roughness in minimum quantity lubrication, including nanoparticles of MWCNT and base fluid ethylene. Singh et al. [55] investigated practically the turning of AISI 304 steel using minimum quantity lubrication and a nanofluid made by mixing Servo Cut base oil with graphene nanoplatelets. RSM was utilized to plan the trials, and testing was done by choosing the orthogonal array L27. ANOVA showed that the concentration of graphene nanoparticles was essential for reducing cutting temperature and surface roughness at higher feed rates and machining cutting speeds.

Joshi et al. [56] experimentally studied the impact of minimum quantity lubrication, dry machining, and minimum quantity lubrication with nanofluid (Al_2O_3 + vegetable oil as base fluid) on the Inconel 600 work material›s surface roughness during a turning operation. They eventually reached the point where a predictive model had been built for it. The results demonstrated that using nanofluid improved the surface quality when compared to machining done under dry machining and minimum quantity lubrication settings. When it came to evaluating surface roughness, Taguchi›s method worked effectively in this investigation. Vishnu et al. [57] experimentally studied a comparison of EN 353 Steel Alloy turning under dry cutting, flooded cutting, and minimum quantity lubrication circumstances. Al_2O_3 nanoparticles were combined with water and stabilizers to produce nanofluids. For the experiment design, the Taguchi method was utilized. Comparisons were made between the dry cutting, flooded cutting, and minimum quantity lubrication conditions; minimum quantity lubrication in machining is a cost-effective, environmentally friendly, and human health-friendly option to complete dry cutting and flood lubricating conditions. It was observed that machining under flooded conditions is superior to dry and minimum quantity lubrication conditions, while the difference between minimum quantity lubrication and flooded conditions is negligible. It was also determined that the impact of lubrication conditions on cutting temperature is more significant than that of tool type, depth of cut, feed rate, and cutting speed.

Pasam et al. [58], experimentally studied developing and testing vegetable oil as a based fluid with micro and nano cutting fluids for turning of AISI 1040 steel in minimum quantity lubrication. Both boric acid (100 nm) and MoS_2 (90 nm) were dispersed in a base fluid as coconut oil. Nanofluids were superior to microfluidics at high cutting rates, but both performed similarly at lower speeds. Furthermore, MoS2 was more effective than boric acid. Ganesan et al. [59] experimentally examined the impacts of nanofluid with copper (50 nm) suspension-based minimum quantity lubrication under dry and wet conditions on H 11 steel turning. Taguchi L18 orthogonal array was used to create experiments, and RSM was utilized to determine the optimal value and mathematical model. Results demonstrated a 66% retardation in tool wear and a 40% retardation in surface roughness. The SEM examination of chip morphology revealed a modest number of notched teeth.

Padmini et al. [60] focused on the performance testing of Molybdenum disulfide-based nanofluids with various nanoparticle inclusions in vegetable oils (coconut, sesame, and canola) as a base fluid for machining AISI 1040 steel in a minimum quantity lubrication environment. Consequently, adding the element of nanoparticles enhances essential nanofluid properties. Cutting temperature, tool wear, cutting force, and surface roughness significantly decreased while 0.5% MoS_2 was blended into coconut, confirming better performance compared to additional lubricating strategies. Sharma et al. [61] studied various concentrations of SiO_2 nanoparticles in vegetable oil and water emulsion base fluid to affect the turning operation of AISI 1040 steel. Dry and minimum quantity lubrication machining procedures were tested in a variety of experiments. It has been demonstrated that dry machining and nanofluid + minimum quantity lubrication reduce cutting force compared to operation in conventional cutting fluids. Furthermore, nanofluid + minimum quantity lubrication retards surface roughness comparable to wet, traditional cutting fluid, and dry machining conditions. Sivalingam et al. [62] improved the machinability of Inconel 718 with a dry and atomized spray cooling environment. In spray cutting fluid was prepared by mixing graphite and molybdenum disulfide in vegetable oil, and experimentation revealed a 17-34% betterment in surface roughness, a decrease in tool wear, an enhancement in tool life, and retardation in machining costs compared to a dry cooling environment.

Das et al. [63] studied how nanofluids like zinc, copper, iron, and aluminium oxide affect turning machining parameters for AISI 4340. Different fluid characteristics were examined. After experiments with minimum quantity lubrication, Al_2O_3 nanofluid was found to have the highest force value, while CuO nanofluid had the lowest force value. After being treated with CuO nanofluid, the flank

area saw the least amount of significant wear in tool and the smoothest surface roughness rating. Venkatesan et al. [64] examined the comparative and performance analysis of hard-to-cut material Hastelloy between dry and minimum quantity lubrication cooling conditions with Physical Vapour Deposition (PWD) and Chemical Vapour Deposition (CVD) coated Tools. In coconut oil as a base fluid, amounts of 0.25 wt% and 0.5 wt% of Hexagonal Boren Nitride (hBN) nanoparticles were added for this study. Ramanan et al. [65] designed an experiment to determine the effect of Al_2O_3 nanofluid with minimum quantity lubrication on the turning of Incoloy 800. Results showed that adding 1 % Al_2O_3 improved the wettability of the cutting fluid, decrement in the cutting force, reduced the surface roughness to a negligible value, and slowed the pace at which the tool wore out in the cradle.

Nikouei et al. [66] experimented with the impact of residual stress on turning the Inconel 718 with different nanoparticle sizes, materials, base fluid, and content. It was founded on a result that TiO_2 nanoparticle-induced fluid machining decreased residual stress from 20.03 % to 11.44 %. For the purpose of providing a comprehensive review of the research works associated with minimum quantity lubrication that contains nanofluid, taking into consideration the materials of the workpiece, the materials of the cutting tool, nanoparticles, size, form, base fluid, and the findings that are compacted with Table 2.

Table 2. Comprehensive overview of the works connected to minimum quantity lubrication that contains nanofluid

Researchers	Workpiece material	Cutting tool material	Nanoparticles types and their size (nm)	Types of base Fluid	Findings	Year
Krishna et al. [39]	AISI 1040 steel	Carbide SNMG 120408	Nanoboric acid (50)	SAE-40 oil and coconut oil	Retard CT, TW, and SR with nanofluids. The better performance showed in coconut oil-based lubricant compared to SAE oil-based lubricant.	2010
Rao et al. [40]	AISI 1040 steel	HSS and cemented carbide	CNT (NM)	NM.	Retard CT and TW up to 2% CNT inclusion, and not much change beyond this range.	2011
Vasu, and Reddy [41]	Inconel 600 alloy	Coated carbide CNMG 1 20 408	Al_2O_3 (NM)	Vegetable oil (Coolube 2210)	Shown that FR is the main parameter that influences SR, DC affects CT and FR, and DC influences CF as well TW among all three parameters. It retard SR, CT, CF, and TW by machining under minimum quantity lubrication nanofluid conditions.	2012
Khandekar et al. [42]	AISI 4340 steel	Uncoated cemented carbide	Al_2O_3 (NM)	Servo cut-S	TW retarded while machining with 1v% nanofluid. When using nanofluid, CF and SR are slowed by 50% and 30%, respectively, as compared to dry and traditional fluid machining.	2012
Amrita et al. [43]	AISI 1040 steel	CNMG 120408, and HSS	Graphite (<80)	Soluble oil	Minimum TW and CF with 0.5wt% nano graphite soluble oil for both tools were obtained. The nanofluid application showed improvement in terms of CT, TW, and CF.	2013
Amrita et al. [44]	AISI 1040 steel	CNMG 120408 TTS	Graphite (80), boric acid (100), MoS_2 (100)	Emulsifier oil	For TW, SR, CF, and MoS_2 showed better wear, surface, and force properties, respectively. For CT, boric acid showed better heat dissipation properties.	2014

Table 2. Comprehensive overview of the works connected to minimum quantity lubrication that contains nanofluid (continued)

Researchers	Workpiece material	Cutting tool material	Nanoparticles types and their size (nm)	Types of base Fluid	Findings	Year
Sayuti et al. [45]	AISI 4140 steel	Sandvik DNMG 150608 PM	SiO_2 (5-15)	Mineral oil	Minimum TW and SR with a 0.5%wt SiO_2 concentration and 60° and 30° nozzle orientation angle, respectively, were obtained.	2014
Shabgard et al. [46]	AISI 4340 steel	DNMG 150604-QM	CuO (NM.)	Soluble cutting fluid	CF and SR were retard by 24% and 49% when machining with a cutting fluid containing 1% CuO instead of soluble oil and dry machining.	2014
Khalil et al. [47]	AISI 10150 mild steel	Carbide	Al_2O_3 (NM.)	Soluble cutting oil	Retard TW when turning with Al_2O_3 and Al_2O_3 + surfactant cutting fluid.	2015
Sharma et al. [48]	AISI. D2 steel	Tungsten carbide (CNMG 120408)	MWCNT (NM)	SAE. 20W40 oil	Showed CT could be decreased with the addition of MWCNTs in cutting fluid. Retardation in SR as retardation in CT and less TW was obtained.	2015
Sharma et al. [49]	AISI 1040 steel	Uncoated cemented carbide	TiO_2 (NM)	Vegetable oil-water emulsion	Retard TW up to 58.1% and 35.85%, and CF up to 62.67% and 34.88% with nanofluid machining compared to dry machining and mist machining, respectively. Nano-cutting fluid machining retard average SR by 47.8%, 39.1%, and 25.5% compared to dry, mist, and wet machining and produced silver colour chips with curl and segmented shapes.	2016
Sharma et al. [50]	AISI 1040 steel	Uncoated cemented carbide	Al_2O_3 (45)	Vegetable oil-water emulsion	Retard TW up to 59.1%, 29.2%, and 28.6%, and CF up to 63.9%, 44.9%, and 5.27% with nanofluid cooling, as compared to machining conditions like dry, mist, and wet, respectively. Nano-cutting fluid machining retard average SR by 47.8%, 29.1%, and 25.5% compared to dry, mist, and wet machining and produced silver colour chips with curl and segmented shapes.	2016
Gupta et al. [51]	Titanium (grade-2) Alloy	Cubic boron nitride	Al_2O_3 (40), MoS_2 (40), Graphite (40)	Vegetable oil	PSO and BFO enhance CF, TW, SR, and CT in graphite-based nanofluids. CF, CT, TW, and SR increase with CS and FR and reduce with cutting fluid.	2016
Paturi et al. [52]	Inconel 718	Carbide SNMG 120408	Tungsten disulfide WS_2 (NM)	Emulsifier oil	Optimal machining parameters were obtained using Taguchi design and ANOVA for SR. It retard SR by 35% with WS_2 lubricant-assisted + minimum quantity lubrication machining compared to minimum quantity lubrication machining.	2016

Table 2. Comprehensive overview of the works connected to minimum quantity lubrication that contains nanofluid (continued)

Researchers	Workpiece material	Cutting tool material	Nanoparticles types and their size (nm)	Types of base Fluid	Findings	Year
Rapeti et al. [53]	AISI 1040 steel,	Coated carbide CNMG 120408 NC6110	MoS_2 (NM)	Coconut oil, sesame oil, and canola oil	ANOVA and ANOM showed that coconut oil+0.5% MoS_2 machining retard CF, CT, TW, and SR. Obtained optimum condition at CS of 40 m/min and FR of 0.14 mm/rev for retardation in CF, CT, and TW, and 100 m/min, and CS value of 100 m/min, and FR value of 0.14 mm/rev for minimum SR.	2016
Padmini et al. [60]	AISI 1040 steel	Coated carbide CNMG 120408 NC6110	MoS_2 (NM)	Coconut oil, sesame oil, and canola oil	Enhanced flash, fire points, specific heat, specific density, thermal conductivity, and heat transfer coefficients were obtained as the percentage of nanoparticles increased. Retard CF, CT. TW and SR at coconut oil+0.5% MoS_2 compared to other cooling conditions.	2016
Patole, and Kulkarni [54]	AISI 4340 steel	Tungsten carbide coated CNMT 090308	MWCNT (15)	Ethylene glycol	Obtained optimized machining parameters using grey relational analysis (GRA). minimum quantity lubrication with nanofluid showed retardation in SR and CF from a comparative analysis of all lubrication techniques.	2017
Sharma et al. [61]	AISI 1040 steel	Uncoated cemented carbide	SiO_2 (10)	Vegetable oil-water emulsion	Nano-cutting fluid machining lowered TW by 57.3% and 34.6% compared to dry and mist machining. Nano-cutting fluid machining retarded average SR by 15%, 19.05%, and 40%. It created smooth silver chips with curl and segmented morphologies.	2017
Singh et al. [55]	AISI 304 steel	Uncoated cemented carbide	Graphene nanoplatelets (30–40)	Servo Cut S oil	Higher value concentration of graphene nanoparticles showed that a higher concentration retard CT at a lower CS, FR, and DC. Retards the SR with a higher value concentration of graphene nanoparticles at higher CS, lower FR, and DC.	2018
Joshi et al. [56]	Inconel 600	Uncoated carbide	Al_2O_3 (NM)	Vegetable oil	Nanofluid was performed dry cutting, and minimum quantity lubrication with vegetable oil cutting. Comparing dry and vegetable oil+minimum quantity lubrication machining, nanofluids+minimum quantity lubrication slowed SR. The hybrid fuzzy controller closely matched experimental data.	2018

Table 2. Comprehensive overview of the works connected to minimum quantity lubrication that contains nanofluid (continued)

Researchers	Workpiece material	Cutting tool material	Nanoparticles types and their size (nm)	Types of base Fluid	Findings	Year
Vishnu et al. [57]	EN353 Alloy Steel	CNMG carbide uncoated, CVD, and PVD	Al_2O_3 (NM)	Water	Obtained best and optimal combination for reducing CT at flooded conditions, CS of 1100 rpm, FR of 0.2 mm/rev, DC of 0.5 mm, and PVD coated tool. Shown better machining condition in flooded condition, but a minimal difference in values between minimum quantity lubrication and flooded need.	2018
Pasam et al. [58]	AISI 1040 steel	Tungsten carbide PVD coated	Boric acid (100), and MoS_2 (90)	Coconut oil	Observed nanofluids performed better than micro fluids at higher CS and retard CF, CT, SR, and TW. MoS_2 performed better than boric acid lubricants. However, microfluidics is far more economical and viable than nanofluids.	2018
Ganesan et al. [59]	H11 steel	Carbide	Copper (50)	Ethylene glycol	It decreased SR by 40% with copper nanofluid machining with FR of 0.1 mm/rev and CS of 209 m/min. Retardation in TW by 66% with copper nanofluid machining was obtained. RSM supported the mathematical model, and copper nanofluids+minimum quantity lubrication enhanced cooling, which resulted in a small number of notched tooth chips.	2018
Sivalingam et al. [62]	Inconel 718	Reinforce ceramic insert	Graphite (30-45), and MoS_2 (10-15)	Vegetable oil	Improvement in SR and TL and retardation in TW and machining coast were obtained compared to the dry cooling environment.	2020
Das et al. [63]	AISI 4340	Uncoated cermet	ZnO, CuO, Fe_2O_3, and Al_2O_3 (NM)	Deionized water	In minimum quantity lubrication nanofluid cooling, the higher value of CF was found with Al_2O_3 and the least with CuO. In contrast to the remaining three nanofluids, the least TW was obtained with CuO. The lowest value of SR resulted in CuO nanofluid.	2020
Venkatesan et al. [64]	Hastelloy X	Cemented Carbide CVD and PVD coated	hBN (30-100)	Coconut oil	Weight concentration, 0.25 % of hBN nanofluid with minimum quantity lubrication PVD carbide coated tool showed retardation in SR, CF, and TW compared to minimum quantity lubrication + CVD carbide coating at 0.25 % and dry PVD carbide coating.	2020
Ramanan et al. [65]	Incoloy 800	CNMG PVD coated	Al_2O_3 (20)	Semi-synthetic soluble oil	minimum quantity lubrication with 1% Al_2O_3 nanoparticles improved wettability and the most negligible value of SR. CF and TW were retard compared to dry, flood, and minimum quantity lubrication strategies.	2021
Nikouei et al. [66]	Inconel 718	Carbide coated	TiO_2 (10, 30, 50)	Water	Nanoparticle application got better results in terms of residual stress average from 20.03 to 11.44%	2021

Significant reviews [1], [11], [48], [67], [68] are focused on nanofluid preparation, nanoparticle concentration, nanofluid with minimum quantity lubrication, and performance of output parameters like machining cutting force, cutting temperature, surface roughness, tool wear, and chip morphology in the machining process, namely; turning, milling, grinding, and drilling with an application of nanofluid + minimum quantity lubrication in compared to conventional cutting fluid with minimum quantity lubrication, dry, and wet lubricant system.

MQL with Hybrid Nano Cutting Fluids (MQLHyNFs)

Sharma et al. [69] performed experimentation for analyzing impact of Al_2O_3 (45 nm) and MoS_2 (30 nm) with a fraction of 9:1 to get dispersed hybrid nanofluid in turning operation of AISI 304 steel. Nanoparticle fractions of 0.25, 0.75, and 1.25 v% were employed. Significant retardation in Fx. Fy, Fz, and Ra were 18.08%, 5.73%, 7.35%, and 2.38%, respectively. Singh et al. [70] studied the simulations of the impacts of incorporating graphene nanoplatelets into Al-based nanofluid at fractions of 0.25, 0.75, and 1.25 v% in the construction of a hybrid nanofluid. With an enhancement in nanoparticle concentration, thermal conductivity, viscosity, and wear were all improved. The results of comparing alumina-based nanofluid and base fluid demonstrate that the hybrid has superior wettability. The results showed a decrease in surface roughness by 20.28 %, a reduction in thrust force value by 17.38 %, a decrease in cutting force value by 9.94 %, and a decrease in feed force by 7.25 %.

Sharma et al. [71] checked the impact of hybrid nanofluids made of Al_2O_3 and GnP in turning AISI 304 steel. The development of Hybrid nanofluids was achieved with varying concentrations of 0.25, 0.75, and 1.25 v%. Hybrid nanofluids had the lowest pin wear and friction coefficient per tribological and wettability testing. Results revealed retardation of 5.29% in cutting temperature and 12.29% in tool wear with nanofluids compared to Al-based nanofluid. Jamil et al. [72], experimentally studied the usefulness of cryogenic cooling (CO_2) and minimum quantity lubricationHF in turning tungsten alloys. Al_2O_3 (30 nm) and MWCNT (10-30 nm) are suspended in vegetable oil base fluid for nanofluid preparation. Experiments were designed using Taguchi's L9 orthogonal array with cutting speed, feed rate, and cooling techniques as design variables. It was seen from the results that mean surface roughness was retard by 8.72%, machining cutting force by 11.8%, and tool life enhanced by 23% with hybrid nanofluids in contrast to CO_2 cooling. Although, retardation of 11.2% in cutting temperature was seen with CO_2 cooling compared to hybrid nanofluid-based minimum quantity lubrication.

Yildirim [73] experimentally investigated the effects of cooling conditions, notably dry, minimum quantity lubrication, cryogenic, nanofluids, and combinations of these approaches on the turning of Inconel 625. Al_2O_3 (70 nm) and hBN (70 nm) were incorporated to the nanofluid in various concentration ratios, nonetheless LN2 worked better for cryogenic cooling. Surface roughness, tool wear, and tool-chip junction temperature were the output parameters. The results indicate that employing the 0.5 v% hBN cooling approach in addition to LN2 was the optimum combination for enhanced manufacturing. Thakur et al. [74] studied the results of turning EN 24 steel using various cooling methods, including minimum quantity lubrication, MQLNF, and minimum quantity lubricationHF. Al_2O_3, CuO, and Al-CuO nanoparticles were added to water at a fraction of 0.2, 0.4, and 0.6wt%, respectively, to create nano-fluids and hybrid nano-fluids. Thermal conductivity of hybrid nanofluids was discovered to be greater than that of either mono nanofluid. The experiments showed that using a hybrid nanofluid + minimum quantity lubrication reduced the coefficient of friction, surface roughness, cutting temperature, and tool wear. Thakur et al. [75], experimentally evaluated The impact of turning EN 24 steel from different cooling methods, minimum quantity lubrication, nanofluids + minimum quantity lubrication, and composite nanofluids + minimum quantity lubrication. Water was enriched with 0.5, 1, and 1.5 percentage per cent of Al_2O_3, CuO, and Al-CuO nanoparticles, respectively. After a comparison investigation, the hybrid nanofluids performed better than the nanofluids. Therefore, two nanofluids› parameters were optimized, and RSM created hybrid nanofluids using ANOVA. According to experimental data and multi-response optimization, the hybrid nanofluid with minimum quantity lubrication had the lowest coefficient of friction, which resulted in a notable reduction in cutting force, cutting temperature, and surface roughness.

Gugulothuet al. [76] experimentally studied sedimentation tests and characterization of nanofluids, which were prepared using CNT (30 nm) and MoS_2 (30 nm) nanoparticles in 0.5, 1, 1.5, 2, 2.5, and 3wt% concentration in sesame oil as a base fluid. These nanofluids were used in machining under minimum quantity lubrication to turn AISI 1040 steel with an uncoated carbide tool. Results revealed retardation in coefficient of friction, thrust force, feed force, main cutting force, and surface roughness at 02 wt% of CNT+MoS_2 hybrid cutting nano-fluids compared to dry and conventional cutting fluids. Furthermore, cutting temperature and tool wear was retard at 3wt% of CNT+MoS_2 hybrid cutting nano-fluids in contrast with dry machining and conventional cutting fluid machining. Zaman et al. [77] studied the influence of parameters: cutting speed, feed rate, depth of cut, nanoparticle concentration, tool type on cutting temperatures, and surface roughness in turning of Ti-6Al-4V alloy under hybrid Al_2O_3-MWCNT nanofluid based minimum quantity lubrication. The Box-Behnken method was implemented in the experiment design. Empirical models were established employing RSM. The temperature improved with cutting speed, depth of cut, and feed rate, consistent with estimations and experimental results. Yet, while the particle proportions enhanced, the temperature declined to its lowest possible point and then expanded even more. The most crucial factor regarding roughness was the machine feed rate.

Sandeep Kumar et al. [78] studied the impact of hybrid nanofluids with Cu-Zn with base fluid as a groundnut oil with an application of minimum quantity lubrication in turning operation of Inconel 718 with a cutting tool fitted with Ti coated insert. In experimentation, input machining parameters were applied under conditions like dry, minimum quantity lubrication including vegetable oil, and minimum quantity lubrication including nanofluid. From investigation, it is founded that retardation in cutting temperature and surface roughness in minimum quantity lubrication + nanofluid with comparison to dry and minimum quantity lubrication + vegetable oil colling conditions.

Sandeep Kumar et al. [79] conducted tests using CuO + ZnO (50:50) hybrid nanoparticles in vegetable oil like palm oil and coconut oil as a base fluid for turning of AISI 1018 steel. In experimentation, the performance of hybrid nano cutting fluids was compared with dry machining process and notable retardation in surface roughness; feed was the main indicating parameter compared to the cut and cutting speed depth. Table 3, presents the thorough analysis of the research works concerning least quantity lubrication in conjunction with hybrid nanofluid based on the material of the workpiece, the cutting tool, the types of nanoparticles, the size, the base fluid, the characterization approach, and the most important discovering.

Table 3. Comprehensive review of the research works concerning minimum quantity lubrication in conjunction with hybrid nanofluid

Researchers	Workpiece material	Cutting tool material	Types of nanoparticles and their size (nm)	Types of base fluid	Char. Tech.	Findings	Year
Sharma et al. [69]	AISI 304 steel	Coated carbide	Al_2O_3 + MoS_2 (45, 30)	Vegetable oil	TEM	Hybrid nanofluid machining reduced thermal conductivity. Al-MoS_2 hybrid nanofluid has lower friction and pin wear than alumina. Al-MoS_2 nanofluid machining over Al_2O_3 nanofluid retarded CF and SR.	2017
Singh et al. [70]	AISI 304 steel	Coated carbide	Al_2O_3 + GnP (45, 11-15)	Servo cut oil	TEM	Thermal conductivity got enhanced by utilizing Al-GnP hybrid and alumina nanofluids. Retardation in the coefficient of friction with an enhancement in nanoparticle concentration was achieved. Hybrid nanofluid machining achieved significant retardation in CF and SR.	2017
Sharma et al. [71]	AISI 304 steel	Coated cemented carbide	Al_2O_3 + GnP (NM)	Conventional fluid	TEM, SEM	Retard the coefficient of friction and pin wear with enhanced nanoparticle concentration for both lubricants, but the hybrid lubricant has a lower coefficient of friction. Retardation in TW and CT with hybrid nanofluid over alumina nanofluid was achieved.	2018
Jamil et al. [72]	Ti–6Al–4V alloy	Coated carbide	Al_2O_3 + MWCNT (30, 10-30)	Vegetable oil	NM	Retard SR, CF, and enhanced TL with hybrid nanofluids+minimum quantity lubrication compared to cryogenic cooling. Retard CT with cryogenic cooling compared to hybrid nanofluids+minimum quantity lubrication cooling.	2019

Table 3. Comprehensive review of the research works concerning minimum quantity lubrication in conjunction with hybrid nanofluid

Researchers	Workpiece material	Cutting tool material	Types of nanoparticles and their size (nm)	Types of base fluid	Char. Tech.	Findings	Year
Yildirim [73]	Inconel 625 superalloy	Coated carbide	hBN + Al_2O_3 (70, 70)	Vegetable oil	SEM	Obtained best results with 0.5 v% hBN when all nano additives were compared as having a higher heat coefficient of hBN than Al_2O_3. CT was retard by 26.77 % and 45.58 % compared to dry machining using liquid nitrogen and minimum quantity lubrication, respectively. Retard SR with 0.5 v% hBN + liquid nitrogen method used, and got lowest TW value with 0.5 v% hBN + minimum quantity lubrication + liquid nitrogen method.	2019
Thakur et al. [74]	EN 24 steel	Tungsten carbide CVD coated	Al_2O_3 + CuO (NM.)	Soluble oil	NM.	When compared to CuO and Al2O3 nanofluids, hybrid nanofluids had the lowest friction coefficient. minimum quantity lubrication and other Al-CuO hybrid nanofluids produced the lowest SR value. Reduce CT and TW using Al-CuO hybrid nanofluids and minimum quantity lubrication turning in comparison to alternative cooling conditions.	2019
Thakur et al. [75]	EN 24 steel	Tungsten carbide CVD coated	Al_2O_3 + CuO (30, 50)	Soluble oil	TEM	Found enhanced thermal conductivity with the addition of nanoparticles. They have significantly retard SR and CT with Al-CuO hybrid nanofluids+minimum quantity lubrication. Retardation in CF due to the lowest coefficient of friction with hybrid nanofluids+minimum quantity lubrication was observed.	2019
Sandeep Kumar et al. [78]	Inconel 718 alloy	TiAlN coated	Cu + Zn (NM)	Groundnut oil	TEM, XRD	SR drop from minimum quantity lubrication tool-tip damage and wear. Hybrid nanofluids enhanced cutting fluid wettability and heat transfer.	2020
Sandeep Kumar et al. [79]	AISI 1018 steel	TiAlN coated	CuO + ZnO (NM)	Coconut oil and Palm oil	NM	Shown retardation in SR with hybrid nanofluids in contrast to conventional dry machining. ANOVA results imply that the major impact for SR is FR, followed by DC and CS.	2020

Table 3. Comprehensive review of the research works concerning minimum quantity lubrication in conjunction with hybrid nanofluid

Researchers	Workpiece material	Cutting tool material	Types of nanoparticles and their size (nm)	Types of base fluid	Char. Tech.	Findings	Year
Gugulothu, and Pasam [76]	AISI 1040 steel	CNMG 120408 uncoated carbide	MWCNT + MoS_2 (30, 30)	Sesame oil	NM	Enhanced thermal conductivity, viscosity, and specific heat with increasing particle concentration. CF was retarded by 32%, 27.3%, SR by 28.5%, and 18.3% at 2wt% of hybrid nanofluids compared to dry and conventional. Retards the CT by 43.4% and 28%, respectively. Retardation in TW by 81.3% and 75% at 3 wt% of hybrid nanofluids compared to dry and conventional.	2020
Zaman et al. [77]	Ti–6Al–4V alloy	Uncoated and TiN-coated carbide	Al_2O_3 + MWCNT (50, 8-15)	Cutting oil	NM	Decreased CT with an enhancement of nanoparticle concentration, CS was the main factor for CT. FR was the most influencing factor for SR, and lower SR was produced with lower DC, but SR was enhanced due to tool vibration.	2020

The concise reviews of minimum quantity lubrication with hybrid nanofluids [80], [81] have been carried out by considering the performance of input variables such as thermal conductivity, specific heat, and friction coefficient, and output parameters such as cutting force, cutting temperature, surface roughness, tool wear, and chip morphology in machining processes with an application of hybrid nanofluids have been the subject of extensive reviews.

Many investigators have studied the various application of minimum quantity lubrication in the turning process with cutting coolant fluid cum lubricants: conventional cutting fluid, vegetable oil (ester oil, aloe vera oil), soluble cutting oil, and emulsion oil. For enhancing process performance in the turning process by improving behaviour in terms of thermal and frictional parameters [82] and reducing health and environmental issues associated with the utilization of conventional cutting fluid, major researchers have studied nanofluid with minimum quantity lubrication. Within this presentation, a comparative analysis of cutting fluid and nano-additives that are relevant is presented in Table 4.

Table 4. Comparative analysis of cutting fluid and nano-additives

Cutting fluid	Conventional cutting fluid	Vegetable oil	Soluble cutting oil	Emulsion oil	Water
Al_2O_3	[42,50]	[41,51,56]	[47]	-	[57]
Boric Acid	-	[39,58,83]	-	[39,44]	-
Graphite	-	[51]	[43]	[44]	-
SiO_2	-	[61]	[45]	-	-
MWCNT	[40]	[54]	-	[48]	-
CuO	-	-	[46]	-	-
TiO_2	-	[49,51]	-	-	-
MoS_2	-	[53,58,60]	-	[44]	-
Tungsten disulfide WS_2	-	-	-	[52]	-
Copper	-	[59]	-	-	-
Graphene Nanoplatelets	-	-	-	[55]	-

Development in nanoscience demands a higher enhancement in nanofluids' heat transfer properties. A few investigators have used hybrid nanofluids with minimum quantity lubrication in turning metals and alloys. According to the study, hybrid nanofluids with different nanoparticles had been with conventional cutting fluid, vegetable oil, and soluble cutting oil as base fluid. Sharma et al. [69] used Al_2O_3 + MoS_2, Singh et al. [70], and Sharma et al. [71] used Al_2O_3 + GnP, Jamil et al. [72], and Zaman et al. [77] used Al_2O_3 + MWCNT, Yildirim [84] used hBN + Al_2O_3, Thakur et al.[74], and [75] used Al_2O_3 + CuO, Sandeep Kumar et al.[78], and [79] used CuO + ZnO, and Gugulothu, and Pasam [76] used MWCNT + MoS_2.

Apart from cutting fluid applications, the mono and hybrid nanofluid can be implemented for various applications, such as electronic component cooling using heat pipes [85], automobile radiators [86], solar water distillation systems [87], solar collectors [88], and energy storage system [89].

Statistical Analysis

A number of cutting fluids for different machining operations are examined in this review study. Performance characteristics, experimental analysis, preparation process, characterization methodologies, and working fluids are among the various aspects of the review work carried out during the last ten years that are highlighted. The cited works in this study are primarily taken from a conference proceeding, notable Elsevier, Taylor & Francis, and Springer publications.

The total number of works published in various categories in the form of nature of work is shown in Figure 3. More research articles on experimental work across multiple cutting fluids have been published throughout the previous decade. The number of articles that have utilized conventional fluid, nanofluid, and hybrid nanofluid as a cutting fluid with minimum quantity lubrication method between the years 2010 and 2021 is displayed in Figure 4. The conclusion that was reached was that the research trends that involve conventional cutting fluid are decreasing with time; the use of nano-fluid as a cutting fluid was trending in the middle of the decade, and the use of hybrid nano-fluid as a cutting fluid is a trend that is currently occurring.

The number of research works that make use of particular kinds of workpiece material while employing a specific kind of cutting fluid is displayed here in Figure 5. As indicated in the figure, most researchers employed AISI 1040 steel as workpiece material, while hybrid nanofluid was employed as cutting fluid.

Figure 5 shows a particular sort of cutting fluid; the number of articles while utilizing a particular kind of tool

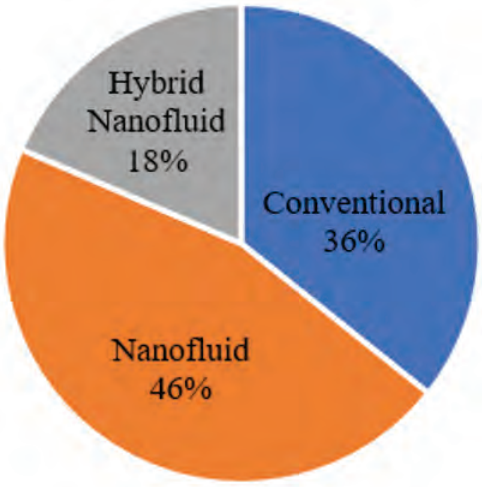


Figure 3. Publications with the nature of work based on cutting fluids.

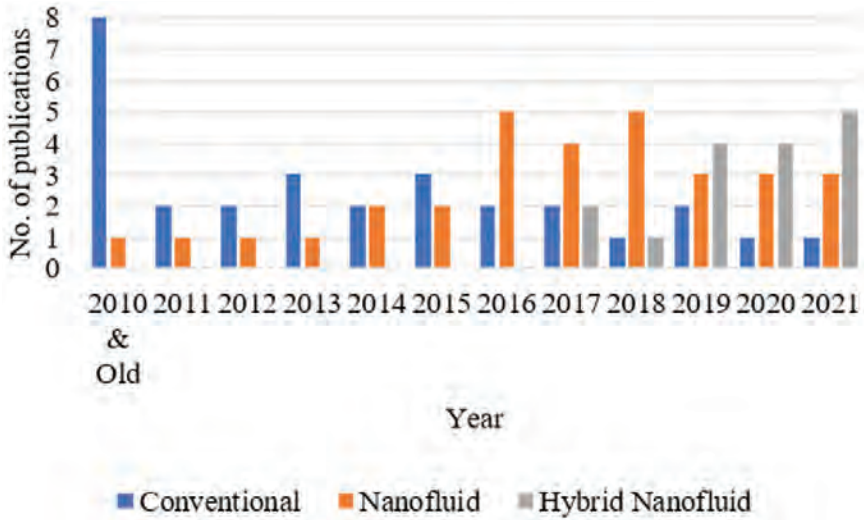


Figure 4. Numbers of publications between 2010 and 2021.

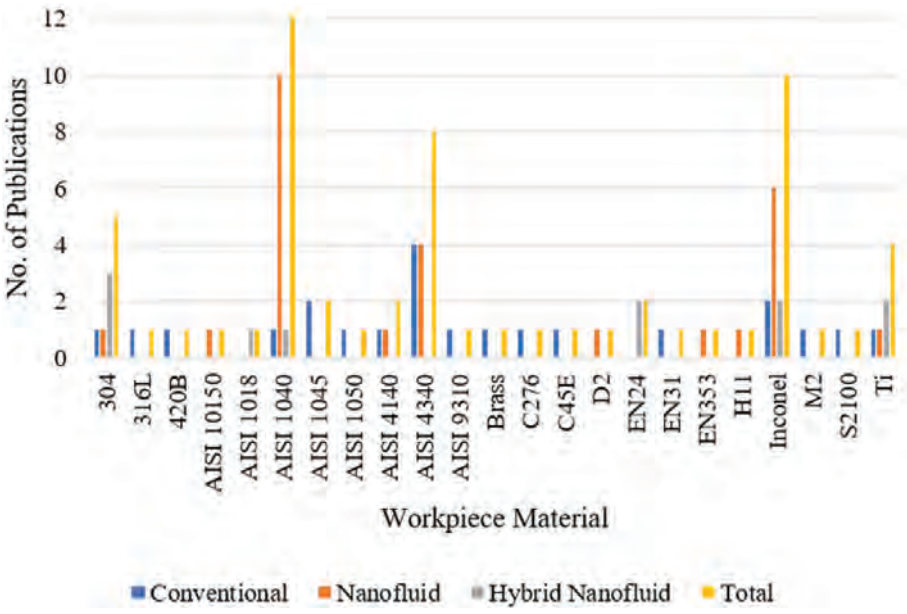


Figure 5. Number of studies and types of workpiece materials.

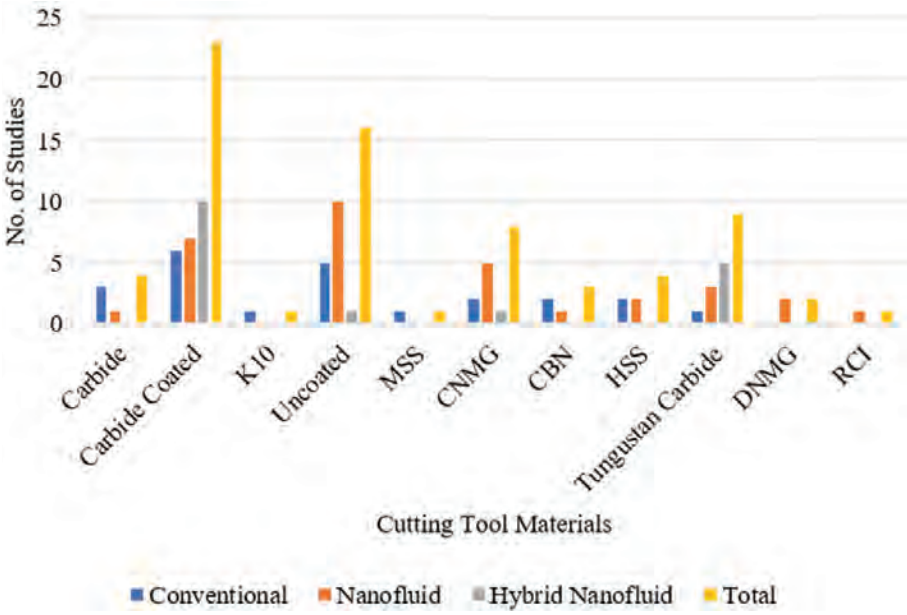


Figure 6. Number of studies with types of tool materials.

material is significant. As can be seen in the picture, an increasing number of researchers have utilized tungsten carbide as a tool material with a carbide coating, while hybrid nanofluid has been used as a cutting fluid.

Figure 6 addresses a variety of studies involving different kinds of tool materials. According to this figure, the majority of researchers used a hybrid nanofluid along with minimum quantity lubrication as a cutting fluid, AISI 1040 steel as a workpiece material, and tungsten carbide as a tool material coated with carbide. Research suggests using a hybrid nanofluid as cutting fluid with AISI 1040 steel as a workpiece material and a tungsten carbide tool coated with carbide for various lathe operations.

Economy and Sustainibility

Aspects in terms of economy and sustainability are incorporated during a critical review of open literature. Significant aspects related to cost and environmental impacts are drawn below:

1. Excellent thermos-physical characteristics lead to a reduction in operational costs. Overall investments got retarded by a cut down in cost [90].
2. Identical heat transport characteristics as conventional working fluids lead to an opportunity to retard the flow while using the minimum quantity lubrication with mono and hybrid nanofluids [90].
3. Conventional cutting fluids are highly hazardous, whereas the minimum quantity lubrication with mono or hybrid nanofluids has a less or null environmental impact in terms of hazard [91].
4. Minimum quantity lubrication, as a cutting fluid, needs less flow than conventional cutting fluid. This leads to less focus on storage, safety, and cleaning of minimum quantity lubrication cutting fluid than conventional ones [92].
5. Minimum quantity lubrication with mono or hybrid nanofluid as a cutting fluid creates a temporary film on the cutting zone interface of the cutting tool and workpiece. It reduces the tool wear and improves the tool life compared to the conventional cutting fluid [93].
6. Minimum quantity lubrication with mono and hybrid nanofluid can reduce wastage and enhance overall productivity in contrast with the conventional cutting fluid [90].
7. Efficiency improvement for heat exchangers, enhancement in compactness, energy recovery, reduction in cutting fluid usage, and improvement in the thermal process are key aspects of nanofluids for practical integration [92,93].

Limitations

After careful consideration of minimum quantity lubrication use along with nanoparticles in the context of performance improvement, a few limiting characteristics have been observed from the part of the literature. The following are the limitations of current work:

1. Utilizing nanofluid for lubrication in minimal quantities increases the initial cost. However, following thorough consideration and a wider range of industrial applications, it is possible to achieve this reduction by producing it in bulk.
2. Nanoparticle size plays a critical influence on jet flow. A greater size might cause the nozzle used in the lowest quantity of lubricating mist to choke. Selecting smaller-sized nanoparticles might lessen the chances of repeated spray nozzle choking.
3. Only a small percentage of nanoparticles possess potentially harmful properties, raising questions about operations› safety. Avoiding the selection of such dangerous particles could resolve safety concerns. Mandatory selection of potentially harmful nanoparticles would result in the implementation of one-of-a-kind and cutting-edge safety measures.
4. Agglomeration is produced, and the stability of the working medium decreases when nanoparticles are added to concentrations greater than 2 wt%. As a result, the utilization of concentrations that are lower than the maximum allowable level will lessen the likelihood of agglomeration and improve the consistency of the working medium.
5. The utilization of surfactant or detergent enhances the stability of the working medium, but it also negatively impacts its thermo-physical properties. It is possible to achieve superior thermo-physical properties of the working medium by avoiding the addition of surfactant or detergent and paying particular attention to stability.

Future Trends

To this point, numerous academics have focused more on nanofluids with minimum quantity lubrication. The below-mentioned points lead to the enhancement of the research in this field:

1. On the other hand, hybrid nanofluids that use a minimum quantity lubrication cooling approach are being investigated to improve turning parameters and boost the efficiency of heat transmission perticularly for turbine blade machining.
2. During the machining (turning) process, the impact that nanofluids with the least quantity of lubrication have on chip shape must be considered. Additionally, hybrid nanofluids with minimum quantity of lubrication should also be considered during gear machining.
3. The purpose of the theoretical and experimental inquiry is to optimize the minimum quantity lubrication operational factors, which include spray distance, flow rate, nozzle distance, air pressure, and nozzle orientation angle, as well as the respective influences that these variables have on turning in micro machining operations.
4. Additional research is still required to optimize the particle size and variable concentration of nanoparticles and their effects on the turning process in terms of tribological parameters, particularly wear and friction during high speed machining on hard-to-machine materials.
5. During turning operations, vibration analysis of a cutting tool should be explored both experimentally and theoretically. This analysis should also investigate the impact that vibration has on cutting parameters during internal boring operations.
6. When using the machining process, it is important to consider the properties of mono and hybrid nanofluids, including their stability, sustainability, and safety in application of profile turning where the operators are very close to machining operations.
7. It was necessary to conduct a thorough analysis of the technological and economic investigations conducted using mono and hybrid nanofluids, respectively, for mass production of a particular job for the machining operation.
8. Predictive and ensemble models, as well as machine learning, can be used to optimize machining parameters such as fluid flow rate or nanoparticle size in Artificial Intelligence and Machine Learning applications.

CONCLUSION

This paper has presented an overview of critical published investigations on applying minimum quantity lubrication with conventional cutting fluid, mono nanofluid, and hybrid nanofluid in turning operation. During the last decade, various experiments shifted the focus of study from traditional juice cutting to nano and hybrid nanofluid due to superior tribological properties that do not affect product quality. A literature survey shows significant improvement in machining parameters with mono and hybrid nanofluid with minimum quantity lubrication. The below-mentioned observations and outcomes are drawn from the following literature:

1. When using minimum quantity lubrication in conjunction with nanofluids and hybrid nanofluids, significant reductions in cutting forces, cutting temperatures, surface roughness, and tool wear can be attained
2. Applying minimum quantity lubrication with nanofluids improves the chip›s morphology in terms of its colour and form.
3. minimum quantity lubrication with hybrid nanofluids significantly improves its machining performance and tribological properties compared to minimum quantity lubrication with nanofluids.
4. The various types of nanoparticles, their particle size, concentration, and the nozzle›s inclination angle each play a crucial role in optimizing heat transfer rate and reducing tool wear.
5. As it tends to develop film between the interface of cutting tool-workpiece and retard friction, Aluminium Oxide (Al_2O_3) is the nanoparticle employed the most in minimum quantity lubrication applications that involve nanofluids and hybrid applications.
6. When comparing minimum quantity lubrication with hybrid nanofluid to minimum quantity lubrication with nanofluid, the coefficient of friction is found to be retarded when there is an enhancement in the nanoparticle concentration in minimum quantity lubrication with hybrid nanofluid.
7. AISI 1040 steel is the most suitable workpiece material for lathe operations when using nanofluid as a cutting fluid.
8. Tungsten carbide-coated tool material is suitable for lathe operation, while hybrid nanofluid + minimum quantity lubrication was used as cutting fluid.
9. The cutting fluid effect on turning parameters has a complex and non-linear research trend. So, it is difficult to predict using a simple mathematical model, simulation, etc.

In addition, theoretical and experimental research is required to comprehend the minimum quantity lubrication employing hybrid nanofluid as a cutting fluid for diverse applications in lathe operations during machining processes.

ABBREVIATIONS

Ag	Silver
Al_2O_3	Aluminum Oxide
AlN	Aluminum Nitride
ANOM	Analysis of Means
ANOVA	Analysis of Variance
Au	Gold
BFO	Bacterial Foraging Optimization
CCD	Central Composite Design
CFsMQL	Conventional Fluids with minimum quantity lubrication
CNT	Carbon Nanotube
CO_2	Cryogenic Cooling
CRSM	Coupling Response Surface Methodology
Cu	Copper
CuO	Copper Oxide
CVD	Chemical Vapour Deposition
EDM	Electro Discharge Machining
Fe	Iron
Fe_2O_3/Fe_3O_4	Iron Oxide
GRA	Gray Rational Analysis
hBN	Hexagon Boron Nitride
LN_2	Liquid Nitrogen
MgO	Magnesium Oxide
MoS_2	Molybdenum Disulfide
MQLNF	minimum quantity lubrication with Nanofluid
MWCNT	Multi-Wall Carbon Nanotube
NFsMQL	Nanofluid with minimum quantity lubrication
NM	Not Mentioned
PSO	Particle Swarm Optimization
PVD	Physical Vapour Deposition
PWD	Physical Vapour Deposition
RSM	Response Surface Methodology
SAE	Society of Automotive Engineers
SDBS	Spectral Database for Organic Compounds
SEM	Scanning Electron Microscopy
SiC	Silicon Carbide
SiN	Silicon Nitride
SiO_2	Silicon Oxide
ST	Surface Topography
TiO_2	Titanium Oxide
WS_2	Tungsten Disulfide
ZnO	Zinc Oxide

Symbol	
CF	Cutting Force (N)
CP	Cutting pressure (bar)
CS	Cutting speed (m/min)
CT	Cutting temperature (°C)
DC	Depth of Cut (mm)
F	Force (N)
FR	Feed Rate (mm/rev)
SR	Surface Roughness (Micron)

TW	Tool Wear (Unitless)
v	Volume Fraction (%)
wt	Weight Fraction (%)

Subscripts

x	Feed
y	Thrust
z	Cutting

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Said Z, Munish G, Hussein H, Neeti A, Mashood AK, Muhammad J, et al. A comprehensive review on minimum quantity lubrication (MQL) in machining processes using nano-cutting fluids. Int J Adv Manuf Technol 2019;105:2057–2086. [CrossRef]

[2] Khan AM. Multi-objective optimization for grinding of AISI D2 steel with Al2O3 wheel under MQL. Materials 2018;11:12269. [CrossRef]

[3] Abhang LB, Hameedullah M. Experimental investigation of minimum quantity lubricants in alloy steel turning. Int J Eng Sci Technol 2010;2:3045–3053.

[4] Sharma VS, Singh G, Sorby K. A review on minimum quantity lubrication for machining processes. Mater Manuf Process 2015;30:935–953. [CrossRef]

[5] Sharma AK, Tiwari AK, Dixit AR. Effects of minimum quantity lubrication (MQL) in machining processes using conventional and nanofluid based cutting fluids: a comprehensive review. J Clean Prod 2016;127:1–18. [CrossRef]

[6] Madanchi N, Kurle D, Winter M, Thiede S, Herrmann C. Energy efficient process chain: the impact of cutting fluid strategies. Procedia CIRP 2015;29:360–365. [CrossRef]

[7] Ajithkumar JP, Xavior MA. Influence of nano lubrication in machining operations-a review. Mater Today Proc 2018;5:11185–11192. [CrossRef]

[8] Okokpujie IP, Bolu CA, Ohunakin OS, Akinlabi ET, Adelekan DS. A review of recent application of machining techniques, based on the phenomena of CNC machining operations. Procedia Manuf 2019;1054–1060. [CrossRef]

[9] Fallenstein F, Aurich JC. CFD based investigation on internal cooling of twist drills. Procedia CIRP 2014;14:293–298. [CrossRef]

[10] Zhang C, Zhang S, Yan X, Zhang Q. Effects of internal cooling channel structures on cutting forces and tool life in side milling of H13 steel under cryogenic minimum quantity lubrication condition. Int J Adv Manuf Technol 2016;83:975–984. [CrossRef]

[11] Shokoohi Y, Shekarian E. Application of nanofluids in machining processes-a review. J Nanosci Technol 2016;2:59–63.

[12] Sharma AK, Tiwari AK, Dixit AR. Progress of nanofluid application in machining: a review. Mater Manuf Processes 2015;7:813–825.

[13] Chavda N, Bumataria R. Effect of particle size and concentration on thermal performance of cylindrical shaped heat pipe using silver-DI water nanofluid. Int J Ambient Energy 2022;4:305–316. [CrossRef]

[14] Bumataria RK, Chavda NK. Current research aspects in mono and hybrid nanofluid based heat pipe technologies. Heliyon 2019;5:e01627. [CrossRef]

[15] Bumataria RK, Chavda NK, Nalbandh AH. Performance evaluation of the cylindrical shaped heat pipe utilizing water-based CuO and ZnO hybrid nanofluids. Energy Sources Part A Recovery Util Environ Effects 2020;47:1832628. [CrossRef]

[16] Dhar NR, Kamruzzaman M, Ahmed M. Effect of minimum quantity lubrication (MQL) on tool wear and surface roughness in turning AISI-4340 steel. J Mater Process Technol 2006;172:299–304. [CrossRef]

[17] Dhar NR, Islam MW, Islam S, Mithu MAH. The influence of minimum quantity of lubrication (MQL) on cutting temperature, chip and dimensional accuracy in turning AISI-1040 steel. J Mater Process Technol 2006;171:93–99. [CrossRef]

[18] Bruni C, Forcellese A, Gabrielli F, Simoncini M. Effect of the lubrication-cooling technique, insert technology and machine bed material on the workpart surface finish and tool wear in finish turning of AISI 420B. Int J Mach Tools Manuf 2006;46:1547–1554. [CrossRef]

[19] Dhar NR, Ahmed MT, Islam S. An experimental investigation on effect of minimum quantity lubrication in machining AISI 1040 steel. Int J Mach Tools Manuf 2007;47:748–753. [CrossRef]

[20] Davim JP, Sreejith PS, Silva J. Turning of brasses using minimum quantity of lubricant (MQL) and flooded lubricant conditions. Mater Manuf Process 2007;22:45–50. [CrossRef]
[21] Khan MMA, Mithu MAH, Dhar NR. Effects of minimum quantity lubrication on turning AISI 9310 alloy steel using vegetable oil-based cutting fluid. J Mater Process Technol 2009;209:5573–5583. [CrossRef]
[22] Hwang YK, Lee CM. Surface roughness and cutting force prediction in MQL and wet turning process of AISI 1045 using design of experiments. J Mech Sci Technol 2010;24:1669–1677. [CrossRef]
[23] Leppert T. Effect of cooling and lubrication conditions on surface topography and turning process of C45 steel. Int J Mach Tools Manuf 2011;51:120–126. [CrossRef]
[24] Che Haron CH, Ghani JA, Kasim MS, Soon TK, Gusri AI, Sulaiman MA. Surface integrity of Inconel 718 under MQL condition. Adv Mater Res 2011;150-151:1667–1672. [CrossRef]
[25] Leppert T. Surface layer properties of AISI 316L steel when turning under dry and with minimum quantity lubrication conditions. Proc Inst Mech Eng B J Eng Manuf 2012;226:617–631. [CrossRef]
[26] Lohar DV, Nanavaty CR. Performance evaluation of minimum quantity lubrication (MQL) using CBN tool during hard turning of AISI 4340 and its comparison with dry and wet turning. Bonfring Int J Ind Eng Manag Sci 2013;3:102–106. [CrossRef]
[27] Hadad M, Sadeghi B. Minimum quantity lubrication-MQL turning of AISI 4140 steel alloy. J Clean Prod 2013;54:332–343. [CrossRef]
[28] Saini A, Dhiman S, Sharma R, Setia S. Experimental estimation and optimization of process parameters under minimum quantity lubrication and dry turning of AISI-4340 with different carbide inserts. J Mech Sci Technol 2014;28:2307–2318. [CrossRef]
[29] Agrawal SM, Patil NG. Experimental study of non-edible vegetable oil as a cutting fluid in machining of M2 steel using MQL. Procedia Manuf 2018;20:207–212. [CrossRef]
[30] Gordana L, Sredanović B, Kramar D, Nedić B, Kopač J. Effects of using MQL technique in metal. In: 13th International Conference on Tribology; 2013. p. 292–301.
[31] Liu ZQ, Xu J, Han S, Chen M. A coupling method of response surfaces (CRSM) for cutting parameters optimization in machining titanium alloy under minimum quantity lubrication (MQL) condition. Int J Precis Eng Manuf 2013;14:693–702. [CrossRef]
[32] Masoudi S, Vafadar A, Hadad M, Jafarian F. Experimental investigation into the effects of nozzle position, workpiece hardness, and tool type in MQL turning of AISI 1045 steel. Mater Manuf Process 2018;33:1011–1019. [CrossRef]
[33] Sivaiah P, Ajay Kumar G, Singh M, Kumar H. Effect of novel hybrid texture tool on turning process performance in MQL machining of Inconel 718 superalloy. Mater Manuf Process 2020;35:61–71. [CrossRef]
[34] Sivaiah P, Guru Prasad M, Singh M, Uma B. Machinability evaluation during machining of AISI 52100 steel with textured tools under minimum quantity lubrication–a comparative study. Mater Manuf Process 2020;35:1761–1768. [CrossRef]
[35] Singh G, Aggarwal V, Singh S. Experimental investigations into machining performance of Hastelloy C-276 in different cooling environments. Mater Manuf Process 2021;36:1789–1799. [CrossRef]
[36] Bhadoria NS, Bartarya G. On the improvement in process performance of ceramic inserts during hard turning in MQL environment. Mater Manuf Process 2021;37:283–293. [CrossRef]
[37] Sivaiah P, Ajay Kumar G, Lakshmi Narasimhamu K, Siva Balaji N. Performance improvement of turning operation during processing of AISI 304 with novel textured tools under minimum quantity lubrication using hybrid optimization technique. Mater Manuf Process 2021;37:693–700. [CrossRef]
[38] Sultana MN, Dhar NR, Zaman PB. A review on different cooling/lubrication techniques in metal cutting. Am J Mech Appl 2019;7:71. [CrossRef]
[39] Krishna PV, Srikant RR, Nageswara Rao D. Experimental investigation on the performance of nanoboric acid suspensions in SAE-40 and coconut oil during turning of AISI 1040 steel. Int J Mach Tools Manuf 2010;50:911–916. [CrossRef]
[40] Rao SN, Satyanarayana B, Venkatasubbaiah K. Experimental estimation of tool wear and cutting temperatures in MQL using cutting fluids with CNT inclusion. Int J Eng Sci Technol 2011;3:2928–2931.
[41] Vasu V, Reddy GPK. Effect of minimum quantity lubrication with Al2O3 nanoparticles on surface roughness, tool wear and temperature dissipation in machining Inconel 600 alloy. In: Proc Inst Mech Eng Part N J Nanoeng Nanosyst; 2012. p. 3–15. [CrossRef]
[42] Khandekar S, Sankar MR, Agnihotri V, Ramkumar J. Nano-cutting fluid for enhancement of metal cutting performance. Mater Manuf Process 2012;27:963–967. [CrossRef]
[43] Amrita M, Srikant RR, Sitaramaraju AV, Prasad MMS, Krishna PV. Experimental investigations on influence of mist cooling using nanofluids on machining parameters in turning AISI 1040 steel. Proc Inst Mech Eng Part J J Eng Tribol 2013;227:1334–1346. [CrossRef]
[44] Amrita M, Shariq SA, Manoj M, Gopal C. Experimental investigation on application of emulsifier oil based nano cutting fluids in metal cutting process. Procedia Eng 2014;97:115–124. [CrossRef]
[45] Sayuti M, Sarhan AAD, Salem F. Novel uses of SiO2 nano-lubrication system in hard turning process

of hardened steel AISI4140 for less tool wear, surface roughness and oil consumption. J Clean Prod 2014;67:265–276. [CrossRef]

[46] Shabgard MR, Zenjanab MJ, Azarafza R. Experimental study on the influence of CuO nanofluid on surface roughness and machining force in turning of AISI 4340 steel. Modares Mech Eng 2014;14:27–33.

[47] Khalil ANM, Ali MAM, Azmi AI. Effect of Al2O3 nanolubricant with SDBS on tool wear during turning process of AISI 1050 with minimal quantity lubricant. Procedia Manuf 2015;2:130–134. [CrossRef]

[48] Sharma P, Sidhu BS, Sharma J. Investigation of effects of nanofluids on turning of AISI D2 steel using minimum quantity lubrication. J Clean Prod 2015;108:72–79. [CrossRef]

[49] Sharma AK, Tiwari AK, Singh RK, Dixit AR. Tribological investigation of TiO2 nanoparticle based cutting fluid in machining under minimum quantity lubrication (MQL). Mater Today Proc 2016;3:2155–2162. [CrossRef]

[50] Sharma AK, Singh RK, Dixit AR, Tiwari AK. Characterization and experimental investigation of Al2O3 nanoparticle based cutting fluid in turning of AISI 1040 steel under minimum quantity lubrication (MQL). Mater Today Proc 2016;3:1899–1906. [CrossRef]

[51] Gupta MK, Sood PK, Sharma VS. Optimization of machining parameters and cutting fluids during nano-fluid based minimum quantity lubrication turning of titanium alloy by using evolutionary techniques. J Clean Prod 2016;135:1276–1288. [CrossRef]

[52] Paturi UMR, Maddu YR, Maruri RR, Narala SKR. Measurement and analysis of surface roughness in WS2 solid lubricant assisted minimum quantity lubrication (MQL) turning of Inconel 718. Procedia CIRP 2016;40:138–143. [CrossRef]

[53] Rapeti P, Pasam VK, Rao Gurram KM, Revuru RS. Performance evaluation of vegetable oil based nano cutting fluids in machining using grey relational analysis-a step towards sustainable manufacturing. J Clean Prod 2016;172:2862–2875. [CrossRef]

[54] Patole PB, Kulkarni VV. Experimental investigation and optimization of cutting parameters with multi-response characteristics in MQL turning of AISI 4340 using nano fluid. Cogent Eng 2017;4:1303956. [CrossRef]

[55] Singh RK, Sharma AK, Dixit AR, Tiwari AK, Pramanik A, Mandal A. Influence of graphene-based nanofluid with minimum quantity lubrication on surface roughness and cutting temperature in turning operation. Mater Today Proc 2018;5:24578–24586. [CrossRef]

[56] Joshi KK, Behera RK, Anurag. Effect of minimum quantity lubrication with Al2O3 nanofluid on surface roughness and its prediction using hybrid fuzzy controller in turning operation of Inconel 600. 2018.

[57] Vishnu AV, Kumar PJ, Venkata Ramana M. Comparison among dry, flooded and MQL conditions in machining of EN 353 steel alloys-an experimental investigation. Mater Today Proc 2018;5:24954–24962. [CrossRef]

[58] Pasam VK, Srikant RR, Gugulothu S. Comparing the performance & viability of nano and microfluids in minimum quantity lubrication for machining AISI1040 steel. Mater Today Proc 2018;5:8016–8024. [CrossRef]

[59] Ganesan K, Naresh Babu M, Santhanakumar M, Muthukrishnan N. Experimental investigation of copper nanofluid based minimum quantity lubrication in turning of H11 steel. J Braz Soc Mech Sci Eng 2018;40:1093–1109. [CrossRef]

[60] Padmini R, Vamsi Krishna P, Krishna Mohana Rao G. Effectiveness of vegetable oil based nanofluids as potential cutting fluids in turning AISI 1040 steel. Tribol Int 2016;94:490–501. [CrossRef]

[61] Sharma AK, Tiwari AK, Dixit AR, Singh RK. Investigation into performance of SiO2 nanoparticle based cutting fluid in machining process. Mater Today Proc 2017;4:133–141. [CrossRef]

[62] Sivalingam V, Zhuoliang Z, Jie S, Baskaran S, Yuvaraj N, Gupta MK, et al. Use of atomized spray cutting fluid technique for the turning of a nickel base superalloy. Mater Manuf Process 2021;36:373–380. [CrossRef]

[63] Das A, Patel SK, Arakha M, Dey A, Biswal BB. Processing of hardened steel by MQL technique using nano cutting fluids. Mater Manuf Process 2021;36:316–328. [CrossRef]

[64] Venkatesan K, Devendiran S, Nishanth Purusotham K, Praveen VS. Study of machinability performance of Hastelloy-X for nanofluids, dry with coated tools. Mater Manuf Process 2020;35:751–761. [CrossRef]

[65] Ramanan KV, Ramesh Babu S, Jebaraj M, Nimel Sworna Ross K. Face turning of Incoloy 800 under MQL and nano-MQL environments. Mater Manuf Process 2021;36:1769–1780. [CrossRef]

[66] Nikouei SM, Razfar MR, Khajehzadeh M. Influence of nanoparticles' size on Inconel 718 machining induced residual stresses. Mater Manuf Process 2021;37:1–10. [CrossRef]

[67] Singh V, Sharma AK, Sahu RK, Katiyar JK. State of the art on sustainable manufacturing using mono/hybrid nano-cutting fluids with minimum quantity lubrication. Mater Manuf Process 2022;11:64. [CrossRef]

[68] Khudaiwala A, Patel RL, Bumataria R. Recent developments in thermal management of light-emitting diodes (LEDS): a review. J Therm Eng 2024;10:517–540. [CrossRef]

[69] Sharma AK, Singh RK, Dixit AR, Tiwari AK. Novel uses of alumina-MoS2 hybrid nanoparticle enriched cutting fluid in hard turning of AISI 304 steel. J Manuf Process 2017;30:467–482. [CrossRef]

[70] Singh RK, Sharma AK, Dixit AR, Tiwari AK, Pramanik A, Mandal A. Performance evaluation of alumina-graphene hybrid nano-cutting fluid in hard turning. J Clean Prod 2017;162:830–845. [CrossRef]
[71] Sharma AK, Tiwari AK, Dixit AR, Singh RK, Singh M. Novel uses of alumina/graphene hybrid nanoparticle additives for improved tribological properties of lubricant in turning operation. Tribol Int 2018;119:99–111. [CrossRef]
[72] Jamil M, Khan AM, Hegab H, Gong L, Mai M, Gupta MK, et al. Effects of hybrid Al2O3-CNT nanofluids and cryogenic cooling on machining of Ti–6Al–4V. Int J Adv Manuf Technol 2019;102:3895–3909. [CrossRef]
[73] Yıldırım ÇV. Experimental comparison of the performance of nanofluids, cryogenic and hybrid cooling in turning of Inconel 625. Tribol Int 2019;137:366–378. [CrossRef]
[74] Thakur A, Manna A, Samir S. Performance evaluation of different environmental conditions on output characteristics during turning of EN-24 steel. Int J Precis Eng Manuf 2019;20:1839–1849. [CrossRef]
[75] Thakur A, Manna A, Samir S. Experimental investigation of nanofluids in minimum quantity lubrication during turning of EN-24 steel. Proc Inst Mech Eng Part J J Eng Tribol 2019;234:712–729. [CrossRef]
[76] Gugulothu S, Pasam VK. Experimental investigation to study the performance of CNT/MoS2 hybrid nanofluid in turning of AISI 1040 steel. Aust J Mech Eng 2020;20:814–824. [CrossRef]
[77] Zaman PB, Tusar MIH, Dhar NR. Selection of appropriate process inputs for turning Ti-6Al-4V alloy under hybrid Al2O3-MWCNT nano-fluid based MQL. Adv Mater Process Technol 2020;8:380–400. [CrossRef]
[78] Kumar SM, Vasu V, Venu Gopal A. Investigation of influence of hybrid nanofluid/MQL on surface roughness in turning Inconel-718. Adv Appl Mech Eng 2020;1137–1145. [CrossRef]
[79] Kumar SM, Murali Krishna V, Kumar M. Investigation on influence of hybrid biodegradable nanofluids (CuO-ZnO) on surface roughness in turning AISI 1018 steel. Mater Today Proceed 2020;24:1570–1575. [CrossRef]
[80] Hemmat Esfe M, Bahiraei M, Mir A. Application of conventional and hybrid nanofluids in different machining processes: A critical review. Adv Colloid Interface Sci 2020;282:102199. [CrossRef]
[81] Junankar AA, Parete SR, Dethe PK, Dhote NR, Gadkar DG, Gadkar DD, et al. A review: enhancement of turning process performance by effective utilization of hybrid nanofluid and MQL. Mater Today Proc 2021;38:44–47. [CrossRef]
[82] Hegab H, Kishawy HA, Darras B. Sustainable cooling and lubrication strategies in machining processes: a comparative study. Procedia Manuf 2019;33:786–793. [CrossRef]
[83] Sodavadia KP, Makwana AH. Experimental investigation on the performance of coconut oil based nano fluid as lubricants during turning of AISI 304 austenitic stainless steel. Int J Adv Mech Eng 2014;4:55–60.
[84] Yıldırım ÇV, Sarıkaya M, Kıvak T, Şirin Ş. The effect of addition of hBN nanoparticles to nanofluid-MQL on tool wear patterns, tool life, roughness and temperature in turning of Ni-based Inconel 625. Tribol Int 2019;134:443–456. [CrossRef]
[85] Bumataria R, Chavda N. Heat load and orientation impacts in cylindrical heat pipes using copper oxide, aluminium oxide, and zinc oxide nanofluids. Int J Ambient Energy 2021;43:6273–6283. [CrossRef]
[86] Jadeja K, Bumataria R, Chavda N. Nanofluid as a coolant in internal combustion engine-a review. Int J Ambient Energy 2022;44:1–65. [CrossRef]
[87] Kayar A. Effect of different aluminium oxide based nanofluid concentrations on the efficiency of solar water desalination system. J Therm Eng 2023;9:61–68. [CrossRef]
[88] Verma SK, Gupta NK. A comprehensive analysis on advances in application of solar collectors considering design, process and working fluid parameters for solar to thermal conversion. Sol Energy 2021;208:1114–1150. [CrossRef]
[89] Rathore KPS, Gupta NK, Yadav D, Shukla SK. Thermal performance of the building envelope integrated with phase change material for thermal energy storage: an updated review. Sustain Cities Soc 2022;79:103. [CrossRef]
[90] Mukherjee S, Wciślik S, Chandra Mishra P, Chaudhuri P. Nanofluids: critical issues, economics and sustainability perspectives. Particuology 2024;87:147–172. [CrossRef]
[91] El Hattab M, Boumhaout M, Oukach S. MHD natural convection in a square enclosure using carbon nanotube-water nanofluid with two isothermal fins. Sigma J Eng Nat Sci 2024;42:1075–1087. [CrossRef]
[92] Sharma R, Aruna D, Mishra MK, Sinha SB. Scrutiny of flow and heat transfer characteristics of hybrid nanofluid passing through a squeezing channel. Sigma J Eng Nat Sci 2024;42:1856–1865. [CrossRef]
[93] Ergüner A, Özkan R, Koca K, Genç MS. Improvement of mechanical behaviour of wind turbine blade using nanofluid-graphene and/or glass fiber in epoxy resin. Sigma J Nat Sci 2021;39:58–69.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2042

Research Article

Effect of window glass area on optimum insulation thickness and environmental analysis

Meral ÖZEL[1,*], Serhat ŞENGÜR[2]

[1]*Department of Mechanical Engineering, Firat University, Elazığ, 23119, Türkiye*
[2]*Technical Vocational School, Firat University, Elazığ, 23119, Türkiye*

ARTICLE INFO

Article history
Received: 16 July 2024
Revised: 04 October 2024
Accepted: 05 March 2025

Keywords:
Environmental Analysis; Insulated Wall; Optimum Thickness of Insulation; Single Glass and Double Glass; Yearly Transmission Loads

ABSTRACT

This research highlights the significance of insulation and glazing area to increase energy efficiency in buildings and reduce fuel consumption and environmental pollution. With the proper combination of window glass area and insulation in the wall, heating and cooling loads and therefore fuel consumption and emissions can be notably reduced. Therefore, the basic goal of this study is to numerically research the impact of window glass area on insulation thickness, fuel consumption and emissions under dynamic thermal conditions for south orientation in the climate of Elazığ, Turkey and to fill knowledge gap in this field. For this goal, the window glass area on the wall which has different insulation thicknesses was enhanced from 0% to 100% with an increase of 10%. The annual heating and cooling loads were computed employing an implicit finite difference procedure taking into consideration that the window is single-glazing and double-glazing. The annual heating loads obtained were used for the optimization of insulation thickness. The results indicate that the heating load in 20 mm insulated wall for single glazing is acquired to be 177.57, 236.79, 317.51 and 398.23 MJ/m^2 year for values 0%, 20%, 40% and 60% of the glazing area, respectively while it is obtained to be 177.57, 114.24, 88.24 and 76.81 MJ/m^2 year for double glass. It is observed that choosing smaller glazing areas in insulated walls containing a single-glazed window results in less heating load. It is also observed that at lower values of insulation thickness, the large double-glazing area on the south facade notably reduces heat loss. It is revealed that the increment in the glazing area for single glass induces an additional heating load, fuel consumption and emissions.

Cite this article as: Özel M, Şengür S. Effect of window glass area on optimum insulation thickness and environmental analysis. Sigma J Eng Nat Sci 2026;44(2):1369–1382.

INTRODUCTION

Today, the energy consumption in the world makes contribution to environmental pollution and greenhouse gas emissions. Industry, construction, agriculture and transportation sectors are the sectors which contribute the highest to energy consumption [1]. The constructions are liable for roughly 40% of yearly energy consumption in the world. Most of this energy consumed is used for

*Corresponding author.
*E-mail address: mozel@firat.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

heating, cooling, air conditioning and lighting [2]. Due to the decrease in energy resources in the world, energy savings should be achieved by reducing the energy consumed in buildings in an effort to ensure thermal comfort. The building envelope is one of the most significant parameters impressing thermal comfort and energy saving [3]. Windows are often the weakest element of buildings in terms of energy savings. About one-third of a building's energy loss is due to windows. But, they are indispensable elements in building applications as they beautify the appearance of buildings and allow daylight to enter. In the past, while it was thought that the size of the windows should be small in terms of thermal comfort, window sizes were increased with double glazing applications. Thus, both thermal comfort and energy savings have improved significantly [4]. Large glazed windows can reduce the use of artificial lighting by allowing more daylight into the space. However, they can also increase the heat gain or loss of the building. This causes an increase in the heating or cooling load and thus energy consumption increases. Therefore, it is necessary to determine the ideal window area by providing a balance between daylight supply and minimum energy consumption [5]. Low-emissivity coated glasses have high reflectance in the near-infrared spectrum and high transmittance in the visible region. They are usually placed on the outer surface of the inner glass of the glazing system and have a low emissivity (approximately 0.1–0.2) to decrease radiative heat transfer to the outside [6,7].

It is seen that there are many works in the literature on optimizing the wall insulation thickness of buildings. In a great many of these works, the degree-day (DD) or degree-hour (DH) method, which is an elementary and rough model, is used [8–25]. Among these studies, Dombayci et al. [11] specified the optimum insulation thickness (OIT) of the external wall by employing the heating DD concept. Çomaklı and Yuksel [12] researched the OIT for Turkey's three coldest cities by employing the DD values. Özkahraman and Bolattürk [13] computed the energy savings by employing porous tuff stone in the exterior walls of buildings. Their study indicated that OIT of tuff stone was found to be 15 cm. Bolattürk [14] computed OIT by employing the DD method for building exterior walls. In another study, Bolattürk [15] used the DH method. Kaynaklı [16] determined yearly heating energy requirement by employing long term measurements. Then, he calculated the OIT by using optimization model for five different types of fuel. Yu et al. [20] analysed the OIT by taking into consideration the effect of solar radiation based on the DD analysis. Shahid et al. [25] analysed the impact of cooling DD value and base temperature on the OIT [23]. Durmaz et al. [26] estimated the OIT for submarines. Özel and Pıhtılı [27] determined the OIT applying to external walls by considering heating and cooling DD values.

Dynamic heat transfer models should be used to obtain more accurate results. Numerical and analytical methods used for this purpose are available in the literature [28–39]. Among these studies, Al-Sanea and Zedan [35] researched the dynamic thermal properties of building walls composed of one, two and three insulation layers with the same thermal mass. Al-Sanea et al. [36] specified the OIT for cavity walls under steady periodic conditions employing the climate conditions of Riyadh. Al-Sanea et al. [37] studied impact of electricity tariff on the OIT in building walls under dynamic thermal conditions. Daouas [38] computed the OIT according to various wall orientations by employing an analytical method in Tunisian climate.

The impact of window glazing on thermal behavior has been examined by several researchers. Zenginis and Kontoleon [40] analyzed the effect of glazing rate, orientations and zone aspect ratio on the heat flux through the surfaces of a building zone. Kontoleon and Zenginis [41] also analyzed the rate of glazing surfaces to the total wall surface for different orientations. They investigated the impact of indoor temperature settings on heat gain or heat loss. Kontoleon and Bikas [42] studied the effect of the percentage of glazed openings and the type of glazing on indoor temperatures and energy efficiency. Shahid and Karimi [43] investigated the building envelope transmittance value with parameters such as orientation, shading devices, projection factor, glass type and window-wall ratio. Ozel and Ozel [44] investigated the thermal performance of walls which are containing windows by taking into consideration different structure materials However, these works do not contain the optimal insulation thickness.

In the literature, it is observed that the heat transfer from the windows is generally neglected when the OIT is specified. There are very few works on the optimization of the insulation thickness of the walls involving the windows. Arıcı and Karabay [45] specified the optimum air layer thickness of double-glazed windows using the DD method. Derradji et al. [46] determined the OIT of external walls containing windows in the region of Algiers, Algeria by using the heating and cooling DDs. Özkan and Onan [47] studied the effect of changing the window and external wall areas on the building's heating energy need and OIT with the P1-P2 method. In these studies, it is seen that the DD method, which is a rough and easy model under static conditions, is used. To get highly accurate conclusions, the heat transmission loads through the wall must be computed by employing a dynamic method based on the implicit finite difference (IFD) process. The impact of glazing area on OIT for different wall directions was previously researched by taking into consideration heating and cooling loads by Ozel [48]. In this study, unlike the previous one, it is aimed to investigate the effects of the glazing area on OIT, fuel consumption and emissions, considering only heating loads. Also, none of the studies in literature include the impact of glass area on fuel consumption under dynamic thermal conditions. This research emphasizes the importance of insulation and glazing area to increase energy efficiency and reduce environmental pollution in buildings. Measures taken to reduce emissions benefit both

human health exposed to air pollution and preventing climate change. Therefore, this research will contribute to the fight against climate change caused by emissions by reducing fuel consumption and emissions. One of the most effective solutions to reduce greenhouse gas emissions against climate change is energy efficiency. With the proper combination of window glass area and insulation in the wall, heating and cooling loads and therefore fuel consumption and emissions can be significantly reduced. Therefore, the basic goal of this study is to fill this information gap and to numerically research the impact of glass area on insulation thickness, fuel consumption and emissions under dynamic thermal conditions for south orientation in the climate of Elazığ, Turkey. The current article also includes heating and cooling loads for each month of the year according to glazing area, type of glazing and wall insulation. For this objective, the wall's glazing area is raised from 0% to 100%, and the annual transmission loads are computed individually for single-glazed and double-glazed windows. Then, the optimization of the insulation thickness and its environmental effects are examined for different GAP values.

MATERIALS AND METHODS

To investigate the effect of window glass area on insulation thickness and fuel consumption, the present analysis is based on the numerical solution of heat transmission through opaque and transparent surfaces, which are given flow chart in Figure 1. Firstly, the heat transmission loads are computed according to rising insulation thickness and percentage of glazing area in case the window is single glazing and double glazing. Secondly, the optimum thicknesses of insulation, energy savings and payback periods for %0, %20, %40 and %60 values of the glazing area are computed. Finally, the impacts of thermal insulation on the environment were investigated for values %0, %20, %40 and %60 of the glazing area.

Heat Transmission Through Opaque Surfaces

A composite wall consisting of layers that have different thermophysical properties is indicated schematically as in Figure 2. The one-dimensional equation related to the transient heat transfer may be described by the following equation:

$$\frac{\partial^2 T_j}{\partial x^2} = \frac{1}{\alpha_j}\frac{\partial T_j}{\partial t} \tag{1}$$

where α represents thermal diffusivity and is expressed as $\alpha=k/(\rho.c)$. ρ, c and k also represent the density, the specific heat and the thermal conductivity, respectively. It is assumed that these thermal properties are constant. It is considered that thermal contact between layers is excellent, and is expressed by the following equations:

$$T_j = T_{j+1} \tag{2}$$

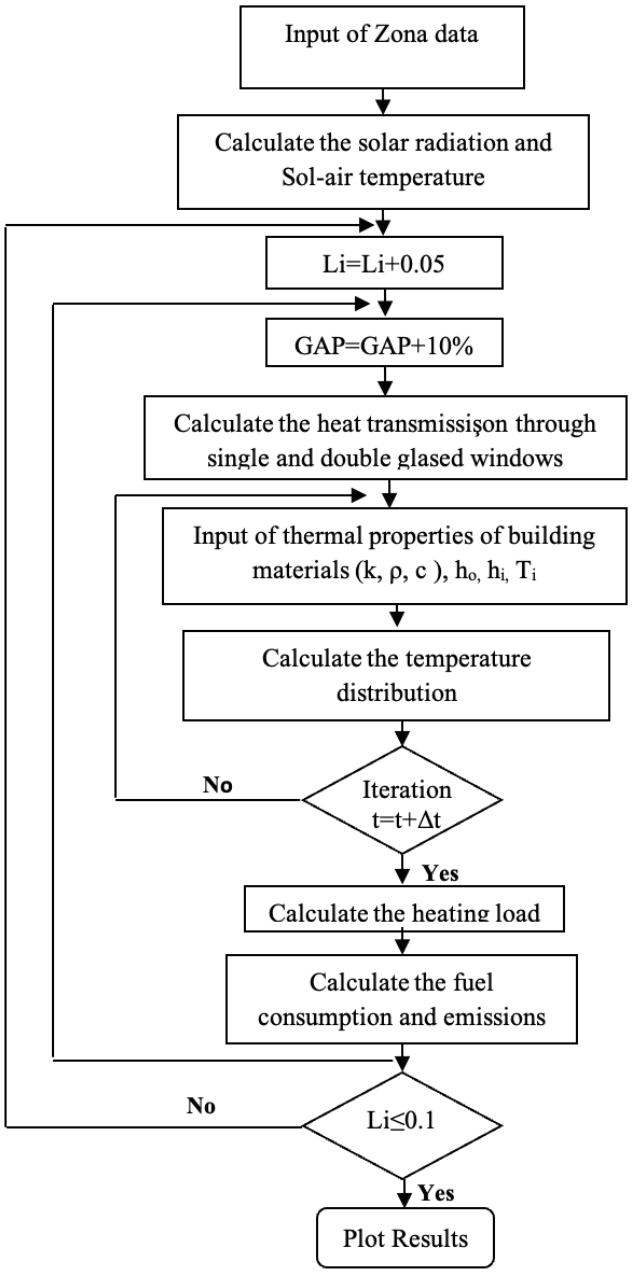


Figure 1. Flow chart.

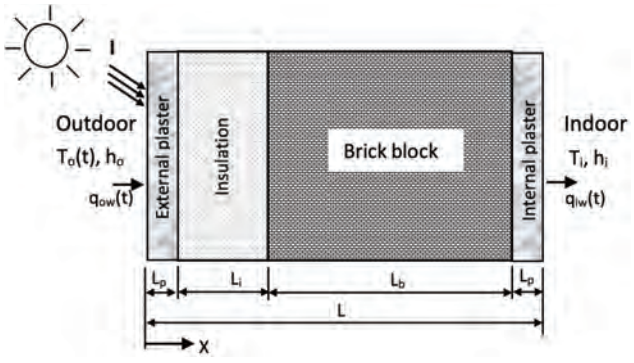


Figure 2. A multilayer composite wall [48] [created by author].

$$k_j \frac{\partial T_j}{\partial x} = k_{j+1} \frac{\partial T_{j+1}}{\partial x} \quad (3)$$

The wall's initial and boundary conditions are expressed as follows [49]:

$$T_j(x,0) = F_j \quad (4)$$

$$-k_1 \left(\frac{\partial T}{\partial x} \right)_{x=0} = h_o (T_o(t) - T_{x=0}) + a_o . I(t) \quad (5)$$

$$-k_N \left(\frac{\partial T}{\partial x} \right)_{x=L} = h_i (T_{x=L} - T_i) \quad (6)$$

The initial condition, F_j represents a uniform temperature that is selected as arbitrary across the wall. h_o and h_i represent external and internal convective coefficients, respectively. T_i denotes the inner air temperature. T_o denotes the outer air temperature. I and a_o represent solar radiation and the outdoor wall surface's solar absorptivity, respectively. The total solar radiation (I) incident upon a surface at any moment can be computed by

$$I = I_D + I_d + I_R \quad (7)$$

Where I_D, I_d and I_R represent the direct radiation, the diffuse sky radiation and the solar radiation reflected upon the surface from surrounding surfaces, respectively. The detailed computing procedures of equation 7 are imputed by reference [50].

Differential equations and boundary conditions have been previously solved in MATLAB by employing the IFD method, and the computing transactions are available in detail in the source given by [51,52]. It has previously been indicated that the numerical solution is in good harmony with the analytical solution [52]. Thus, the temperature distribution along the thickness of the wall was obtained by the numerical solution. The instantaneous heat transmission is computed following as:

$$q_{iw} = h_i (T_{x=L} - T_i) \quad (8)$$

Heat Transmission Through Transparent Surfaces

The solar radiation and the temperature difference between inner and outer thermal environments cause heat transmission through window glass. Direct, diffuse and reflected solar radiations can come to the outer surface of the window. Some of these radiations are transmitted directly through the glass, some are reflected and some are absorbed. As shown in Figure 3a, heat transfer through a single-glazed window is given by Threlkeld [49].

$$q_{isg} = F_s \tau_D I_D + \tau_d I_d + \tau_R I_R + (U/h_o)(F_s a_D I_D + a_d I_d + a_R I_R) + U(T_o - T_i) \quad (9)$$

$$U = 1/[(1/h_i) + (1/h_o)] \quad (10)$$

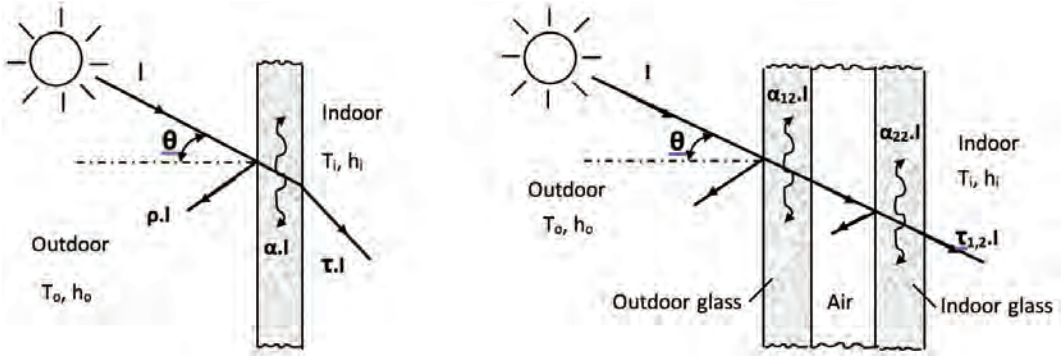
In the above equation, F_s represents the window surface's sunlit fraction. The heat transfer through the double-glazed window shown in Figure 3b can be computed as [53–55]:

$$q_{idg} = F_s \tau_{(1,2)D} I_D + \tau_{(1,2)d} I_d + \tau_{(1,2)R} I_R + (U/h_o)[F_s \alpha_{(1of2)D} I_D + \alpha_{(1of2)d} I_d + \alpha_{(1of2)R} I_R] + U[(1/h_o) + (1/h_a)](F_s \alpha_{(2of2)D} I_D + \alpha_{(2of2)d} I_d + \alpha_{(2of2)R} I_R) + U(T_o - T_i) \quad (11)$$

$$U = 1/[(1/h_i) + (1/h_a) + (1/h_o)] \quad (12)$$

In the above equation, ha represents the heat transfer coefficient of the air gap between the glass sheets. The transmittance, reflectivity and absorption for a single glass layer are computed as follows [49,55–57]:

$$\tau = [(1-r)^2 a]/[1-r^2 a^2] \quad (13)$$



(a) (b)

Figure 3. Heat Transmission from window with a) single glass b) double glass [48] [created by author].

$$\rho = r + r[(1-r)^2 a^2]/[1-r^2 a^2] \quad (14)$$

$$\alpha = 1 - r - [(1-r)^2 a]/[1-ra] \quad (15)$$

In the above equation, r represents the component reflectivity and may be found from the Frensel relations as follows:

$$r = \frac{1}{2}\left[\frac{\sin^2(\theta-\theta')}{\sin^2(\theta+\theta')} + \frac{\tan^2(\theta-\theta')}{\tan^2(\theta+\theta')} \right] \quad (16)$$

Where θ denotes the incidence angle of the sun's rays. Its detailed calculation procedures are available in the reference [49]. θ' represents the refraction angle and is computed as follows:

$$\theta' = \sin^{-1}(\sin\theta / n) \quad (17)$$

In the above equation, n represents the refraction index of the glass. The absorption coefficient symbolized by a is computed as follows:

$$a = \exp\left(-K.L/\sqrt{1-(\sin^2\theta/n^2)}\right) \quad (18)$$

In the above equation, K denotes the extinction coefficient. L is the thickness of the glass. Transmissivity and reflectivity for double-glazed windows have been shown by Parmelee as following [49]:

$$\tau_{1,2} = (\tau_1\tau_2)/(1-\rho_1\rho_2) \quad (19)$$

$$\rho_{1,2} = \rho_1 + (\tau_1^2\rho_2)/(1-\rho_1\rho_2) \quad (20)$$

Where subscript 1 represents the first glass sheet while subscript 2 represents the second glass sheet. Also, the solar absorption in each sheet of glass is as follows:

$$\alpha_{1of2} = [1-(\rho_1+\tau_1)][1-\rho_2(\rho_1-\tau_1)]/[1-\rho_1\rho_2] \quad (21)$$

$$\alpha_{2of2} = [1-(\rho_2+\tau_2)]\tau_1/[1-\rho_1\rho_2] \quad (22)$$

Daily and Yearly Total Heating and Cooling Transmission Loads

The research is carried out on the fifteenth day of every month of the year. Firstly, the instantaneous heat loads through the wall and window are determined as above. Daily transmission loads for the wall and window are computed by integrating instant load over a one-day (24 h) as follows:

$$Q_i = \int_0^{24h} q_i dt \quad (23)$$

Also, yearly total heating and cooling loads are acquired individually by adding the daily heating and cooling loads for the winter and summer periods.

The Glazing Area Percentage (GAP)

GAP is determined by accounting for heat transfer through external walls and windows. The window frame area is not taken into consideration in this analysis. When only the glass area is taken into account, *GAP* is calculated as [42,44]:

$$GAP = \frac{A_g}{A_t}.100 = \frac{A_g}{A_g + A_w}.100 \quad (24)$$

In the above equation, A_g and A_w represent glass and wall areas, respectively. Using this definition, the heating or cooling transmission loads can be computed for single and double glass with increasing glazing area percentages from 0% to 100% with a 10% increment as follows:

$$\underset{or\ gain}{Q_{loss}} = Q_{ig}.GAP + Q_{iw}.[1-GAP] \quad (25)$$

Where Q_{iw} and Q_{ig} represent the daily total loads of wall and window (for single or double-glazed), respectively.

Cost Analysis for Optimum Thickness of Insulation

Most of the studies available in the literature have calculated the OIT by taking something into consideration the heat gains and losses from only external walls. In this investigation, the OIT was computed by paying regard to the heat transmission not only through the exterior walls but also through the windows. The annual cost of energy for heating per unit area, **(*CA*)** is computed as follows [33]:

$$C_A = (Q_{loss}.C_F)/(H_u.\eta_s) \quad (26)$$

where Q_{loss} is the annual heat loss of the insulated wall as regards glass area percentage (kWh/m^2). C_F, H_{uand} η_s represent the fuel's cost (\$/kg), the fuel's lower heating value (kJ/kg) and the heating system's efficiency, respectively. The overall cost is determined as follows:

$$C_t = PWF.C_A + C_{ins}.L_{ins} \quad (27)$$

where C_{ins} and L_{ins} denote the insulation material's cost per unit volume (\$/m^3) and the insulation thickness (m), respectively. Besides, *PWF* represents the Present Worth Factor. *PWF* value for the *N* years lifetime is calculated as follıws [33,48]:

$$PWF = [(1+r)^N - 1]/[r(1+r)^N], \quad \begin{Bmatrix} i>g & r=(i-g)/(1+g) \\ i<g & r=(g-i)/(1+i) \end{Bmatrix} \quad (28)$$

$$PWF = N/(1+i), \qquad i = g \quad (29)$$

The energy savings per unit area are determined by the divergence between the total cost of an uninsulated wall and the total cost when the wall was insulated to the optimum thickness. It may be computed as follows [33,48]:

$$S_e = \left(\frac{((Q_{loss})_{un} - Q_{loss}).C_F}{H_u.\eta_s} \right) PWF - C_{ins}.(L_{ins})_{opt} \quad (30)$$

Where $(Q_{loss})_{un}$ represents the annual total heat loss in the wall without insulation as regards the percentage of glass area (kWh/m^2). Annual energy saving (A_s) is computed as follows [33]:

$$A_s = \frac{((Q_{loss})_{un} - Q_{loss}).C_F}{H_u.\eta_s} \quad (31)$$

The payback period (p_b) may be computed as below [28,42]:

$$PWF(p_b) = \frac{C_i.(L_i)_{opt}}{A_s}, \quad p_b = \frac{\ln\left(1 - r(C_i.(L_i)_{opt} / A_S)\right)}{\ln\left(1/(1+r)\right)} \quad (32)$$

The Fuel Consumption and Emissions

The combustion equation of fuel is as follows [33,58]:

$$C_a H_b O_d S_e N_f + \varepsilon.X(O_2 + 3.76N_2) \rightarrow aCO_2 + \left(\frac{b}{2}\right)H_2O + (\varepsilon - 1)XO_2 + eSO_2 + YN_2 \quad (33)$$

If the above chemical reaction is balanced based on the principle of conservation of mass, the constants specified by X and Y are determined as follows:

$$X = a + e + (b/4) - (d/2), \quad Y = (f/2) + 3.76.\varepsilon.X \quad (34)$$

The CO_2 and NO_2 emission rates of combustion products which are result from the combustion of 1 kg of fuel may be computed as follows [33]:

$$M_{CO_2} = (aCO_2)/M \equiv kg\ CO_2 / kg\ fuel \quad (35)$$

$$M_{SO_2} = (eSO_2)/M \equiv kg\ SO_2 / kg\ fuel \quad (36)$$

M denotes fuel mol weight and may be computed as follows:

$$M = 12a + b + 16d + 32e + 14f \quad kg/kmol \quad (37)$$

Total CO_2 and SO_2 emissions are determined based on annual fuel consumption as follows:

$$M_{CO_2} = \frac{44a}{M} M_F, \quad M_{SO_2} = \frac{64e}{M} M_F \quad (38)$$

The annual fuel consumption is computed as follows [33]:

$$M_F = Q_{loss} / H_u.\eta_s \quad (39)$$

RESULTS AND DISCUSSION

In this research, the impact of glass area on insulation thickness and fuel consumption was numerically investigated for a south-oriented wall in Elazig province which is located in Turkey's Eastern Anatolia zone and has a latitude of 38.41°N and a longitude of 39.14°E. The wall structure consists of external plaster which has 20 mm thickness, an insulation layer which has variable thickness, brick block which has 200 mm thickness and internal plaster which has 20 mm thickness. The insulation thickness is increased from 5 mm to 100 mm at every 5 mm interval. The windows consist of single glass with a thickness of 3 mm and double glass with air gaps, each layer of which is 3 mm thick. For this aim, the window glass area on the wall which has different insulation thicknesses is risen from 0% to 100% with a ratio of 10%, taking into account the single glazing and double glazing. Yearly heat loads through the wall were numerically obtained under dynamic thermal conditions. Thermophysical characteristics of materials which are used in the structure of the wall are imputed in Table 1 [28]. It is assumed that the variation of thermal properties of materials given in Table1 is negligible. Yearly heating loads through the opaque wall and window glass are employed as input data in the cost analysis to specify the OIT over the 20-year life of the building. The parameters used for OIT and environmental analysis are imputed in Table 2.

Table 1. Thermophysical properties of building materials, and cost of insulation [48] [created by author]

Material	k (W/m K)	ρ (kg/m^3)	c (J/kg K)	C_1 ($/m^3)
Brick block	0.620	1800	840	—
Extruded polystyrene	0.029	35	1213	304.23
Cement plaster	0.720	1865	840	—

Table 3. Indoor design temperatures selected for the representative day of each month [48] [created by author]

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Ti (°C)	20	20	20	22	23	23	23	23	23	22	20	20

Table 2. The parameters used in calculations [48] [created by author]

Parameter	Value
C_r	0.3777 $/kg
H_u	29.307*10^6 J/kg
η_s	% 65
Chemical formula of Coal [58]	$C_{7.078}H_{5.149}O_{0.517}S_{0.01}N_{0.086}$
Interest rate, i	% 8.75
Inflation rate, g	% 7.50

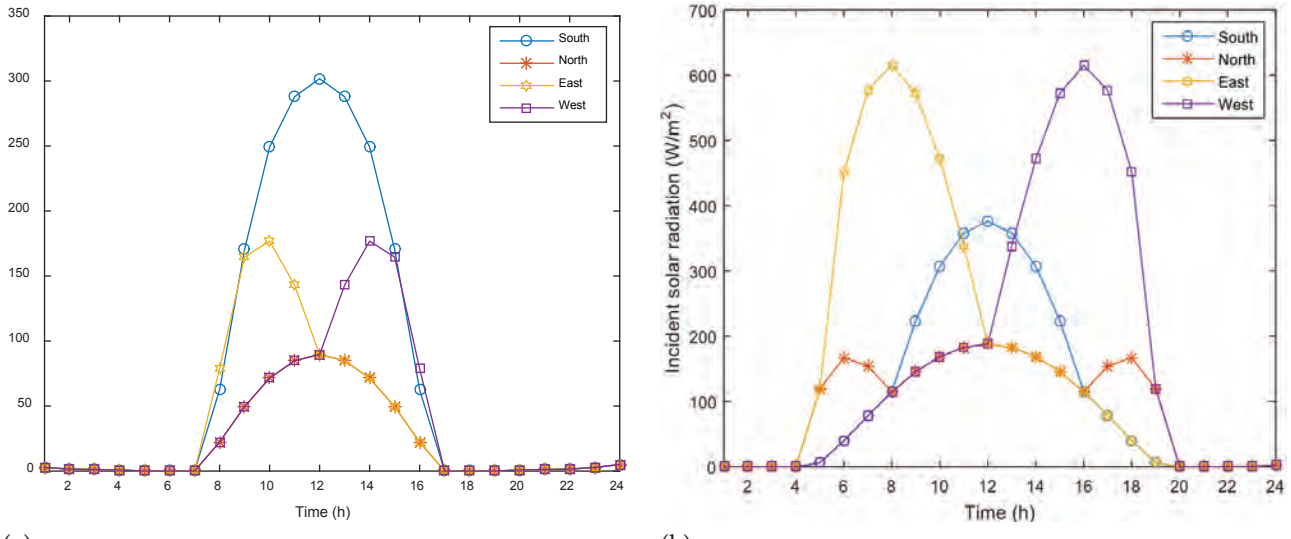
The external environment temperatures of Elazig were ensured from the meteorological data of the years 2010-2020 [59]. As in Table 3, fixed inside air temperatures were chosen for each month's 15th day. The solar absorption of the opaque wall was chosen to be 0.8. The combined convection coefficients for the outer and inner wall surfaces were assumed as 22 and 9 W/m^{2}K, respectively [31]. In this work, clear plate glass was chosen as the glass type. For this type of glass, the refractive index (n) was received as 1.526 and the extinction coefficient (K) was taken as 6.85 m-1. It is assumed that the sunlit fraction (Fs) was 1. The thickness of the air gap between the double glazing was chosen as 6 mm. The heat transfer coefficient of the air gap was received to be 5.56 W/m^{2}K [53].

In this study, a Matlab program was comprehensively developed to perform numerical calculation. With the help of this program, heat transmission loads through walls and windows, OIT, fuel consumption and emissions are computed depending on many parameters.

Figure 4(a) and (b) show the incident solar radiation's hourly variation in all wall directions for Jan. 15 and Jul. 15, respectively. It is observed that in summer, the incident solar radiation's highest value is received as 615.36 W/m^2 at 8:00 hour in the east direction and at 16:00 hour in the west direction while in winter, it is acquired to be 301.77 W/m^2 at noon (12:00 hour) in the south direction. On the other part, the incident solar radiation's highest value for Jul. 15 and Jan. 15 occurs at 8:00 and 10:00 hours, respectively in the east direction and at 16:00 and 14:00 hours, respectively in the west direction while it is acquired at noon (12:00 hour) in the south and north directions. It is obvious that in summer, the south facade does not receive the maximum radiation because the sun is too high in the sky, while in winter, it receives the maximum radiation because the sun is at a much lower altitude [60]. These results were found to be in good agreement with previously obtained results [7,60].

Figure 5 presents the incident solar radiation's highest values for all months in different orientations. It is revealed that the east and west directions give maximum solar radiation in July while the south orientation gives in September [48]. It is also revealed that the lowest solar radiation is acquired in the north orientation for all months. Particularly, south walls provide large solar heat gain and eliminate heat losses in winter, while in summer the solar heat gain is relatively small and limits unwanted gain. The most desirable external wall orientation is south and therefore, the south direction is selected in this study.

In case the wall is insulated, as the GAP is increasing, heating and cooling transmission loads are increasing for single glass. However, as the GAP is increasing, the heating transmission load on the wall consisting of a double-glazed window is increasing for January month while it is decreasing for February month. In addition, as the GAP



(a) (b)

Figure 4. Hourly variation of incident solar radiation for all wall orientations: (a) in January 15 (b) in July 15.

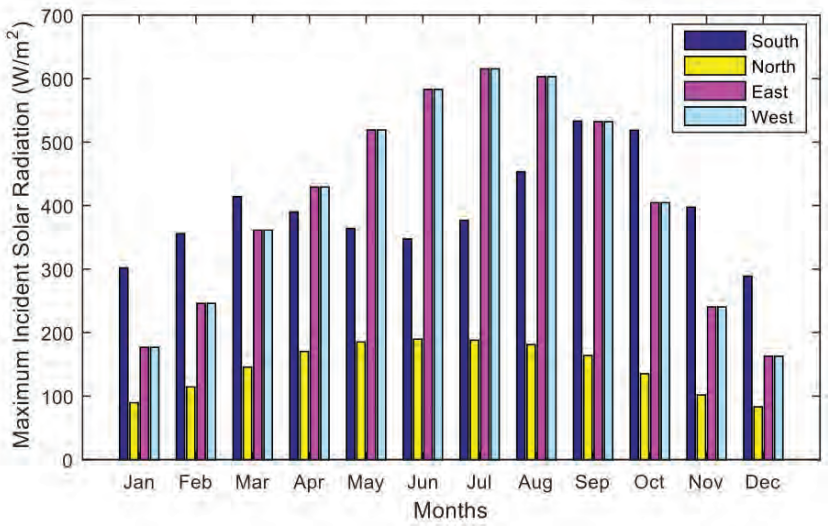


Figure 5. Maximum values of incident solar radiation according to months for all wall orientations [48] [created by author].

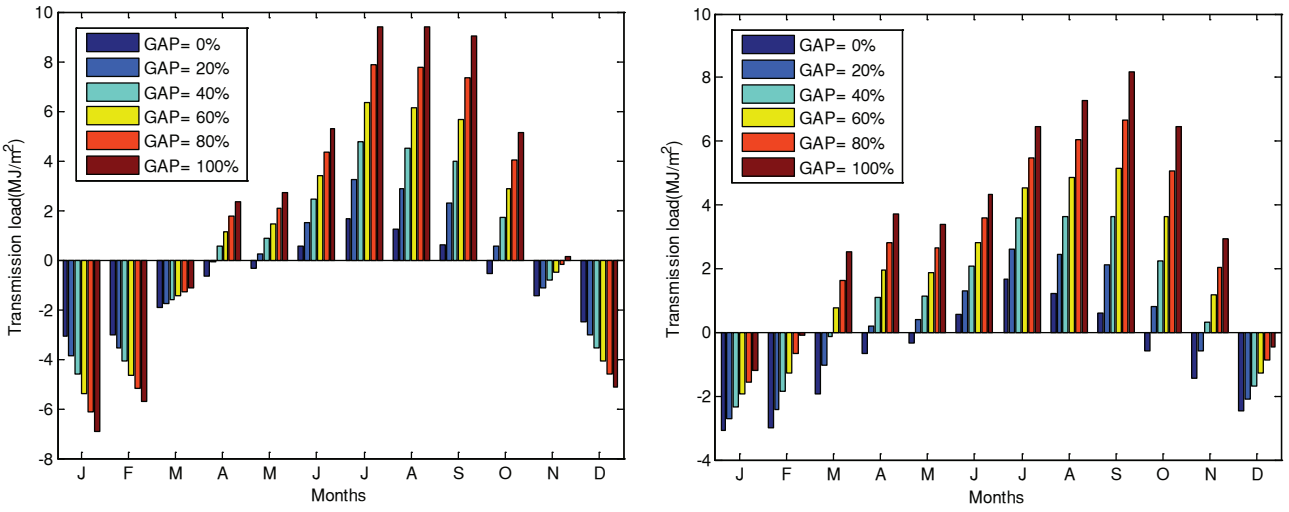
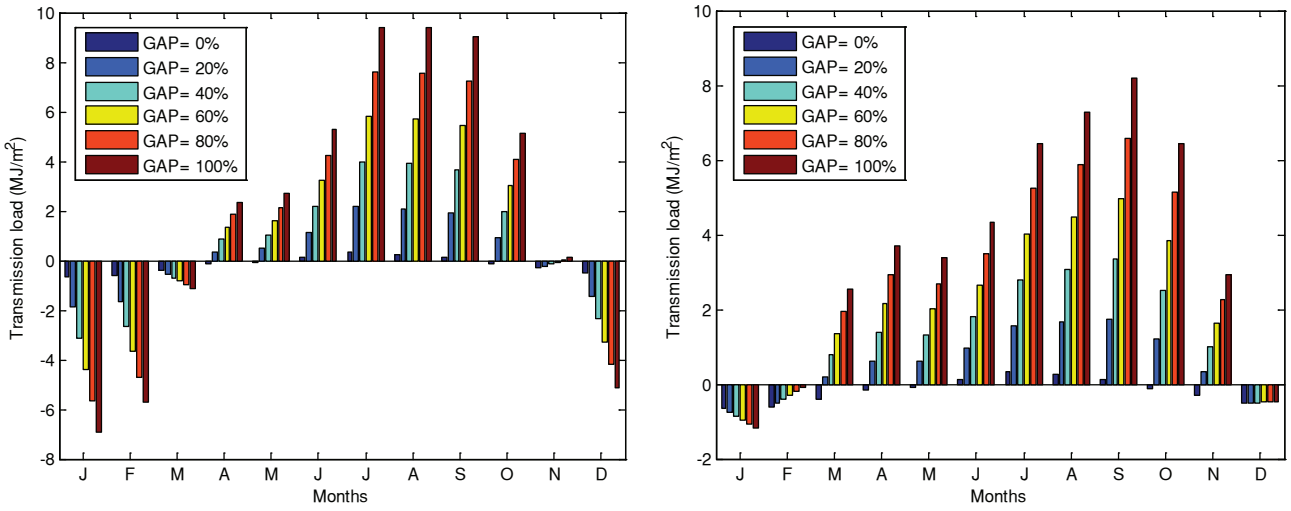


Figure 6. Daily total cooling and heating transmission loads for the 15th day of each month with respect to different GAP values in uninsulated wall for a) single glass and b) double glass.

is increasing, the cooling transmission load is increasing for the walls consisting of single and double-glazed windows.

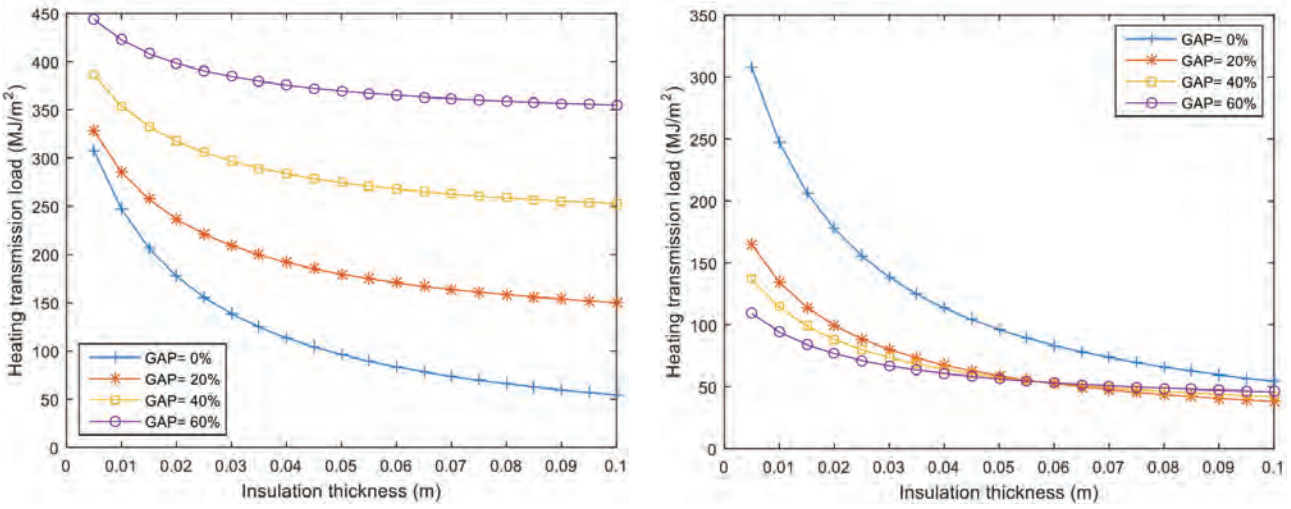
Figure 8(a) and (b) show the trends of annual heating loads with respect to insulation thickness for various GAP values in the wall consisting of single-glazed and double-glazed windows, respectively. The annual heating load is determined considering the months in which the heating load occurs. It is revealed that heating transmission loads for all GAP values decline as the insulation thickness rises. This decline in the heating load of wall consisting of a single-glazed window is acquired to be 82.25%, 54.19%, 34.55% and 20% for 0%, 20%, 40% and 60% values of GAP, respectively. Whereas, this decrease for wall consisting of a double-glazed window is acquired to be 82.25%, 76.86%, 69.29% and 57.89%. It is seen that double glazing provides a greater reduction in heating load. Additionally, the results show that at larger glazing ratios, the insulation thickness has less impact on the heating load. This result is consistent with previously reported using P1-P2 method under different climate conditions [47].

The heating load in 20 mm insulated wall for single glazing is acquired to be 177.57, 236.79, 317.51 and 398.23 MJ/m^2 year for values 0%, 20%, 40% and 60% of the GAP, respectively while it is obtained to be 177.57, 114.24, 88.24



(a) (b)

Figure 7. Daily total cooling and heating transmission loads for the 15th day of each month with respect to different GAP values in 6 cm insulated wall for a) single glass and b) double glass.



(a) (b)

Figure 8. Yearly heating transmission loads according to increasing insulation thickness and glazing area percentage (GAP) for a) single glass and b) double glass.

and 76.81 MJ/m^2 year for double glass. On the other hand, The heating load in 60 mm insulated wall for single glazing is acquired to be 83.49, 170.63, 267.89 and 365.15 MJ/m2 year for values 0%, 20%, 40% and 60% of the GAP, respectively while it is obtained to be 83.49, 52.58, 52.92 and 53.26 MJ/m2 year for double glass. It is observed that in south facing wall, as GAP rises, the heating loads for single glazing increase. The results show that increasing the glazing area for single glazing results in an additional heating load. This conclusion is consistent with conclusion obtained under the environmental conditions of the Greek region [41]. It is seen that when GAP rises, the heating transmission load for double glazing declines at insulation thickness's smaller values while it increases at bigger values of insulation thickness. It is revealed that the heating loads for different GAP values of a 50 mm insulated wall consisting of a double-glazed window approach each other and are almost the same for insulation thicknesses from 55 mm to 65 mm. In high insulation thicknesses, it is revealed that for double glazing, the heating load rises as the GAP rises, as in single glazing. It is revealed that for an insulated wall consisting of a double-glazed window, the heating loads are significantly reduced for different GAP values when compared to single-glazed. Using double-glazed instead of

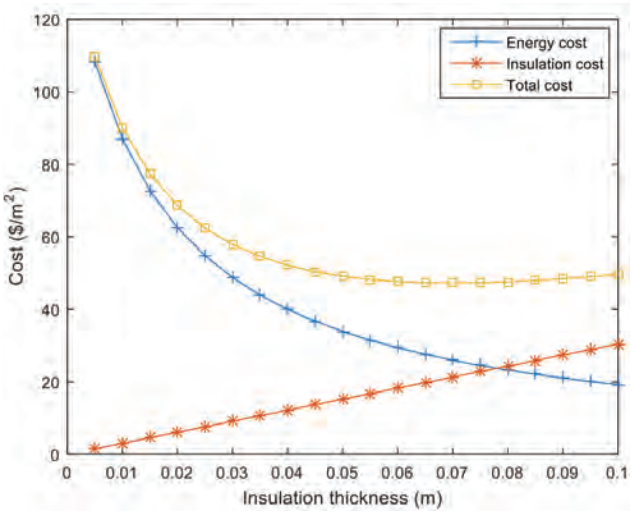


Figure 9. Variation of costs to insulation thickness for GAP=0 value in a south-oriented wall with single glass.

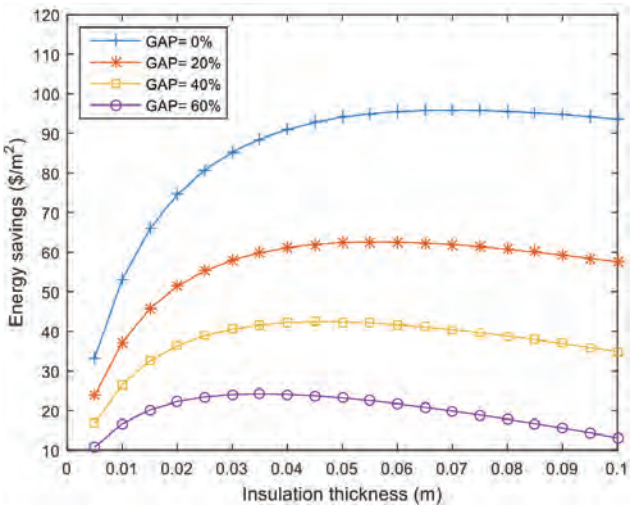


Figure 10. Effect of insulation thickness on energy savings for different GAP values in a south oriented wall with single glass.

single-glazed in a 50 mm insulated wall reduces the heating load by 67.17%, 78.98% and 84.72% for 20%, 40% and 60% GAP values, respectively.

For south orientation, it is observed that In winter, double glazing of windows increases the heat gain compared to single glazing while in summer, double glazing of windows reduces heat gain as seen in Reference [60]. This situation, explained above for winter conditions, can be attributed to the fact that double glazing increases heat gain compared to single glazing in winter. It is revealed that the larger double-glazing area on the south facade significantly reduces heat loss in winter.

Figure 9 demonstrates the trends of costs to insulation thickness for GAP=0% in a wall in the south direction. It is revealed that as the thickness of insulation rises, the cost of insulation linearly rises while the cost of energy decreases. The sum of insulation cost and energy cost gives the total cost. The insulation thickness where the total cost is minimum gives OIT. The impression of insulation thickness on energy savings for various GAP values of a wall in the south direction is shown in Figure 10. It is revealed that the energy saving decreases as GAP increases.

The optimum thicknesses of insulation, energy savings and payback periods for various values of the GAP according to type of glazing are illustrated in Table 4. It is revealed that when the GAP rises, OIT and energy-saving decline for south wall while the payback period rises. The conclusions show that the GAP values have a notable effect on OIT, energy saving and payback period. It is revealed that OITs for single glazing are acquired as 70, 55, 45 and 35 mm for 0%, 20%, 40% and 60% values of the GAP, respectively while they are acquired as 70, 45, 35 and 25 mm for double glazing. It is seen that the OIT decreases as GAP increases for both glass types. This result is consistent with those acquired by other studies using the DD method [46,47] under different climate conditions. Also, the similar conclusions are obtained by using both heating and cooling transmission loads [48]. For example, it is seen in Reference 46 that the OIT for a wall including a double-glazed window is 23 mm for 10% and 10 mm for 50% by using cooling DD in Algiers.

Table 4. The OIT, energy savings and payback periods for different values of GAP

Glazing Type	GAP	OIT (m)	Energy savings ($/m²)	Payback period (years)
Single glazing	0%	0.070	95.90	3.30
Single glazing	20%	0.055	62.56	3.85
Single glazing	40%	0.045	42.39	4.47
Single glazing	60%	0.035	24.23	5.63
Double glazing	0%	0.070	95.90	3.30
Double glazing	20%	0.045	56.88	3.53
Double glazing	40%	0.035	28.09	5.05
Double glazing	60%	0.025	14.50	6.37

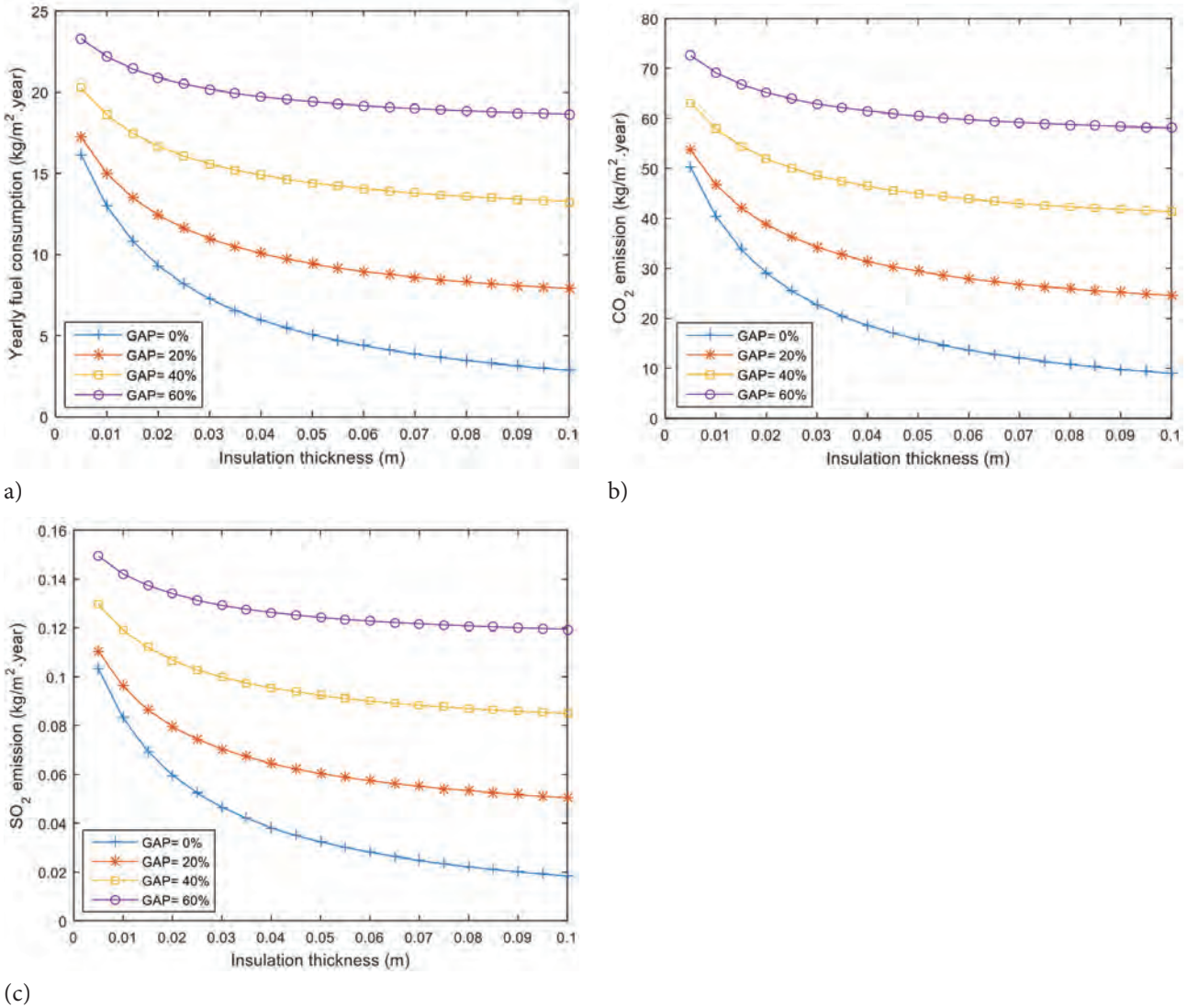


Figure 11. Variations of the yearly fuel consumption and, CO_2 and SO_2 emissions according to insulation thickness for different GAP values in a south oriented wall with single glass.

The thermal insulation's environmental impacts are also researched for different values of the GAP. Figure 11(a-c) shows fuel consumption and emissions trends according to insulation thickness and various GAP values for the single glazing in the south direction. It is revealed that the annual fuel consumption and emissions decrease for all GAP values depending on increasing insulation thickness. This decline in the wall consisting of a single-glazed window is acquired as 82.25%, 54.19%, 34.55% and 20% for 0%, 20%, 40% and 60% of the glazing area, respectively. It is obvious that this decrease is the same as the decrease in heating load, as seen in Figure 7a. The results show that if the wall consisting of a single-glazed window is insulated with the OIT, fuel consumption and emissions are reduced by 76.02%, 46.73%, 27.77% and 14.46% for values 0%, 20%, 40% and 60% of the glass area, respectively. However, it is seen that fuel consumption and emissions increase as GAP rises. For example, at 6 cm insulation thickness, yearly fuel consumption was obtained as 4.3826, 8.9574, 14.0630 and 19.1686 kg/m^2.year for 0%, 20%, 40% and 60% of the glazing area, respectively. It is seen that the increment in the glazing area for single glass induces additional heating, fuel consumption and emissions. It is observed that insulation is more effective in the lower glazing areas. These results acquired are found to be in good agreement with previously reported using P1-P2 method under different climate conditions [47].

CONCLUSION

In this work, the effect of window glass area on insulation thickness and fuel consumption is researched numerically for the south orientation under dynamic thermal conditions in the climate of Elazığ, Turkey. For this goal, the

window glass area on the wall which has different insulation thicknesses is enhanced by %10 from %0 to %100. The results showed that as the glazing area percentage increases, the heating transmission load for single glazing increases in all wall insulation thicknesses. It is seen that choosing smaller glazing areas in insulated walls containing a single-glazed window results in less heating load.

The results also showed that as the glazing area percentage rises, the heating transmission load for double glazing declines at minor values of the insulation thickness and becomes almost the same for approximately 6 cm insulation thickness. It is observed that at lower values of insulation thickness, the large double-glazing area on the south facade significantly reduces heat loss in winter from the point of energy efficiency.

It is seen that the increment in the glazing area for single glass induces additional heating, additional fuel consumption and emissions. Besides, it is obvious that double glazing can provide a greater reduction in terms of insulation thickness, heating load, fuel consumption and emissions.

The conclusions demonstrate that the glazing area and type of glazing have a notable effect on heating and cooling loads, the optimum insulation thickness, fuel consumption and emissions.

As a result, it appears that with the appropriate combination of design parameters such as window glass area, glazing type and wall insulation, heating and cooling loads and therefore fuel consumption and emissions can be substantially reduced. The reduction in heating and cooling loads results in lower building operating costs. Therefore, the results acquired in this work will be helpful from the point of energy efficiency in the selection of glass type, the glazing area and insulation thickness when designing building exterior walls. In addition, this work will contribute to the fight against climate change caused by emissions by reducing fuel consumption and emissions. Besides, both heat loss and fuel consumption; and hence emissions, can be reduced by changing the window glass of existing buildings to double glazing and applying insulation to the walls from the outside.

It is revealed that some findings acquired in this work are supported by the results that were acquired in different climate zones in the literature. The same methodology can be replicated to other kinds of buildings (residential or commercial). Additionally, the study can be also expanded for different climatic regions, different building orientations, different insulation materials and advanced glazing materials using the same methodology, and presented in another study.

Since this study focuses on the effect of wall insulation and glazing area on heating and cooling loads and hence fuel consumption and emissions, the daylight factor is not taken into account in this study. The effect of artificial lighting on energy consumption can also be investigated in another study.

AUTHORSHIP CONTRIBUTIONS

Meral Özel: Investigation, methodology, software, analyzing and interpreting the data, discussion, writing original draft, review and editing.

Serhat Şengür: Investigation, analyzing and interpreting the data, discussion, writing original draft, review and editing.

DATA AVAILABILITY STATEMENT

The authors confirm that data supporting the findings of this study are available within the article.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Pargana N, Pinheiro MD, Silvestre JD, Brito J. Comparative environmental life cycle assessment of thermal insulation materials of buildings. Energy Build 2014;82:466–481. [CrossRef]

[2] Omer AM. Energy, environment and sustainable development. Renew Sustain Energy Rev 2008;12:2265–2300. [CrossRef]

[3] Oral GK. Appropriate window type concerning energy consumption for heating. Energy Build 2000;32:95–100. [CrossRef]

[4] Karabay H, Arıcı M. Multiple pane window applications in various climatic regions of Turkey. Energy Build 2012;45:67–71. [CrossRef]

[5] Ghisi E, Tinker J. Optimising energy consumption in offices as a function of window area and room size. 7th International IBPSA Conference. Rio de Janeiro, Brazil; 2001 p. 13–15.

[6] Tsai DC, Chang ZC, Kuo BH, Chen EC, Huang YL, Hsieh TJ, et al. Thermal stability and optical properties of low emissivity multilayer coatings for energy-saving applications. Ceram Int 2020;46:7991–7997. [CrossRef]

[7] Kontoleon KJ. Dynamic thermal circuit modelling with distribution of internal solar radiation on varying façade orientations. Energy Build 2012;47:139–150. [CrossRef]

[8] Hasan A. Optimizing insulation thickness for buildings using life cycle cost. Appl Energy 1999;63:115–124. [CrossRef]

[9]Dombayci OA. The environmental impact of optimum insulation thickness for external walls of buildings. Build Environ 2007;42:3855–3859. [CrossRef]
[10] Sisman N, Kahya E, Aras N, Aras H. Determination of optimum insulation thicknesses of the external walls and roof (ceiling) for Turkey ' s different degree-day regions. Energy Policy 2007;35:5151–5155. [CrossRef]
[11] Dombayci OA, Golcu M, Pancar Y. Optimization of insulation thickness for external walls using different energy-sources. Appl Energy 2006;83:921–928. [CrossRef]
[12] Comakli K, Yuksel B. Optimum insulation thickness of external walls for energy saving. Appl Therm Eng 2003;23:473–479. [CrossRef]
[13] Ozkahraman HT, Bolatturk A. The use of tuff stone cladding in buildings for energy conservation. Constr Build Mater 2006;20:435–440. [CrossRef]
[14] Bolatturk A. Determination of optimum insulation thickness for building walls with respect to various fuels and climate zones in Turkey. Appl Therm Eng 2006;26:1301–1309. [CrossRef]
[15] Bolatturk A. Optimum insulation thicknesses for building walls with respect to cooling and heating degree-hours in the warmest zone of Turkey. Build Environ 2008;43:1055–1064. [CrossRef]
[16] Kaynakli O. A study on residential heating energy requirement and optimum insulation thickness. Renew Energy 2008;33:1164–1172. [CrossRef]
[17] Erturk M. A new approach to calculate the energy saving per unit area and emission per person in exterıor wall of building using different insulation materıals and air gap. J Fac Eng Archit Gazi Univ 2016;31:395–406.
[18] Erturk M. A new model for exergetic optimum insulation thickness. Int J Exergy 2017;22:309–330. [CrossRef]
[19] Erturk M, Kecebas A. Prediction of the effect of insulation thickness and emission on heating energy requirements of cities in the future. Sustain Cities Soc 2021;75:103270. [CrossRef]
[20] Yu J, Yang C, Tian L, Liao D. A study on optimum insulation thicknesses of external walls in hot summer and cold winter zone of China. Appl Energy 2009;86:2520–2529. [CrossRef]
[21] Mahlia TMI, Taufiq BN, Masjuki HH. Correlation between thermal conductivity and the thickness of selected insulation materials for building wall. Energy Build 2007;39:182–187. [CrossRef]
[22] Kaynakli O. A review of the economical and optimum thermal insulation thickness for building applications. Renew Sustain Energy Rev 2012;16:415–425. [CrossRef]
[23] Kurt H. The usage of air gap in the composite wall for energy saving and air pollution. Environ Prog Sustain Energy 2011;30:450–458. [CrossRef]
[24] Mahlia TMI, Iqbal A. Cost benefits analysis and emission reductions of optimum thickness and air gaps for selected insulation materials for building walls in Maldives. Energy 2010;35:2242–2250. [CrossRef]
[25] Shahid M, Karimi MN, Mishra AK. Optimum insulation thickness for external building walls for different climate zone in India. J Therm Eng 2021;10:1198–1211. [CrossRef]
[26] Durmaz S, Colak AB, Mercan H, Dalkilic AS. Determination of optimum insulation thickness in submarines 2023;9:1655–1666. [CrossRef]
[27] Ozel M, Pıhtılı K. Determination of optimum insulation thickness by using heating and cooling degree-day values. J Eng Nat Sci 2008;26:191–197.
[28] Ozel M. Effect of wall orientation on the optimum insulation thickness by using a dynamic method. Appl Energy 2011;88:2429–2435. [CrossRef]
[29] Ozel M. Thermal performance and optimum insulation thickness of building walls with different structure materials. Appl Therm Eng 2011;31:3854–3863. [CrossRef]
[30] Ozel M. The influence of exterior surface solar absorptivity on thermal characteristics and optimum insulation thickness. Renew Energy 2012;39:347–355. [CrossRef]
[31] Ozel M. Cost analysis for optimum thicknesses and environmental impacts of different insulation materials. Energy Build 2012;49:552–559. [CrossRef]
[32] Ozel M. Determination of optimum insulation thickness based on cooling transmission load for building walls in a hot climate. Energy Convers Manag 2013;66:106–114. [CrossRef]
[33] Ozel M. Thermal, economical and environmental analysis of insulated building walls in a cold climate. Energy Convers Manag 2013;76:674–684. [CrossRef]
[34] Al-Sanea SA, Zedan MF. Optimum insulation thickness for building walls in a hot-dry climate. Int J Ambient Energy 2002;23:115–126. [CrossRef]
[35] Al-Sanea SA, Zedan MF. Improving thermal performance of building walls by optimizing insulation layer distribution and thickness for same thermal mass. Appl Energy 2011;88:3113–124. [CrossRef]
[36] Al-Sanea SA, Zedan MF, Al-Ajlan SA, Abdul Hadi AS. Heat transfer characteristics and optimum insulation thickness for cavity walls. J Therm Envel Build Sci 2003;26:285–307. [CrossRef]
[37] Al-Sanea SA, Zedan MF, Al-Ajlan SA. Effect of electricity tariff on the optimum insulation-thickness in building walls as determined by a dynamic heat-transfer model. Appl Energy 2005;82:313–30. [CrossRef]
[38] Daouas N. A study on optimum insulation thickness in walls and energy savings in Tunisian buildings based on analytical calculation of cooling and heating transmission loads. Appl Energy 2011;88:156–164. [CrossRef]

[39] Daouas N, Hassen Z, Aissia H Ben. Analytical periodic solution for the study of thermal performance and optimum insulation thickness of building walls in Tunisia. Appl Therm Eng 2010;30:319–326. [CrossRef]

[40] Zenginis DG, Kontoleon KJ. Influence of orientation, glazing proportion and zone aspect ratio on the thermal performance of buildings during the winter period. Environ Sci Pollut Res 2018;25:26736–26746. [CrossRef]

[41] Kontoleon KJ, Zenginis DG. Analysing Heat Flows Through Building Zones in Aspect of their Orientation and Glazing Proportion, under Varying Conditions. Procedia Environ Sci 2017;38:348–355. [CrossRef]

[42] Kontoleon KJ, Bikas DK. Modeling the influence of glazed openings percentage and type of glazing on the thermal zone behavior. Energy Build 2002;34:389–399. [CrossRef]

[43] Shahıd M, Karımı MN. Optimization of energy transmittance through building envelope for hot dry climate. J Therm Eng 2022;8:595–605. [CrossRef]

[44] Ozel M, Ozel C. Effect of window-to-wall-area ratio on thermal performance of building wall materials in Elazığ, Turkey. PLoS One 2020;15:1–14. [CrossRef]

[45] Arici M, Karabay H. Determination of optimum thickness of double-glazed windows for the climatic regions of Turkey 2010;42:1773–1778. [CrossRef]

[46] Derradji L, Imessad K, Amara M, Boudali Errebai F. A study on residential energy requirement and the effect of the glazing on the optimum insulation thickness. Appl Therm Eng 2017;112:975–985. [CrossRef]

[47] Ozkan DB, Onan C. Optimization of insulation thickness for different glazing areas in buildings for various climatic regions in Turkey 2011;88:1331–1342. [CrossRef]

[48] Ozel M. Influence of glazing area on optimum thickness of insulation for different wall orientations. Appl Therm Eng 2019;147:770–780. [CrossRef]

[49] Threlkeld JL. Thermal Environment Engineering. 3rd ed. London: Pearson; 1998.

[50] Duffie JA, Beckman WA. Solar engineering of thermal processes. 2nd ed. New York: John Wiley & Sons; 1991.

[51] Ozel M, Pihtili K. Optimum location and distribution of insulation layers on building walls with various orientations 2007;42:3051–3059. [CrossRef]

[52] Ozel M, Pihtili K. Investigation of the most suitable location of insulation applying on building roof from maximum load levelling point of view. Build Environ 2007;42:2360–2368. [CrossRef]

[53] Davies MG. Useful solar gains through a south-facing window in the UK climate. Build Environ 1980;15:253–272. [CrossRef]

[54] Kontoleon KJ, Bikas DK. The influence of the zone's indoor temperature settings on the cooling/heating loads for fixed and controlled ventilation. Build Environ 2006;41:75–86. [CrossRef]

[55] Ozel M. Investigation of methods for minimizing heat gain and losses in the buildings with periodical temperature variation on the wall surface. Firat Univ Nat Sci Inst 2003.

[56] Ozel M, Pihtili K. Investigation of effect of wall insulation thickness on window areas. J Fac Eng Archit Gazi Univ 2008;23:655–662.

[57] Kılıç A, Ozturk A. Güneş Enerjisi. İstanbul: Kipaş Dağıtımcılık; 1983.

[58] Yildiz A, Gurlek G, Erkek M, Ozbalta N. Economical and environmental analyses of thermal insulation thickness in buildings. J Therm Sci Technol 2008;28:25–34.

[59] State Meteorological Station. Records for weather data. Turkey, 1997-2007.

[60] Ozel M. Impact of glazing area on the thermal performance of buildings. Int J Ambient Energy 2022;43:2039–2055. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2043

Research Article

The role of building information modeling in enhancing public-private partnership project outcomes: Insights from Jordan and Malaysia

Ali Al-SHAMAYLEH[1,*], Sharifah ZAKARIA[1], Nur RAHIM[2]

[1]*Universiti Sains Malaysia, 11800, Malaysia*
[2]*Universiti Malaysia Perlis, 02600, Malaysia*

ARTICLE INFO

Article history
Received: 26 July 2024
Revised: 03 October 2024
Accepted: 17 December 2024

Keywords:
Building Information Modeling (BIM); Public-Private Partnership (PPP) Projects; Project Performance

ABSTRACT

Public-Private Partnerships (PPP) are crucial for infrastructure development in developing countries. PPP projects frequently face challenges, sych as inefficiencies in project management and collaboration. Building Information Modeling is capable to address these challenges, by enhancing collaboration and improving project efficiency. This study aims to evaluate the impact of BIM on PPP projects in Jordan and Malaysia. PPPs in devoloping countries play a pivotal role in infrastructure development. The study uses a questionnaire approach to achieve the objectives. A total of 1100 questionnaires were distributed to experts from public, private, and educational institutions, 441 valid responses received, which represents a 40.1% response rate. The questionnaire covered five aspects, starting with demographic data, then followed by the key aspects of BIM adoption, which are administrative, sustainability, management, tools framework, and productivity. The data was analyzed using SPSS software to assess BIM's influence on PPP project performance. The results show that BIM adoption could significantly improve performance in areas such as document management, interface management, energy efficiency, public service quality, and organizational structures, with specific increases in productivity and sustainability metrics. The experts identify that the adoption has positive impact on PPP projects in both Jordan and Malaysia. The study provides an imperical evidence of the potential of success PPP projects if and when BIM is adopted. The results shows improvements in document management, effective interface management, energy efficiency, quality and innovation of public services, and organizational structures. The study also contributes in inriching the body of literature on the context on BIM and PPPs, and showcasing the applications developting countries.

Cite this article as: Al-Shamayleh A, Zakaria S, Rahim N. The role of building information modeling in enhancing public-private partnership project outcomes: Insights from Jordan and Malaysia. Sigma J Eng Nat Sci 2026;44(2):1383–1398.

***Corresponding author.**
*E-mail address: ali.be@student.usm.my

This paper was recommended for publication in revised form by Regional Editor Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Public-private partnership (PPP) investments play a crucial role in driving economic growth in developing Asian countries [1]. Jordan and Malaysia have turned to PPPs to address countries' growing infrastructure needs and resource constraints [2,3]. Despite their strategic importance, PPPs encounter numerous challenges in these developing nations. In Jordan, these challenges include risk factors related to the transfer phase, organizational issues, financing, project management, and feasibility studies [4]. In Malaysia, the challenges include inflation, interest rate risks, currency fluctuations, technical expertise, and limitations in financial capacity [5].

Jordan's Public-Private Partnership (PPP) framework was established with Law No. 31 of 2014. İn 2015 Jordan created PPP by-laws council and strategy chaired by the Prime Minister [6]. Serveral Projects under the PPP procurement method are uderway in Jordan, Such as Bus Rapid Transit (BRT) Project between Amman and Zarqa, the National Broadband Network (NBN) to support digital transformation, and the King Hussein Bridge Land Border Crossing Terminal. Jordan initiated a group of projects such as the construction of 14 public schools, then a Non-revenue water project to reduce losses of water in the capital city. One more major project that to improve key roads in the country through users chargers [7].

The economic development in malaysia started in the early 1980s. The goverement engaged the private sector expertise as a national development effort [8]. However, in the early 80s, significant progress on funding PPP initiatives in the Ninth and Tenth Malaysia plans. The main goal for these plans were to improve development strategies [9]. Despite the benefits of PPP in Malaysia, obligations and critisism imposed the policies of the malaysian governement [10].

Some challenges faces the implementation of BIM in PPP Projects. Some the challenges include technological obstical, some expamples are the complexity and interoperability between different platforms [11,12].

On the other hand, BIM application faces legal and contractual challenges, such as conflection in responsibilities [13]. Another challenge is the cultural resistenace from stakeholders accustomed to traditional work, which makes communication gaps feels bigger and hinder effective collaboratios [14].

BIM helps PPP projects work smoother, by identifying risks early which provide time to mitigate the risk early in the project life. Furthermore, BIM reduces and cuts costs[15], modulate construction to address poor communications [16]. Support cost tracking standarized KPIs [17], and improve performance management systems [18]. Prior studies have explored various aspects of BIM in PPP projects, BIM impact on key performance indicators (KPIs) and risk management, but the focus if the literature mainly is on developed economies or a wider regional perspective. Prior studies lacks comprehensive assessment of the role of BIM in managing PPP projects and improving results in Jordan and Malaysia.

Building Information Modeling (BIM) is an innovative procedure that fundamentally changes the architecture, engineering and construction sectors by improving conceptualization, design, construction and overall management of constructed spaces [19–21].

BIM is seen as beneficial in the construction inducstry dues to the collaboration improvement, as well as minizing errors, reduces costs, and improving efficincy [22,23]. BIM allows somther data exchanges, that mitigates risks, and reduces projects setbacks and overruns. Although some studies have explored aspects of BIM adoption in PPP projects within developing countries, a comprehensive evaluation of its implications, especially in Jordan and Malaysia, is still lacking.

Recognizing the growing interest in BIM to improve PPP projects, various studies have explored its potential benefits [24]. For example, [12] conducted a study in Sweden, proposing a BIM implementation plan that identifies major drivers and barriers to the adoption of BIM, although it does not specifically address cost reduction. On the contrary, [15] provided a quantitative analysis from Iran, demonstrating that BIM can improve performance by 28.9% on average through improved key performance indicators (KPIs), utilizing a Copula Bayesian Network for causal analysis. Furthermore, [25] focused on China, examining how contractual flexibility in BIM-enabled PPP projects can significantly impact cost reduction during the procurement and construction phases. While these studies illuminate the potential performance benefits of BIM in PPP projects, they also underscore the need for comprehensive strategies for BIM's potential benefits.

The recent studies have investigated the role of BIM in managing risks of PPP projects. [26] explored the integration of BIM and the Integrated Project Delivery (IPD) model in characteristic Chinese towns, demonstrated that this integration enhanced collaboration and reduced risks in construction projects. The same study highlighted the need to address a list of challenges related to these towns and the potential drawbacks of the BIM-based IPD model. Another study implemented in Malaysia by [11], identified six significant risk factors in PPP projects and provided insights and lessons from the UK context to improve BIM implementation in Malaysia. They used questionnaires and expert interviews to uncover critical challenges that need to be addressed in this context [14] analysed critical PPP risk factors and BIM functions, and proposed a comprehensive BIM-based risk management framework. That study identified seven critical risk factors and eight BIM functions, and noted the limited scope of critical factors and BIM n-D functions. The findings provided valuable insights into integrating BIM into PPP projects.

Studies show that BIM save energy, makes projects more envirnmently friendly, which in turn improves the operational effectiveness, and boost the productivity. Furthermore, BIM adoption has the potential to reduce CO2 emissions. These benefits suggests that BIM has the ability to support sustainable development 8, 9, 11, and 13, which advances overall construction sustainability [27]. BIM tools facilitate the improvement of building designs to substantially reduce energy consumption, leading to potential savings of up to 46% in overall energy usage [28]. Depending on BIM in Value for Money (VFM) evaluations of PPP projects could result in more scientific and accurate assessments. This also could help the long-term growth of such efforts [29]. The use of BIM in PPP projects add a high value in improving decision making, operational management, and collaborative efforts, resulting in more and more sustainable environment for all memers of the ecosystem [26].

When you look at the challenges that prevent companies in Jordan and Malaysia from using Building Information Modeling (BIM), you find a lot of obstacles. In Jordan, the main problems are that it's expensive to train people and buy the software, many people don't know much about BIM, and there aren't clear rules to follow. Also, companies have to spend a lot of money upfront, they don't get enough support from their leaders, and it's hard for them to get the right technology and frameworks. In Malaysia, it's also tough because people don't have many chances to learn about BIM, the laws are not helpful, and different systems can't work well together, which makes it hard to use BIM. All these issues make it difficult for companies in both countries to start using BIM. It's clear that both Jordan and Malaysia face unique challenges when it comes to adopting BIM [30]. If companies and governments can work together to address these issues, it might be easier for them to use BIM and improve their construction projects. This could involve creating more training programs, making clearer rules, and investing in better technology to support BIM adoption. By doing so, they can help their construction industries become more efficient and competitive [31].

A lot of researchers were motivated to investigate KPIs to evaluate the adoption of BIM in the construction industry from different point of views [32] exploreded the impact of BIM technology on safety KPIs in construction projects in Slovakia, Slovenia, and Croatia [33] analyzed the ROI value of BIM in an architectural company, emphasized on cost control improvements and workflow optimization [34] identified the primary stakeholder groups that add value from BIM adoption, including design units, facility management, contractors, and owners. In another study, [29] provided evaluation criteria for innovative constructon projects, showing that BIM improves material usage control and minimizes initial material consumption.

In our previous research [17], we identified and selected key performance indicators (KPI) related to administrative, sustainability, and management aspects for PPP projects. A framework was established for the current analysis, that allowed a comprehensive examination of the multidimensional impacts of BIM on PPP performance. Supplementary Table 1 shows the overall KPIs utilized to evaluate BIM performance in PPP Projects.

The mentioned studies emphasize the diverse benefits of BIM in PPP projects and stress the necessity for strategic implementation, quantitative analysis, and adaptable contractual agreements. Even with these insights, there is still a lack of understanding of how BIM affects cost savings in PPP projects in countries like Jordan and Malaysia. This research aims to fill that gap. BIM technology is not being used enough in many countries, especially in Jordan [35,36] and Malaysia [37,38], where it seems to be used in a random and poorly integrated way. This research investigates the influence of BIM on PPP projects within these nations, thereby addressing the existing gap.

Reduce in cost, which also means to see in to a reduction of delay, we see an improvement in quality, we see an increase in productivity, we also put forth that which is sustainable has been increased and we have better risk transfer. We put forth these points to in hope of convincing public and private sector players to adopt BIM into wider use. The current research investigates how BIM influences the success of PPP projects to foster greater adoption and to achieve more successful project benifits. It also aims to fill this gap by systematically examining the impact of BIM on the success of PPP projects in Jordan and Malaysia.

MATERIALS AND METHODS

The current study adopts a descriptive quantitative approach to systematically analyze experts' perceptions of the potential impact of BIM adoption on PPP projects. BIM and PPP projects experts were targeted in Jordan and Malaysia. The study sample included professionals of the public and private sectors, and educational institutions experts, providing a comprehensive view of industry practices and policy implications. A pilot study was conducted with 30 respondents prior to the main survey distribution in order to validated the questionnaire.

The study was conducted during 05-2023 to 01-2024, using an electronic questionnaires distributed using email, in person, and phone calls. We ensured informed consent from each participant prior to data collection. 1,100 questionnaires were distributed and 441 were responded with 40% response rate . The questionnaire was carefully designed and divided into two main sections: general demographic questions and specialized questions. The demographic section collected general information such as: gender, age, education, experience, and employment sector.

We utilized a systematic method to insure comprehenisve and reliable questionnaire design. The first step was reviewing existing literature on topics that is related to BIM and PPP, we searhced through Google Scholar, Scopus, and Web of Science databases. The aim was to idenfity key variables and themes. Followed by organizing the data into Excel spread sheed to systematically categorize the data.

Upon identifying recurring themes across multiple studies, we consolidated these findings and categorized them into five main areas: administrative aspects, sustainability, management, tools and frameworks, and productivity. This categorization formed the basis for drafting the first version of the questionnaire. Supplementary Table 1 in the appendix shows the KPI factors selected from the literature.

For the main section, we developed questions to cover five key aspects of BIM adoption in PPP projects: Administrative (3 items), Productivity (10 items, covering manpower, machine, and materials), Management (11 items, addressing time, cost, and quality), Tools Framework (11 items, focusing on people, process, and technology), Sustainability (10 items, including environmental, economic, and social factors). We used a Likert scale from 1 to 5 for responses. Table 1 shows the distribution of the questions across different aspects.

To assess if the data is appropriate for factor analysis, we calculated Kaiser-Meyer-Olkin (KMO) measure for all question groups of the questionnaire. The KMO values range from 0 to 1. values closer to 1 indicating that the data are suitable for factor analysis. Table 1 shows that the KMO values for all aspects meet acceptable criteria.

The aim of this study is to assss the overall performance of utilizing BIM in the performance of PPP projects. We analyzed the data using SPSS software (version 25), employing a variety of statistical tests and descriptive analyzes to understand BIM's impact on PPP projects. In order to do so, we calculated means, standaard deviations, frequencies. Furthermore, we summarized the central tendencies, variablity, and distribution.

The next step was to validate and refine the final questionnaire. We started with conducting a pilot study with the preminary questionnaire with a 30 respondents. The main questionnaire after validating was distributed for around six months period. The distribution took place from May 2023 to January 2024. The distribution was done by email, in person, and through phone calls. Before distribution we ensured informed consent from each participant prior to data collection.

Questionnaire Design

To ensure the questionnaire validity and effectiveness, feedback from three experts who specialize in questionnaire design was collected and the related modifications were implemented. Three additional experts were consulted to enhance the questionnaire quality further. This process helped to address potential issues and improve the clarity and relevance. After that, a pilot study was conducted for 30 respondents to evaluate the validity and reliability. Analysis of this pilot study was used to make final adjustments to the questionnaire, ensuring that it accurately captured the intended data. The final questionnaire form is shown in Supplementary Table 2.

We distributed the finalized questionnaire to a larger sample and received 441 responses of 1100 distributed. The collected data were analyzed using SPSS, applying various statistical tests to ensure the data reliability. This process enabled to confirm that the questionnaire effectively measured the relevant constructs and provided valuable insights for the study. (Fig. 1) explains the research process framework that is adopted in this research. A quantitative research approach was employed.

Table 1. Distribution of questions in the questionnaire and the results of the Kaiser-Meyer-Olkin test

Category	Indicators	No of Questions	KMO
Administrative	Administrative	3	0.696
Productivity	Manpower	5	0.856
	Machine	2	0.500
	Material	3	0.696
Management	Time	3	0.711
	Cost	4	0.812
	Quality	4	0.800
Tools Framework	People	4	0.802
	Process	5	0.701
	Technology	2	0.500
Sustainability	Environment	3	0.690
	Economy	3	0.705
	Social	4	0.816

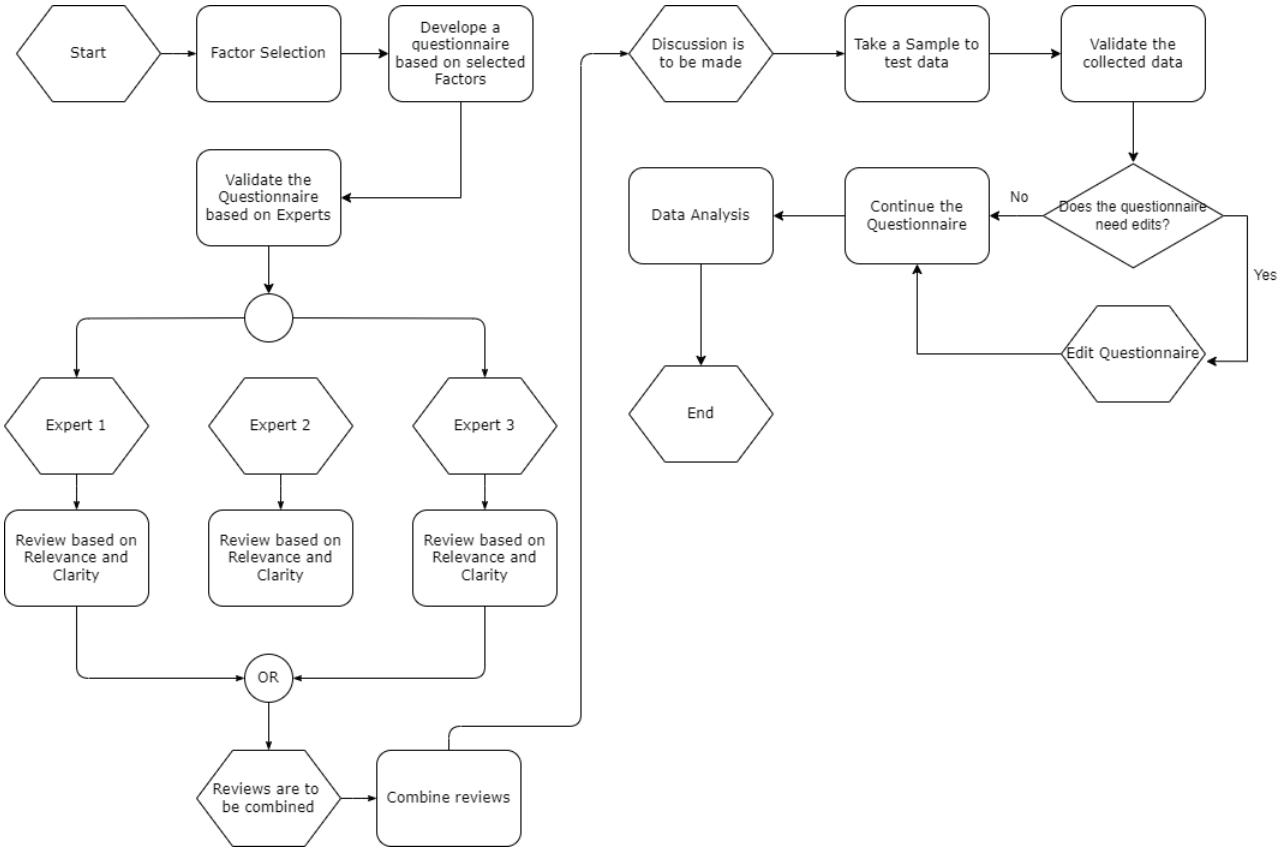


Figure 1. Research method framework.

RESULTS AND DISCUSSION

Demographic Characterstices of the Respondents

We faced significant challenges in obtaining participants with experience in PPP and BIM projects. 1,100 questionnaires have been distributed to individuals who involved in projects in Jordan and Malaysia's for both public and private sectors. To colect data, face-to-face interviews, emails, and social media platforms using Google Forms were included. The survey achieved a notable response rate, garnering 441 valid responses, which equal about 40% of the total distributed.

The research sample were selected based on their PPP and BIM projects expirience within the public and private sectors in Jordan and Malaysia. The process prioritized individuals with expertise in project management, engineering, and construction. The samples' qualifications ranged from Bachelor's to Doctorates degrees, shows diverse affiliations, including government bodies, private engineering organizatins and academic institutions as well. A purposive sampling approach was used to ensure representation for both public and private sectors, as outlined in Table 2.

The respondents screening revealed a presence of the private sector, which comprised 79.1%, while the public sector was only 20.9% of the total respondents. A higher representation of males (73.5%) than females (26.5%) was presented, reflecting broader workforce trends. Educational qualifications ranged from Bachelor's to Doctorate degrees, with 73.9% holding a Bachelor's degree, 21.3% possessing a Master's degree, and 4.8% having attained a Ph.D., indicating a diverse educational background. The responses covered a wide range of professional experience, with 1-5 years and 21-25 years being the most common, illustrating a varied pool of participants with differing levels of expertise. The demographic characteristics of the respondents are detailed in Table 2.

Descriptive Analysis

The data shown in Table 3 indicates a detailed analysis of the questionnaire responses, the questionnaire covered various dimensions, including administrative, sustainability, tools framework, management, and productivity factors. The level of agreement or disagreement with each factor.

We evaluated factors across administrative, sustainability, tools framework, management, and productivity aspects of performance. The tools framework category was ranked first, scoring 3.94±0.772, underscoring the importance of document management and clash detection, followed by productivity factors with a score of 3.93±0.783, emphasizing effective interface management and experience gains. However, sustainability factors came in third and rated

Table 2. Demographic characteristics of respondents

Background information	Parameter	Frequency	Percent
Sector	Private	349	79.1
	Public	92	20.9
	Total	441	100
Highest Academic Qualifications	Bachelor's degree	326	73.9
	Master's degree	94	21.3
	Doctoral Degree	21	4.8
	Total	441	100
Gender	Male	324	73.5
	Female	117	26.5
	Total	441	100
Years of experience	1-5	123	27.9
	6-10	68	15.4
	11-15	85	19.3
	16-20	39	8.8
	20-25	126	28.6
	Total	441	100

Table 3. Detailed analysis of the responses to the questionnaire

Measurement	S.Dis (1)	Dis. (2)	N. (3)	Ag. (4)	S.Ag (5)	Total %	Mean±SD	Rank
1- Administrative Factors					Agree		3.83±0.935	5
Organization structure	27	18	71	144	181	441	3.98±1.135	1
	6.1	4.1	16.1	32.7	41.0	100.0		
Risk of adverse changes in law, policy, or regulations	18	33	87	162	141	441	3.85±1.079	2
	4.1	7.5	19.7	36.7	32.0	100.0		
Risk of force majeure	32	40	95	157	117	441	3.65±1.174	3
	7.3	9.1	21.5	35.6	26.5	100.0		
2- Productivity Factors					Agree		3.93±0.783	2
Experience and knowledge gains	19	16	56	170	180	441	4.08±1.032	2
	4.3	3.6	12.7	38.5	40.8	100.0		
Staff training	16	24	65	168	168	441	4.02±1.037	4
	3.6	5.4	14.7	38.1	38.1	100.0		
effective interface management	12	21	62	167	179	441	4.09±0.988	1
	2.7	4.8	14.1	37.9	40.6	100.0		
Use of construction waste	11	45	97	147	141	441	3.82±1.069	8
	2.5	10.2	22.0	33.3	32.0	100.0		
Employment opportunities	26	39	92	141	143	441	3.76±1.168	10
	5.9	8.8	20.9	32.0	32.4	100.0		
Specialized expertise	14	23	74	150	180	441	4.04±1.035	3
	3.2	5.2	16.8	34.0	40.8	100.0		
Technical staff resources	17	24	80	171	149	441	3.93±1.040	5
	3.9	5.4	18.1	38.8	33.8	100.0		
Equipment condition and number	13	32	104	169	123	441	3.81±1.018	9
	2.9	7.3	23.6	38.3	27.9	100.0		

Table 3. Detailed analysis of the responses to the questionnaire (continued)

Measurement	S.Dis (1)	Dis. (2)	N. (3)	Ag. (4)	S.Ag (5)	Total %	Mean±SD	Rank
Use of innovative materials	12	29	93	173	134	441	3.88±1.005	7
	2.7	6.6	21.1	39.2	30.4	100.0		
Material condition and number	12	27	91	176	135	441	3.90±0.997	6
	2.7	6.1	20.6	39.9	30.6	100.0		
3- Tools Framework Factors					Agree		3.94±0.772	1
Requirements of stakeholders/goals	14	24	89	176	138	441	3.91±1.006	7
	3.2	5.4	20.2	39.9	31.3	100.0		
Change management - resistance to change	16	41	117	155	112	441	3.69±1.061	11
	3.6	9.3	26.5	35.1	25.4	100.0		
Information and knowledge sharing	15	30	75	181	140	441	3.91±1.030	6
	3.4	6.8	17.0	41.0	31.7	100.0		
Effective safety management: no. of security incidents	19	43	80	155	144	441	3.82±1.121	10
	4.3	9.8	18.1	35.1	32.7	100.0		
Document management	10	23	67	139	202	441	4.13±1.004	1
	2.3	5.2	15.2	31.5	45.8	100.0		
Updated timeline	16	22	65	162	176	441	4.04±1.036	3
	3.6	5.0	14.7	36.7	39.9	100.0		
Clash detection	12	27	63	155	184	441	4.07±1.023	2
	2.7	6.1	14.3	35.1	41.7	100.0		
Transferring risk to the private sector	15	29	100	170	127	441	3.83±1.028	9
	3.4	6.6	22.7	38.5	28.8	100.0		
Procedures for settling claims and disputes	12	31	83	177	138	441	3.90±1.010	8
	2.7	7.0	18.8	40.1	31.3	100.0		
Detail and transparent level of tendering procedure	13	28	72	159	169	441	4.00±1.034	4
	2.9	6.3	16.3	36.1	38.3	100.0		
Life-cycle management ability	11	29	71	169	161	441	4.00±1.008	5
	2.5	6.6	16.1	38.3	36.5	100.0		
4- Management Factors					Agree		3.88±0.781	4
On-time or earlier project completion	14	33	81	162	151	441	3.91±1.052	3
	3.2	7.5	18.4	36.7	34.2	100.0		
Schedule variance	16	30	81	182	132	441	3.87±1.033	9
	3.6	6.8	18.4	41.3	29.9	100.0		
Investment life cycle cost and return	15	31	84	173	138	441	3.88±1.039	6
	3.4	7.0	19.0	39.2	31.3	100.0		
Cost management	17	27	85	170	142	441	3.89±1.047	5
	3.9	6.1	19.3	38.5	32.2	100.0		
Cost variance	16	29	85	175	136	441	3.88±1.038	8
	3.6	6.6	19.3	39.7	30.8	100.0		
Completion/time delay	18	27	77	187	132	441	3.88±1.036	7
	4.1	6.1	17.5	42.4	29.9	100.0		
Construction cost overrun	18	33	93	180	117	441	3.78±1.048	11
	4.1	7.5	21.1	40.8	26.5	100.0		
Change in scope	16	25	102	186	112	441	3.80±1.000	10
	3.6	5.7	23.1	42.2	25.4	100.0		
Quality and innovation public service	11	41	68	173	148	441	3.92±1.040	1
	2.5	9.3	15.4	39.2	33.6	100.0		

Table 3. Detailed analysis of the responses to the questionnaire (continued)

Measurement	S.Dis (1)	Dis. (2)	N. (3)	Ag. (4)	S.Ag (5)	Total %	Mean±SD	Rank
The third-party assessment results	13	28	76	188	136	441	3.92±0.998	2
	2.9	6.3	17.2	42.6	30.8	100.0		
Contract management and control	13	29	83	175	141	441	3.91±1.015	4
	2.9	6.6	18.8	39.7	32.0	100.0		
5- Sustainability Factors					Agree		3.90±0.774	3
Environmental risk	13	32	96	181	119	441	3.82±1.006	9
	2.9	7.3	21.8	41.0	27.0	100.0		
Energy efficiency	7	34	73	176	151	441	3.98±0.980	1
	1.6	7.7	16.6	39.9	34.2	100.0		
Reduction of pollutant discharge	19	34	105	162	121	441	3.75±1.072	10
	4.3	7.7	23.8	36.7	27.4	100.0		
Economic development	13	40	79	178	131	441	3.85±1.041	8
	3	9	18	40	30	100		
Value for money (VFM)	8.0	28.0	94.0	173.0	138.0	441.0	3.92±0.969	7
	2	6	21	39	31	100		
Feasibility analysis	12.0	28.0	88.0	168.0	145.0	441.0	3.92±1.012	6
	3	6	20	38	33	100		
Government reputation and improvement	17.0	31.0	67.0	172.0	154.0	441.0	3.94±1.062	4
	4	7	15	39	35	100		
Trust and respect	11	28	80	179	143	441	3.94±0.991	3
	2.5	6.3	18.1	40.6	32.4	100.0		
Market interest and opportunity	11	34	74	178	144	441	3.93±1.012	5
	2.5	7.7	16.8	40.4	32.7	100.0		
Public interest	12	24	82	173	150	441	3.96±0.995	2
	2.7	5.4	18.6	39.2	34.0	100.0		

3.90±0.774, highlighting energy efficiency and government reputation, while emphasizing pollutant discharge and economic development less. Management factors were ranked fourth of all aspects, with a score of 3.88±0.781, prioritizing quality in public service and third-party assessments, but noting challenges in cost overruns and scope changes. Lastly, administrative factors, with a mean score of 3.83±0.935, highlighted organizational structure as the most critical, followed by legal and policy change risks.

Summary of Key Dimensions

The survey results indicate that respondents reported that BIM adoption could have a positive influence on PPP projects. Most of the participants agreed or strongly agreed on categories like technology, people, manpower, social, economy, quality, and administrative aspects, as illustrated by the significant portions of light blue and dark blue bars. Only a few respondents indicated strongly disagree or disagree, suggesting that negative perceptions are minimal. Some participants chose neutral responses, indicating a degree of ambivalence. In general, the data strongly endorse the positive influence of BIM on these factors. (Fig. 2) shows the survey results in detail.

However, our study ranked various project management dimensions based on the levels of agreement levels. The tools framework factors were deemed to be the most critical, followed by the sustainability and productivity factors. Management factors ranked fourth, highlighting their importance, but slightly less critical than the top three. Administrative factors were ranked the fifth, meaning that they are considered essential but not as immediately impactful as the other dimensions. These rankings reflect a consensus on the effect of BIM adoption on various aspects of performance in PPP projects. Table 4 represents a highlighted summary of the consensus on the agreement on various dimensions of the impact of BIM adoption in PPP projects on different factors.

Correlation Analysis

Linear correlation was used to explore the relationship between BIM implementation (dependent variable) and various performance indicators (independent variables). Pearson's correlation coefficient, ranges from -1 to 1,

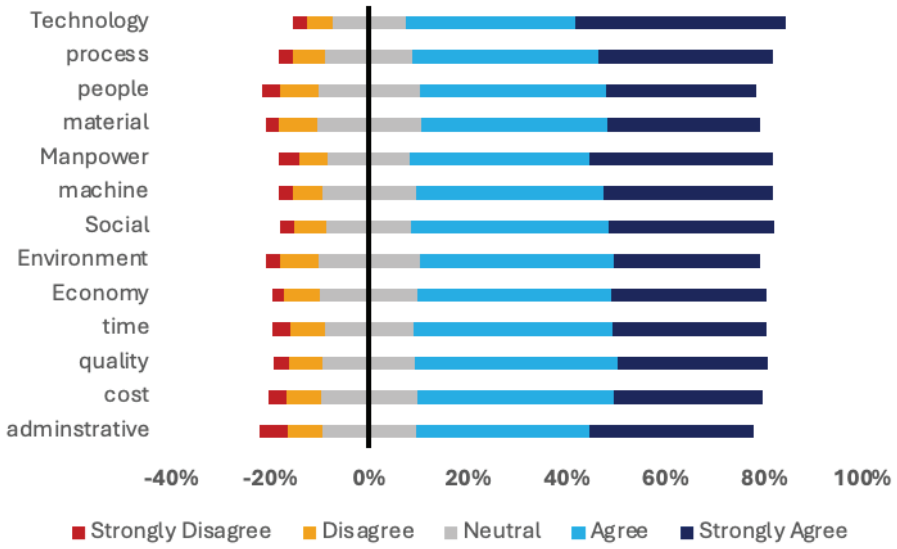


Figure 2. The survey detailed results.

Table 4. Summary of key dimensions

Dimension	Mean	SD	Rank	Scale
Tools framework factors	3.94	0.772	1	Agree
Productivity factors	3.93	0.783	2	Agree
Sustainability factors	3.90	0.774	3	Agree
Management factors	3.88	0.781	4	Agree
Administrative factors	3.83	0.935	5	Agree

Table 5. Correlation analysis of the questionnaire

Project performance factors	Overall performance (pearson correlation coefficient value)
Tools framework factors	0.948**
Management factors	0.938**
Productivity factors	0.937**
Sustainability factors	0.922**
Administrative factors	0.711**

**: Correlation is significant at the 0.01 level (2-tailed).

determines the degree and direction of these relationships. A positive value means a good relationship, while a negative value means a bad relationship. Table 5 shows that all of the factors are strongly related to each other. The factors that make up the tools framework have the strongest relationship with BIM implementation at 0.948. Next are the management factors at 0.938, the productivity factors at 0.937, the sustainability factors at 0.922, and the administrative factors at 0.711. These findings indicate that respondents believe that greater BIM adoption may improve performance in tools framework, management, sustainability, and productivity. Even though administrative factors are positively linked to BIM implementation, they are much weaker, which means the relationship is only moderate. The findings indicate that the adoption of BIM favorably influences administrative factors, albeit to a lesser extent than other variables.

Reliability Test (Cronbach's Alpha Coefficient)

Our study demonstrated a high internal consistency across the questionnaire, indicated by a Cronbach alpha

Table 6. Cronbach's Alpha coefficient for all the dimensions of the study

Dimension	Cronbach's Alpha	Interpretation	N of Items	Valid
Administrative factors	0.769	Acceptable	3	441
Productivity factors	0.915	Excellent	10	
Tools Framework factors	0.921	Excellent	11	
Management factors	0.925	Excellent	11	
Sustainability factors	0.920	Excellent	10	

of 0.975 for all items. We collected a total of 441 valid responses. Examining individual dimensions, we found that with ten items, productivity factors achieved an excellent reliability score of 0.915. Tools framework factors (11 items) and management factors (11 items) also exhibited excellent reliability, with Cronbach's alpha values of 0.921 and 0.925, respectively. Sustainability factors showed excellent reliability, with a Cronbach alpha of 0.920 over ten items. With three items, administrative factors had an acceptable reliability score of 0.769.

These values represent the internal consistency within each dimension. The administrative factors dimension shows an acceptable level of reliability, while sustainability factors, tools framework factors, management factors, and productivity factors exhibit excellent levels of reliability. Table 6 shows the Cronbach's Alpha coefficient for all the study dimensions.

The results showed a stronge positive correlation between the tools framework factors and BIM implementation. Management, sustainability, and productivity factors also demonstrated high positive correlations. The results underscore the critical role in enhancing project benifits for multiple dimensions.

The current research results align with the most of previous studies on BIM adoption in PPP projects. [15] demonstrated that BIM improves the performance of construction projects by an average of 28.9%, especially when it comes to feasibility studies, financial and cost performance, and risk management. While [39] stressed the importance of helping people make decisions that are worth the money in PPP projects. [12] proposed a BIM implementation plan to enhance collaboration, corroborating our assertion that successful BIM adoption markedly improves project outcomes.

This study offers valuable insights while also acknowledging specific limitations. The emphasis on Jordan and Malaysia, along with the absence of distinction among project types, may influence the generalizability of the results. It is acknowledged that the questionnaire was not widely distributed and that the use of a 5-point Likert scale may have led to biased responses. The results are interesting, but they shouldn't be used to make generalizations about other situations.

Future research on BIM adoption in PPP projects may include sector-specific analyses of various PPP procurement methods, which are crucial for comprehending BIM's impact on their efficacy. Looking at different levels of BIM maturity could help us understand the benefits better. Comparing regions, like countries with different levels of economic development, will help us understand how geography and culture affect things. Examining the amalgamation of BIM with nascent technologies can elucidate forthcoming trends in the construction sector. To improve future studies, it would be helpful to use both qualitative and quantitative data collection methods, like interviews, case studies, and surveys. This would give a full picture of BIM's effects.

CONCLUSION

With BIM the benefits of collaboration (with partners) on construction projects are enormous. It helps in organizing all the documents, prevent errors from taking place, and keeps schedules and bids transparent. Which naturally optimizes the entire process as everyone is at the same level and has the right talent. And there's an environmental side to it too: by conserving energy, the community gets added value. It makes managers less prone to misunderstanding or legal conflict. BIM is a great tool to help streamline and execute construction projects smoothly.

In order to encourage the usage of BIM by a greater number of organizations, both government and private business need to collaborate. They can accomplish this by offering training, establishing nationwide standards and incentives. They also need to ensure BIM is used from beginning to end on every project. This is one of the initial types of research to demonstrate how BIM really works in developing countries. This indicates that BIM can lead to improved performance, sustainability and service delivery. BIM aids companies to get better in a lot of things and it comes important for the success of their projects.

It recommends fostering innovation through supportive regulations, mandating BIM in large projects, strengthening collaboration with technology providers, encouraging knowledge sharing, and developing skilled professionals to ensure effective and sustainable adoption.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Atapattu PJ. Infrastructure Investment with Public-Private Partnership (PPP) and Economic Growth in Developing Countries in Asia. Vidyodaya J Manag 2019;5. [CrossRef]

[2] Ismail S, Mohamad R, Said JM. Performance indicators for lifecycle process of public private partnership (PPP) projects in Malaysia. Built Environ Proj Asset Manag 2022;12:704–718. [CrossRef]

[3] Mistarihi A, Hutchings K, Shacklock A. Differing Opinions Do Not Spoil Friendships: Managing Public–Private Partnership (Ppp) Infrastructure Projects in Jordan. Public Adm Dev 2013;33:371–388. [CrossRef]

[4] Almuhisen M, Gokcekus H, Celik T. Developing an AHP Risk Assessment Model for PPP Projects in Jordan. J Eng Res 2021. [CrossRef]

[5] Razi PZ, Ali MI, Ramli NI. Incorporation of Risk Index for Risk Response and Risk Mitigation Strategies of Public-Private Partnership (PPP) Housing Construction Project in Malaysia. IOP Conf Ser: Mater Sci Eng 2020;712:012031. [CrossRef]

[6] Biygautane M, Jarrar Y. Public-private partnerships and achievement of public value: The experience of Jordan's Queen Alia International Airport. J Infrastruct Policy Dev 2023;7:1936. [CrossRef]

[7] PPP Unit. PPP Projects Registered in the National Registry of Public Investment Projects - Public–Private Partnership Unit. Available at: https://pppu.gov.jo/EN/List/Projects_. Accessed on Apr 28 2026.

[8] Ismail S. Driving forces for implementation of public private partnerships (PPP) in Malaysia and a comparison with the United Kingdom. J Econ Admin Sci 2014;30:82–95. [CrossRef]

[9] Ahmad U, Ibrahim Y, Abu Bakar A. Malaysian public private partnership. Acad Account Financ Stud J 2018;22.

[10] Abdul AFC, Sarif SM, Mat NHN. Public Private Partnership and Financial Implication to Government: A Case Study from Malaysia. Inf Manag Bus Rev 2024;16:372–384. [CrossRef]

[11] Abdullah Habib SNH, Ismail S, Syed Khuzzan SM. Risk Factors Towards Public-Private Partnerships (Ppp) Projects Implementing Building Information Modelling (Bim) in the United Kingdom (Uk): A Lesson Learnt For Malaysia. Plan Malays 2020;18. [CrossRef]

[12] Deniz GO, Lokhandwala Z. Building Information Modeling (BIM) Implementation in Public–Private Partnership (PPP) Projects. Eng Proc 2023;40. [CrossRef]

[13] Laishram B. Building information modeling in Public Private Partnership Projects–perspectives and hurdles. International Conference on Structural Engineering, Construction & Management. Sri Lanka; 2011.

[14] Song YH. Building Information Modelling Integration into Public-Private Partnership (Ppp) Risk Management. Univ Technol Malalysia 2019.

[15] Ghorbany S, Noorzai E, Yousefi S. BIM-based solution to enhance the performance of public-private partnership construction projects using copula bayesian network. Expert Syst Appl 2023;216:119501. [CrossRef]

[16] Cui B. BIM - based modular integrated construction risk management. Apll Comput Eng 2024;70:8–13. [CrossRef]

[17] Al-Shamayleh A, Zakaria S, Rahim N, Totanji DA. A Political, Sustainable, and Management Building Information Modelling (BIM) Key Performance Indicators (KPIs) for Public-Private Partnership (PPP) Projects. Int J Adv Eng Manag Sci 2024;10:153–61. [CrossRef]

[18] Yuan J, Li X, Ke Y, Xu W, Xu Z, Skibnewski M. Developing a building information modeling–based performance management system for public–private partnerships. E 2020;27:1727–1762. [CrossRef]

[19] George WK, Sunday SE, Pandey D, Kabari AD, Pandey BK, George AS, et al. Building Information Modeling: Harnessing Innovations for Professional Proficiency. In: El Kacimi Y, Alaoui K, editors. Advances in Civil and Industrial Engineering, IGI Global; 2024, p. 305–331. [CrossRef]

[20] Spirkova D, Stepanek P. Building Information Modeling (BIM) In Effective Management of Development Projects and In the Administration of Housing Companies. Commun Int Proc 2023. [CrossRef]

[21] Tang L, Wu Z, Wen Y, Chen C, Tang S, Trofimova P. Building Information Modeling and Building Performance Optimization. Elsevier 2024;750–764. [CrossRef]
[22] Adhikari S, Roca AK. Impact of BIM Technology on the Construction Project Management Process. EasyChair 2024;948–938. [CrossRef]
[23] Shao M, Wang F. Analyzing the Application of BIM Technology in Construction Project Management. J Civil Transp Eng 2024;1:16–20. [CrossRef]
[24] Al-Shamayleh A, Zakaria S, Rahim N, Totanji DA. Building information modeling for Public Private Partnership Projects A literature review and future research directions. Int J Adv Eng Manag Sci 2024;10:138–152. [CrossRef]
[25] Xu Y, Chong H-Y, Chi M. Impact of Contractual Flexibility on BIM-Enabled PPP Project Performance during the Construction Phase. J Infrastruct Syst 2022;28:04021057. [CrossRef]
[26] Shyng JH. Based on BIM intelligent technology and IPD collaborative management model applied to characteristic towns with PPP mode. J Infras Policy Dev 2024;8. [CrossRef]
[27] Silva AD, Cardoso AM. Advancing BIM and Sustainability with Coopetition: Evidence from the Portuguese Stone Industry. Apll Syst Innov 2024. [CrossRef]
[28] Mughal S, Khoso AR, Najeeb H, Noor Khan MS, Hussain Ali T, Hussain Khahro S. Green Retrofitting of Building using BIM-based Sustainability Optimization. J Kejuruterr 2024;36:179–189. [CrossRef]
[29] Jiang X, Ouyang W, Dai X, Hu K, Liu Y, Xia B. A Framework of BIM-Based Value for Money (VFM) Evaluation Methods of PPP Projects at Decision-Making Stages. Springer 2021;797–807. [CrossRef]
[30] Hyarat E, Hyarat T, Al Kuisi M. Barriers to the Implementation of Building Information Modeling among Jordanian AEC Companies. Buildings 2022;12:150. [CrossRef]
[31] Wong SY, Gray J. Barriers to implementing Building Information Modelling (BIM) in the Malaysian construction industry. IOP Conf Ser: Mater Sci Eng 2019;495:012002. [CrossRef]
[32] Mandičák T, Spišáková M. BIM Technology and Impact on Safety KPIs in Construction Projects in Slovakia, Slovenia, and Croatia. Springer 2024;242–250. [CrossRef]
[33] Lechhab N, Iordanova I, Forgues D. Evaluation of the Return on Investment of BIM—The Case of an Architectural Firm. Springer 2023;431–443. [CrossRef]
[34] Toan NQ, Tam NV, Diep TN, Anh PX. Adoption Of Building Information Modeling In The Construction Project Life Cycle: Benefits for Stakeholders. Alex Eng J 2022;7:56–71. [CrossRef]
[35] Ala'a A, Zeeshan A. Crucial barriers of building information modelling (BIM) in the Jordanian construction industry. Global J Eng Tech Adv 2020;3:20–30. [CrossRef]
[36] Matarneh R, Hamed S. Barriers to the Adoption of Building Information Modeling in the Jordanian Building Industry. Open J Civ Eng 2017;7:325–335. [CrossRef]
[37] Othman I, Al-Ashmori YY, Rahmawati Y, Mugahed Amran YH, Al-Bared MAM. The level of Building Information Modelling (BIM) Implementation in Malaysia. Ain Shams Eng J 2021;12:455–463. [CrossRef]
[38] Wong SY, Gray J. Barriers to implementing Building Information Modelling (BIM) in the Malaysian construction industry. IOP Conf Ser: Mater Sci Eng 2019;495:012002. [CrossRef]
[39] Ren G, Li H. BIM Based Value for Money Assessment in Public-Private Partnership. Springer 2017;51–62. [CrossRef]
[40] Aghimien D, Aigbavboa C, Oke A. Critical success factors for digital partnering of construction organisations - A Delphi study. Eng Constr Archit Manag 2020;27:3171–188. [CrossRef]
[41] Cheng M, Liu G, Xu Y. Can joint-contract functions promote PPP project sustainability performance? A moderated mediation model. Eng Constr Archit Manag 2021;28:2667–2689. [CrossRef]
[42] Doloi H. Understanding Impacts of Time and Cost Related Construction Risks on Operational Performance of PPP Projects. Int J Strateg Prop Manag 2012;16:316–337. [CrossRef]
[43] Ke Y, Wang S, Chan APC, Lam PTI. Preferred risk allocation in China's public–private partnership (PPP) projects. Int J Proj Manag 2010;28:482–492. [CrossRef]
[44] Hoeft M, Pieper M, Eriksson K, Bargstädt H-J. Toward Life Cycle Sustainability in Infrastructure: The Role of Automation and Robotics in PPP Projects. Sustainability 2021;13:3779. [CrossRef]
[45] Li H, Xia Q, Wang L, Ma Y. Sustainability assessment of urban water environment treatment public-private partnership projects using fuzzy logic. J Eng Design Technol 2020;18:1251–1267. [CrossRef]
[46] Chen H, Zhang L, Wu X. Performance risk assessment in public–private partnership projects based on adaptive fuzzy cognitive map. Appl Soft Comput 2020;93:106413. [CrossRef]
[47] Liang Y, Wang H. Sustainable Performance Measurements for Public–Private Partnership Projects: Empirical Evidence from China. Sustainability 2019;11:3653. [CrossRef]
[48] Wang Y, Liang Y, Li C, Zhang X. Operation Performance Evaluation of Urban Rail Transit PPP Projects: Based on Best Worst Method and Large-Scale Group Evaluation Technology. Adv Civ Eng 2021;2021:1–15. [CrossRef]
[49] Hossain M, Guest R, Smith C. Performance indicators of public private partnership in Bangladesh: An implication for developing countries. Int J Product

Perform Manag 2019;68:46–68. [CrossRef]
[50] Xu Z, Wang X, Xiao Y, Yuan J. Modeling and performance evaluation of PPP projects utilizing IFC extension and enhanced matter-element method. Eng Constr Archit Manag 2020;27:1763–1794. [CrossRef]
[51] Eadie R, Browne M, Odeyinka H, McKeown C, McNiff S. BIM implementation throughout the UK construction project lifecycle: An analysis. Autom Constr 2013;36:145–151. [CrossRef]
[52] Porwal A, Hewage KN. Building Information Modeling (BIM) partnering framework for public construction projects. Autom Constr 2013;31:204–214. [CrossRef]
[53] Deng Z, Yin Y, Yu X. Research on Construction of Life Cycle Supervision Platform for PPP Project in Sponge City Based on BIM. DEStech 2017. [CrossRef]
[54] Budayan C, Okudan O, Dikmen I. Identification and prioritization of stage-level KPIs for BOT projects – evidence from Turkey. Int J Mod Phys B 2020;13:1311–1337. [CrossRef]
[55] Love PED, Liu J, Matthews J, Sing CP, Smith J. Future proofing PPPs: Life-cycle performance measurement and Building Information Modelling. Autom Constr 2015;56:26–35. [CrossRef]
[56] Olawumi TO, Chan DWM. Green-building information modelling (Green-BIM) assessment framework for evaluating sustainability performance of building projects: a case of Nigeria. Archit Eng Des Manag 2021;17:458–477. [CrossRef]
[57] Chan DWM, Olawumi TO, Ho AML. Perceived benefits of and barriers to Building Information Modelling (BIM) implementation in construction: The case of Hong Kong. J Build Eng 2019;25:100764. [CrossRef]
[58] Liang Y, Jia H. Key Success Indicators for PPP Projects: Evidence from Hong Kong. Adv Civ Eng 2018;2018:1–10. [CrossRef]
[59] Ren G. Knowledge management in PPP decision making concerning value for money assessment. Cardiff Univ 2019.
[60] Du J, Wu H, Zhao X. Critical Factors on the Capital Structure of Public–Private Partnership Projects: A Sustainability Perspective. Sustainability 2018;10:2066. [CrossRef]
[61] Ma G, Jia J, Ding J, Shang S, Jiang S. Interpretive Structural Model Based Factor Analysis of BIM Adoption in Chinese Construction Organizations. Sustainability 2019;11:1982. [CrossRef]
[62] Song S, Zhang C, Marks E. Effectiveness and Practicability Analysis of BIM Adoption in the AEC Industry. Comput Civ Eng 2022;530–537. [CrossRef]

SUPPLEMENTARY APPENDIX

Supplementary Table 1. KPIs for evaluating BIM in PPP projects

Category	Indicator	Subset	Reference
Administrative	Administrative	Risk of Adverse changes in law, policy or regulations	[14,39–43]
		Organization structure	[42,43]
		Risk of Force Majeure	[44,45]
Sustainability	Manpower	Experience and knowledge gains	[11,46–48]
		employment opportunities	[45,48–50]
		Specialized Expertise	[49,51,52]
		Technical Staff Resources	[41,50,52]
		Staff training	[11,18]
	Machine	Equipment condition and number	[18,39,50,53]
		effective interface management	[48,54,55]
	Material	Use of construction waste	[18,39,44,45,56]
		Material condition and number	[18,39,50,53]
		Use of innovative materials	[44,45]
Management	Time	on-time or earlier project completion	[41,50,57,58]
		Completion/Time Delay	[39,42,59]
		schedule variance	[46,50]
	Cost	Investment life cycle Cost and return	[11,41,44,45,47–52,53–55,57–59]
		Construction Cost overrun	[39,60]
		Cost management	[46,47]
		cost variance	[47,50]
	Quality	contract management and control	[39,40,48–50,52,54,54]
		quality and innovation public service	[47–49,58,62]
		The third-party assessment results	[18,48]
		Change in scope	[41,42]
Tools Framework	People	effective Safety management: no. of security incidents	[18,39,41,48,50,54,55,57]
		Change management - resistance to change	[11,52,61]
		Information and knowledge sharing	[18,57,61]
		Requirements of Stakeholders/Goals	[39,41]
	Process	procedures for settling claims and disputes	[25,41,42,49,55]
		Detail and transparent level of tendering procedure	[39,52,54,55]
		Clash Detection	[57,62]
		transferring risk to the private sector	[54,58]
		Life-cycle management ability	[49,60]
	Technology	Document management	[52,53,56,62]
		updated Timeline	[18,50]
Productivity	Environment	Environmental risk	[18,55,60]
		Energy efficiency (e.g., reduction of energy consumption and use of renewable energy resources)	[44,45,56]
		Reduction of pollutant discharge	[44,45]
	Economy	Value for Money (VFM)	[18,39,48,49,54,55]
		Feasibility analysis	[39,49,52,54,55]
		Economic development	[41,47]
	Social	market interest and opportunity	[39,47,54,59]
		Government reputation and improvement	[41,47]
		Trust and respect	[44,49]
		Public interest	[49,60]

Supplementary Table 2. The overall questionnaire form

1. Adopting BIM in PPP Projects can enhance and modify organizational structures to better align with the project›s requirements.
2. Adopting BIM in PPP Projects can effectively decrease and alleviate the risk of unfavorable law, policy, or regulation alterations.
3. Adopting BIM in PPP Projects can reduce force majeure risks.
4. BIM adoption in PPP projects will improve the overall experience and knowledge gains due to BIM integration.
5. BIM adoption in PPP projects can improve the quality of staff training by introducing interdisciplinary training and the operability of information.
6. BIM adoption in PPP projects will make it easier to get information by using effective interface management.
7. Adopting BIM in PPP Projects facilitates the efficient assessment and utilization of waste materials.
8. Adopting BIM in PPP Projects can generate employment opportunities or productive jobs by meeting the growing knowledge and demand for such expertise.
9. BIM adoption in PPP projects will produce more specialized expertise when compared to traditional PPP procurement.
10. BIM adoption in PPP projects will attract more qualified technical staff resources.
11. BIM adoption in PPP projects has vast resource allocation potential by keeping records of Machines› conditions and numbers.
12. Better use of innovative materials through design and implementation can be achieved by adopting BIM in PPP projects.
13. BIM adoption in PPP projects has vast resource allocation potential by keeping records of Materials conditions and numbers.
14. BIM adoption in PPP Projects will better meet stakeholders› diverse requirements and Goals in performance management.
15. BIM adoption in PPP Projects can reduce resistance to change among team members and stakeholders.
16. The adoption of BIM in PPP Projects promotes the effective sharing of information and knowledge and an interoperability framework.
17. BIM adoption in PPP Projects is more effective in Safety management on construction sites and can reduce security incidents.
18. Data and documents used in PPP projects that adopt BIM have everything sorted in digital folders and archives to be quickly and effectively created, managed, and shared.
19. BIM Adoption in PPP projects makes timeline updates quicker and more efficient, including design, construction, and in-use operational information.
20. BIM Adoption in PPP projects has the potential to utilize smart technology for better coordination and Clash Detection.
21. BIM adoption in PPP Projects helps transfer the boundaries of risk and insurability to the private sector, leveraging risk mitigation and prevention affordability.
22. BIM adoption in PPP Projects simplifies claims management by identifying potential claims, information consistency, data storage, and improved collaboration.
23. BIM adoption in PPP Projects can help create a more detailed and transparent level of tendering procedure due to the integration and clearness of the project.
24. BIM adoption in PPP Projects has the tools for smoother and more efficient Life-cycle management throughout projects› life cycle stages.
25. BIM adoption in PPP Projects increases efficiency and thus increases the chances of on-time or earlier project completion.
26. BIM adoption in PPP Projects can notify early warnings on schedule indicators, resulting in fewer schedule variances.
27. BIM adoption in PPP Projects will result in a better investment Cost and return throughout the project›s life cycle.
28. BIM adoption in PPP projects improves cost management by estimating the total price reasonably based on the specific expenses required for each part of the project quantity.
29. BIM adoption in PPP Projects can notify early warnings on cost indicators and result in fewer cost variances.
30. BIM adoption in PPP Projects increases efficiency and thus reduces the chances of project time delays.
31. Adopting BIM in PPP construction Projects can significantly contribute to cost overrun mitigation.
32. BIM adoption in PPP projects helps meet the score or requirements, making a change in scope (scope creep) less likely to happen.
33. BIM adoption in PPP Projects can test and try different design alternatives to improve public service quality and innovation.
34. By adopting BIM in PPP projects, the third-party assessment process can be made more efficient and streamlined, leading to improved results and higher quality outcomes.
35. BIM Adopted in PPP projects helps in ensuring that the terms and conditions contained within the contract are adhered to and that all party's contractual obligations are met satisfactorily.
36. BIM adoption in PPP projects can allow designers to run an automated assessment of the environmental impact of their design choices immediately.
37. BIM adoption in PPP projects significantly impacts increasing efficiency during a project›s design and construction phases.
38. BIM adoption in PPP projects can help create business models that reduce the overall pollutant discharge by enabling environmental performance in building design and operation.

39. BIM adoption in PPP projects has the potential for better, more affordable, and quicker project alternatives to help economic development.

40. BIM adoption in PPP projects can achieve better Value for Money (VfM) than traditional PPP procurement.

41. BIM adoption in PPP projects creates a more detailed and accurate feasibility analysis reflecting the project life cycle stages.

42. Adopting BIM in PPP projects can facilitate more effective project completion and enhance project improvements, which, in turn, can bolster the reputation of government agencies involved in such projects.

43. Adopting BIM in PPP projects holds the potential to cultivate greater trust and respect among project teams and stakeholders.

44. Adopting BIM in PPP projects facilitates fulfilling market needs and realizing project opportunities by enhancing the comprehension of project details through BIM implementation.

45. The adoption of BIM in PPP projects has the potential to generate greater public interest in more advanced and complex projects.

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1399–1418, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2044

Research Article

Assessing the impact of natural language processing-driven clustering on engineering students' growth in mentoring programs

Pankaj BELDAR[1,*]

[1]*Department of Mechanical Engineering, K.K. Wagh Institute of Engineering Education and Research, Savitribai Phule Pune University, Maharashtra, 411007, India*

ARTICLE INFO

Article history
Received: 12 November 2024
Revised: 19 February 2025
Accepted: 07 March 2025

Keywords:
Clustering; Engineering; Guidance; Mentor

ABSTRACT

Although they help to build engineers, the mentoring programs hardly are reviewed number-wisdom, based on what learners say. Through that gap comes this work - based upon the voices of 3,362 pupils stretched over a ten year span of disciplines, year by year of study. Only by filtering comments with the help of such tools as sentiment decoding and group-finding algorithms, patterns could be observed. There were six distinct filaments that emerged: project assistance and grants advice, on-campus assistance and career guidance, personal development and assistance with research issues. Where opinions were floating freely, numbers took form. The score of 0.7985 on the silhouette score was a confirmation of the optimal number of clusters, which was also supported by the measures used, such as WCSS, Calinski-Harabasz, and Davies-Bouldin. Positive sentiment constituted almost 84.6 percent of the answers - confirming that mentorship helps in the development of learners. This is a good look at numbers to get a clear understanding of how students are going through these programs, which points to practical steps schools can implement. Rather than basing the analysis on intuition, the use of detailed analysis introduces new approaches to evaluate mentoring arrangements, closing gaps left unanswered in previous studies and informing future changes with evidence.

Cite this article as: Beldar P. Assessing the impact of natural language processing-driven clustering on engineering students' growth in mentoring programs. Sigma J Eng Nat Sci 2026;44(2):1399–1418.

INTRODUCTION

Mentoring is exceptional in most learning facilities and in particular in engineering studies where demands are fast accumulated. In the process of creating future engineers, the path is more than coursework - coaching takes place in the face of actual challenges, including personal challenges. Since these degree paths are pressuring in terms of time, energy, and concentration, the presence of someone who has been through it all assists the learners to remain grounded. Lectures are not what lead to success; it is lifelong progress that is intertwined with the human relationship. The workload feels heavy at times, layered with expectations few anticipate before starting. Mentoring gives tailored assistance so learners can manage tough situations while gaining stronger

*Corresponding author.
*E-mail address: prbeldar@kkwagh.edu.in

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

grades, greater self-assurance, yet better preparation for jobs. Research shows mentoring addresses the educational and subjective needs of engineering students to assist with their success as well as provide them an avenue to stay in the field. An integral component of acquiring engineering ability comes down to a continual advising relationship. With the courses considered some of the most difficult in educational institutions, students in many cases bear a major workload of course work, challenging research tasks and group tasks. With good mentoring comes crucial advising, support which has a significant impact on student success in college also at subsequent stages of development.

The need for Mentoring for Engineering Students

Tackling challenging coursework can take its toll on engineering students. When mentoring occurs, the eyes are opened - to how people execute tasks, manage their time, day in and day out. Rather than attempting to get through challenges on their own, the supported learners in this study benefit from strategies, support. Research indicates it helps improve marks, while reducing the risk of early withdrawal [1]. Typically, engineering pupils seek advice when looking for internships and jobs. As experienced travellers, mentors know about changing job demands. Through interactions, students better understand making industry connections. Befriending advice, options for development are less daunting. The dialogues increase the chance of achieving goals [2]. Employing high pressure, engineering study can at times make students feel isolated or stressed [3]. But support from mentors helps - to reduce nerves, boost self-confidence, as speaking helps to change the perspective. Although learning is intimate, it's enhanced by human interactions. As such, self-confidence is not just based on acquired knowledge but on having a champion [4]. Having someone listen makes the struggle worthwhile, able to be sustained. That quiet encouragement? It's there when you need it. Often, engineering education mentors do much more than advise and cheer up their students [5]. Typically, they are experienced teachers, professional engineers, or senior peers who have compete information to provide guidance through the roadblocks of education, profession and personal challenges. Their understanding of understanding of fundamental engineering principles and potential careers also allows specific advising within practical contexts [6].

Mentor Responsibilities

Guidance and support

Diving in right away, mentors help students to set specific learning and career goals. Advice often relates to choosing certain courses, assignments or career options. One contribution they make is to offer words of wisdom during times of great stress. Dealing with significant deadlines, support can help. But tackling challenging designs is less daunting when you know you don't have to go it alone. Help is sometimes subtle - more listening than talking Bad times go better with a buddy.

Role modeling

Engineers gain ethics, problem solving skills, from mentors. Through role modeling, students explore assumptions of product designs. During project challenges, learner engineers' mentors solve problems step by step. Such observations transform apprentices' responses to uncertainty in a project. Witnessing innovative solutions boosts apprentices' faith in systematic thinking.

Feedback and evaluation

Mentor feedback can guide students to make improvements to their work, such as a paper, project or presentation. Opportunities for learning rely on understanding, and so advice can be provided on what is going well as well as what needs to change. Examples in practice could be critiques and discussions of engineering summaries or feedback following fieldwork experiences.

Networking and opportunities

Mentored students are introduced to professionals often during events such as conferences. Practical experience is the part exhilarating hands-on work comes from contacts. Agname drop might help access lab experience few can get.. Training places sometimes are filled by a timely self-assertion.

A starting point may be through looking at text feedback from 3,362 students in different stages and groups of engineering. Instead of simply tallies, analyses using sentiment detection reveal insights about learning support - particularly working on skill for their work, training for jobs and help on campus - that draw deductions from experiences. As these comments reveal points of help, these findings support Vygotsky's notion that learning progresses when people can receive assistance just in time. What's interesting is how student feedback reflects the journey from 'being' to independence - through mentoring.

This approach using clustering methods, reflects on continuing changes and patterns in students' opinions about mentoring. Student expectations vary with time in clustered responses. Results relate to Bandura - learning influenced by contact with others and mentors. Groupings create patterns - semester-to-semester differences.

Understanding improves and engagement increases, focus is sharpened. They learn: academically and for careers due to guidance.

This research differs as it influences more effective mentoring strategies to match evolving engineering learners as their knowledge and confidence grow, leading to better learning and greater outcomes. Mentoring has long been a highlight in the growth of future engineers at school. Studies show performance and learning are linked to such programs, as careers get a break start and students gain confidence. Reflections on previous research show how some types of mentoring relationships work. Patterns of types of support emerge (some by students, some by faculty) and these affect outcomes. Identity and race can change who and how much.

Retention (staying enrolled) and degree completion tend to improve with advising programs. The following synthesises insights about these relationships from a series of studies? Our learning is often observational Bandura put it like this - we learn by watching. A teacher's example is more important than instruction. Advice is more effective when it's based on experience. Attitude changes are subtle, through habituation. A good role model provides a good example. Mentors' actions are more important than words. Learning comes by observing, as well as doing. Good habits are learnt through contact with talent. Behaviours are spread by example. Students learn a lot from closely watching mentors, so they develop working habits, and how to solve problems and communicate them [7]. Patterns of feedback explored here reveal practices that enhance the impact of mentorship in practical learning - to mimic good role models. Input from an expert enables places for learning beyond independent work alone - a key element of Vygotsky's development zones theory. By gradual support and guidance, mentors allow students to stretch to more complex tasks that they couldn't do alone. Thoughts like age-old career planning or project steering demonstrate how we can use small cues to gradually build trust and confidence in students' growing independence [8]. While bringing together such models, the research reveals mentoring aids student learning via social (observational) learning and support [9], facilitating course and career building practices [10]. The Peugeot Center Model was investigated by Dodson and colleagues in 2022 and involves integrating mentoring within relevant engineering challenges, improving students' applied skills, cross-cultural skills and engagement. Engaging in industry-relevant scenarios, the model places theory and practice in context, fostering critical and design thinking.

Continuing support led to greater engagement, and higher grades. Rather than individual contact, ongoing relationships made the difference. Agholor, with de Nalda and Bárcena three years ago launched the idea of shaping the engineers of the future with structured mentoring in higher education settings. By creating opportunities for mentoring, skill development thrives; both technical and interpersonal skill building occur. Where human identity meets mentoring programs, there are differences; ethnicity and gender take on new meanings in mentoring. In 2017 Aikens and her peers examined the effect of mentoring on undergraduate research outcomes - racial minorities had access to less mentoring. Given this imbalance, culturally-sensitive mentoring techniques would seem to be needed for equitable experiences [11]. Jumping forward to 2021, Tuladhar's team explored what shapes the paths of African American students in engineering, spotlighting informal networks and visible role models within mentorship. These aspects, it turned out, played a key part in helping underrepresented learners feel they truly belonged while also doing well academically [12]. Then there is Dennehy and Dasgupta's work from 2017: their focus fell on how having women peers as mentors influenced whether female students stayed in engineering programs. Early peer mentoring may enhance how women experience academia, boosting their chances of staying in engineering programs [13]. Research into such support systems shows consistent links to better student results. A 2010 analysis by Budny, Paul, and Newborg found mentored first-year engineers performed stronger academically, sticking with their studies more often [14]. Work led by Still maker and colleagues in 2020 looked at whether instructor gender affects guidance offered to women pursuing engineering. Female faculty members often help female students thrive - especially within engineering, where men are more numerous [15]. Not long ago, Gehr and colleagues showed how guidance from professors shapes graduate learning in engineering. At one public university, close advising ties boosted both scholarly output and fulfilment among learners [16]. Supportive connections between teachers and pupils appear linked to better persistence in tough degree paths. Mendez's team pointed out that meaningful bonds involving Black instructors and their mentees matter deeply. One reason these connections matter is how they turn teacher insights into better outcomes for learners, especially those from marginalized backgrounds. During the pandemic, Saw and colleagues in 2023 looked at gaps in guidance and emotional well-being among American undergraduates. What stood out was that strong mentorship softened the blow of isolation and stress on grades and mood, showing its value when conditions get tough [17]. In another case, Rayford's team two years later found younger engineering majors gained both socially and scholastically through mentors close in age. Evidence points to advising bonds lifting involvement, persistence, completion - especially where support runs thin [18].

This work looks into how NLP-driven grouping methods affect mentorship within engineering learning, bringing practical uses for teachers, school leaders, corporate collaborators, and counsellors. Results reveal ways to boost academic performance, keep students enrolled longer, and prepare them better for jobs - all while aiding emotional well-being and encouraging varied participation. Such discoveries may shape teaching plans, rules, and course design over time. While centered on engineering classrooms, the approaches apply just as well elsewhere - across fields or regions - for those aiming to lift education standards using strong mentoring frameworks.

MATERIALS AND METHODS

This paper follows a systematic approach to the research topic of the role of mentoring in engineering students, based on the written feedback obtained during several academic semesters. According to Figure 1, the process starts with the group assignment of mentees, and periodical guidance meetings. Then, the feedback is gathered with the help of feedback forms that will be completed by the participants. These are then subjected to the pattern detection process through machine learning. The overall effectiveness is assessed as the last step, with regard to the tendency of clustered responses.

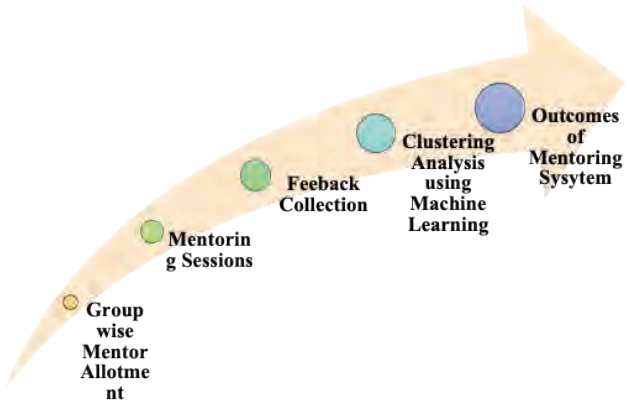


Figure 1. Methodology.

Group-Wise Mentor Allotment

Purpose: To make sure that, mechanical engineering students in India will be provided by personalized guidance through the mentors during their academic course, their academic performance, their readiness to enter the career, and their overall development will be improved. By using systematic scheduling whereby certain time is allocated to mentoring, faculty can balance between mentoring and their academic life since the time allocated to mentoring is at a single time and as such, it does not interfere with teaching, research and administrative activities. Peer learning and collaboration can be achieved by group mentoring sessions that are able to make the most of time. Co-mentoring or teaching assistant models of collaborative mentoring assist in distributing the workload and offering different perspectives. Also, incorporation of mentoring with academic work, including coursework or research projects, will make sure that mentoring does not interfere with other duties, but instead, it supplements them. The support of the institutions such as the acknowledgment of mentoring in workload calculations and the provision of professional development opportunities is also important. The strategies will help the faculty to strike a balance between mentoring and professional obligations.

Mentor Selection

Objective: Look for mentors that can teach mechanical engineering students and industry pros are hired. Selected candidates must be able to articulate well. Experience with industry is critical. Encouraging communication is favored. Students are reviewed before admitted. Matching is done to student needs. Long-term involvement is encouraged. Good mentoring is key to success. Roles are defined in each partnership.

Look for those with deep knowledge of the discipline - well-experienced academics, company engineers with experience, or highly productive senior undergraduates. Examples are faculty members heavily engaged in mechanical design courses, practitioners who work in the areas of automotive or flight technologies, or undergraduates who perform well in fundamental courses. One practice is to look through past experience in research projects, followed by looking through record of mentoring given in past projects. Look past titles when selecting to mentor; participation more important than position. Typically, those already supportive of helping others are selected.

Look at credentials first - mentors face scrutiny via experience, training, background screening, references, and depth of understanding of a subject, communication, ability to provide feedback, time availability, cultural compatibility, and various development activities

A mentor should have a good grasp of mechanical engineering so, a university lecturer who is an expert in heat and liquid flow - good. Or perhaps someone out of the car industry who has tackled problems in industry for years.

For mentors, specific knowledge of mechanical engineering is most important. Mentors should understand key concepts and how to apply them. A robotics expert could be suitable. Or a lecturer in structural analysis might be also suitable. Out of classroom experience enhances their worth.

Being available is a feature of mentoring. It is crucial the mentor's calendar is open for discussion. An interest in helping forms the basis of the mentor-mentee relationship. Involvement rings in showing leadership, listening well, responding well. Time reflects a commitment to learn. Encouragement takes little actions, done consistently. Advice is best delivered as a mental organically given service, rather than receiving a service by obligation.

Following identification of potential mentors, the next step is selection processes. Those selected then prepare by attending workshops. This involves focusing not only on theory but practical methods of mentoring. Awareness of what learners are challenged by is a focal point of the program. Techniques for providing feedback are role-played and practiced so they're instinctive when delivered to learners. Creations of positive environments are practiced during the program. Training continues only until participants are confident in certain aspects.

Industry Professional Allotment Process

1. Determining Mentor-Student Ratio:

Typically, an industry specialist caters to four or five students. This is a good balance - attention can be focused, but also doesn't overwhelm the guide.

The number of industry professionals required will be determined by student numbers. If the number of students is 100, we need approximately 20 - 25 industry experts? On the other hand, much fewer are needed for smaller class sizes and more are needed for larger class sizes.

2. Finding Professionals to Work With:

Alumni can help in locating professionals to mentor others. Recent graduates in employment, such as aerospace, automotive or production get approached through university-based social networks. This contact is made using platforms like LinkedIn, as well as graduate societies.

Collaborating with firms that are engaged with the mechanical engineering program - as employers for student placements, visiting lecturers or engaged in joint research projects - helps identify potential mentors. These may stem from on-going development projects in which mentoring roles arise. Companies already involved with the department generally provide academically minded people. In projects identified, needs lead to appropriate personnel. Links forged in work-based learning stages lead to mentoring relationships. Engagement in practical work creates avenues to mentoring relationships. Collaboration towards common aims assists identifying mentors.

Collaboration with organizations such as ASME leads to those working in mechanical engineering. These connections develop organically and mostly through working towards a common goal. Engagement generally occurs at conferences and meetings. Networks facilitate access typically denied estranged. Project creation can be casual, stemming from interest. Trust is developed gradually over time. Knowledge is shared profusely in projects. Multi-year involvement is preferred to one-offs.

3. Getting Industry Insurance Professionals to Get Involved:

There is an interest for pros to participate as it enhances skills like team work and communicating. Participation expands networks - between organizations, as well as professionals. Visibility is gained, too, with insights into trends in research. Experience is also developed, along the way, changing the way people make contributions.

Thanks come in the form of public thanks, including on the internet. Being involved could result in a later collaboration on studies. Participants may get written acknowledgement.

Because of busy schedules, experts can decide their frequency of participation - every few weeks or monthly - known as juries, online. Scheduling is flexible per individual, as needs and schedules vary.

Session Length and Format

Each meeting typically lasts half an hour to 40 minutes so there's time for discussion, and so the students aren't overloaded.

Prior to each meeting, a theme is assigned. This includes topics to be discussed. This ensures that things will run smoothly. Topics listed in advance at a meeting provide a flow of discussion. Understanding is achieved through knowing what will be discussed. Attention is maintained with an outline. Planning promotes better meetings

- Academic Advising (e.g., course and exam preparation)
- Employment Advice (e.g., job-related skills, interviews)
- Personal Development (e.g., technical, soft and management skills)
- Personal Well-being (e.g. stress management, workload)

Session Planning and Flexibility

These are meetings prearranged early in the semester, via the university's systems or e-mail notices. These are scheduled rather than coinciding, throughout the term.

If changes occur, the mentor and student can rearrange the meeting to proceed with the plan - this enables the project to proceed regardless of changes to study plans or other commitments.

First Year Student Timeline Example

August: Orientation and welcoming to the mentoring program

Meetings are bi-weekly between September and December to help with settling into the year, as well as develop skills

Half-year meetings are held in January to inform goals. Targets emerge around this period, shaping future. Mid-year reviews impact improvement, shaping next steps. Plans unfold for the next semester; plans are also impacted. Reflection times mark month's path forward

February - May: Ongoing twice-weekly sessions, with new focus on electives and revision

Evaluation and improvement

After each session, we collect feedback - this helps to evaluate how effective a session was, and inform what's coming up (adjusting dates and content). Participants' feedback accompanies adjustments to improve sessions, without aiming to stick to pre-planned structures.

Documentation

Digital recording of adjustments, such as collaborative documents or guidance systems can speed up tracking the learner's progress and minimizing paper. These systems helps updates to go live across devices, eliminating re-entry. Progress reports are easier to access as they are stored electronically. Techniques like this shift paper to simplicity. Data can more easily get from teacher to students.

The focus in on the next step. Thorough records draw attention to positive outcomes rather than a heap of details. This allows mentors to focus on supporting people, rather than paperwork.

Meeting minutes

There are meeting summaries created after each meeting; brief reports to keep you up to date. Objectives, agenda items and action items are accompanied by details. When developing a group project idea, between a mentor and student for instance, the summary will contain the objectives and anticipated outcomes. Everyone's roles are distributed accordingly, with dates if necessary. Progress reports might be added later.

Session outcomes

Progress notes taken frequent meetings assist with monitoring progress and efficacy. Reflections from discussions inform future decision making about inadequate support in the future.

Example:

A few mechanical engineering second years have advisors with expertise in their targeted fields. These guiders and learners meet bi-weekly. In one such meeting, the learner mentions being unsure about what electives might best lead to a career in robotics. Advice is given; not just on what subjects to study but also extra-curricular pursuits. Gaining experience from projects might be recommended. Knowledge emerges in a two-way dialogue - decisions emerge without formulas. Knowledge also evolves, as a result of authentic talk and practice. Objectives are discussed - how to raise grades in certain subjects are gradually developed. There's a record - simple, straightforward: the learner's intention is on paper. Additional tutoring could occur; commitment to a team building robots might form.

Feedback collection

Students' feedback drives good mentoring outcomes for engineering students. It focuses on students' program experiences. We take their opinions into account when evaluating support.

1) Objective

The goal of collecting feedback is to assess the mentoring program. The first assessments are made through student feedback about halfway down the term. A crucial element is to identify potential improvements. Further reviews follow at the end of the year. These provide a better view of impacts on learning. Information at various times supports learning checks as well as judgements.

2) Survey Design

The feedback surveys ask about the different stages of mentoring, but also mentors' communication style. Areas examined include the helpfulness of advice shared. Surveys indicate how satisfied mentees are through fixed-answer questions that blend elements of a scale and comment boxes. Instead of monomeric measures, a mixture of numbers and comment is used. Each approach has slight variations for each program stage. What's impressive is the non-overlapping collection of information about timings, clarity and supports

For example, numbers are used: for instance, a rating of one to five: one being strongly disagree, five being strongly agree, on a trait. The responses measure attitudes towards specific issues, becoming increasingly positive. Each numerical position represents an opinion, well organized but simple. Ratings concern aspects being studied, with a clear pattern of numbers. This converts opinions into this information, so they can be compared

- Mentor's accessibility and inter-personal skills.
- The mentor's advising ability.
- The mentors' relevance and helpfulness.
- Overall experience with the mentor.

3) Feedback Timing and Multiple Feedback Points

As the year unfolds, opportunities for feedback are garnered from a variety of phases. This ensures that feedback is delivered with some time to process, overridden by feedback over time. Stages provide differing perspectives. Responses help us understand how things are moving on even months later. Thinking small, often, avoids impact

Immediately, we use brief check-in responses to each mentoring meeting - to check for problem areas. Meanwhile, students can anonymously communicate in a secure online environment.

In feedback, mid-term provides an opportunity to identify issues. If sent mid-way through semester 1, students can work on any adjustment before term-end. This assessment occurs as part of the mentorship process, and encourages ongoing development. Any problems noted can be addressed later. It's reasonable to do it right now - it will be useful in the near future. Students may indicate issues with scheduling time for meetings with the mentor and the mentoring relationship may change accordingly. Occasionally there are timing problems, and the strategy is changed a little to optimize the process. Schedule availability frequently has an impact on the guidance cycle, and requires adjustments from participants. As access to them is sometimes problematic, an occasional shift may be needed. The wave changes a bit when the days in between them are missing from the calendar.

Deviations may arise as a student might identify difficulties in scheduling time to meet their mentor, which will affect the mentoring. Occasionally there are timing problems, and the strategy is changed a little to optimize the process. Schedule availability frequently has an impact on the guidance cycle, and requires adjustments from participants. As access to them is sometimes problematic, an occasional shift may be needed. The wave changes a bit when the days in between them are missing from the calendar.

4) Data Analysis and Interpretation

By analyzing the survey's numbers, we identify common themes in student satisfaction. Ratings are analyzed using a statistical technique to find the success of mentoring. Scores are analyzed by grouping by frequency of response. Common ideas are identified by calculation. Patterns ensure when scores are ranked. Patterns stand out and can be finds representative. Any duplication is indicative of outcomes holding across students.

The system begins with unfiltered comments and NLP to analyses emotional attributes in responses. Rather than have a single person review the work, multiple reviewers can review the texts independently. To ascertain agreement, their decisions are compared for their agreement rate. This minimizes bias while helping to achieve an unbiased conclusion.

Each step in seeking feedback is ethical, including disclosing details about the feedback session before people participate; letting people choose freely to participate and ensuring people can leave if they choose. Trust and safety ensue when learners' protected honest opinions are valued in an environment of their choosing.

Clustering Analysis Using Machine Learning

Objective:

The objective of clustering analysis using machine learning is to uncover underlying patterns and themes in student feedback. By grouping similar feedback responses, the analysis helps to identify common experiences, sentiments, and areas that require attention in the mentoring process.

Feature extraction

- TF-IDF Vectorization: This technique converts the preprocessed text data into numerical features that can be used by machine learning algorithms [19].
- Term Frequency (TF): Measures how frequently a word appears in a document.
- Inverse Document Frequency (IDF): Measures how unique a word is across all documents [20]
- TF-IDF Score: Combines TF and IDF to give a score that highlights important words that are frequent in a particular document but rare across others.
- Example: The phrase "mentor support helpful" in one feedback might have a high TF-IDF score for "mentor" and "helpful," indicating their importance.

Clustering

- Analysis Algorithm: Many clustering algorithms are considered, frequently the simplest and most effective (K-means) algorithm.
- K-means Clustering: This algorithm divides the data into K clusters, such that each feedback is allocated to a cluster it is closest to in terms of features [21].
- Alternative Algorithms: Other algorithms such as Hierarchical Clustering or DBSCAN might be used, depending on the data [22].
- Finding the Right Number of Clusters:
- Silhouette Score: Calculates how closely related an object is to the cluster it is a part of relative to other clusters. The higher the score the clearer the clusters [23].
- Elbow Method: A line graph is plotted based on the within-cluster sum of squares for various numbers of clusters, and the "elbow" where the decrease begins to diminish is determined, optimal number of clusters.
- For instance, following K-means clustering you might find that 5 is the optimal number of clusters to separate feedback responses[24].

Cluster Analysis

- New Themes: Each cluster is analysed to find the most common words and phrases that constitute each cluster, identifying new themes in the feedback
- Example: A cluster might be based around "academic support" and phrases such as "goal," "study plan" and "course selection" may be common.
- Sentiment Analysis: The feedback in each cluster can be analysed to determine the overall sentiment (positive, negative, neutral) [25].
- Example: "Mentor availability" might show a combination of positive and negative sentiments, depending on whether students could reach their mentors or not.
- Identify Student Experiences: Patterns of student experiences are identified, providing opportunities to improve a program.
- Example: A cluster based on "communication skills" could reveal students value communicative transparency, but think mentors could be faster in responding.

Evaluation of the Mentor

1. Beginning with gather information from the learner, reviews attempt to gauge mentor effectiveness in their support of the learner. Next, observations of mentoring practice help to identify skills performing. Occasionally this includes a mentors' reflections. Overall, using a combination of these methods highlights mentor performance
2. Students rate their mentor immediately after mentoring talks. But they are asked specific questions - about presentation, support, and helpfulness. But answers demonstrate if mentors really disappoint or not. But, it offers a view of how mentoring advice matches the real-life needs. While short, reviews act on lack of connection, effectiveness.
3. The assessment starts with meeting frequency, followed by punctuality. Mentors also must be available in between meetings. Discipline is demonstrated with these actions, and commitment over time. Most important is 'being there' - over and over again without pressure.
4. The benefits of mentorship are measured in students' improved academic scores, skills and judgement skills. Improvements in these areas can be a measure of mentorship. The result is this cycle - help results in progress. Over the long term, learning styles indicate greater learning. Success is subtle, it's in the results.
5. Yet another lens is to have fellow colleagues or supervisors observe mentoring sessions. At times, program managers gather feedback on what mentors are doing. Programs with this type of oversight bring consistency to training. Loops with senior staff can identify issues. Another source of feedback can be from formal checkpoints. Areas needing attention are often pinpointed through these assessments. Voice from several sources promotes equitable judgements of impact.
6. As review may identify gaps, mentors could also attend follow-up training to strengthen their mentoring skills - ensuring sustained development. What's important is ability to adjust follow-up support.

RESULTS AND DISCUSSION

When it comes to numbers of clusters, six is the most prominent given its silhouette value of 0.7985 - the highest observed. This is close to 'one', signaling compact clusters and large gaps between them. Rather than blurring together or breaking apart, the data fall nicely into these six parts. The closer the value is to the top of the -1 to 1 range, the more

certain we can be the structure is real. Segments of six seem to be consistent with naturally occurring dissimilarities in responses. High silhouette values demonstrate close bunching and good separation; so items within one group are similar, and distinct from items in other groups. With six divisions, we can explore what these groups are defined by; what are the top themes and concerns associated with the groups? Such findings suggest improvements to better align strategies for mentoring with what students live. The future lies in revisiting the prominent terms in each cluster and summarizing the insights gained, then making changes based on the data.

WCSS (Within-Cluster Sum of Squares)

With increasing cluster count, WCSS drops steadily - a predictable pattern. When plotted using the elbow technique, a slowing decline should become visible at some point. Yet here, the drop keeps going without clear levelling off, making that approach less useful. Instead, six stands out as the most suitable number, supported by evidence in Figure 2.

Silhouette Score

One way to judge cluster quality is by checking how tightly grouped each point sits within its assigned set. Points standing apart from neighbouring groups tend to yield stronger scores. With a value reaching 0.077932, six clusters pull ahead in performance. This peak appears when exactly half a dozen groupings are used. Seen clearly in Figure 3, this number marks the top result across tested options.

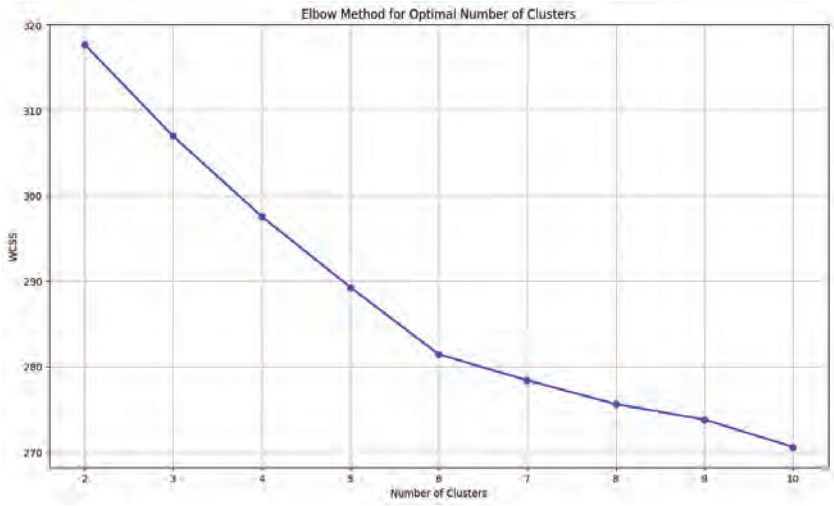


Figure 2. Optimum clusters by elbow method.

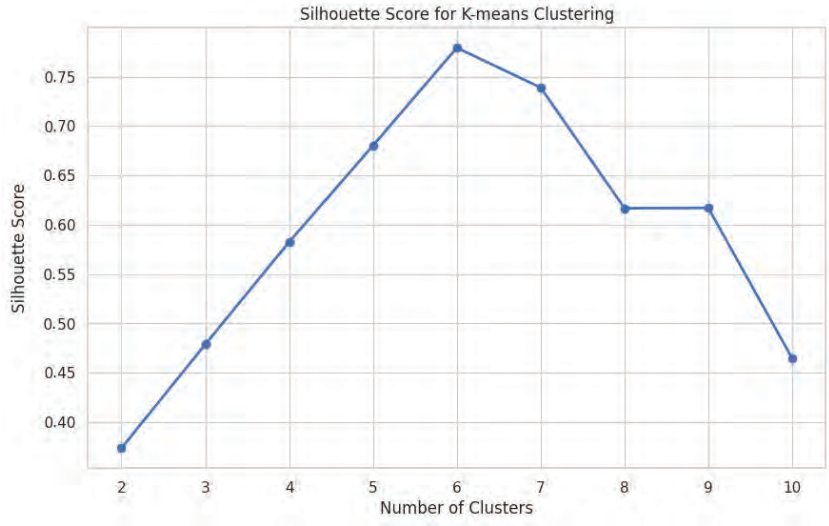


Figure 3. Cluster wise Silhouette Score for K Means Clustering.

Davies-Bouldin Index

Starting at 3.397363, the smallest Davies-Bouldin score appears when there are six groups. This index compares each cluster to its closest counterpart, averaging how much they resemble one another. Lower numbers here point to stronger separation between sets. Because the minimum occurs at six clusters, that count fits best under this method. See Figure 4 for the trend across different counts.

Calinski-Harabasz Index

Looking at cluster quality, the Calinski-Harabasz Index compares spread between groups against spread inside them. A greater number suggests stronger distinction. Peaking at 0.992897 when six segments form, it signals a clearer divide than other counts - this peak appears clearly in Figure 5.

Looking at how responses grouped together helps uncover what students really talk about when they mention mentorship. Each group reveals something distinct, shaped by its most common terms:

Cluster 1: Professional Development

Keywords: internship, guidance, helped, practical, company, mentoring, advice, industry, career, find

In the case of this group, there is a focus on job-related work life, job-based mentorship, and growth. The most critical concepts are that learners should find value in the advice that relates them to real work when undertaking placements or roles associated with particular areas. The use of words like company, career and industry is frequent and preparation to working environments is of importance in this case. Coaching assistance also evidently assists to

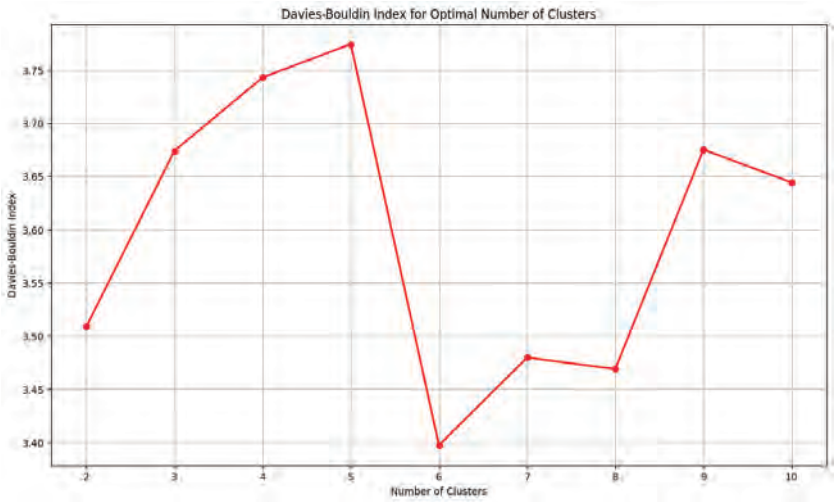


Figure 4. Cluster wise Davies-Bouldin Index.

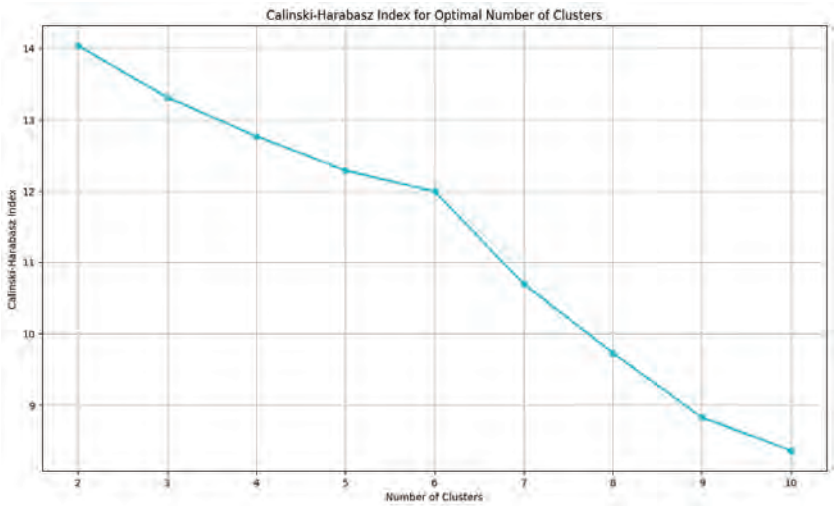


Figure 5. Cluster wise Calinski-Harabasz Index.

develop future decisions. Figure 6 shows a visual representation of these patterns with popular terms included in Cluster 1.

Figure 6. Word Cloud Cluster 1.

Cluster 2: Campus and Community Support

Keywords: cultural, support, campus, feel, mentoring, helped, made, community, activity, event

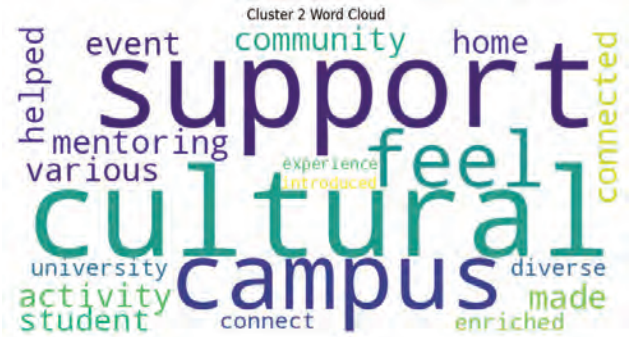


Figure 7. Word Cloud for Cluster 2.

Figures 7 shows a word cloud for Cluster 2

For Cluster 2 mentoring is indeed important to the students on campus. With words such as cultural, community and event used most often, it seems clear the relationships of mentoring help to shape social relations amongst students. Such exchange is open to further engagement than just from a consideration of grades. When there's support, it seems that the students are drawn more into participation. Mentoring streams, for its part, is a subtle element in pervading the ordinary college life. The Word Cloud of Cluster 1 (Figure 7) gives more information.

Cluster 3: Project and Feedback Focus

Keywords: project, provided, mentoring, feedback, session, mentor, continuous, helped, valuable, improved

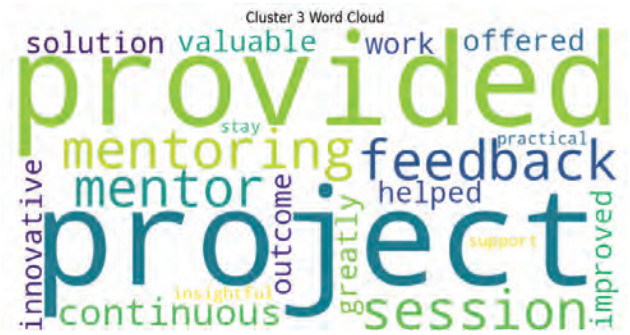


Figure 8. Word Cloud for Cluster 3.

Looking at Cluster 3, mentoring is one of the important issues regarding project work and making progress. The pattern of keywords in the data results in feedback and frequent workshops. The high number of words such as project, feedback, better suggests learners will benefit from targeted and repetitive instructions. The words are spread as a visual depiction (known as word cloud) of this group across the board in Figure 8.

Cluster 4: Research and Resource Support

Keywords: research, mentoring, provided, helping, helped, develop, needed, resource, methodology, crucial

Figure 9. Word Cloud for Cluster 4.

Cluster 4 is focused on the mentoring in terms of research in academia and research skills support. Words research, methodology and resource are quite common - these refer to the mentor's being engaged in the student's research project. Mentor advice matters - guidance has an impact on the study design and work with data of students. This is shown by evidence when we look at word patterns expressed in replies. We can see a visualisation in Figure 9 and this shows some overlap in words for Cluster 4.

Cluster 5: Scholarship and Application Guidance

Keywords: scholarship, application, advice, clear, successful, helped, mentoring, detailed, led, process

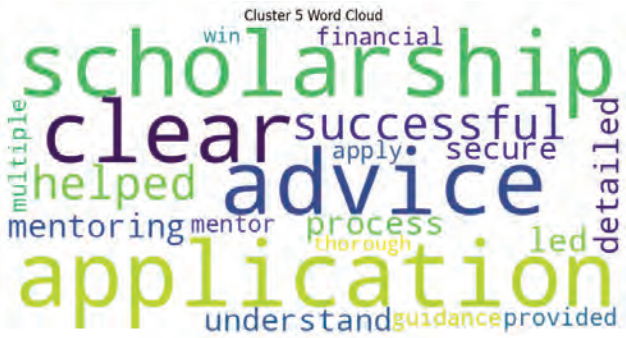


Figure 10. Word Cloud for Cluster 5.

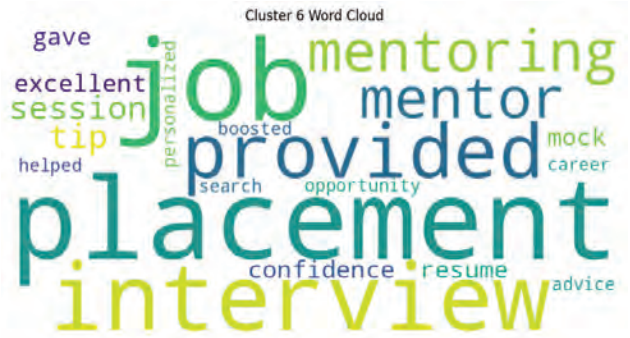


Figure 11. Word Cloud for Cluster 6.

In this cluster, the emphasis is on an approach based on scholarship assistance and follow up with the processes. These terms - scholarship, apply (how to) or advice - are used to highlight how students can benefit from support in applying for scholarships. Most prominent is support where mentors explain and tell learners the specific facts and details. The Word Cloud representation (see Figure 10) of Cluster 5 is shown here.

Cluster 6: Placement and Job Preparation

Keywords: placement, job, interview, provided, mentoring, mentor, session, tip, confidence, resume

Looking at Cluster 6 - it is about finding a job and managing interviews. Care is linked to stronger levels of confidence and better pieces of job applications - words like employment, interview, resume and confidence. Dialogues seem to imply a focus on ways of world-preparedness rather than advising. In this case, the preparation to work will be a dominant motive as it establishes the modes of introducing students to work. We can see the influence of mentors in the way the students will put their CVs to the employer. For example, in Figure 11 the Word Cloud associated with Cluster 6 is shown.

Sentiment Analysis

Most student feedback comes across as clearly approving, based on an analysis of responses from 3,362 individuals. Positive remarks take up 84.56%, standing out when set beside the 8.77% labeled neutral and just 8.67% marked negative. Seen together, reactions form a pattern favoring satisfaction with how mentoring unfolds. Scattered throughout, quiet reservations remain visible but do not shift the general trend. What emerges most distinctly is a tilt toward affirmation rather than doubt. Visualization of this emotional breakdown appears later as Figure 12.

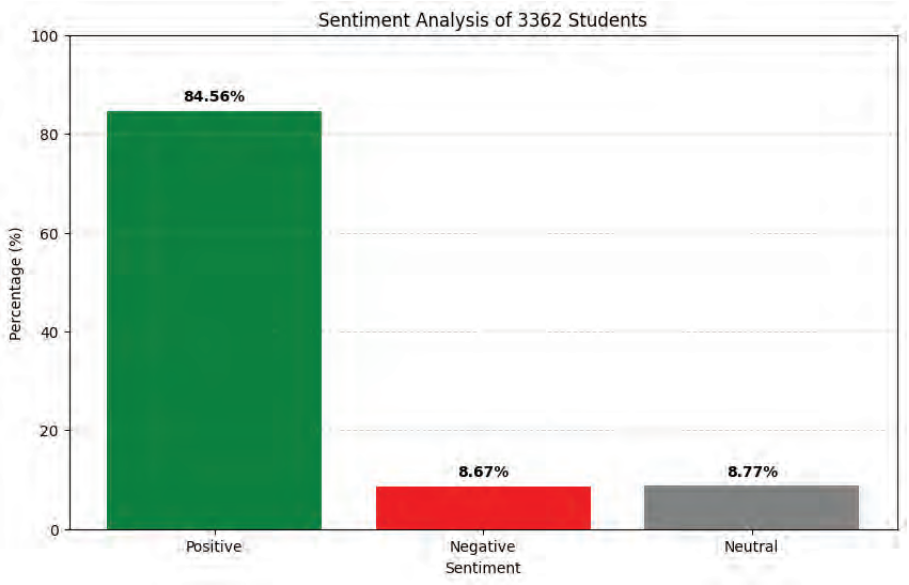


Figure 12. Sentiment Analysis.

Word Network

Figure 13, 14, 15, 16, 17 and 18 shows the word network for cluster 1, 2, 3, 4, 5 and 6

Cluster 1: Internship and Mentoring

Key Words:

o internship, helped, mentoring, guidance, advice, practical, company, find, career, industry

The cluster focuses on the practical aspects of internships and mentoring. Internship and mentoring are central nodes, with frequent connections to words like helped and guidance.

Cluster 2: Campus Culture and Support

Key Words:

Cultural, mentoring, support, campus, helped, feel, made, activity, event, community

Cluster 3: Project and Feedback

Key Words:

project, mentor, mentoring, provided, helped, session, feedback, valuable, improved, continuous

Cluster 4: Research and Resources

Key Words: Research, mentoring, helped, provided, develop, helping, needed, resource, methodology, crucial

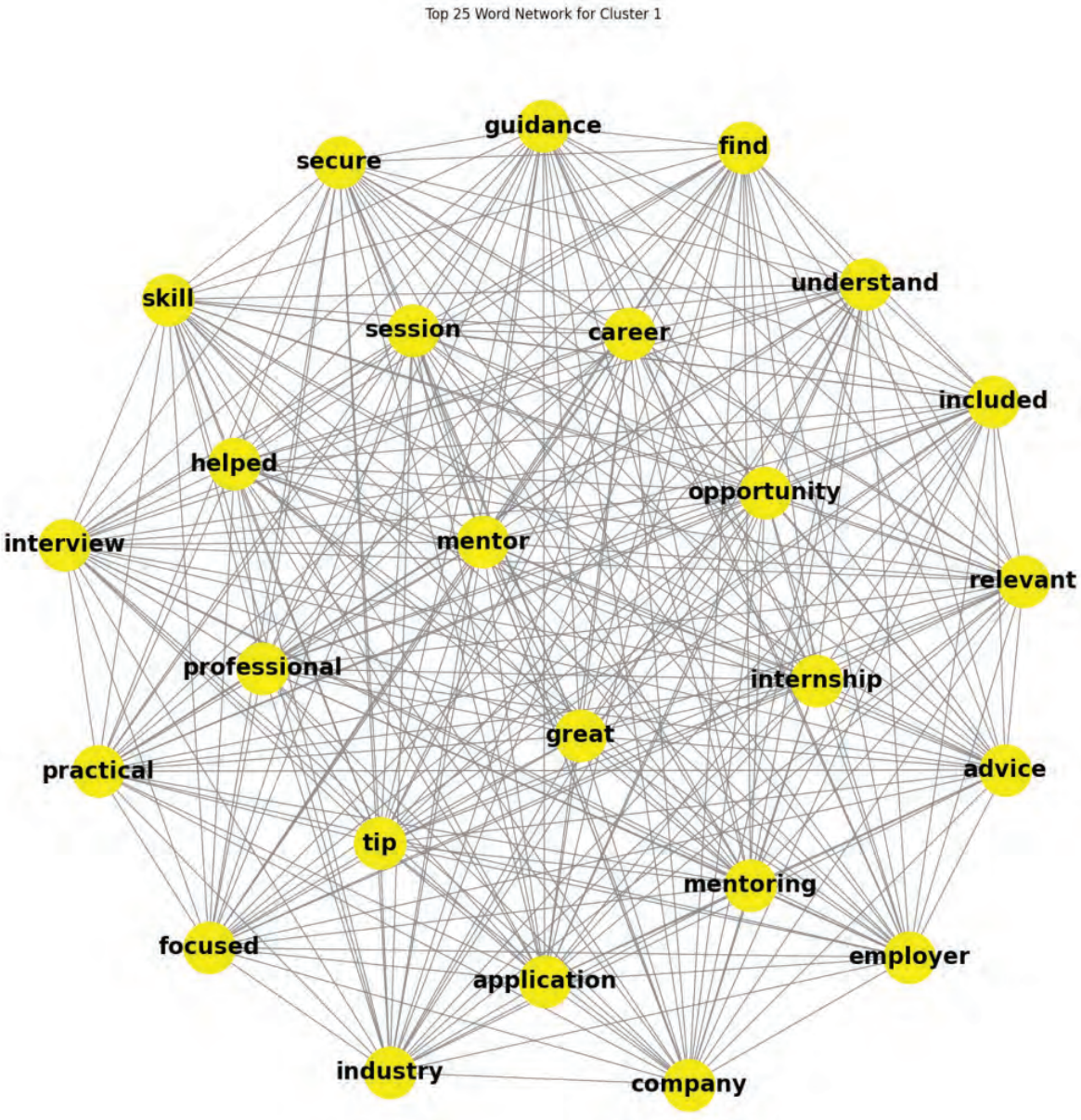


Figure 13. Word network for top 25 words of cluster 1.

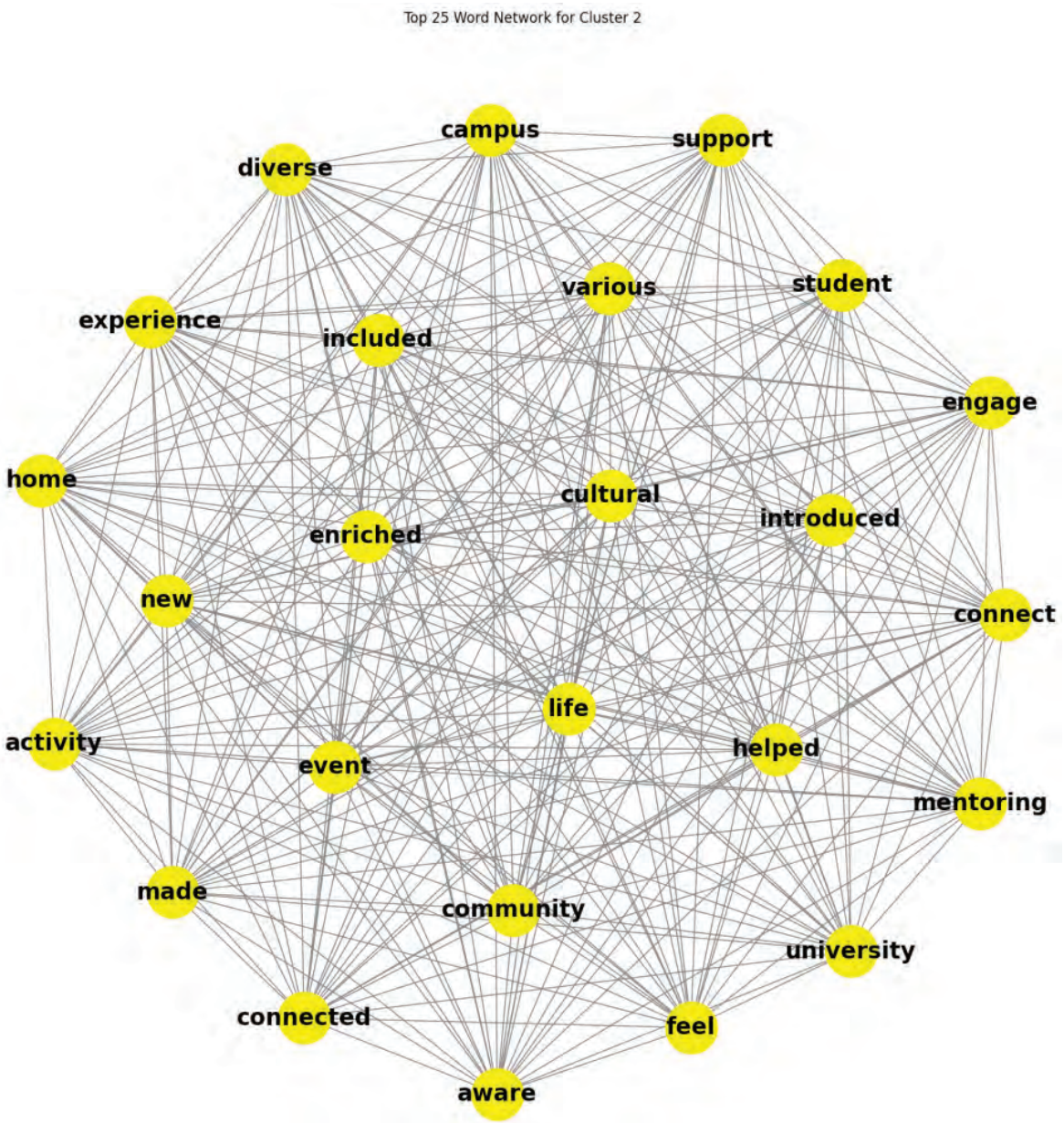


Figure 14. Word network for top 25 words of cluster 2.

Cluster 5: Scholarships and Applications

Key Words: Scholarship, mentoring, application, helped, advice, led, successful, clear, process, detailed

Cluster 6: Placement and Career Preparation

Key Words: Placement, mentor, mentoring, provided, job, interview, session, tip, confidence, resume

Hierarchical Clustering Analysis

And a closer examination of the feedback data via hierarchical clustering reveals six distinct clusters. The dendrogram, in Figure 19, illustrates how each response relates to others based on the similarity. At the point the connected lines cross the horizontal line - at a similarity value of 3 - that is where the clusters split. This divides the branches into six branches. Each grouping is distinct - based on different patterns. These divisions suggests six different themes in the feedback. Opinions and experience fall along these topics; not arbitrarily but in these forms. Rather than being amorphous, responses are drawn around concepts. Clusters form in nature. Six themes emerge from respondents' responses.

To begin with all the data points are individually grouped. Then similar groups merge, based on proximity (in straight-line distance). What you see at the top of the

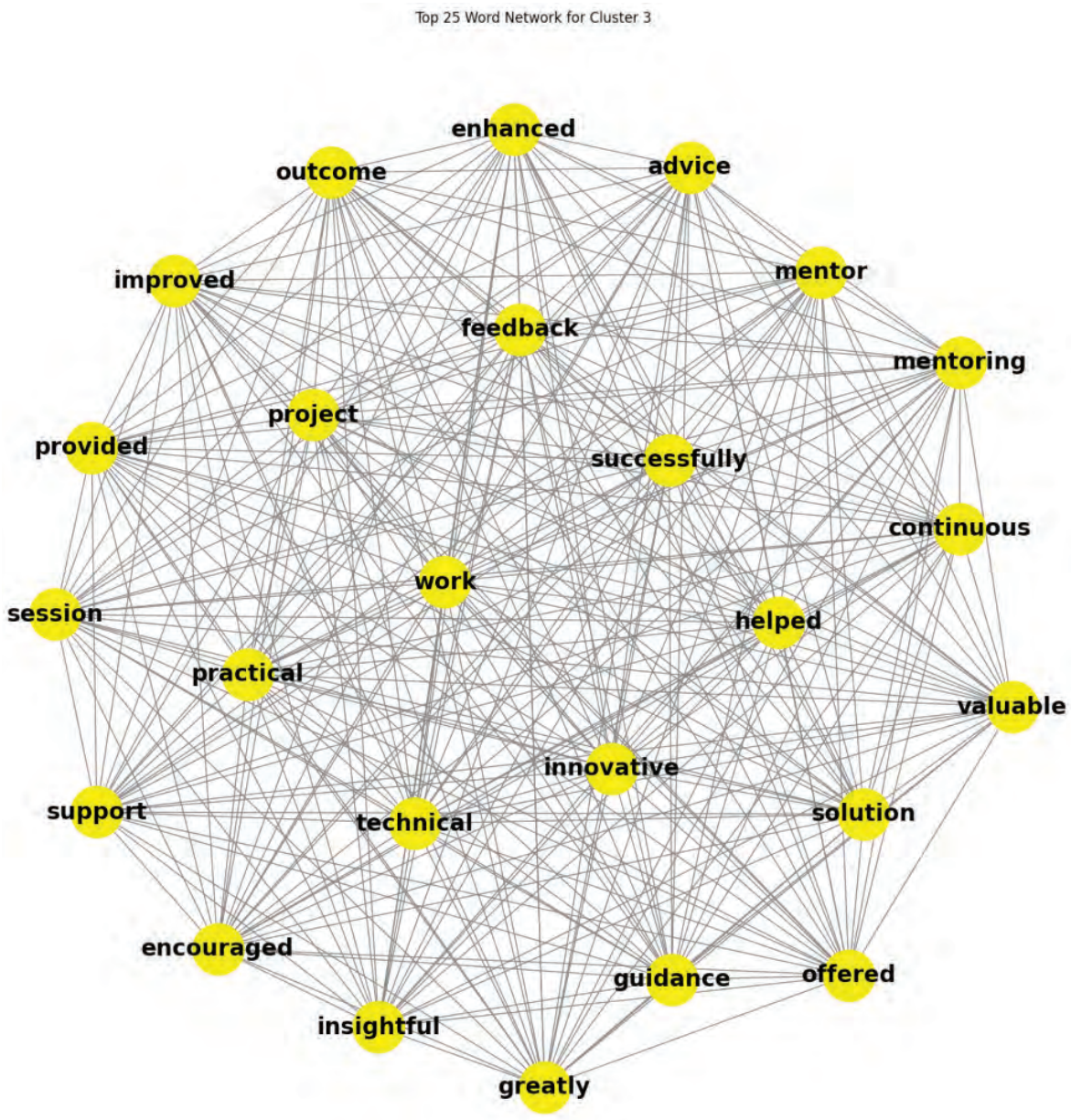


Figure 15. Word network for top 25 words of cluster 3.

branch diagram indicates how far apart the groups were to start with. When the distances between groups are larger than a certain threshold, they no longer join - the line here indicates where divisions are resolved into groups. But the silhouette score was 0.741 - less than K-means' 0.7985. With further investigation, feedback patterns were less consistent for members of different groups.

Table 1 shows the Optimum mentoring clusters. Table 2 shows the mentoring sessions and number of students benefited throughout the 4 years of engineering.

This approach works well when finding oddly shaped groups. Still, too many dots ended up labeled as outliers. Because of that, pieces scattered instead of forming whole units - silhouette stood at 0.689. Themes within each group felt looser than those from K-means splits.

Through 1756 internships, students gain exposure to linking academic work with the workplace. Their mentors support them when selecting placements, rehearsing their interview skills or getting used to the workplace. It provides access to build confidence and skills valued by employers. Practical assignments during these times make them ready for their future jobs. Knowledge is gained progressively in various business settings.

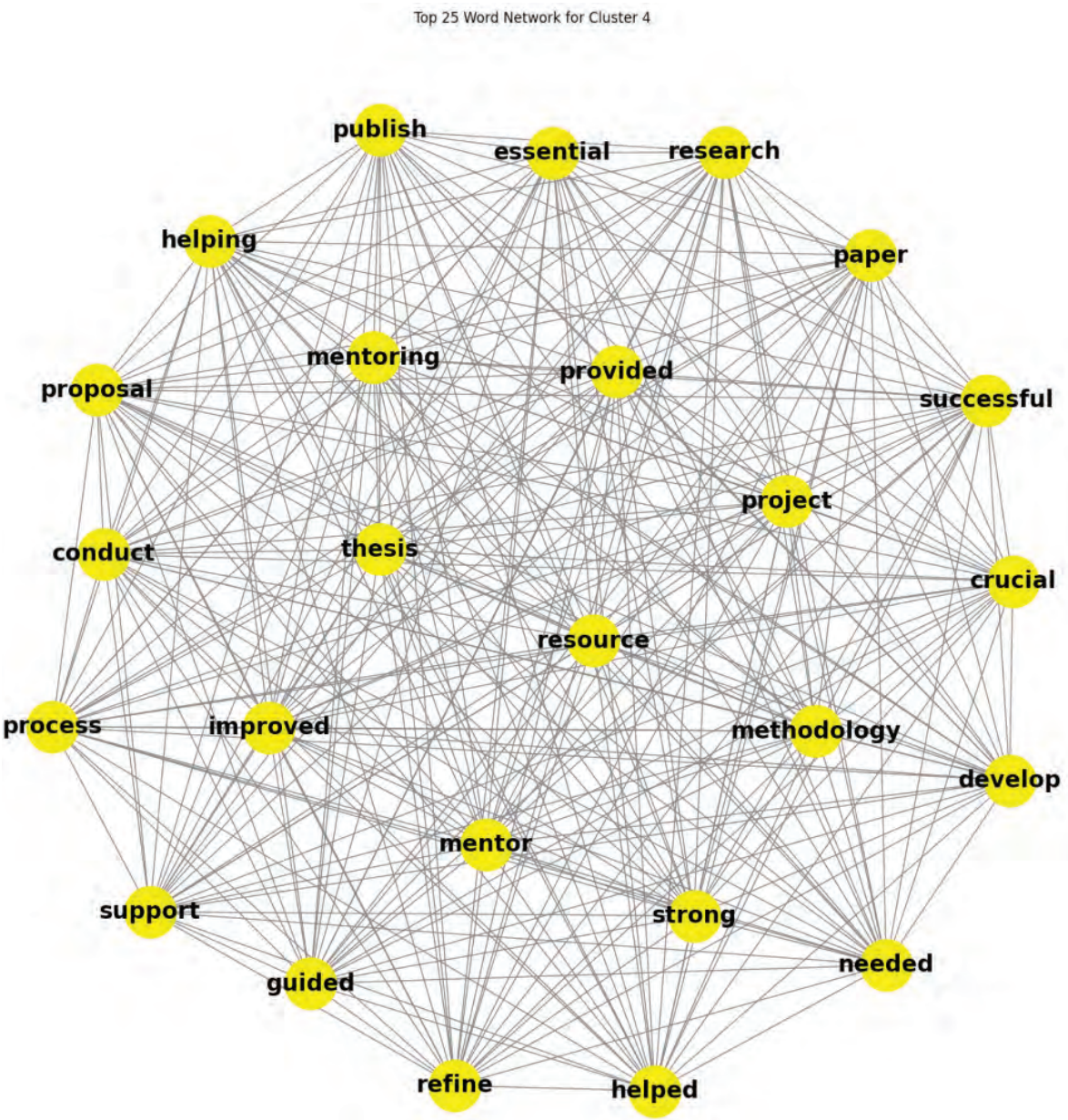


Figure 16. Word network for top 25 words of cluster 4.

Given 1127 students' placements, entry into employment role is a significant transition. Mentors help students to apply for jobs, and so influence how candidates present their capabilities. Some help with CVs; others, interview questions - all guided uniquely. With practice comes better performance. It is success that is low key - placement rates remain consistent. Counselling is adapted to the individual not standardized. Students' progress into work in their own time. Encouragement from input helps make working transitions smooth. Practice makes perfect, advice is advice. When a large number of students are placed, mentors are hard at work.

A smaller group, 563 learners have Higher Education Counseling. These get individuals into postgraduate training, grants and academics. Support is through mentors who advise on appropriate courses, support applications and keep time tables. This doesn't happen as frequently as other support. But such personal support is critical as learners advance to new learning settings.

A quarter of all the mentored learners - 259 - received scholarship support. Given funding relieves financial stress, this lets them focus even more on academics. The mentoring supports people to find suitable awards, develops applications and explains complicated processes. Awards'

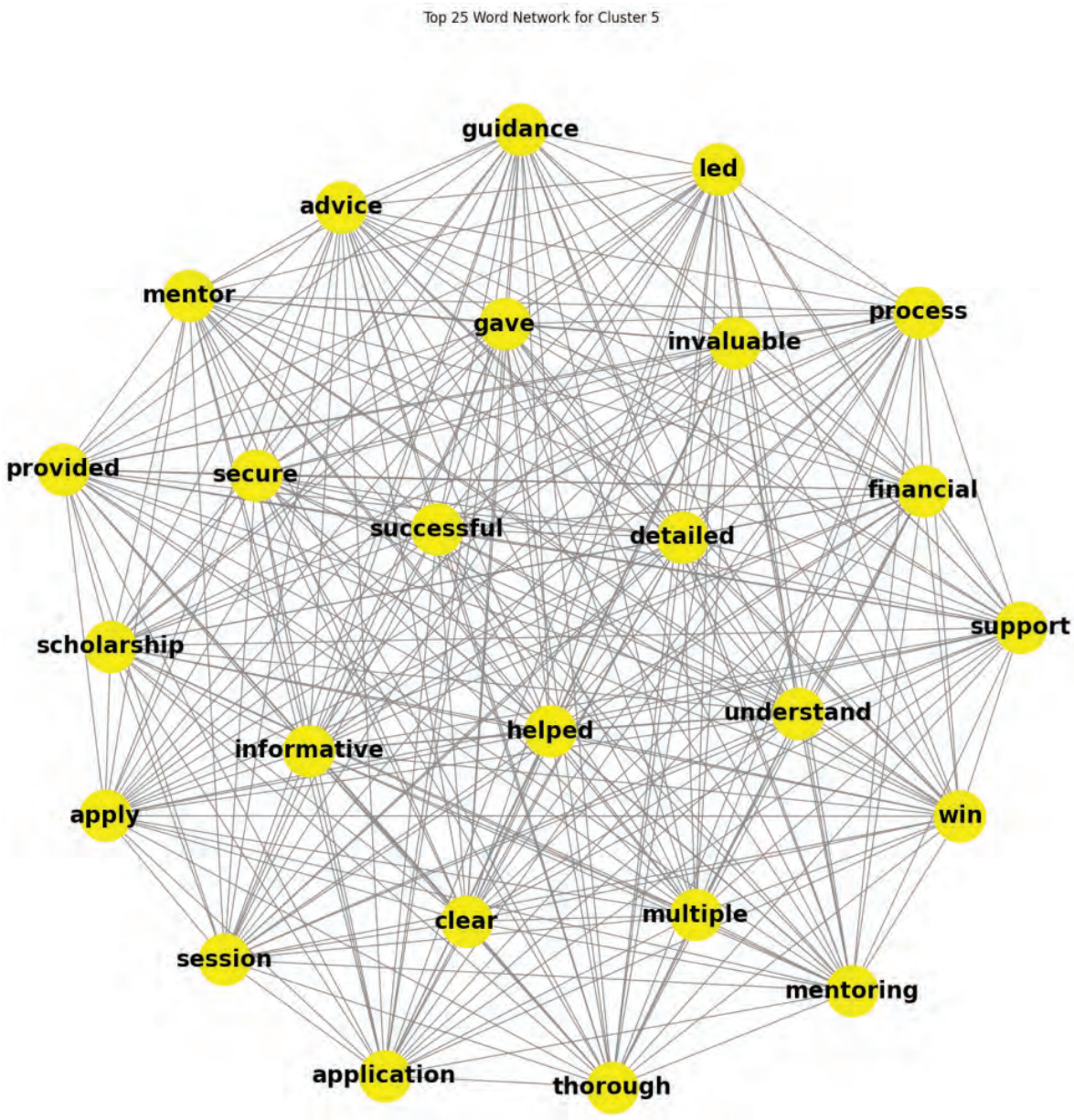


Figure 17. Word network for top 25 words of cluster 5.

success is covertly based on mentors' effectiveness in linking students to resources. Unsure of where to get funds, learning steads its pace.

A number of 35 students have published in prominent journals. With professors' help, these students investigate novel, worth-while questions. One reason they were successful was they received support early in the process. Setbacks provided opportunities to consult about study design and analyses. Others went through drafts before submitting their proposal. Despite modest participation, the results seem promising in contributing to academic productivity. Through such circuits, autonomy in learning to think is grown. The merit of tailored mentorship stands out in terms of its influence to academia.

Acquiring the mentoring experience, 254 students won awards in project competitions ranging from national to global. This demonstrates guidance building on new ideas, teamwork and problem-solving. Students work as a team with constant mentoring. As design and development takes place, mentors guide on how to improve the product. Public recognition demonstrates linking learning to practice. Through support, participants practice and refine skills to where they are tested. Success doesn't come easily

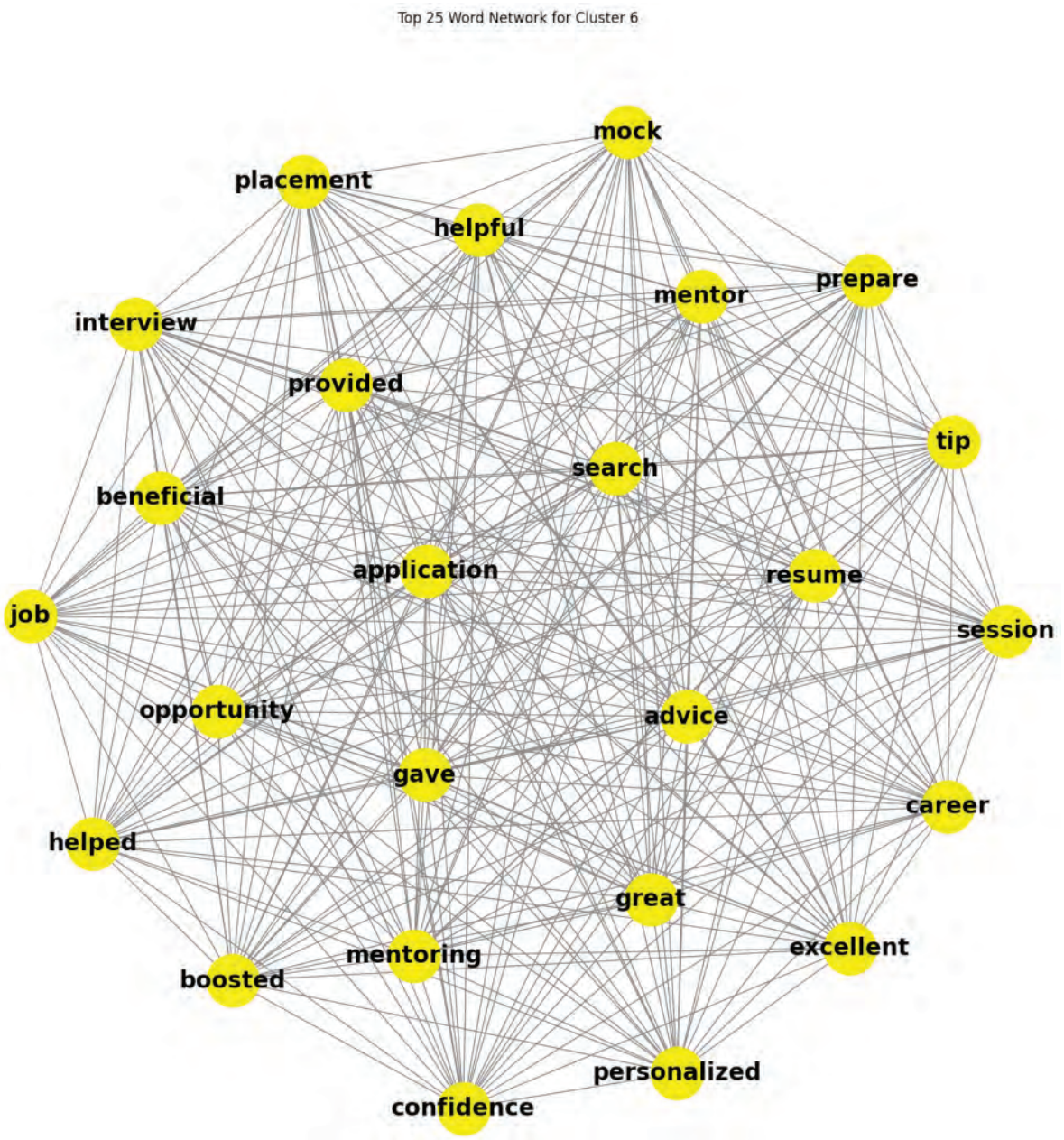


Figure 18. Word network for top 25 words of cluster 6.

but is the culmination of many days of trial and improvement. Globally, success like this demonstrates excellent preparation and confidence.

While effective, supporting students in engineering is tricky. Time and effort are critical, Akili noted in 2014 - guidance is most effective when undiluted, but not if time-limited [24]. Sometimes it's not smooth sailing: online advice altered situations during the pandemic, as Speer, Lyon and Johnson found later [25]. Remoteness offered flexibility, but eroded people's connectedness. What worked for some was "far away" to others Research on mentoring in engineering education shows its importance to student success. Mentoring from peers and/or teachers, and/or more real-life project work, leads to improved grades, increased student confidence and industry readiness, and retention, especially of disadvantaged students. Despite this, there are challenges with maintaining support that caters to all students' needs. Future research needs to explore new mentoring strategies, and their impacts on learning outcomes.

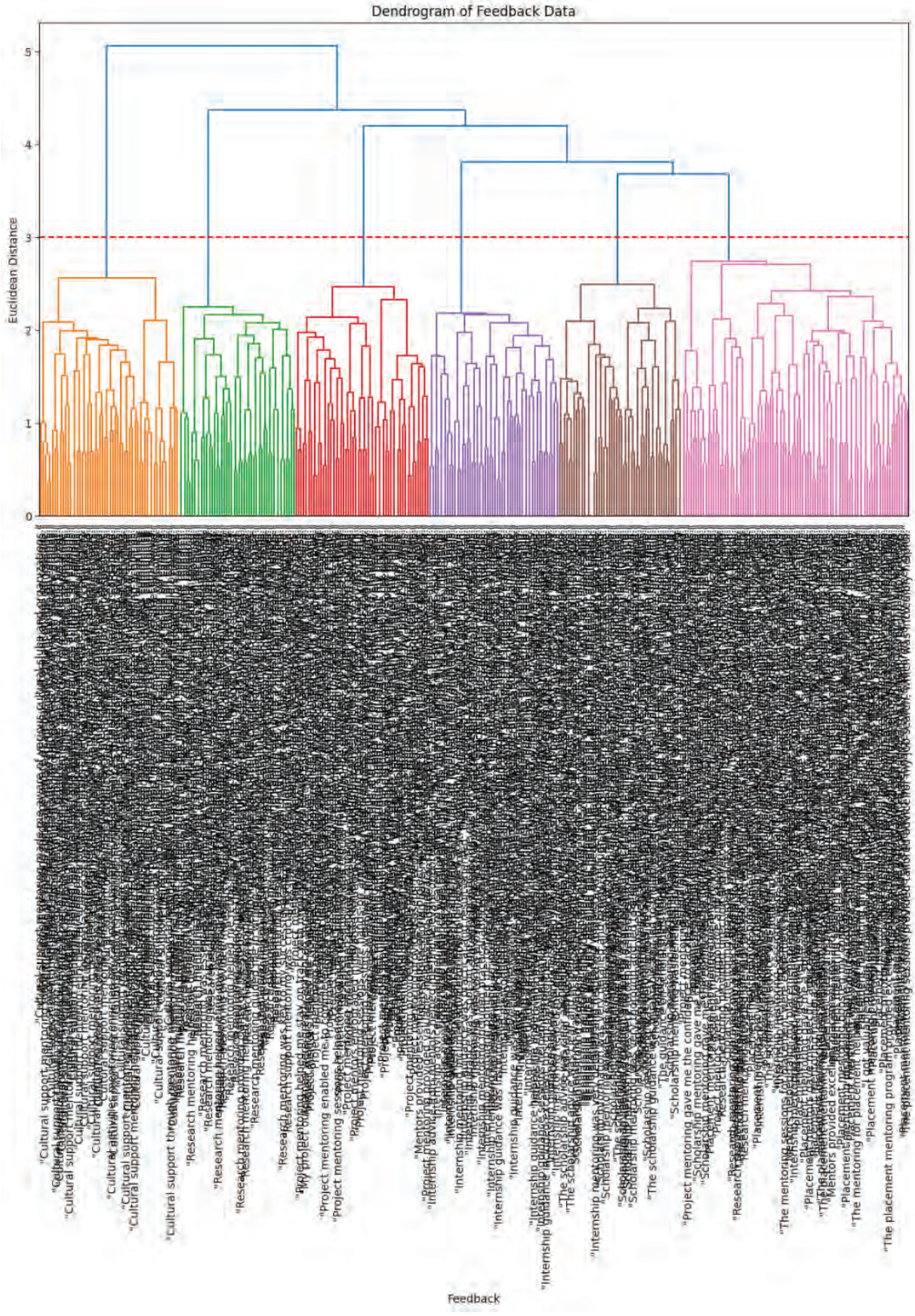


Figure 19. Dendogram for finding out optimum clusters.

Table 1. Optimum mentoring clusters

Sr.No.	Method	Evaluation Criteria	Number of optimum clusters
1	K Means	Davies-Bouldin Index, Calinski-Harabasz Index, and Within-Cluster Sum of Squares (WCSS).	6
2	DBSCAN	dendrogram	6
3	Hierarchical Clustering Analysis	dendrogram	6

Table 2. Mentoring sessions and number of students benefited

Sr. No.	Guidance through Mentoring Sessions	Number of students benefited
1	Internship	1756
2	Placement opportunities	1127
3	Higher education counseling	563
4	Cultural program award	786
5	Various scholarship schemes	259
6	Publications in reputed journals	352
7	Patents/ Copyrights	225
8	Participation in national/ international events	263
9	Awards in national/ international project competition other than academics	254

CONCLUSION

A six-cluster solution of the mentoring text responses was identified through K-means analysis revealing key themes. Supported by a silhouette score of 0.7985, and the Davies-Bouldin, Calinski-Harabasz and WCSS scores. The number of groups was varied - six was chosen because it demonstrated greater cohesion between members of the groups, and greater separations between groups. From the experience of mentoring, these clusters resonate with some of the important aspects known about the program. This provides some basis to choose this grouping. One approach to understanding mentoring is through these six groupings, which cover different factors such as skill development, employability skills, resources offered by the university, comments and feedback on work and study support as well as award advice. Beginning with word clustering and real feedback, clusters held together well - the different categories remained distinct, and made sense. There was the following sentiment analysis.

- Positive feedback: 84.56%
- Neutral feedback: 8.66%
- Negative feedback: 8.77%

There were positive feedbacks from most of the members, suggesting the mentoring program is effective. But a few identified problems that need to be addressed, particularly with regard to personalised support rather than "one size fits all".

Limitations and Future Directions:

We found results more or less consistent across methods. However, some ideas still end up in several clusters - not a failure, but an indication of the diversity of students' views.

A future direction is to try better grouping techniques such as topic modelling (LDA) or embedding (BERT). Using these, groups might better capture student experiences. Rather than a simple list, insights could emerge that reveal true differences in mentor interactions. And complexities in student experience may be highlighted with finer sorting of responses. These fine-tunes might encompass information ignored by more simple models. This can be achieved by fitting response characteristics to appropriate methods.

A further step up is to check a range of sentiment tools against methods of cross-validation - confirming these results. A composite view of mentoring feedback obtained through quantitative data, valid observations and sentiment provides greater clarity. The findings here suggest changes that have the potential to impact on mentoring. When strategies align to insights, programs have a greater chance of growing in responses to diverse student needs. Growth means more than form; it stems from support in line with individual journeys.

REFERENCES

[1] Peña-Martín JP, González-Parada E, García-Berdonés C, Pérez-Rodríguez EJ. Mentoring program for students newly enrolled in an engineering degree. Multidiscip J Educ Soc Technol Sci 2016;3:30-48. [CrossRef]

[2] Santhanamari G, Deepa M, Susithra N, Reba P. Establishing a constructive mentoring scheme for engineering students: A case study. J Eng Educ Transform 2022;35:303-311. [CrossRef]

[3] García-Laso A, Montalvo C, Martín DA, Presa L, Parra JL, Costafreda JL. Mentoring in the School of Mines and Energy to cover student needs: From the newly enrolled to the near-graduates. Sustainability 2022;14:3156. [CrossRef]

[4] Martin JP, Suresh DE, Jensen PA. Using collaborative autoethnography to investigate mentoring relationships for novice engineering education researchers. Int J STEM Educ 2024;11:13. [CrossRef]

[5] Agholor D, de Nalda ÁL, Bárcena NS. Mentoring future engineers in higher education: A descriptive study using a developed conceptual framework. Production 2017;27:e20162207. [CrossRef]

[6] Vesilind PA. Mentoring engineering students: Turning pebbles into diamonds. J Eng Educ 2001;90:407-411. [CrossRef]

[7] Dodson KH, Baugh D, Roland A, Edmonds S, York HP. The Peugeot Center model and mentoring explored through a case study of the design and installation of a potable water system in Guatemala with ADICAY. Adv Eng Educ 2022;10:1-30. [CrossRef]

[8] Aikens ML, Robertson MM, Sadselia S, Watkins K, Evans M, Runyon CR, et al. Race and gender differences in undergraduate research mentoring structures and research outcomes. CBE Life Sci Educ 2017;16. [CrossRef]

[9] Tuladhar A, Queener C, Mondisa JL, Okwudire C. Informal community spaces, mentoring and representation: Unpacking factors that influence African American engineering undergraduates. Int J Mentor Coach Educ 2021;10. [Epub ahead of print] doi: 10.1108/IJMCE-06-2020-0032 [CrossRef]

[10] Dennehy TC, Dasgupta N. Female peer mentors early in college increase women's positive academic experiences and retention in engineering. Proc Natl Acad Sci U S A 2017;114:5964-5969. [CrossRef]

[11] Budny D, Paul CA, Newborg BB. Impact of peer mentoring on freshmen engineering students. J STEM Educ Innov Res 2010;11:9-24.

[12] Stillmaker K, Oka LG, Plascencia JG, Schwartz-Doyle CC, Lor K. Investigating the role of faculty gender in mentoring female engineering students for success. In: ASEE Annual Conference and Exposition, Conference Proceedings. American Society for Engineering Education, 2020.

[13] Gehr RE, Garcia E, Swaine S, McBride S, Rispoli JV, Brinton CG. Assessment and support of advisor-student mentoring for graduate engineering students at a land-grant institution. In: ASEE Annual Conference and Exposition, Conference Proceedings. American Society for Engineering Education, 2023.

[14] Mendez SL, Tygret J, White JC, Gosha K. Developing deeper student mentoring relationships: Black engineering faculty translating their mentee experiences to students. In: ASEE Annual Conference and Exposition, Conference Proceedings. American Society for Engineering Education, 2022.

[15] Saw GK, Chang CN, Lin S, Hernandez PR, Culbertson R. Disparities in mentoring and mental health problems of U.S. college students in science and engineering during the COVID-19 pandemic. Educ Psychol 2023;43:1-22. [CrossRef]

[16] Rayford T, et al. Educational enrichment: The benefits of near-peer mentoring for undergraduate engineering students. In: ASEE Annual Conference and Exposition, Conference Proceedings. 2022.

[17] Mu W, Lim KH, Liu J, Karunasekera S, Falzon L, Harwood A. A clustering-based topic model using word networks and word embeddings. J Big Data 2022;9. [CrossRef]

[18] Arafa A, El-Fishawy N, Badawy M, Radad M. RN-SMOTE: Reduced noise SMOTE based on DBSCAN for enhancing imbalanced data classification. J King Saud Univ Comput Inf Sci 2022;34. [CrossRef]

[19] Silva KD, Jönsson D, Demmer RT. A combined strategy of feature selection and machine learning to identify predictors of prediabetes. J Am Med Inform Assoc 2020;27. [CrossRef]

[20] Bhardwaj AS, Deep A, Veeramani D, Zhou S. A custom word embedding model for clustering of maintenance records. IEEE Trans Ind Inform 2022;18. [CrossRef]

[21] Irawaty I, Andreswari R, Pramesti D. Vectorizer comparison for sentiment analysis on social media YouTube: A case study. In: 2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020. 2020. [CrossRef]

[22] Marczyk VR, Recamonde-Mendoza M, Maia AL, Goemann IM. Classification of thyroid tumors based on DNA methylation patterns. Thyroid 2023;33. [CrossRef]

[23] Akili W. Mentoring engineering students: Challenges and potential rewards. In: ASEE Annual Conference and Exposition, Conference Proceedings. 2014.

[24] Speer JE, Lyon M, Johnson J. Gains and losses in virtual mentorship: A descriptive case study of undergraduate mentees and graduate mentors in STEM research during the COVID-19 pandemic. CBE Life Sci Educ 2021;20. [CrossRef]

[25] Berna AG, Gizem G, Nuri Can B. Turkish sentiment analysis: A comprehensive review. Sigma J Eng Nat Sci 2024;42.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.000135

Research Article

Enhancing dynamic performance of turbine-generator systems with optimal SVC control using hybrid CSA-GWO technique in series-compensated networks

Sangeeta BAMBA[1,*], SK GUPTA[1]

[1]*Department of Electrical Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Sonipat, Haryana, 131001, India*

ARTICLE INFO

Article history
Received: 22 November 2024
Revised: 15 January 2025
Accepted: 19 February 2025

Keywords:
Hybrid Cuckoo Search Algorithm-Gray Wolf Optimization (CSA-GWO) Technique; Static Var Compensator (SVC); Sub Synchronous Resonance (SSR)

ABSTRACT

This paper presents a comprehensive study on enhancing the dynamic performance of a turbine-generator set connected to an infinite busbar, with a primary focus on mitigating sub-synchronous resonance in a series-compensated transmission network. SSR is a critical phenomenon that threatens turbine-generator systems by inducing harmful oscillations in generator rotor shafts, potentially leading to severe damage. To address this issue, the study employs the IEEE first benchmark model, without natural damping, and proposes a robust solution involving a Static Var Compensator coupled with an optimal controller optimized through a hybrid Cuckoo Search Algorithm-Gray Wolf Optimization approach, benchmarked against the Bacterial Foraging Optimization Algorithm. The study considers zero inherent damping to simulate the most severe conditions, while natural damping is included for comparative analysis. Eigenvalue analysis and MATLAB Power System Blockset simulations demonstrate significant improvements in system damping characteristics, with eigenvalues shifting towards the left half-plane and a notable reduction in peak rotor shaft oscillations compared to BFOA-based control. Time-domain analysis confirms the system›s effectiveness, with the hybrid CSA-GWO technique achieving superior damping, minimal peak time variation, and the best improvement in settling time and overshoot, especially at higher series compensation levels. The study concludes that the hybrid CSA-GWO-based controller not only enhances the system›s response to disturbances but also outperforms over BFOA Algorithm, offering faster, more effective mitigation of SSR and significantly improving the dynamic performance of the turbine-generator system.

Cite this article as: Bamba S, Gupta SK. Enhancing dynamic performance of turbine-generator systems with optimal SVC control using hybrid CSA-GWO technique in series-compensated networks. Sigma J Eng Nat Sci 2026;44(2):1419–1439.

***Corresponding author.**
*E-mail address: 20001902005sangeetabamba@dcrustm.org

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

The rapid expansion of power systems and the integration of renewable energy sources have introduced new complexities into grid dynamics, particularly in mitigating adverse effects such as sub-synchronous resonance (SSR). SSR is an electrical resonance condition that occurs when the natural frequency of mechanical oscillations in turbine-generators coincides with sub-synchronous frequencies in the electrical network. This phenomenon can lead to significant system instability and potentially catastrophic damage to turbine-generator shafts. Therefore, effective SSR mitigation strategies have become crucial in maintaining power system stability and preventing mechanical failures.

SSR was first identified in the 1970s with the integration of series capacitors in transmission lines, a practice aimed at enhancing power transfer capability [1]. When mechanical and electrical resonant frequencies align, turbine-generator shafts experience sustained oscillations, resulting in severe mechanical stresses[2]. Notable incidents like the Navajo and Mohave projects[3] highlighted the destructive potential of SSR in real-world applications [4]. Over time, various mitigation strategies have been proposed, with substantial advancements in technology and research contributing to improved system stability.

One of the most notable innovations in SSR mitigation is the introduction of Flexible AC Transmission Systems (FACTS). FACTS devices such as Static VAR Compensators (SVC) and Static Synchronous Compensators (STATCOM) offer advanced solutions to enhance grid stability and damp power oscillations, including SSR. Padiyar [5] emphasized the role of FACTS controllers in addressing grid stability issues. However, the increasing complexity of modern grids, especially with the growing penetration of renewable energy sources, has created a demand for more sophisticated approaches that combine traditional methods with optimization algorithms and advanced control techniques.

The advent of FACTS technology revolutionized SSR mitigation by offering dynamic solutions for power system stability. FACTS devices like SVCs were initially explored by Narendra Kumar and Dave [6]for damping SSR, demonstrating their effectiveness in enhancing system stability. Similarly, Padiyar and Prabhu[7] investigated the use of STATCOM for SSR damping, further demonstrating the potential of FACTS devices in managing grid dynamics.

As renewable energy sources become increasingly integrated into power grids, new challenges have emerged in SSR mitigation. Recent studies have explored the interaction between renewable energy systems and SSR. For instance, Prada et al. [8] examined the role of Type-2 wind turbines in damping SSR, study [9,10] proposed intelligent STATCOM control strategies to enhance stability during torsional oscillations in systems with high renewable penetration.

Despite their effectiveness, traditional FACTS controllers rely on fixed control parameters, which limits their adaptability to dynamic grid conditions. As power systems evolve, the increasing penetration of renewable energy sources[11] introduces unpredictable variations, requiring advanced control strategies that enhance the responsiveness and efficiency of SSR mitigation[12]

To overcome the limitations of conventional SSR mitigation strategies, researchers have increasingly turned to soft computing and optimization techniques. Metaheuristic algorithms, such as Cuckoo Search Optimization (CSO) [8], Grey Wolf Optimizer (GWO) [13], and Quantum-behaved Particle Swarm Optimization (QPSO) [14], have been successfully applied to enhance FACTS-based damping controllers. These algorithms offer improved controller tuning, real-time adaptation, and enhanced SSR suppression [15].

Additionally, hybrid optimization approaches combining multiple algorithms—such as Cuckoo Search Algorithm with Grey Wolf Optimizer (CSA-GWO)—have demonstrated superior performance in mitigating SSR. These techniques leverage the strengths of individual algorithms, optimizing STATCOM and SVC controllers to effectively damp sub-synchronous oscillations. Recent studies, including Bamba and Gupta [11], have explored intelligent STATCOM control strategies for enhancing system stability, particularly in grids with high renewable penetration.

Several challenges complicate SSR analysis and mitigation in modern power systems. First, the increased complexity of grid dynamics, driven by renewable energy integration and enhanced interconnectivity, makes it difficult to predict SSR behavior accurately. Second, real-time optimization of SSR mitigation strategies requires significant computational power and efficient algorithms capable of adapting to changing grid conditions. Finally, ensuring the interoperability of FACTS devices and SSR damping controllers [16] with various grid components, including renewable energy sources, is essential to maintaining overall system stability.

This study aims to bridge the research gap by proposing an advanced FACTS-based SSR mitigation strategy, leveraging hybrid optimization techniques to improve controller adaptability, dynamic stability, and overall power system resilience.

This paper aims to address these challenges by proposing advanced SSR mitigation strategies that combine FACTS technology with optimization algorithms and machine learning techniques, offering a comprehensive solution for modern power grids. Following reference, Table 1 provides a comparative analysis between the proposed work and the existing cutting-edge research.

Table 1. Comparison of SSR mitigation strategies and control approaches

Reference	System Type	Compensation Method	Controller Gain Selection	Key Points
[2][**1990**]	Series-Compensated Systems	SSR damping with series compensation	Not specified	Foundational study on SSR and its impact on turbine-generator interactions.
[3,17], [**1997**]	Navajo Project	SSR analysis and mitigation	Not specified	Detailed mitigation strategies from the Navajo project.
[17],[**2002**]	General Power Systems	SVC-based damping	Optimized using auxiliary control	Explored SSR damping through SVC and auxiliary control mechanisms.
[7], [**2006**]	Series-Compensated Systems	STATCOM for SSR damping	Optimized for system stability	Demonstrated the effectiveness of STATCOM in mitigating SSR.
[10],[**2023**]	Multi-source Systems	Cuckoo Search Optimization for AGC	Optimized using Cuckoo Search	Applied modern optimization techniques for improved SSR damping.
[13], [**2023**]	Series-Compensated Power Systems	GWO-based control for SSR mitigation	Optimized using Grey Wolf Optimizer	Applied modern optimization techniques for improved SSR damping.
[9], [**2024**]	Series-Compensated Power Systems		STATCOM-based PI Controller.	Damp torsional oscillations.

Novelty and its Main Contribution

This paper aims to contribute to the field by:

- The paper proposes the use of a Static Var Compensator (SVC) as an effective device to damp electromechanical oscillations in turbine-generator systems. SVC improves the system›s voltage stability and enhances its overall response to dynamic disturbances.
- In addition to this, study presents an optimal control design for SVC using the CSA-GWO technique, which uses all the system states as feedback to stabilize the output to handle higher levels of series compensation. The hybrid optimization approach leads to a controller design that minimizes SSR risks while enhancing the dynamic performance of the turbine-generator system.
- The study introduces a hybrid optimization approach that combines the strengths of Cuckoo Search Algorithm (CSA) and Gray Wolf Optimization (GWO) to optimize the control parameters of optimal SVC in mitigating SSR.
- The proposed CSA-GWO-based controller is benchmarked against the Bacterial Foraging Optimization Algorithm (BFOA), demonstrating superior performance in terms of convergence speed and system stability.
- The effectiveness of the proposed solution is validated through eigenvalue analysis and MATLAB simulations, providing a detailed comparison of system performance under different levels of damping i.e with and without natural damping.

The manuscript is organized as follows: Section 2 presents the system overview, including the IEEE First Benchmark Model. Section 3 outlines the mathematical model, while Section 4 explores optimal control theory. Section 5 introduces the Hybrid CSA-GWO technique for enhanced system performance. Finally, the Results and Discussion Section 6 provides an in-depth analysis of the findings.

SYSTEM OVERVIEW

The system under consideration consists of several key components. First, it includes a turbine-generator system that is represented by the IEEE first benchmark model, which is commonly used for studying Sub-Synchronous Resonance (SSR) and its effects. Additionally, the system features a series-compensated transmission network, which enhances power transfer capability but also introduces the risk of SSR by potentially exciting sub synchronous oscillations. To address these challenges, a Static Var Compensator (SVC) is employed for reactive power compensation and voltage control, helping to stabilize the system under varying operating conditions. Furthermore, an optimal controller is designed using a hybrid CSA-GWO (Cuckoo Search Algorithm - Gray Wolf Optimization) technique, which enhances overall system stability by effectively damping SSR and improving dynamic performance.

IEEE First Benchmark Model

The IEEE first benchmark model, widely used for SSR studies, comprises several key components as shown in Figure 1. It includes a three-phase, salient-pole synchronous generator, and a mechanical system that features the generator rotor, low-pressure turbine, high-pressure turbine, and

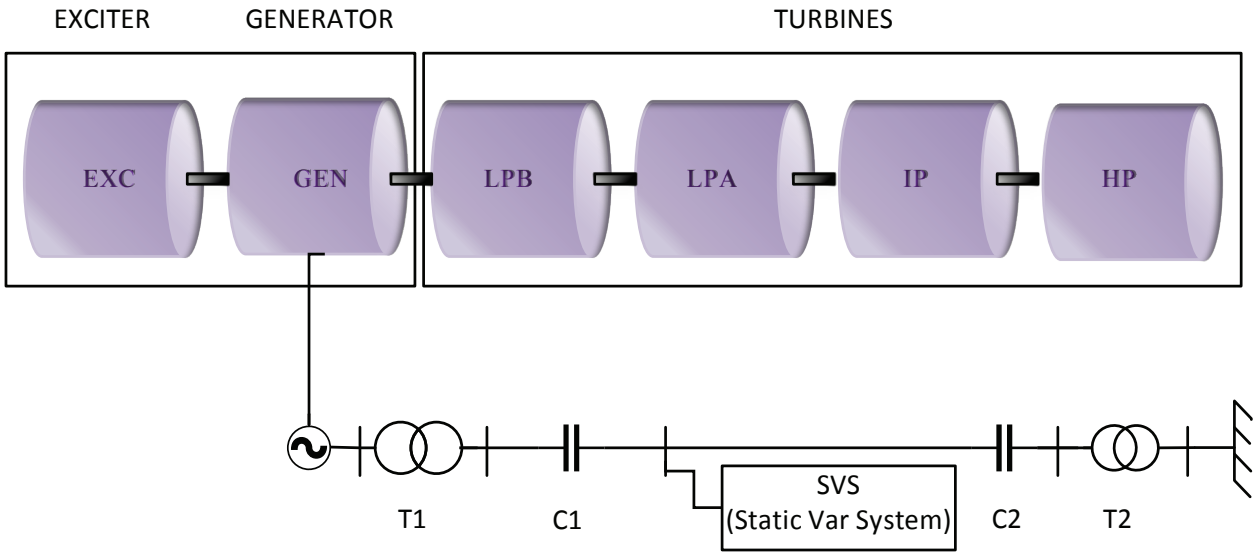


Figure 1. Modified IEEE first benchmark model with svc.

exciter, all modeled as multi-mass shafts. Additionally, the model incorporates an excitation system that regulates terminal voltage and reactive power using a voltage reference signal. Furthermore, it features a series-compensated network that introduces capacitive compensation to enhance power transfer, which can excite SSR modes.

MATHEMATICAL MODELING

Generator

A generalized system model for a synchronous machine, as depicted in Figure 2, was considered. To address the challenges associated with dynamic saliency, the concept of dummy coils is utilized, as discussed by [5]. The generator model comprises a 'field winding, f' with a 'damper winding, h' along the d-axis, and 'two damper windings, g and k' along the q-axis[18].

The rotor flux linkages, denoted as 'Ψ', associated with various windings, lead to the following rotor equations:

$$\frac{d\psi_f}{dt} = a_1.\psi_f + a_2\psi_h + b_1 v_f + b_2 i_d \tag{1}$$

$$\frac{d\psi_f}{dt} = a_3.\psi_f + a_4\psi_h + b_3 i_d \tag{2}$$

$$\frac{d\psi_f}{dt} = a_5.\psi_g + a_6\psi_k + b_4 v_g + b_5 i_q \tag{3}$$

$$\frac{d\psi_f}{dt} = a_7.\psi_g + a_8\psi_k + b_6 i_q \tag{4}$$

Here, (V_f) represents the field excitation voltage. These equations are linearized and transformed into the D-Q frame of reference, as described by Ramsaw and Padiyar (1973). The resulting state space model is as follows:

$$\begin{aligned}
&\dot{X}_R = [A_R]X_R + [B_{R1}]U_{R1} + B_{R2}[U_{R2}] + [B_{R3}]U_{R3} \\
&X_R = \begin{bmatrix}\Delta\psi_f & \Delta\psi_f & \Delta\psi_f & \Delta\psi_k\end{bmatrix}^t \\
&U_{R1} = \begin{bmatrix}\Delta\delta & \Delta\omega\end{bmatrix}^t \\
&U_{R2} = \begin{bmatrix}\Delta V_F\end{bmatrix} \\
&U_{R3} = \begin{bmatrix}\Delta i_D & \Delta i_Q\end{bmatrix}^t \\
&Y_{R1} = [C_{R1}]X_R + [D_{R1}]U_{R1} \\
&Y_{R2} = [C_{R2}]X_R + [D_{R2}]U_{R2} + [D_{R3}]U_{R3} + [D_{R3}]U_{R3} \\
&Y_{R1} = \begin{bmatrix}\Delta I_D \\ \Delta I_Q\end{bmatrix}, Y_{R1} = \begin{bmatrix}\Delta I_D \\ \Delta I_Q\end{bmatrix}
\end{aligned} \tag{5}$$

Mechanical System

The mechanical system, illustrated in Figure 3, is represented by the spring-mass model. The fundamental equations governing this system, along with the state and output equations, are detailed as follows:

$$\frac{d\omega_1}{dt} = \left[\frac{-(D_{11}+D_{12})(\omega_1-\omega_0) + D_{12}(\omega_2-\omega_0) - K_{12}(\delta_1-\delta_2) + T_{M1}}{M_1}\right] \tag{6}$$

$$\frac{d\omega_2}{dt} = \left[\frac{D_{12}(\omega_1-\omega_0) - (D_{12}+D_{22}+D_{23})(\omega_2-\omega_0) + D_{23}(\omega_3-\omega_0) - K_{12}(\delta_2-\delta_1) - K_{23}(\delta_2-\delta_3) + T_{M2}}{M_2}\right] \tag{7}$$

$$\frac{d\omega_3}{dt} = \left[\frac{D_{23}(\omega_2-\omega_0) - (D_{23}+D_{33}+D_{34})(\omega_3-\omega_0) + D_{34}(\omega_4-\omega_0) - K_{23}(\delta_3-\delta_2) - K_{34}(\delta_3-\delta_4) + T_{M3}}{M_3}\right] \tag{8}$$

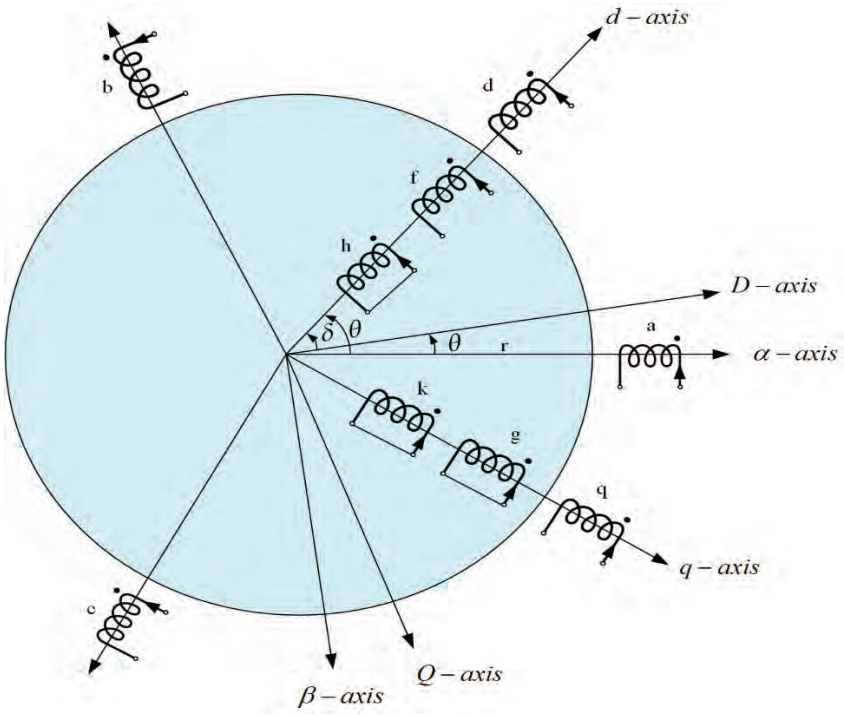


Figure 2. Generator model with four damping winding.

$$\frac{d\omega_4}{dt}=\left[\frac{D_{34}(\omega_3-\omega_0)-(D_{34}+D_{44}+D_{45})(\omega_4-\omega_0)+D_{45}(\omega_5-\omega_0)-K_{34}(\delta_4-\delta_3)-K_{45}(\delta_4-\delta_5)+T_{M4}}{M_4}\right] \quad (9)$$

$$\frac{d\omega_5}{dt}=\left[\frac{D_{45}(\omega_4-\omega_5)-(D_{56}+D_{45}+D_{55})(\omega_5-\omega_0)+D_{56}(\omega_6-\omega_0)-K_{45}(\delta_5-\delta_4)-K_{56}(\delta_5-\delta_6)+T_E}{M_5}\right] \quad (10)$$

$$\frac{d\omega_6}{dt}=\left[\frac{-(D_{56}+D_{66})(\omega_6-\omega_0)+D_{56}(\omega_5-\omega_0)-K_{56}(\delta_6-\delta_5)}{M_6}\right] \quad (11)$$

$$\omega_i=\frac{d\delta_i}{dt}\text{ , Where i= 1 to 6}$$

$$T_E=-X_D{}''\left(i_D I_Q-i_Q I_D\right) \quad (12)$$

In Figure 3, the variables (i) and (i) represent the angular position and velocity of the i[th] mass of various shaft segments, respectively, where (i) ranges from 1 to 6.

By linearizing these equations, we derive the state space model as follows:

$$\Delta T_{M1}=\Delta T_{M2}=\Delta T_{M3}=\Delta T_{M4}=0$$

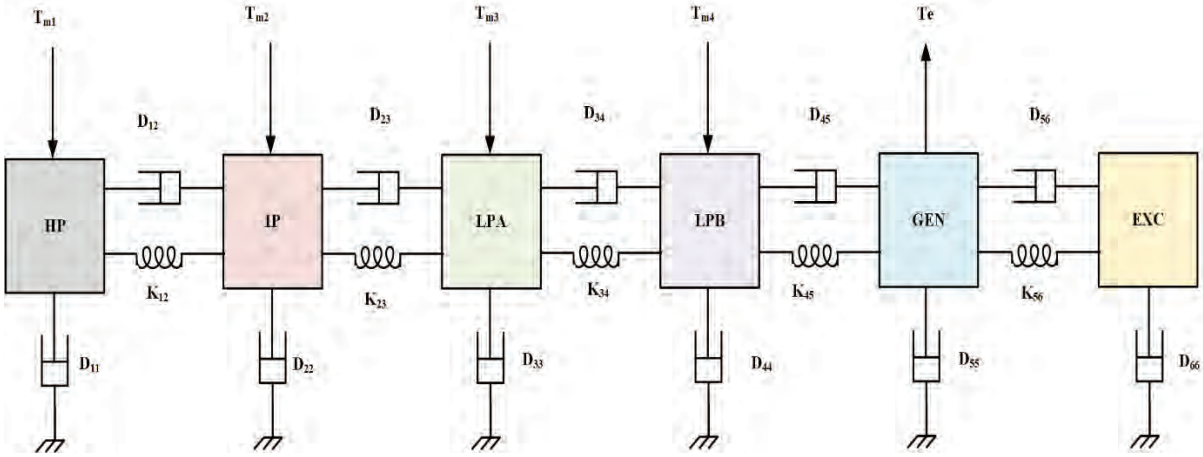


Figure 3. Six spring mass model of mechanical system.

Given that the governor system operates with significant time constants

$$\dot{X}_m = [A_M]X_M + [B_{M1}]U_{M1} + [B_{M2}]U_{M2} \quad (13)$$

$$Y_M = [C_M]X_M \quad (14)$$

Where, $X_M = [\Delta\delta_1 \quad \Delta\delta_2 \quad \Delta\delta_3 \quad \Delta\delta_4 \quad \Delta\delta_5 \quad \Delta\delta_6 \quad \Delta\omega_1 \quad \Delta\omega_2 \quad \Delta\omega_3 \quad \Delta\omega_4 \quad \Delta\omega_5 \quad \Delta\omega_6]$

$$Y_M = [\Delta\delta_5 \quad \Delta\omega_5]^t$$

$$U_{M1} = [\Delta I_D \quad \Delta I_Q]^t$$

$$U_{M2} = [\Delta i_D \quad \Delta i_Q]^t$$

Excitation System

The IEEE type-1 model, depicted in Figure 4, illustrates the excitation system. In this model, (Vg) denotes the generator terminal voltage, while (S_E) represents the saturation function. The excitation system is characterized by the following set of equations:

$$\frac{dv_F}{dt} = \left[\frac{-(K_E + S_E)}{T_E}\right]v_F + \frac{v_R}{T_E} \quad (15)$$

$$\frac{dv_s}{dt} = -\left[\frac{K_F(K_E + S_E)}{(T_E T_F)}\right]v_F + \frac{v_s}{T_E} + \frac{K_F v_R}{T_E T_F} \quad (16)$$

$$\frac{dv_R}{dt} = \frac{K_A V_{ref}}{T_A} - \frac{K_A v_s}{T_A} - \frac{K_A v_G}{T_A} - \frac{v_R}{T_A} \quad (17)$$

$$S_E = A_E . e^{(B_E V_F)} \quad (18)$$

$$\dot{X}_E = [A_E]X_E + [B_E]U_E \quad (19)$$

$$Y_E = [C_E]X_E \quad (20)$$

The resulting linearized state and output equations are as follows:

$$X_E = [\Delta v_F \quad \Delta v_S \quad \Delta v_R]^t \quad (21)$$

$$Y_E = [\Delta v_F] \quad (22)$$

$$U_E = [\Delta V_g]^t \quad (23)$$

Network

The transmission line modeled using lumped parameters in a π-circuit configuration. The network is depicted by its α-axis equivalent circuit, which is identical to the positive sequence network. The governing equations for the α-axis and π-network representation are derived as follows:

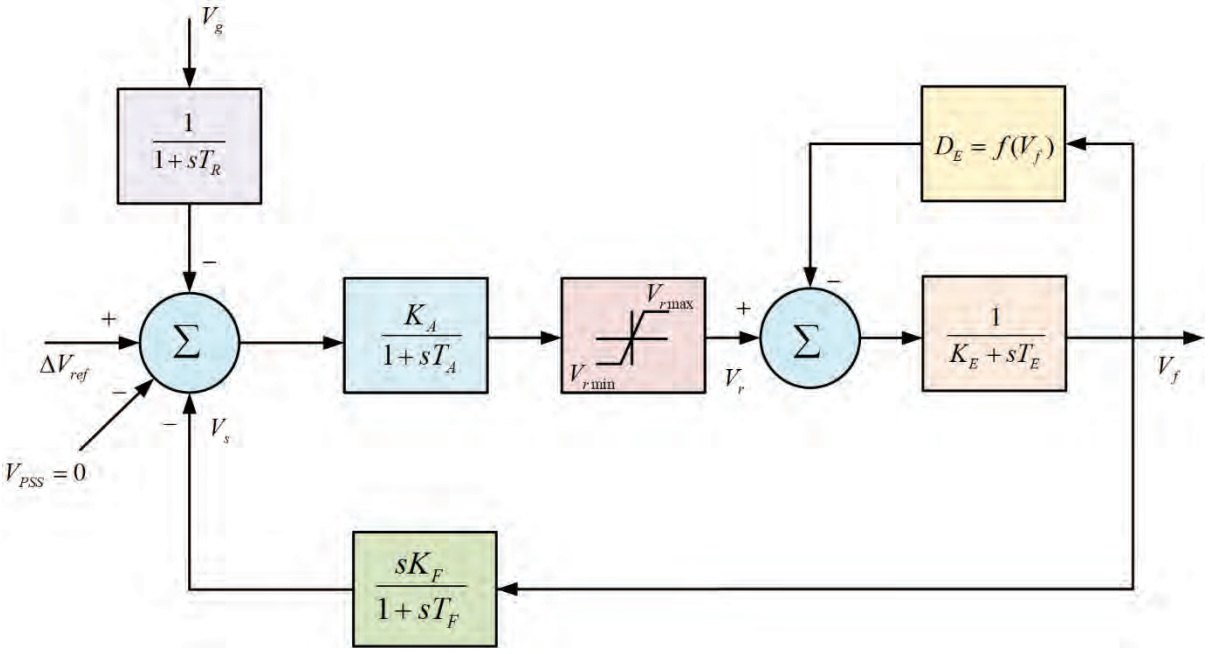


Figure 4. IEEE type-1 excitation system.

$$\frac{di_{1a}}{dt} = \frac{v_{2\alpha}}{L_{T2}} - \frac{v_{1\alpha}}{L_{T2}} - \frac{v_{3\alpha}}{L_{T2}} \quad (24)$$

$$\frac{di_{2a}}{dt} = -i_{2\alpha}\left[\frac{R_r}{L_r}\right] - \frac{v_{2\alpha}}{L_r} - \frac{v_{3\alpha}}{L_r} \quad (25)$$

$$\frac{di_{4a}}{dt} = -i_{4\alpha}\left[\frac{R_s}{L_s}\right] - \frac{v_{3\alpha}}{L_s} - \frac{v_{4\alpha}}{L_s} \quad (26)$$

$$\frac{di_a}{dt} = -\frac{R_a i_\alpha}{L_A} - \frac{v_{4\alpha}}{L_A} - \left(\frac{L''_d}{L_A}\right)\left[\frac{dI_\alpha}{dt}\right] - \frac{v_{6\alpha}}{L_A} \quad (27)$$

$$\frac{dv_{2\alpha}}{dt} = -\left(\frac{1}{C_r}\right)i_{2\alpha} - \left(\frac{1}{C_r}\right)i_{1\alpha} \quad (28)$$

$$\frac{dv_{3\alpha}}{dt} = -\frac{i_{2\alpha}}{C_n} - \frac{i_{3\alpha}}{C_n} + \frac{i_{4\alpha}}{C_n} \quad (29)$$

$$\frac{dv_{4\alpha}}{dt} = -\frac{i_\alpha}{C} - \frac{i_{4\alpha}}{C_n} \quad (30)$$

$$\frac{dv_{5\alpha}}{dt} = -\frac{i_{1\alpha}}{C_{se2}} \quad (31)$$

$$\frac{dv_{6\alpha}}{dt} = -\frac{i_\alpha}{C_{se1}} \quad (32)$$

$$v_{ga} = R_\alpha i_\alpha + L''_D\left(I'_\alpha + I'_\alpha\right) \quad (33)$$

$$v_{ga} = R_\alpha i_\beta + L''_D\left(I'_\beta + I'_\beta\right) \quad (34)$$

$$L_A = L_{T1} + L''_D \quad (35)$$

$$L_s = L_r = \frac{L}{2} \quad (36)$$

$$C_s = C_r = \frac{C}{2} \quad (37)$$

$$R_s = R_r = \frac{R}{2} \quad (38)$$

The line inductance and resistance are denoted by (L) and (R), respectively. In a similar manner, equations can be formulated for the β-network. These α-β network equations are then transformed into the DQ frame of reference. By linearizing these differential equations around the operating point, we obtain the state space model equations for the transmission network system.

$$\dot{X}_N = [A_N]X_N + [B_{N1}]U_{N1} + [B_{N2}]U_{N2} + [B_{N3}]U_{N3} \quad (39)$$

$$Y_{N1} = [C_{N1}]X_N + [D_{N1}]U_{N1} + [D_{N2}]U_{N2} + [D_{N3}]U_{N3} \quad (40)$$

$$Y_{N2} = [C_{N2}]X_N \quad (41)$$

$$Y_{N3} = [C_{N3}]X_N \quad (42)$$

Where,

ΔV_{4D} ΔV_{5D} ΔV_{6D} Δi_{1Q} Δi_{2Q} Δi_{4Q} Δi_Q ΔV_{2Q} ΔV_{3Q} ΔV_{4Q} ΔV_{5Q}]

$$U_{N1} = \begin{bmatrix}\Delta i_{3D}\\ \Delta i_{3Q}\end{bmatrix}, U_{N2} = \begin{bmatrix}\Delta I_D\\ \Delta I_Q\end{bmatrix}, U_{N3} = \begin{bmatrix}\dot{\Delta} I_D\\ \dot{\Delta} I_Q\end{bmatrix},$$

$$Y_{N1} = \begin{bmatrix}\Delta V_{gD}\\ \Delta V_{gQ}\end{bmatrix}, Y_{N2} = \begin{bmatrix}\Delta i_D\\ \Delta i_Q\end{bmatrix}, Y_{N3} = \begin{bmatrix}\Delta V_{3D}\\ \Delta V_{3Q}\end{bmatrix}$$

The matrices AN, BN1, BN2, BN3, CN1, DN1, DN2, DN3, CN2, and CN3 are derived from the works of Anderson et al. (1990) and Padiyar (2009)

Static Var Compensator

The Static Var Compensator (SVC) is a shunt device used for reactive power compensation and voltage regulation in power systems. It comprises a Thyristor-Controlled Reactor (TCR) and a fixed capacitor, regulated by a Proportional-Integral (PI) controller for stable voltage control. The linearized SVC control model includes TCR transients and delays from the controller and measurement units, accounting for thyristor switching behavior and signal processing times as shown in Figure 5. TCR currents are expressed in α-axis and β -axis components, simplifying the three-phase system dynamics. These features are key for maintaining voltage stability in dynamic power environments

The currents entering the TCR from the network along the α and β axes are expressed as follows:

$$L_s\frac{di_{3\alpha}}{dt} + R_s i_{3\alpha} = V_{3\alpha} \quad (43)$$

$$L_s\frac{di_{3\beta}}{dt} + R_s i_{3\beta} = V_{3\beta} \quad (44)$$

By applying Kron's transformation to the equations, which involves converting them into the D-Q frame of reference and subsequently linearizing them around the operating point, we obtain the following results:

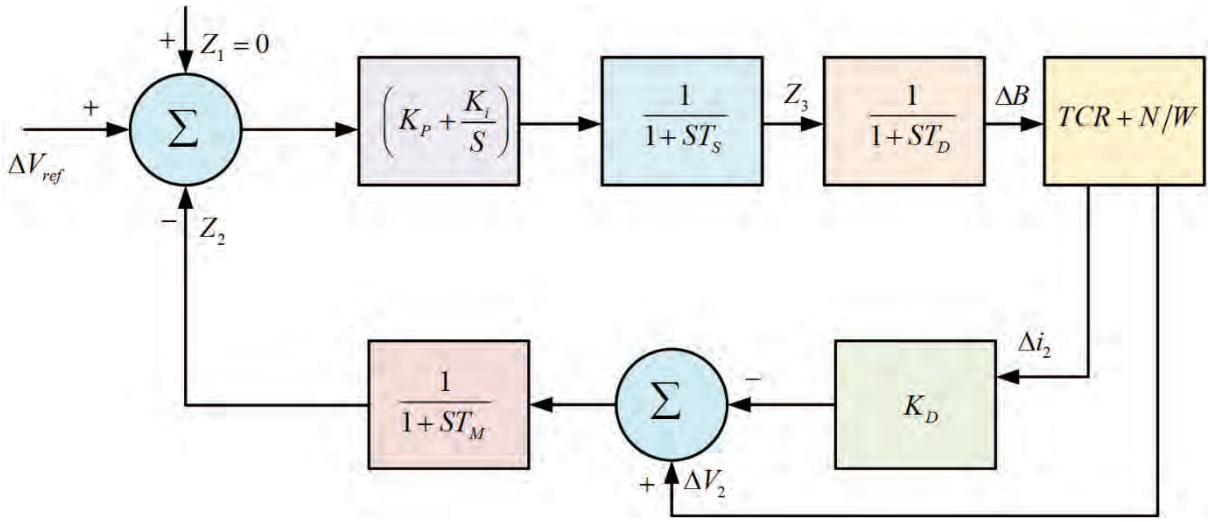


Figure 5. SVS control model.

$$\begin{bmatrix}\Delta i_{3D}\\ \Delta i_{3Q}\end{bmatrix} = \begin{bmatrix}-R_s/L_s & -w_0\\ -w_0 & -R_s/L_s\end{bmatrix}\begin{bmatrix}\Delta i_{3D}\\ \Delta i_{3Q}\end{bmatrix} + \begin{bmatrix}w_0B_0 & 0\\ 0 & w_0B_0\end{bmatrix}\begin{bmatrix}\Delta v_{3D}\\ \Delta v_{3Q}\end{bmatrix} + w_0\begin{bmatrix}\Delta v_{3D0}\\ \Delta v_{3Q0}\end{bmatrix}\Delta B \quad (45)$$

Here, $Q = \frac{\omega_0 L_S}{R_S}$ and

$B = \frac{1}{\omega_0 L_S}$

In this context, (Ls) and (Rs) denote the inductance and resistance of the TCR, respectively. The linearized state and output equations for the SVS model are formulated as follows:

$$\dot{X}_S = [A_S]X_S + [B_{S1}]U_{S1} + [B_{S2}]U_{S2} + [B_{S3}]U_{S3} \quad (46)$$

The additional equations derived from the SVS control model are as follows:

$$\frac{d\Delta B}{dt} = \frac{Z_3}{T_D} - \frac{\Delta B_3}{T_D} \quad (47)$$

$$\frac{dZ_2}{dt} = 1/T_M\left[\Delta V_3 - Z_2 + K_D\Delta i_3\right] \quad (48)$$

$$\frac{dZ_3}{dt} = -1/T_S\left[Z_3 - K_1Z_1 + K_P\Delta V_{ref} - K_PZ_2 + K_P\Delta v_F\right] \quad (49)$$

where ΔV_{ref} represents the 'reference voltage perturbation', ΔV_f is the output of the 'auxiliary controller', and ΔV_3 and Δi_3 denote the 'incremental magnitudes of the SVS bus voltage and TCR current', respectively. The 'SVS bus voltage' V_3 is expressed in terms of its DQ-axis components as follows:

$$V_2^2 = V_{2D}^2 + V_{2Q}^2 \quad (50)$$

The magnitude of the TCR currents is expressed as follows:

$$i_2^2 = i_{2D}^2 + i_{2Q}^2 \quad (51)$$

The output equations derived are as follows:

$$Y_S = [C_S]X_S + [D_S]U_{S1} \quad (52)$$

$$X_S = \left[\Delta i_{3D} \quad \Delta i_{3Q} \quad Z_1 \quad Z_2 \quad Z_3 \quad \Delta B\right]^t$$

$$Y_S = \left[\Delta i_{3Q} \quad \Delta i_{3Q}\right]^t$$

$$U_{S2} = \left[\Delta V_{ref}\right]$$

$$U_{S3} = \left[\Delta V_f\right]$$

The matrices AS, BS1, BS2, BS3, CS, and DS are derived from the works of Anderson et al. (1990) and Padiyar (2009). The overall system model is shown in Figure 6.

Optimal Control Theory

The comprehensive system model, which integrates the generator model, mechanical model, excitation system, network model, and SVS model, can be represented using "optimal control theory".

$$\dot{X} = AX(t) + Bu(t) + Fw \quad (53)$$

Above equation is utilized in optimal control theory (Kothari and Nagrath, 2007), and its Performance Index (PI) is given below:

$$PI = \frac{1}{2}\int_0^\infty (X'QX' + RU') \quad (54)$$

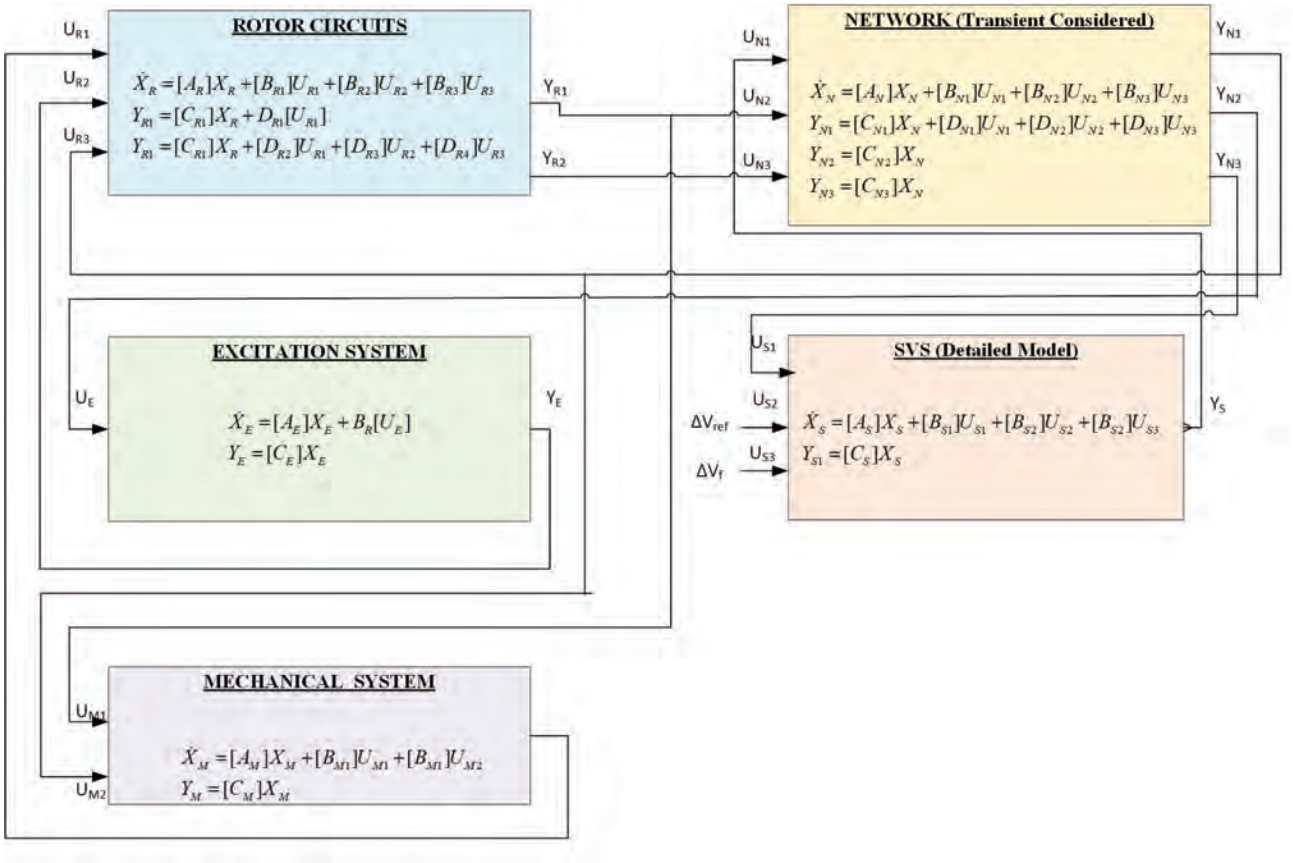


Figure 6. System model.

Let (Q) be a symmetrical matrix, (R) (or (K_I)) also a symmetrical matrix, and (K) the feedback matrix. By applying the reduced order Riccati matrix equation:

$$A^T P + PA - PBR^{-1}B^T P + Q = 0 \tag{55}$$

The acceptable solution for (K) is the one that ensures the system remains stable

$$X' = AX' + B(-KX') = (A - BK)X' \tag{56}$$

To ensure stability, all the eigenvalues of the matrix (A - BK) must have negative real parts. The control law is typically designed to minimize this cost, ensuring system stability and optimal performance. However, solving this optimization problem for the control gain matrix K is challenging, especially for nonlinear or high-dimensional systems. This is where metaheuristic optimization methods like CS-GWO come into play.

Hybrid CSA-GWO Technique

The Purposed (algorithm combines the cuckoo search (CS)[19] for its excellent global search capabilities with the grey wolf optimization (GWO), known for its efficient local search. This hybridization ensures better convergence speed and solution quality compared to the individual algorithms

Grey Wolf Optimization (GWO)

Grey Wolf Optimization mimics the social hierarchy and hunting mechanism of grey wolves. The GWO algorithm categorizes wolves into four types: alpha (leader i.e m), beta (n), delta(o), and omega (follower wolves). The wolves' positions are updated based on the best solutions found (m, n and o wolves), guiding the search toward optimal solutions.

Mathematical Model

An outline of the Grey Wolf Optimizer (GWO) algorithm follows the presentation of the mathematical models for tracking, surrounding, attacking, and social hierarchy in this section.

Social Hierarchy

In the mathematical model of the social structure of wolves in the GWO, the strongest and most resilient solution is represented as alpha (m). Conversely, the solutions ranked second and third are denoted as beta (n) and delta (o) correspondingly. Other candidate solutions are denoted as omega (x). The optimization process is guided by m, n, and o, with the remaining wolves following their lead.

Encircling Prey

Grey wolves exhibit encircling behavior during hunting. The mathematical representation of this behavior is described by the following equations:

$$\vec{D}_1 = \left|\vec{C}_1.\vec{X}_Q(k) - \vec{X}(k)\right| \quad (57)$$

$$\vec{X}(k+1) = \vec{X}_Q(k) - \vec{A}_1.\vec{D}_1 \quad (58)$$

In this context, the variable k represents the current iteration, whereas $\vec{A}_1$ and $\vec{C}_1$ are 'coefficient vectors'. $\vec{X}_Q$ refers to the 'position vectors of the prey', while $\vec{X}$ represents the 'position vector of a grey wolf'. The vectors $\vec{A}_1$ and $\vec{C}_1$ are calculated using specific formulas involving random vectors.

$$\vec{A}_1 = 2\vec{a}.\vec{r}_1 - \vec{a} \quad (59)$$

$$\vec{C}_1 = 2.\vec{r}_2 \quad (60)$$

Hunting

Grey wolves, guided by alpha, exhibit hunting behavior. To simulate this behavior mathematically, the positions of the m (alpha), n (beta), and o (delta) wolves are considered more informed about the potential prey location. The positions of other wolves, including the omegas, are then updated based on these informed positions:

$$\vec{D}_m = \left|\vec{C}_a.\vec{X}_m - \vec{X}\right| \quad (61)$$

$$\vec{D}_n = \left|\vec{C}_b.\vec{X}_n - \vec{X}\right| \quad (62)$$

$$\vec{D}_o = \left|\vec{C}_c.\vec{X}_o - \vec{X}\right| \quad (63)$$

$$X_a = \vec{X}_m - \vec{A}_a.(\vec{D}_m) \quad (64)$$

$$\vec{X}_b = \vec{X}_n - \vec{A}_b.(\vec{D}_n) \quad (65)$$

$$\vec{X}_c = \vec{X}_o - \vec{A}_c.(\vec{D}_o) \quad (66)$$

$$\vec{X}(k+1) = \frac{\vec{X}_m + \vec{X}_n + \vec{X}_o}{3} \quad (67)$$

This approach ensures that the search agents, except for the top three, update their positions according to the more knowledgeable alpha, beta, and delta wolves. These mathematical formulations and algorithms capture the essence of the GWO, where social hierarchy, encircling prey, and hunting behaviors are translated into mathematical models for optimization purposes.

CUCKOO SEARCH ALGORITHM (CSA)

The natural world is home to more than a thousand bird species, many of which share characteristics and habits. As an example, all bird mothers build nests in safe areas and deposit eggs in a variety of forms to ward off predators. Mother birds, for instance, construct nests in safe areas to shield their young from predators and lay eggs in a variety of designs. The cuckoo is a well-known brood parasite, renowned for its deceptive strategy. A female cuckoo will remove one of the host's eggs and replace it with her own, mimicking the appearance of the host's eggs to avoid detection. This ensures that the cuckoo egg hatches first, and the cuckoo chick often evicts the host's eggs from the nest, securing all resources for itself. This action served as inspiration for Yang and Deb's Cuckoo Search Algorithm (CSA) [20]. It is a productive optimization technique that strikes a balance between regional and global exploration. CSA is based on three main rules:

1. Each cuckoo deposits a single egg in a nest chosen at random.
2. Nests of superior quality, containing high-grade eggs, are passed down to future generations.
3. There is a set number of host nests and the probability that a host bird finds an alien egg is Ba ∈ (0, 1). When a host bird's eggs are found, it can either drop them or leave the nest to start a new one. The number of available host nests stays the same, and the chance that a host bird will lay an egg laid by an alien is Ba, where Ba can take on values between 0 and 1. The host bird will either leave the nest or drop the egg once it's found. Then it will start building a new nest. The Pseudo code of the CS-GWO algorithm is shown in Figure 7.

The algorithm can be succinctly summarized in the subsequent steps:

1. Objective function f(Z), where Z = (z1, z2……,zd)T
2. Establish the starting population of n host nests
3. While loop, until a halt criterion is satisfied(e.g., maximum iterations or generations):
   - Obtain a cuckoo through random movements following Lévy flights.
   - Assess the suitability of its fitness in terms of K_x and K_y.
   - Select a random nest y from the set n.
   - If $K_x > K_y$, substitute y with the revised solution.
   - Discard a portion B_a of the least favorable nests and construct new ones.
   - Retain the optimal options and arrange them in order of preference.

The Lévy flight mechanism is key to CSA, as it enables a balance between exploration and exploitation:

$$z_{t+1}^i = z_t^i + \alpha \oplus Lévy(\lambda) \quad (68)$$

where α>0 is the step size, and ⊕ denotes entry wise multiplication. Lévy flights are characterized by:

$$Lév \sim u = t^{-\lambda}, \quad 1 < \lambda \leq 3 \quad (69)$$

```
Initialize the grey wolf population Xa(a=1,2,3,...n)
Initialize a, A, C,and pa
Calculate the fitness of each search agent (wolf)
Xm-the best search agent
Xn=the second best search agent
Xo=the third best search agent
While (t<Max number of iterations)
    for each search agent
       Update the position of current search agent by equation (40)
    end
Update a, A, and C
Calculate the fitness of all search agents
Update Xm , Xn and Xo
Update the wolf's position by equation (66).
   if random number> pa then
random change the wolf's position.
Compute and update the fitness function
t=t+1
  end while
return best solution Xm
```

Figure 7. Pseudo code of CS-GWO algorithm.

Table 2. Comparative Analysis of CSA, GWO, and Hybrid CSA-GWO for Optimization Efficiency

Feature	Cuckoo Search Algorithm (CSA)	Grey Wolf Optimizer (GWO)	Hybrid CSA-GWO
Inspiration	Brood parasitism of cuckoo birds	Leadership hierarchy of grey wolves	Combination of CSA and GWO principles
Exploration	Strong due to Lévy flight	Moderate due to leader-based search	Enhanced due to hybridization
Exploitation	Moderate	Strong due to adaptive leadership	Balanced through adaptive switching
Convergence Speed	Can be slow in large search spaces	Faster due to hierarchical structure	Improved due to hybrid approach
Optimization Accuracy	Good for global search	Good for local refinement	High accuracy with adaptive tuning
Application in SSR Mitigation	Effective but may require tuning	Effective in dynamic conditions	Superior due to adaptive and efficient tuning

The flow chart for CSA-GWO is shown in Figure 8. This process ensures that some solutions are explored globally, while others exploit the local area around the current best solution.

The performance comparison of CSA, GWO, and the hybrid CSA-GWO algorithm is summarized in Table 2, highlighting their optimization efficiency and key characteristics. The Hybrid Cuckoo Search and Grey Wolf Optimization (CS-GWO) is a powerful optimization technique for solving complex control system problems, such as designing optimal controllers. By combining the exploration capabilities of CS with the exploitation strengths of GWO, this hybrid algorithm is well-suited for finding optimal control gain matrices that minimize the system›s performance cost. The hybrid CS-GWO approach demonstrates fast convergence, robustness, and high solution quality, making it a valuable tool in modern control system design.

RESULTS AND DISCUSSION

Case A : System without Natural Damping

For a system delivering (Pg = 800) MW, the eigenvalues were calculated under various scenarios: with and without

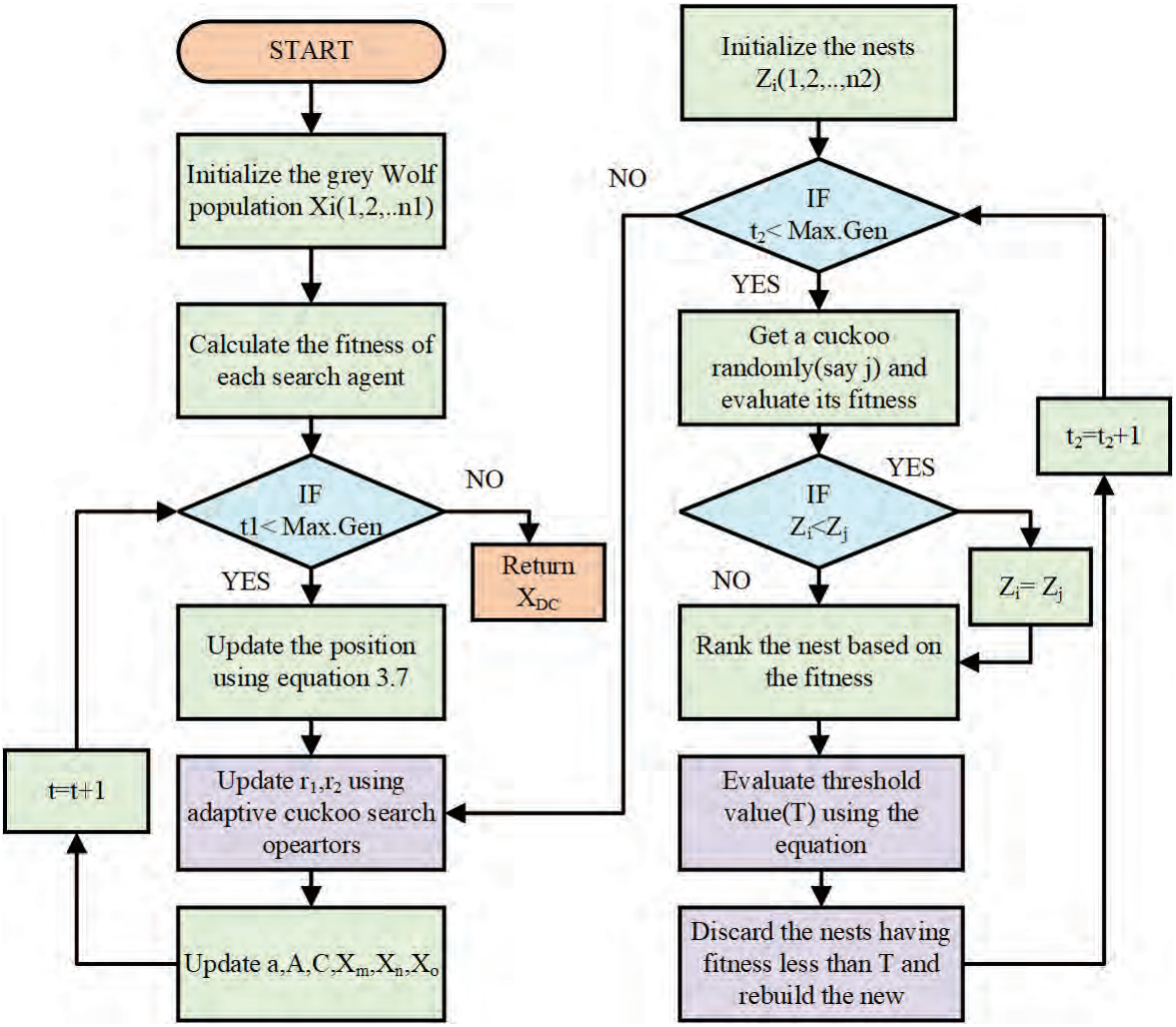


Figure 8. Flowchart of cuckoo-grey wolf proposed algorithm.

the application of optimal control, with zero natural damping, all within a series-compensated network and 10% change in electric torque and 10% change in reference voltage. As shown in Table 3 compares the system eigenvalues under different control strategies—SVS only, optimal controller, BFOA-based controller, and the proposed hybrid CSA-GWO controller at 50% compensation without natural damping. The results show that the hybrid CSA-GWO controller achieves the most significant leftward eigenvalue shift, enhancing system damping and stability. Key improvements include reduced oscillations in generator modes, better network stability, and superior performance of SVC and AVR, leading to more effective mitigation of SSR and improved turbine-generator dynamic performance.

Figure 9 depicts the mechanical rotor angle (MECH-DELTA (RAD)), showcasing the rotor›s dynamic response. Figure 10 presents the HP-IP Torque (P.U), highlighting the torque variations in the high-pressure and intermediate-pressure turbine stages. Figure 11 shows the SVC Bus Voltage, demonstrating the voltage regulation at the bus controlled by the Static Var Compensator (SVC). Lastly, Figure 12 illustrates the SVC Susceptance (P.U), reflecting the changes in the SVC›s reactive power compensation to maintain voltage stability.

The transient response characteristics of the system under investigation are evaluated based on Table 4, which presents the time-domain analysis (T.D.A) for the rotor angle curve without natural damping. A comparative assessment is made between the SVS-only system and the optimal controller, BFO-optimized controller, and hybrid CSA-GWO technique to determine their effectiveness in improving system stability.

For rise time, all three control strategies result in an increase compared to SVS-only (0.59301 s). The optimal controller alone shows an increase of 41.8%, while the BFO-optimized controller and hybrid CSA-GWO method exhibit increases of 40.3% and 37.4%, respectively. Despite this, the controlled systems demonstrate better performance in other critical stability measures.

Table 3. System eigenvalue for case A at 50% compensation without natural damping

Mode	With SVS Only	With optimal controller	With optimal controller & BFO [18]	With optimal controller &Hybrid CS-GWO (Proposed study)
Mode 5	7.0129e-05±298.18i	-0.42252±298.18i	-0.42258± 298.18i	-0.5±298.18i
Mode 4	-0.11323±203.01i	-6.5595±202.91i	-6.5668±202.91i	-6.812±202.91i
Mode 3	-0.089727±160.77i	-7.9616±160.76i	-7.9655±160.76i	-8.0±160.76i
Mode 2	-0.018756±127.06i	-5.8125±128.31i	-5.8132±128.31i	-5.92±128.31i
Mode 1	0.0306±99.6767i	-7.4304±98.8561i	-7.4306±98.8562i	-7.491±98.8562i
Mode 0	-2.0265±13.6598i	-2.9213±12.6086i	-2.9214±12.6052i	–3.13±12.6052i
Network	-3.2878±3497.1i	-3.2882±3497.1i	-3.3192±3497.1i	-4.1950±000.0i
	-3.2880±2868.8i	-3.2888±2868.8i	-3.3484+2868.8i	-3.459±2868.8i
	-13.911±2528.9i	-13.913±2528.9i	-14.029±2528.9i	-14.57±2528.9i
	-15.861±1900.7i	-15.867±1900.7i	-16.133±1900.7i	-17.012±1900.7i
	-12.490±1120.8i	-18.919±1121.4i	-50.065± 1125.3i	-50.98±1125.3i
	-18.929±493.31i	-58.218±498.67i	-66.235±493.65i	-67.01±493.65i
	-11.854±328.38i	-371.39±366.05i	-1585.1± 453.73i	-1589.42±453.73i
	-3.3187±315.07i	-11.839+328.30i	-11.839±328.30i	-12.49±328.30i
	-11.698±299.62i	-11.815 ±299.63i	-11.815±299.64i	-12.45±299.64i
Generator system	-32.3668	-51.4780	-51.4780	-52.623
	-24.8779	-30.4181	-30.4182	-31.578
	-1.9448	-5.7939	+5.7939	±5.7i
SVC	-545.08	-625.19	-804.95	-810.45
	± 72.704i	-447.95	-417.13	-419.74
	-55.4864	-165.53	-169.20	-170.68
	±92.8421i	±351.39i	±333.78i	±333.0i
	-2.9118	-127.40	-127.71	-130.0
	±310.64i	-1.7614	-1.7816	-1.865
AVR	-26. 1307	-0. 14624	-0.14624	-0.14624
	±24.1266i	±314.16i	±314.16i	±314.16i
	-0.6407	-0.9066	-0.9066	-0.9066
	±1.2882i	±1.2882i	±1.2747i	±1.298i

Table 4. T.D.A for the rotor angle curve without natural damping

MECH-DELTA (rad.)	With SVS only	With optimal controller	With optimal controller and BFO	With optimal controller and Hybrid CS-GWO
Rise time	0.59301 s	0.84108 s	0.83252 s	0.81430 s
Settling time	24.99 s	4.5629 s	3.3641 s	2.3732 s
Over-shoot	32.563	18.83	17.968	17.461
Under-shoot	0	0	0	0
Peak time	1.62 s	1.88 s	1.86 s	1.73 s

A significant improvement is observed in settling time across all controllers compared to SVS-only (24.99 s). The hybrid CSA-GWO technique achieves the most improvement with a change of 90.5%, followed by the BFO-optimized controller at 86.5% and the optimal controller alone at 81.7%. This substantial enhancement highlights the controllers› effectiveness in enhancing system damping and stability.

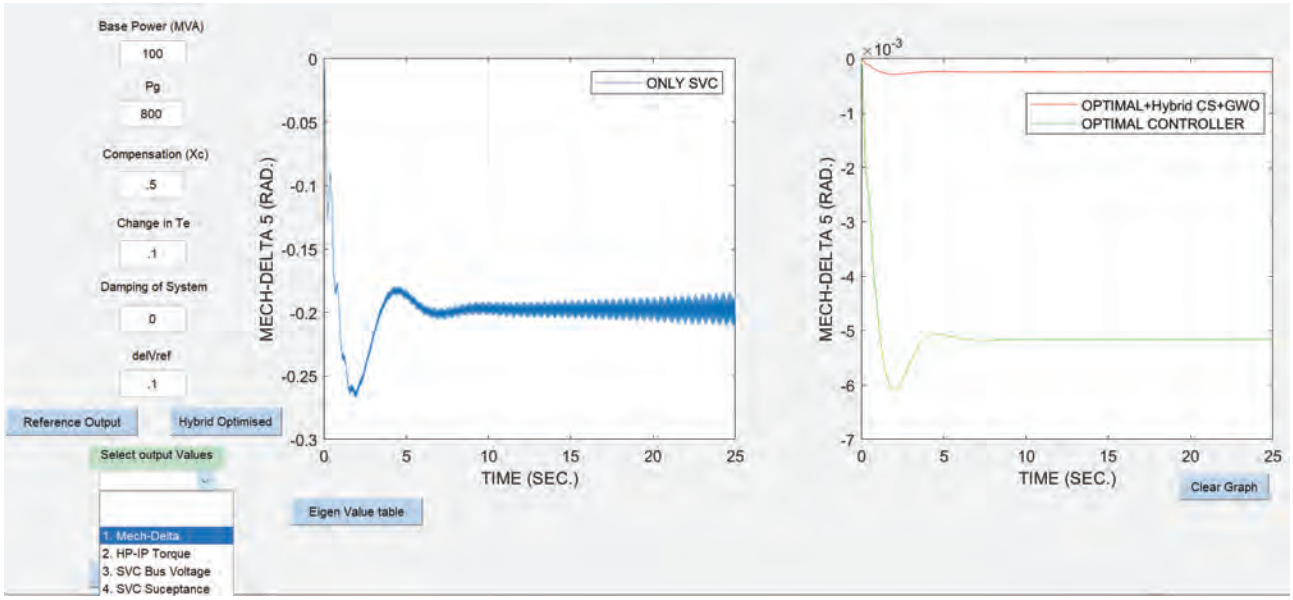


Figure 9. Mech-delta (rad) of case A.

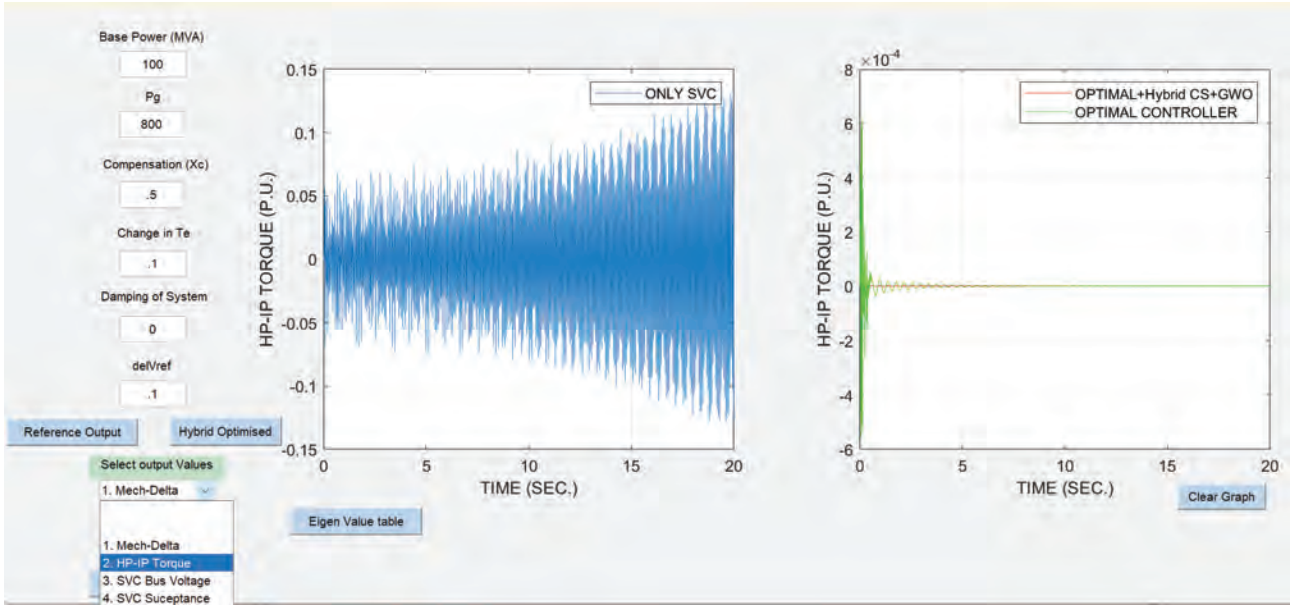


Figure 10. HP-IP torque (P.U) of case A.

In terms of overshoot, the hybrid CSA-GWO method exhibits the most improvement (46.4%) over the SVS-only case (32.563), followed by BFO-optimized (44.8%) and the optimal controller alone (42.2%). A lower overshoot indicates a more controlled system response with minimal excessive fluctuations.

For peak time, the hybrid CSA-GWO method achieves the smallest increase at 6.8%, making it the best performer in this aspect. In contrast, the BFO-optimized controller shows an increase of 14.8%, while the optimal controller alone exhibits a 16.0% increase compared to SVS-only (1.62 s). The controlled systems maintain zero undershoot across all cases, ensuring no negative oscillations or system instability.

The hybrid CSA-GWO technique outperforms all other methods, achieving the best improvement in settling time and overshoot with minimal peak time variation. While the BFO-optimized and optimal controllers show strong performance, hybrid CSA-GWO remains the most effective for system stability and transient response

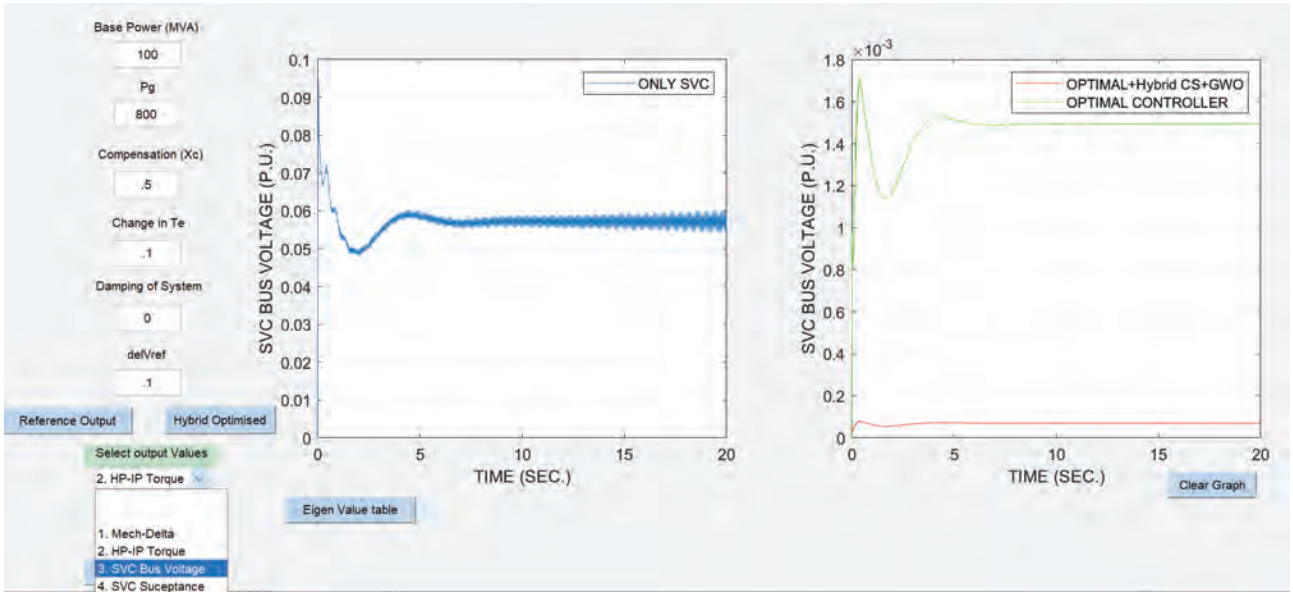


Figure 11. SVC bus voltage of case A.

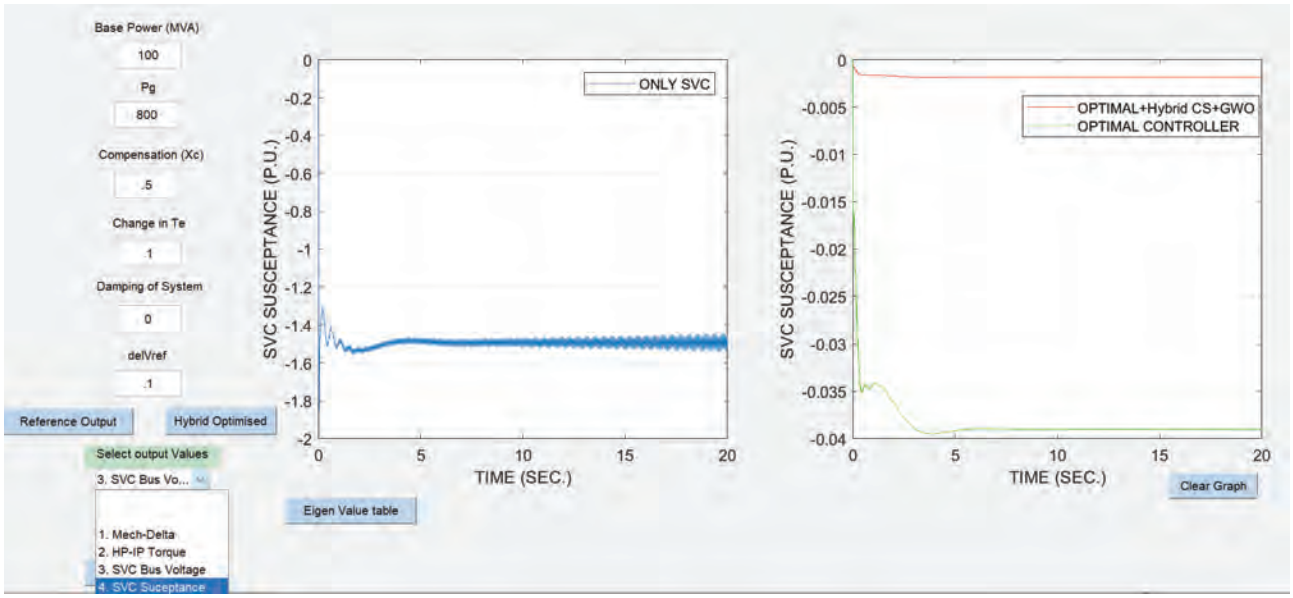


Figure 12. SVC susceptance (P.U) of case A.

Case B: System with Natural Damping

The system›s response is illustrated after a 10% change in electric torque and a 10% change in reference voltage, and analyzed under 50% compensation while considering inherent damping effects (D = 0.0322). Table 5 presents eigenvalues, assessing stability, while Figure 13 and Figure 14 illustrate rotor angle and HP-IP torque dynamics, respectively. Voltage stability at the SVC bus is depicted in Figures 15 and 16, offering insights into system reliability..

Table 5 compares eigenvalues for different control strategies SVS only, optimal controller, BFOA-based controller, and the proposed hybrid CSA-GWO controller. The results indicate that the hybrid CSA-GWO controller shifts eigenvalues further left in the complex plane, significantly improving system damping. This leads to reduced oscillations in critical modes, enhanced network stability, and better performance of the SVC and AVR, ensuring more effective mitigation of SSR and improved dynamic performance of the turbine-generator system.

Table 5. System eigenvalue for case B at 50% compensation with natural damping (D=0.0322)

Mode	With SVS Only	With optimal controller	With optimal controller & BFO [18]	With optimal controller &Hybrid CS-GWO (Proposed study)
Mode 5	7.0129e-05±298.18i	-0.42252±298.18i	-4.4484±298.14i	-4.4484± 294.14i
Mode 4	-0.11323±203.01i	-6.5595 ± 202.91i	-7.1784±202.88i	-7.1789±202.88i
Mode 3	-0.089727±160.77i	-7.9616 ±160.76i	-8.1888±160.65i	-8.1892 ± 160.65i
Mode 2	-0.018756±127.06i	-5.8125 ±128.31i	-11.868±127.73i	-11.868±124.73i
Mode 1	0.0306±99.6767i	-7.4304±98.8561i	-7.5498±99.0542i	-7.5499±99.0542i
Mode 0	-2.0265±13.6598i	-2.9213±12.6086i	-3.6804±12.5053i	-3.6806±12.5049i
Network	-3.2878±3497.1i	-3.2882±3497.1i	-3.3143±3497.1i	-3.3133 ±3497.1i
	-3.2880±2868.8i	-3.2888±2868.8i	-3.3402±2868.8i	-3.3377 ±2868.8i
	-13.911±2528.9i	-13.913±2528.9i	-14.015±2528.9i	-14.981 ±2528.9i
	-15.861± 1900.7i	-15.867±1900.7i	-16.112±1900.7i	-17.012 ±1900.7i
	-12.490±1120.8i	-18.919±1121.4i	-49.502 ± 1125.3i	-72.502 ± 1130.3i
	-18.929± 493.31i	-58.218±498.67i	-66.222±493.67i	-68.322±493.67i
	-11.854± 328.38i	-371.39±366.05i	-1450.8± 563.85i	-1758.8 ± 726.85i
	-3.3187±315.07i	-11.839+328.30i	-11.839±328.30i	-11.839 ±321.30i
	-11.698±299.62i	-11.815±299.63i	-11.815±299.63i	-11.815±299.63i
Generator system	-32.3668	-51.4780	-51.1758	-51.1758
	-24.8779	-30.4181	-30.4163	-30.4163
	-1.9448	-5.7939	-5.7792	-5.7792
SVC	-545.08	-625.19	-800.79	-823.74
	± 72.704i	-447.95	-417.20	-402.60
	-55.4864	-165.53	-169.22	-169.22
	±92.8421i	±351.39i	±333.82i	±333.82i
	-2.9118	-127.40	-127.84	-127.92
	±310.64i	-1.7614	-1.7742	-1.7934
AVR	-26. 1307	-0. 14624	-0.14624	-0.14624
	±24.1266i	±314.16i	±314.16i	±314.16i
	-0.6407	-0.9066	-0.8997	-0.8997
	±1.2882i	±1.2882i	±1.2762i	±1.2631i

Table 6 provides a time-domain analysis (T.D.A) of the rotor angle curve, evaluating key performance metrics. For rise time, all controllers exhibit an increase compared to the SVS-only system (0.61326 s). The optimal controller shows a 38.4% increase, while the BFO-optimized controller and hybrid CSA-GWO method result in increases of 37.2% and 46.4%, respectively.

Examining settling time, the hybrid CSA-GWO method achieves the highest improvement with a 60.1% change, followed by the BFO-optimized controller at 36.6% and the optimal controller alone at 12.7%, compared to the SVS-only system (5.3236 s).

For overshoot, the hybrid CSA-GWO technique exhibits the most improvement with a 47.1% change, while the BFO-optimized controller and optimal controller achieve changes of 44.6% and 41.9%, respectively, over the SVS-only case (32.715).

Considering peak time, the hybrid CSA-GWO method results in the smallest increase at 4.7%, making it the most effective in this aspect. In comparison, the BFO-optimized controller and optimal controller show increases of 11.9% and 14.3%, respectively, over the SVS-only system (1.68 s).

The hybrid CSA-GWO technique emerges as the most effective control strategy, offering the highest improvement in key stability parameters. It excels in settling time and overshoot while ensuring minimal deviation in peak time, making it the best choice for enhancing overall system performance.

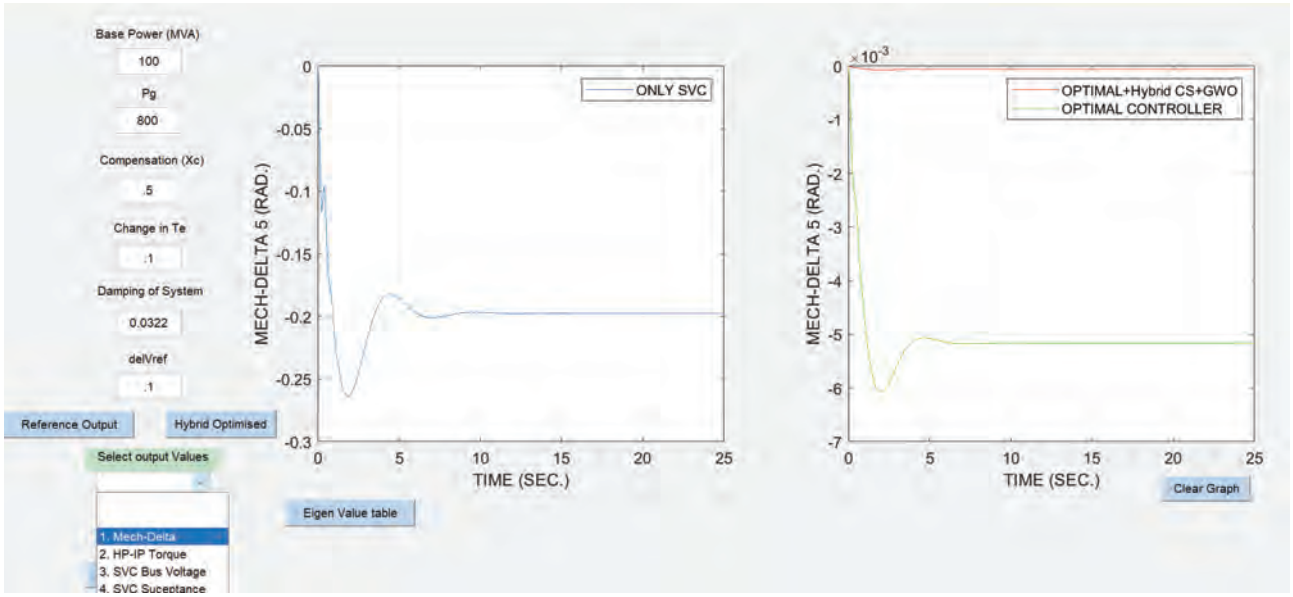


Figure 13. Mech-delta (RAD) of case B.

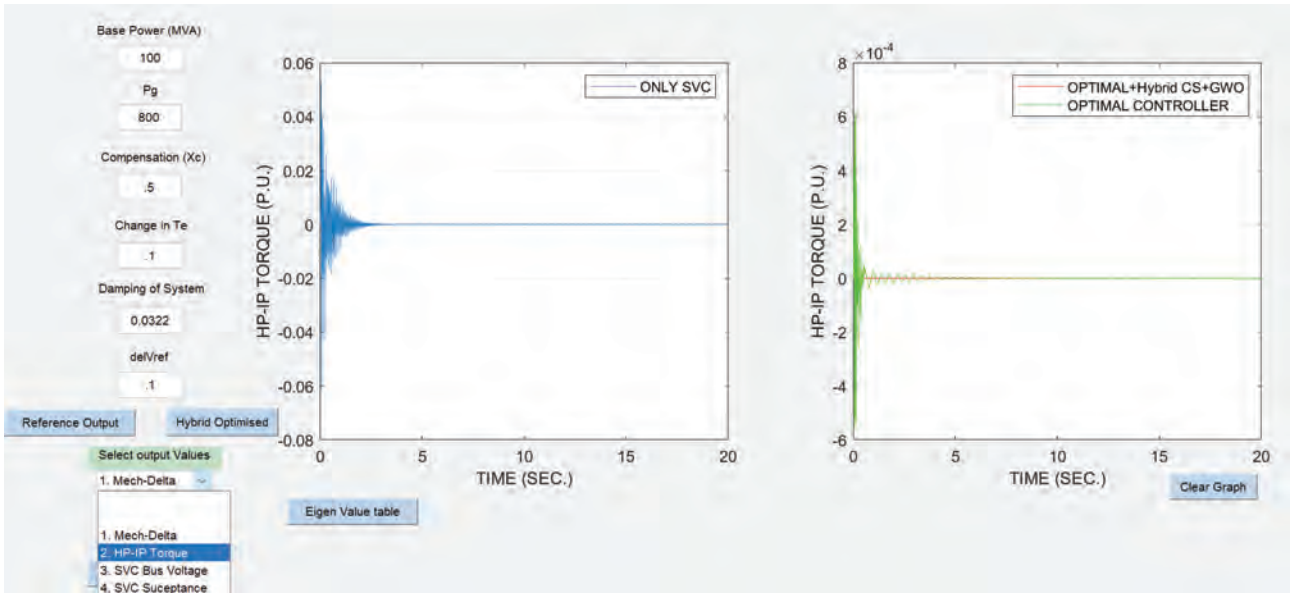


Figure 14. HP-IP torque (P.U) of case B.

Table 6. T.D.A for the rotor angle curve with natural damping

MECH-DELTA (rad.)	With SVS only	With optimal controller	With optimal controller and BFO	With optimal controller and Hybrid CS-GWO
Rise time	0.61326 s	0.84882 s	0.84142 s	0.89732 s
Settling time	5.3236s	4.6479 s	3.3735 s	2.1246 s
Over-shoot	32.715	18.993	18.11	17.325
Under-shoot	0	0	0	0
Peak time	1.68 s	1.92 s	1.88 s	1.76 s

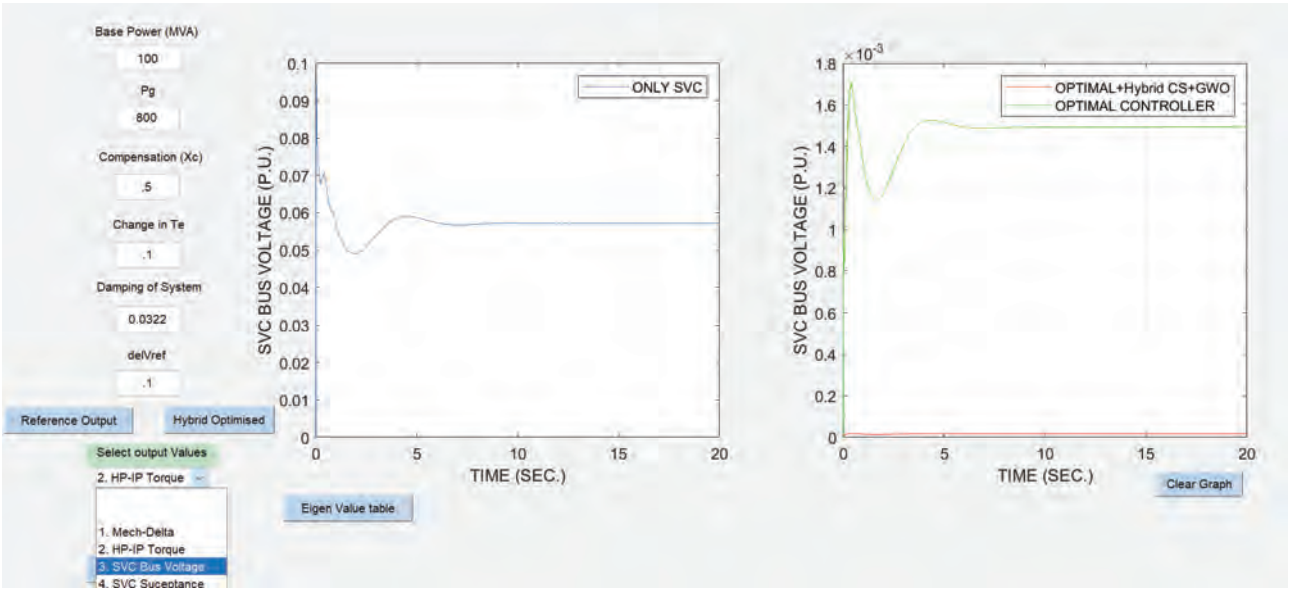


Figure 15. SVC bus voltage of case B.

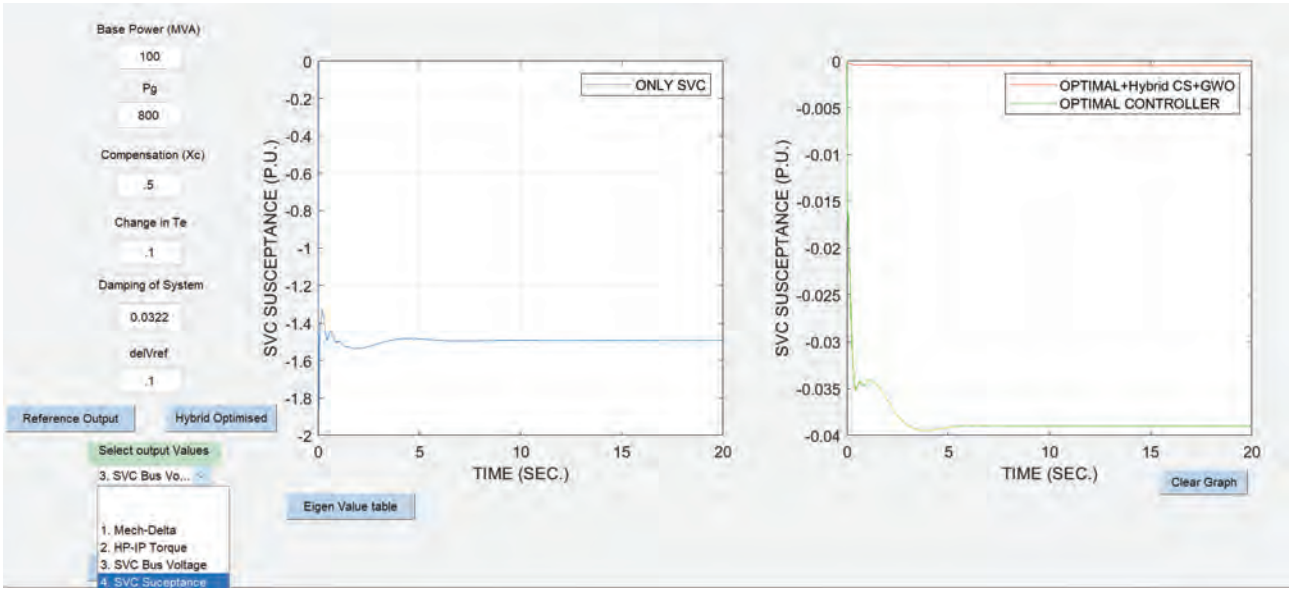


Figure 16. SVC bus voltage of case B.

CONCLUSION

In this paper, an investigation is conducted to examine the effectiveness of a Hybrid Cuckoo Search Algorithm-Grey Wolf Optimization (CSA-GWO)-based control strategy for Static Var Compensator (SVC) in mitigating sub-synchronous resonance (SSR) and enhancing the stability of turbine-generator systems in series-compensated networks. Using the IEEE First Benchmark Model, the proposed Hybrid CSA-GWO algorithm proves to be more effective than BFO, achieving faster damping of harmful oscillations, superior transient stability, and enhanced steady-state performance. Fluctuations in electric torque and reference voltage intensify SSR, increasing control complexity and system instability. The study evaluates both natural damping and zero damping cases, with zero damping included to analyze the system under more severe conditions. Eigenvalue analysis and time-domain simulations confirm that the Hybrid CSA-GWO method consistently outperforms BFO, proving to be the most effective solution for mitigating SSR, improving transient response, and ensuring grid resilience.

Future Scope

Building on this research, the following key areas offer promising directions for further enhancement of SSR mitigation and power system stability.

1. Real-Time Implementation: Validate the CSA-GWO-based controller in real-world power systems using hardware-in-the-loop (HIL) simulations and prototype testing.
2. Adaptive Control Strategies: Develop adaptive and self-tuning controllers that dynamically adjust based on real-time system conditions for improved SSR mitigation.
3. Multi-Machine System Analysis: Extend the study to multi-machine power systems to evaluate effectiveness in complex network scenarios.
4. Integration with Renewable Energy Sources: Investigate the impact of renewable energy sources like wind and solar on SSR mitigation and overall system stability.
5. Machine Learning Applications: Implement AI/ML-based predictive control for real-time SSR detection and mitigation, enhancing system resilience.

NOMENCLATURE

Ψ_f	Flux associated with coil 'f'
Ψ_h	Flux associated with coil 'h'
Ψ_g	Flux associated with coil g
Ψ_k	Flux associated with coil 'k'
i_d and i_q	are the d-axis and q-axis components of the machine terminal currents
M_i	Inertia of ith mass
Di	Damping of ith mass
K_{ij}	Spring constants of the shaft elements between mass i and j
Di	Damping of shaft segments between mass i and j
δ_{M2}	Angular position of HP turbine
δ_{M3}	Angular position of IP turbine
δ_{M4}	Angular position of LPA turbine
δ_{M5}	Angular position of LPB turbine
δ_{M6}	Angular position of GEN
δ_{M7}	Angular position of exciter
ω_1	Angular velocity of HP turbine
ω_2	Angular velocity of IP turbine
ω_3	Angular velocity of LPA turbine
ω_4	Angular velocity of LPB turbine
ω_5	Angular velocity of LPB turbine
S_E	Saturation function
V_F	Excitation voltage of gen.
V_S	Stabilizing feedback signal voltage
V_f	Excitation system regulator's output voltage
L_{T1}	Inductance of sending end transformer
L_{T2}	Inductance of receiving end transformer
Ra	Armature resistance of the synchronous generator
I	Dependent current source equivalent of generator
V_g	Vs generator terminal voltage
lg	Generator terminal current
i_{gd}	Gen. terminal current associated with direct axis
i_{gq}	Gen. terminal current as sociated with quadratur axis
Y_{shr}	Shunt admittance at receiving end
Y_{shs}	Shunt admittance at sending end
C_{ses}	Series capacitance at sending end
C_{ser}	Series capacitance at receiving end
V_1	Infinite bus voltage
i_1	Incoming current of infinite bus
V_i	Voltage at bus represented by its subscripts (i= 1 to 5)
V_6	Voltage at sending end series capacitor
V_{5RCD}	Voltage at receiving end series capacitor
i_I	Current in the different sections of network
i_3	TCR current
i_x	Net current injected by SVC
ix'	Current through shunt capacitance at SVC bus
C_n	Fixed capacitance of SVC
Z_1	Regulator block input
Z_2	Measurement unit output
Z_3	Regulator block output of SVS
B	Susceptance offered by SVS
K_{iD}	Constant of SVC system associated with i3D
K_{iQ}	Constant of SVC system associated with i3Q
K_{vD}	Constant of SVC system associated with V3D
K_{vQ}	Constant of SVC system associated with V3D
X_e	Capacitive-reactance of compensating capacitor
X_{se1}	Series compensation reactance at sending end
X_{se2}	Series compensation reactance at sending end
δ,R,L	Power angle ,resistance, inductance for Line
P,Q,V,θ	Real power, reactive power, voltage,bus voltage angle at different buses.
T_E	electrical torque
K_P,K_I	Proportional, integral gains of SVS voltage controller
G	Gain of SVS
K_D	Slope of control characteristics
T_S, T_D,T_M	FIring delay,thyristor delay, measurement time constant
V_F	Auxiliary control signals
I_D,I_Q	D and Q axis components of la currents of the 'generator
x_1	Stator leakage reactance
x_d, x_d', x_d''	Direct axis synchronous, transients and sub transients reactance, respectively
x_q, x_q', x_q''	Quadrature axis synchronous transients and sub transients reactance, respectively
T_{do}',T_{do}''	Direct axis transients and sub transients open circuit time constants, respectively.
T_{qo}',T_{qo}''	Quadrature axis transients and sub transients open circuit time constants, respectively.
I_g	Generator current
HP	High pressure turbine
IP	Intermediate pressure turbine
LPA	Low pressure turbine A

LPB	Low pressure turbine B
GEN	Generator
EXC	Exciter
P_E	Electrical power output from generator
P_M	Mechanical power input to the generator

Subscripts:

R, M, E, N	rotor circuit, mechanical, excitation, network
f, h, g, k	field and amortisseur circuits
D. Q, 0	D-Q-0 axis variable (stator frame of reference)
d, q, 0 : d-q-0	axis variable (rotor frame of reference)

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Farmer RG, Schwalb AL, Katz E. Navajo project report on subsynchronous resonance analysis and solutions. IEEE Xplore 1977;96:1226–1232. [CrossRef]

[2] Anderson PM, Agrawal BL, Ness JEV. Sub-Synchronous Resonance in Power System. 1st ed. New York: IEEE Press; 1990.

[3] Farmer RG, Schwalb AL, Katz E. Navajo project report on subsynchronous resonance analysis and solutions. IEEE Xplore 1977;96:1226–1232. [CrossRef]

[4] Walker DN, Bowler CEJ, Jackson RL, Hodges DA.Results of subsynchronous resonance test at Mohave. IEEE Xplore 1975;94:1878–1889. [CrossRef]

[5] Padiyar K. Facts Controllers in Power Transmission and Distribution. New Delhi, India:New Age International Publishers 2007.

[6] Kumar N, Dave MP. Application of an auxiliary controlled static var system for damping subsynchronous resonance in power systems. Electr Power Syst Res 1996;37:189–201. [CrossRef]

[7] Padiyar KR, Prabhu N. Design and performance evaluation of subsynchronous damping controller with STATCOM. IEEE Xplore 2006;21:1398–1405. [CrossRef]

[8] Prada M, David FM, Garcia JLD, Muljadi E, Singh M, Bellmunt OG, et al., Contribution of type-2 wind turbines to sub-synchronous resonance damping. Int J Electr Power Energy Syst 2014;55:714–722. [CrossRef]

[9] Bamba S, Gupta SK. Enhancing power system stability through intelligent STATCOM control strategies in torsional oscillation environments. Sigma J Eng Nat Sci 2024. [CrossRef]

[10] Wadhwa K, Gupta SK. Analysing the Design and Performance of a Controller for AGC in Multi-Source Systems Using Cuckoo Search Optimization. IEEE Xplore 2023;1–5. [CrossRef]

[11] Arockiaraj S, Manikandan BV, Alagammal S, Bhavani R. Sub synchronous resonance analysis of inverter-based wind and solar farms using genetic widow optimization. J Power Electron 2024;24:1516–1526. [CrossRef]

[12] Wang L, Lai JT, Li TY, Tseng CC, Mokhlis H, Chua KH. Suppression of Subsynchronous Resonance in a Hybrid Steam-Turbine Generator/Offshore Wind Farm System Using a Static Synchronous Compensator with a Vanadium Redox Flow Battery. IEEE Xplore 2025;1–8. [CrossRef]

[13] Bamba S, Gupta SK. Sub-Synchronous Resonance Mitigation in Series-Compensated Power Systems Using a GWO-Based Optimal Controller. IEEE Xplore 2023. [CrossRef]

[14] Soeprijanto A, Abdillah M. Subsynchronous Resonance Damping Using SMES Optimized By Quantum Behaved Particle Swarm Optimization. 3rd International Conference on Advances in Computing, Control, and Telecommunication Technologies. Jakarta, Indonesia; 2011.

[15] Shivashanker K, Janaki M. Mitigating Torsional Interactions with MMC Connected to Turbine-Generator Using a Novel Subsynchronous Damping Controller. Int J Intell Eng Syst 2025;18: 1275–1290. [CrossRef]

[16] Thota K, Velpula S, Basetti V. A scientometric analysis on DFIG-based wind energy conversion system research trends. Discov Appl Sci 2024;7:7. [CrossRef]

[17] Gupta SK, Gupta AK, Kumar N. Damping subsynchronous resonance in power systems. Inst Eng Technol 2002;149:679–688. [CrossRef]

[18] Gupta SK. Analysis and mitigation of sub-synchronous resonance in power system with svc and using bfo based optimal controller. Int J Recent Sci Res 2014;1:39–44.

[19] Amin MN, Soliman MA, Hasanien HM, Abdelaziz AY. Hybrid CSA-GWO Algorithm-Based Optimal Control Strategy for Efficient Operation of Variable-Speed Wind Generators. Springer 2021;227–245. [CrossRef]

[20] Abraham A. 2009 World Congress on Nature & Biologically Inspired Computing. IEEE 2009.

APPENDIX A.

SVS DATA (quality factor = 100))			
K_I (integral gain in PI controller)	1200sec	T_S (Thyristor delay time constant)	0.005
T_M (Measuring time constant)	0.0024sec	K_D (current droop)	0.01
K_P (Proportional gain in PI controller)	-1	T_D (thyristor avg. dead time)	0.001667

Network Impedances in Per Unit on the Generator MVA Base (100 MVA}, Voltage =400KV, Freq=50Hz			
R	0.012525	XL	0.1197 P.U for whole line
B	3.552	No of conductors	2
Length of line	600km	Area of conductors	0.4inch2
Reactance of line	0.034 ohm/phase/km	Surge impedance	296ohm
Resistance of line	0.0325 ohm/phase/km	Natural load	540MW

Shaft inertia and spring constant for first bench mark model in per unit on Machine base			
Inertia	Inertia Constant H in Sec	Shaft Section	Spring Constant K in pu T/rad
HP Turbine	0.1033586	HP-IP	25.772
IP Turbine	0.1731106	IP-LPB	46.635
LPA Turbine	0.9553691	LPA-LPB	69.478
LPB Turbine	0.9837909	LPB-Gen	94.605
Generator	0.9663006	Gen-Exc	3.768
Exciter	0.03880697		

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00103

Research Article

Exploring transient potential steps on rotating disk electrodes with steady-state concentration profiles in single electron transfer reactions

V. SREELATHA DEVI[1,*], A. KAVITHA[2], M L. SURESH[3], D. VIJAYABALAN[2], T. LOGANATHAN[4], R. MANIMUTHUKUMAR[5]

[1]*Department of Mathematics, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, 602105, India*
[2]*Department of mathematics, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamilnadu, 600085, India*
[3]*Department of Mathematics, Veltech Rangarajan Dr.Sakunthala R&D Institute of Science and Technology, Chennai, 600062, India*
[4]*Department of Mathematics, S.A. Engineering College, Thiruverkadu, Chennai, Tamilnadu, 600077, India*
[5]*Department of English, Easwari Engineering College, Ramapuram, Chennai,Tamilnadu, 600089, India*

ARTICLE INFO

Article history
Received: 22 January 2025
Revised: 22 March 2025
Accepted: 15 June 2025

Keywords:
Background Current Rejection; Closed Form Solution; Potential Step; Rotating Disk Electrode; Transient Solution

ABSTRACT

This study develops a theoretical model to analyze transient currents in one-electron transfer reactions over rotating disk electrodes for quasi-reversible and irreversible processes. The model employs the nearest diffusion layer approximation, separation of variables, variational iteration, and two-point pade approximation methods to derive concentration profiles. The techniques not only yield better precision and reliability of electrochemical measurements through considerations of transient currents and background current rejection in solid-state electrochemical systems, thereby improving the precision and interpretation of experiments devoted to rotating disk electrodes. Analytical expressions for the concentrations of oxidized and reduced species are obtained, enabling the examination of how parameters operate on current during the transition between activation-controlled currents. The steady-state background current has also solutions given in closed form. Lastly, we came to the conclusion that numerical simulation is employed to explain and differentiate between reversible and quasi-reversible electrode processes.

Cite this article as: Sreelatha Devi V, Kavitha A, Suresh ML, Vijayabalan D, Loganathan T, Manimuthukumar R. Exploring transient potential steps on rotating disk electrodes with steady-state concentrationprofilesinsingleelectrontransferreactions.SigmaJEngNatSci2026;44(2):1440–1455.

INTRODUCTION

Electrochemical processes are essential in various fields, including environmental remediation, energy conversion, and storage. Understanding how electrochemical systems change over time is important for improving efficiency and making basic reaction mechanisms clearer. Looking at the steps that can happen on rotating disk electrodes (RDEs) gives us important new information about how electrons and mass move across electrode surfaces. We can fully

***Corresponding author.**
*E-mail address: sreelatu@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

understand single electron transfer reactions at the electrode-electrolyte interface when these studies are used with steady-state beginning concentration profiles. The present work investigates the exciting subject of transient solutions to potential steps at RDEs, shedding light on the complicated connections between reaction kinetics, diffusion, and convection. Our purpose is to clarify through theoretical studies and experimental proof. Hydrodynamic modulation (HMRDE) is often used to control the background currents on the rotating disc electrode (RDE). Research was done by Sonke and his colleagues [1, 2]. Lock-in amplifiers and a theoretical framework are usually used to deal with the system's nonlinearities. Vielstichet al. [3-6]. Conversely, we will devise a theoretical framework for prospective step techniques. The goal of this unique theory is to lessen the effects of background current. It is used in systems where electrochemical solid-state reactions have happened. Our method differs from standard methods in that it emphasizes potential stages, which offer unique advantages when applied to electrochemical systems. By giving us a better understanding of how solid-state reactions work, potential step approaches can better block the background current. Our goal is to make measurements more accurate and consistent in systems that use electrochemical solid-state processes by building a strong theoretical base for these methods. Another major goal is to develop a theory for kinetic transients at the rotating disk electrode (RDE) to study the typical characteristics of one-electron transfer processes. It is especially important to pay attention to important kinetic variables like the charge transfer coefficient and the standard rate constant (k_0). To be more specific, we are looking at a more complicated and probably more accurate method, which is to find the solution to the convection-diffusion equation in a semi-infinite interval. Chen, P. et al. [6-9]. The advanced technique offers profound insights into the kinetic behaviour at the RDE, facilitating the evaluation of electron transfer processes and their associated parameters. The simplest way to solve these transients is to use the Nernst diffuse layer model to turn the convective diffusion equation into a finite interval diffusion equation. Then, with the help of Saranya K. et al. [10-14], we find transient solutions. To do the inverse Laplace transformation and put the transformation back into the time domain, Nguye.T has to figure out the terms and development of the growth and series expansions of Laplace transformation [15-18]. We are using the Nernst diffuse layer model and the method of separation of variables together because they are both well-known ways to solve heat and diffusion equations at a short time scale [19]. This is a mix of a simple but useful way to deal with kinetic transients in electrochemical systems. In a possible step experiment without convection, quasi-reversible systems act like reversible systems when this reversible system happens eventually. Because there is no convection at first, the short-term response of a Rotating Disc Electrode (RDE) may look like a possible step at a stationary electrode. After this first kinetic transient, the RDE's surface concentration can reach a steady state. Based on what Stoll, M. D. et al. [2, 20] showed, it is likely that it is possible to guess how long the predicted transients in RDE dimensions will last. To incorporate kinetics, initial boundary value problems (IBVP) must be addressed using authentic (electrochemical) kinetic boundary conditions, specifically Robin boundary conditions. Although numerical methods have successfully addressed significantly more complex systems, it appears that neither the irreversible nor the quasi-reversible system has been subjected to analytical investigation regarding the potential step of the RDE (Balasubramanian et al. [21]). Chang emphasises the steady-state initial concentration profiles in the study of one-electron transfer. Y et al. [22-24] looked at the temporary solutions of the possible actions at rotating disc electrodes. Their study showed that steady-state concentration profiles are a good way to describe these systems and help us understand how they change over time. This gives us new information about how electron transfer processes work. Liu. H et al. looked at how rotating disc electrodes behave in steady state and how they change with temperature [25-27]. They discovered that the impact of the heat variables significantly influences steady-state behaviour, thereby improving the control and optimisation of electrochemical processes. Evanko et al. investigated the behaviour of rotating disc electrodes under conditions of exceedingly rapid comproportionating reactions [28-29]. They discovered that the limiting current ratios for two versus one electron are the same for both spinning disc electrode voltammetry and sampling current voltammetry. Examining their concentration profile helped make sense of these data and learn more about fast electrochemical reactions. Murray, R.W. et al. [30] conducted a quasi-one-dimensional steady-state investigation of multi-ion rotating disc electrode electrochemical systems. Their research, which considered homogeneous reactions, convection, migration, and diffusion, resulted in a comprehensive model that illustrates the intricate interconnections among these processes. Henderson, L. K., et al. [31] say that these research papers help us understand how spinning disc electrodes act in different situations so that we can make better predictions and take effective steps to improve electrochemical systems. Additionally, Zhu X et al. [32] have done a steady-state computational study of the current density distributions in axisymmetric systems with multi-ion electrolyte systems, focusing on the rotating disc electrode [33, 34]. The Journal of Electroanalytical Chemistry published a study by Doherty, M. F., et al. [35] that showed how useful the numerical model is for optimising rotating disc electrode applications. This is because the model does a good job of showing the current density distributions [36]. When used in many different settings, these works greatly improve our understanding and ability to predict how rotating disc electrode conduct will behave. This makes it easier to optimise electrochemical systems, according to Lopez, R. L., et al. [37]. To make sure that the RDE stays

stable after each step, it is common to measure it using slow staircase linear sweep voltammetry at different rotation speeds. By changing our mathematical model, we are trying to fit and repeat these measured transient sequences. With this change, a new theory will be created that will stop at the step's steady state and start at the steady state of the applied potential. In their study, Mosayeb Gholinia et al. [38] looked into how heat moves in a two-phase rotating system with a magnetic field while nanofluid flows through it. The basic partial differential equations of mass, momentum, and energy are made easier by turning them into nonlinear ordinary differential equations. Akbari-Ganji's Method can be used to solve these equations. The Galerkin method was used by Mohammad Reza Zeinolabedini and others [39] to study the free vibrations of a rotating disc. MertolTufkec et al. [40] made a simpler model of the surge and stall when there is a distributed force on a rotor disc system. It is a basic and conceptual framework for future research in the field. [41]. After that, the two real parameters affect the properties of the modified Bernstein-type operator's approximation. Some of the immediate effects are an error estimate for functions in Lipschitz-type spaces, an error estimate for the ordinary and Ditzian-Totik modulus of smoothness, and the Voronovka-type asymptotic theorem. There is also a discussion about how close these operators can be to a Kirov and Popova type. The study wants to look at the power-law nanofluid flow of titanium dioxide on water across a small needle [42]. Also, because it was a full case study of quadratic, nonlinear, and linear radiation, heat radiation was added and looked at. This [43] study looks at how important the activation energy is for starting a chemical reaction in a two-dimensional micropolar fluid laminar flow across a semi-infinite inclined plate. When modelling how heat moves through a fluid, we use convective thermal boundary conditions and radiation.

Literature Review

Electrochemical processes are important in a lot of areas, from cleaning up the environment to changing and storing energy. It is essential to comprehend the transient behavior of electrochemical systems to maximize their efficiency and clarify basic reaction mechanisms. One way to learn more about the kinetics of electron transfer and mass transport at electrode surfaces is to investigate possible steps on rotating disk electrodes (RDEs). Benjamin Levich formulated the foundational theory that elucidates mass transport phenomena at a rotating disk electrode (RDE) at the Institute of Electrochemistry within the Academy of Sciences of the USSR. Levich articulated the theory in his seminal work, Physiochemical Hydrodynamics, which was first published in Russian in 1952. A decade later, Levich's book was translated from Russian to English, leading to the RDE gaining greater recognition among Western scholars. During the early 1960s, Stanley Bruckenstein at the University of Minnesota, along with his students Dennis Johnson and Duane Napp, collaborated with Ronnie Bell at Oxford University and his student John Albery on the study of rotating electrodes. The rotating disk electrode has become a well-known tool for studying how electrochemical reactions happen over time, building on the work of earlier generations. Laminar flow at a rotating disk electrode makes it easy for material to move from the bulk solution to the surface of the electrode. Although the bulk solution at a distance from the electrode is effectively mixed due to the convection caused by rotation, the region of the solution closer to the electrode surface tends to move in sync with the electrode's rotation. Therefore, when considering the solution from the perspective of the rotating electrode surface, it appears to be relatively motionless. These studies give a thorough understanding of single electron transfer reactions at the electrode-electrolyte interface when combined with steady-state starting concentration profiles. Electrochemical methods [44, 45] are popular chemical characterization techniques used in the chemical industry, building construction, safeguarding the environment, and damage tests. This is owing to the low cost of measurement equipment and the availability of standard processes. Aside from its potential application in chemical analysis, electrochemistry is significant in the operation of solar cells, batteries, and electrolyzes that convert electrical to chemical energy, such as hydrogen synthesis by water electrolysis or CO_2 reductive recycling. Electrochemistry is currently the most promising approach for temporarily storing renewable energy. Accurate knowledge of the mechanism of electrochemical reactions is essential for better understanding and future development of these technologies. The rotating ring-disk electrode (RRDE) is a popular and widely used technique for researching electrode response mechanisms [46-52]. Frumkin, Levich, and their coworkers came up with the first RRDE or generator-collector electrode assembly in 1959. [53]. When using an RRDE, we use forced convection caused by rotation to make sure that electroactive species created at any one potential on the disc move to the surrounding ring, where a new reaction can happen at a new potential. This changes the measured ring current. The ability to independently change the potentials of the two electrodes of the RRDE makes it a powerful tool for finding and understanding the actions of intermediates that may not be stable. RRDEs are popular for several reasons: A solid theory has existed for a long time; the RRDE provides well-regulated hydrodynamics; the mass transport rate can be easily adjusted by varying the rotation speed; the electrodes are easily cleaned and polished; and the RRDE is housed in a vessel that can be easily de-aerated, and temperature controlled. [54] To use RRDEs to identify electrode reaction products, the following approaches are most used: - The disk is kept at a constant potential, whereas the ring is kept at a variable potential where the disk reaction products are expected to transport electrons. A current-potential curve forms at the disk, while the ring potential remains constant. The disk is held at the potential where the reaction of interest occurs, and a current-potential curve is

recorded at the ring to capture disk reaction products with different redox properties. The strategies outlined above share a commonality: at least one of the two electrodes must remain unaltered throughout the experiment. The potential of both electrodes should not be impeded by any theoretical constraints, enabling the researcher to modify the two independently and simultaneously. But any potentiostat, even the ones that most instrument suppliers sell, should be able to control the potential of two working electrodes at the same time with waveforms that change over time. This is usually only possible with the software that runs the instruments. Unfortunately, most of the commercial software that is available today does not support any of the processes that are not part of the standard mentioned above. The development of electrochemical measuring systems that allow for dynamic and simultaneous control of two working electrodes, particularly the dynamic control of the potential applied to the RRDE disc and ring electrodes, is a recent achievement. The mechanics of electrochemical processes have been shown to be examined through the "dual dynamic voltammetry" (DDV) technique. A rotating disc electrode (RDE) setup is a common way to test the electrochemical properties of new OER catalysts [55, 56]. The RDE is a well-known and commercially available electrochemical technique that is also experimentally complex. To determine the catalyst's activity, quantifiable metrics such as the onset potential (the ability to achieve a particular current density) are typically utilised. The principle of onset potential is not widely accepted, making it challenging to compare data from different authors regarding various materials. To address the problem of definition, there are many attempts to standardise and measure OER catalysts [57, 58-63]. There is no universally accepted method for determining comparability. Also, RDE studies are often only about materials that can be electrodeposited on the supporting electrode or powder samples. The development of the catalyst film is especially important for the second part, and even small changes in the experiment could have a big effect on how well the RDE measurements work.The rotating disk electrode (RDE) is a conventional hydrodynamic electrochemical method that employs controlled convective mass transfer. The working electrode is rotated, causing a consistent laminar flow of analyte to the active surface [64, 65]. The rates of mass transfer are not only regulated but also easily altered, allowing for a wide range of reaction time scales and the investigation of kinetic and mechanistic information [65-67]. RDEs typically consist of an insulation (PTFE) layer surrounding an active material disk (Pt, Ni, Cu, Au, Fe, Si, CdS, GaAs, glassy carbon, or graphite). Other materials have been utilised, including boron-doped diamonds [25]. Single-atom catalysts have attracted interest because they can connect homogeneous and heterogeneous catalysis in a way that is both highly active and selective [69-71]. Lu et al.'s study shows how the electronic interactions between single metal atoms and their support can be used to improve catalytic performance. This agrees with earlier studies that showed how important EMSI is in catalytic processes. For example, a study on hybrid organic-inorganic perovskites showed that adding single Pt atoms to perovskite structures can improve their catalytic properties because the metal atoms and the support material interact with each other in a way that is good for the reaction [72]. The authors come up with a new way to use the ring electrode in an RRDE setup as a "shield" for the current. Keeping the ring electrode at the same potential as the disc electrode changes the distribution of the electric field, which makes the current flow more evenly across the disc surface. This setup reduces edge effects and potential gradients that usually mess up kinetic measurements in resistive media. The study shows that this shielding method works by using both computer simulations and real-world tests to show that it does what it says it does [73, 74]. The authors probably used finite element modelling to simulate the electric field and current distribution, and then they used experimental electrochemical measurements to check their theoretical predictions. The enhanced uniformity in current distribution increases the dependability of kinetic parameters derived from methodologies such as Tafel analysis and Koutecký–Levich plots [75]. For making hydrogen in a way that is good for the environment, the Oxygen Evolution Reaction is an important part of water electrolysis. Metal-based catalysts, especially those with precious metals like Ir and Ru, have been used for a long time because they have good d-electron configurations. But because these metals are rare and expensive, we need to find effective alternatives that don't use metals. Researchers have looked into carbon materials that don't contain metals, but their catalytic activity is often limited because there aren't many good ways to change the p-band structures of non-metal atoms. This research advances prior initiatives aimed at modifying the electronic structure of electrocatalysts to enhance OER performance. Conventional methods have concentrated on doping, alloying, and generating oxygen vacancies in metal-based catalysts [76]. Many biological and chemical systems, such as photosynthesis and respiration, depend on proton-coupled electron transfer processes. It is important to know if these reactions happen in steps (sequential proton and electron transfers) or all at once (simultaneous proton and electron transfers) in order to make energy conversion systems that work well. The researchers used high-pressure pump-probe spectroscopy to study the excited-state PCET of a ruthenium complex. They could see changes in the pathways and rates of reactions by changing the pressure and concentrations of buffers or quenchers [77]. Electrochemical Impedance Spectroscopy (EIS) is a common method for studying how fast and how electrochemical reactions happen. People usually use equivalent circuit models to make sense of EIS data. These models need to make some assumptions about the system first. The DRT method provides a model-free alternative by deconvoluting impedance spectra into a distribution of relaxation times, enabling the

identification of individual electrochemical processes without reliance on predefined models. The authors utilised the DRT method on EIS data derived from RDE experiments concerning the ORR in alkaline environments. They looked at the resulting impedance spectra to find different relaxation processes by changing things like oxygen saturation levels, rotation rates, and current densities [78]. The Rotating Disc Electrode (RDE) is a widely used electrochemical technique and tool for studying how fast electrochemical reactions happen when the conditions for mass transport are well-controlled. A rotating disc electrode (RDE) has a flat disc electrode (usually made of glassy carbon, platinum, or gold) that is attached to a shaft that turns. When placed in an electrolyte and spun at a controlled speed, the rotation causes a laminar flow that brings reactants to the electrode surface and keeps products from building up. The authors utilised a combination of RDE voltammetry and differential electrochemical mass spectrometry in a flow cell setup to investigate BOR on Pt/C, Au/C, and AuPt/C thin-film electrodes. This method made it possible to measure both electrochemical activity and the formation of gaseous products at the same time. This gave us information about reaction pathways and intermediate species [79]. In electrochemical studies, rotating disc electrodes are very important for controlling mass transport and studying reaction kinetics. But commercial RDE systems are often too expensive for most people, which makes them hard to get, especially in schools and labs that don't have a lot of money. There are some cheap designs, but they usually need parts that have been precision-machined, which makes things even harder. The prototype showed clear and repeatable electrochemical data, which proved that it could be used for research and teaching. Standard electrochemical experiments were used to test how well the system worked, which showed that it could be used for a variety of analytical tasks [80]. Rotating disc electrode-optical emission spectrometry (OES) is a common way to find trace elements in oils. This is important for things like diagnosing machinery wear and checking the quality of lubricants. But in most RDE-OES setups, the upper rod electrode stays still while the lower disc electrode moves. This setup causes uneven wear on the electrodes, inconsistent arc gaps, and changing spectral intensities, which makes it harder to get accurate and reproducible results. Huang et al.'s method directly addresses this issue by redesigning the electrode configuration, providing a viable solution that improves the durability of the electrodes and the dependability of the analytical outcomes [81].

The paper is given in this order: We give mathematical formulas for rotating disc electrodes with steady-state concentration profiles in single electron transfer operations in Section 2. Using the variational iteration method, Section 3 looks into the concentrations of species in analytical solutions. Section 4 talks about analytical and numerical solutions, using graphs and table values to do so. The proposed theoretical model culminates in a definitive system for examining transient currents in one-electron transfer reactions on rotating disc electrodes for substances undergoing quasi-reversible and irreversible processes, as illustrated in Section 5.

MATHEMATICAL FORMULATIONS OF THE PROB

We start with steady-state RDE concentration profiles and use the separation of variables method to solve the equations for a potential step. A diffusion equation that is limited by the Nernst diffusion layer comes close to the convection-diffusion equation. We don't consider the second layer's charging or any possible infrared losses in any situation. We also agree that the diffusion coefficients for the reduction (Red) and oxidation (OX) species are the same. The reversible scenario doesn't consider how fast the reaction really is; instead, it thinks that the system is instantly in electrochemical balance. It is helpful to discuss when reversibility is obtained within the potential step framework that is when heterogeneous kinetics cannot be detected. This will never be the case for the mathematically treated system since we can always expand it to lower times before hitting kinetic limits. Though measurements are restricted to periods after the double-layer charge has deteriorated, we are unable to sample at a rate greater than the data capture unit's maximum sampling rate. A reversible theory that has been extended to linear initial circumstances may help achieve that goal. It may also be applied to moderately fast kinetics, where surface concentration relaxation is still discernible but occurs during the phase of the reaction where convection is not yet influencing the process (Fig. 1). Strictly speaking, though, we must consider reversible and quasi-reversible systems have different steady-state surface concentrations, which prevents a one-to-one match with the potential step theory [44]. We compare the reversible and irreversible systems in the following section, which addresses this issue. The reversible situation is first. Since we begin in the steady state of a prior phase, we have a system in which the initial OX and Red concentrations are a mirror copy of one another, equipoised by their majority concentrations. This means that throughout the step, the surface concentrations of OX and Red remain constant and are equivalent to their corresponding reversible steady-state concentrations. After that, the OX and Red systems can be handled independently, and the current can be found using either gradient.

$$i = nFAD\left(\left[\frac{2(C_{s,init} - C_{s,ss})}{L}\right]\sum_{n=1}^{\infty} e^{\left(\frac{C_b - C_{s,ss}}{L}\right) - \left(\frac{n^2\pi^2 D}{L^2}t\right)}\right) \quad (1)$$

The symbol C_b denotes the total concentration of a given species in a solution. The steady-state concentration, denoted $C_{(s,ss)}$ as is, is calculated using the Nernst equation and bulk concentrations. The term $C_{(s,init)}$ refers to the species' initial concentration before any changes occur. Finally, L denotes the thickness of the Nernst diffusion layer, which

is a region close to the electrode surface where diffusion dominates mass transfer [6-8].

$$L = 1.61\, D^{\left(\frac{1}{3}\right)}\omega^{-\left(\frac{1}{2}\right)}V^{\left(\frac{1}{6}\right)} \tag{2}$$

For the equation $C_{s,init} = C_b$, it reduces to the known formula with $i_{ss} = nFAD(C_b\text{-}C_{s,ss})/L$:

$$R(t) = \frac{i(t)}{i_{ss}} = 1 + \sum_{n=1}^{\infty} e^{-\left(\frac{n^2\pi^2 D}{L^2}t\right)} \tag{3}$$

While formula (3) works well for long-time scales, a different series is better suited for shorter times. The Laplace transform method is used to get this alternative series, which is written as a sum of negative exponential terms. This new series is better because it converges more quickly, which means it gets to the right level of accuracy more quickly for calculations that need to be done quickly [15, 18].

$$i(t) = \left(\frac{nFAD^{\left(\frac{1}{2}\right)}C_b}{\pi^{\left(\frac{1}{2}\right)}t^{\left(\frac{1}{2}\right)}}\right) = \left[1 + 2\sum_{n=1}^{\infty} e^{-\left(\frac{n^2\pi^2 D}{L^2}t\right)}\right] \tag{4}$$

The infinite series solution shows good agreement with the results obtained from the finite element method (FEM) (Fig. 2) simulation of the convection-diffusion equation. However, a relative inaccuracy of up to 10% (Fig. 3) is observed. This level of discrepancy is significant and only occurs in short times. For intermediate time scales, the agreement between the two methods is much better. We look at reaction rates (kinetics) using fixed-rate constants (k values), but we only care about finding out how the concentration of a single species changes.

$$\left(\frac{\partial_c(x,t)}{\partial t}\right) = \left(D\frac{\partial^2 c(x,t)}{\partial x^2}\right)_{x[0,L],t\geq 0} \tag{5}$$

$$\begin{aligned} &c(x,t=0) = c_b + \left(\frac{c_s, init - c_b}{L}\right) \\ &\left(D\frac{(L,t)}{dx}\right) + kc(L,t) = 0 \\ &c(0,t) = c_b \end{aligned} \tag{6}$$

Using the following dimensionless variables

$$\chi = \frac{x}{L}, T = \frac{Dt}{L^2}, C_1 = \frac{c}{c_b}, L = 1.61\, D^{\frac{1}{3}}\omega^{-\left(\frac{1}{2}\right)}v^{\frac{1}{6}} \tag{7}$$

The equation (5) becomes

$$\left(\frac{\partial C_1}{\partial T}\right) = \left(\frac{\partial^2 C_1}{\partial \chi^2}\right) \tag{8}$$

The boundary conditions are

$$\begin{aligned} &C_1(\chi,0) = 1 + \alpha x \\ &C_1(0,T) = 1 \\ &\left(\frac{\partial C_1}{\partial \chi}\right) + bC_1 = 0\ ,\ \text{When} x = 1 \end{aligned} \tag{9}$$

Concentrations of the species in analytical solutions using variational iteration method

$$C_1(\chi,T) = (1+\alpha x)\left[1 - 7(\alpha^2 T) + \frac{5(\alpha^2 T)^2}{2} - \frac{(\alpha^2 T)^3}{6}\right] \tag{10}$$

We can write the equation (10) as follows.

$$C_1(\chi,T) = (1+\alpha x)\left[\sum_{n=1}^{3} 1 + b_n(\alpha^2 T)^n\right] \tag{11}$$

When,

$$b_1 = -7,\ b_2 = \frac{5}{2},\ b_3 = \frac{-1}{6} \tag{12}$$

We construct the two-point Pade approximation such that in equation (10) the coefficient of equation (11) is reproduced namely

$$C_1(\chi,T) = \left(\frac{p_0 + p_1(\alpha T) + p_2(\alpha T)^2}{1 + q_0(\alpha T) + q_1(\alpha T)^2}\right) \tag{13}$$

Where, $p_0 = 1$, $p_1 = q_0$, $p_2 = b_2 + b_1\, q_0 + q_1$

$$q_1 = \left(\frac{b_1 b_3}{b_1 b_3 - (b_2)^2}\right), q_2 = \left(\frac{(b_3)^2}{(b_2)^2 - b_1 b_3}\right) \tag{14}$$

The numerical values of p_0, p_1, p_2, q_0, q_1 are

$$\begin{aligned} &p_0 = 1,\ p_1 = -7.0243, p_2 = 2.6755,\ q_0 = -0.0243, \\ &q_1 = 0.0054492 \end{aligned} \tag{15}$$

Consequently,

$$C_1(\chi,T) = (e^{ax})\left[\frac{1 - 7.0243(a^2T) + 2.6755(a^2T)^2}{1 - 0.0243(a^2T) + 0.0054492(a^2T)^2}\right] \tag{16}$$

Now the current becomes.

$$j = \frac{i}{nFAk} = a\left[\frac{1 - 7.0243(a^2T) + 2.6755(a^2T)^2}{1 - 0.0243(a^2T) + 0.0054492(a^2T)^2}\right] \tag{17}$$

RESULTS AND DISCUSSION

Figure 1 illustrates the bulk concentration C_1 (χ, T) of a species vs. dimensionless distance X. In the graph, multiple lines are seen, one for each value, as indicated in the parameter that most likely determines the profile of concentration. Corresponding to different values of the parameter,

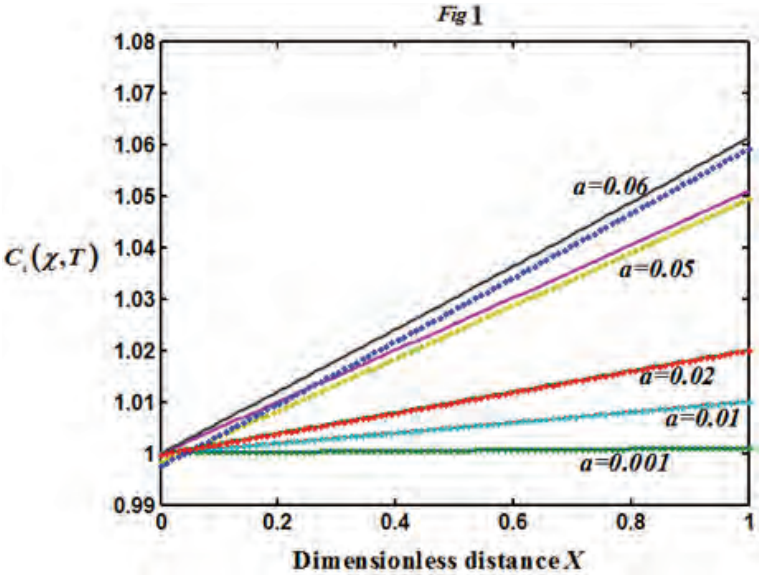


Figure 1. Comparison of concentration of C_1 for values of T= 0.1.

five values have been shown: The values displayed are 0.001, 0.01, 0.02, 0.05, and 0.06. As the value increases from 0.001 to 0.06, the slope of the lines increases. With lower values (0.001 and 0.01), the concentration profiles straighten out even more, showing a nearly horizontal line on the y-axis, which means there are minimal changes in the sum of concentrations along the distance (on the X). The concentration profiles become more sloped as X goes down. This means that changes in concentration with distance are bigger for higher amounts of the gold label (0.05 and 0.06). This concentration changes more over the two-dimensional distance X as the value goes up. This makes sense from a physical point of view, as it could be related to a material property (reaction rate, diffusion coefficient).

The analytical solution is depicted by the dotted lines, while the numerical solution is shown by the solid lines. The concentration of a certain component as a function of dimensionless distance (X) is given in Figure 2. There are several lines in the graph, each equivalent to a different value of the dimensionless time T. If we increase T from 5 to 12,

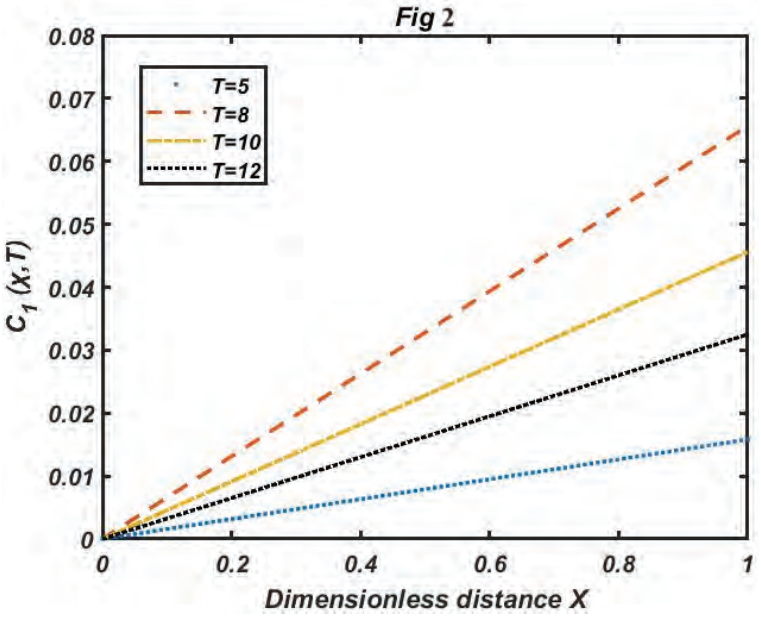


Figure 2. Plot of concentration profiles of C_1 versus dimensionless distance X calculated using equation (16) for values a = 0.001 and various values of time parameter T in linear transition.

the slope of the lines will get larger. Then the concentration profile is flat, as it indicates that the concentration changes gradually with distance for lower values of T (e.g., 5). The other consequence of the profile represented in the graph at higher values of T (at T = 12) is still a steeper profile, which indicates greater variation of concentration with distance. Larger values of T due to a wider change in concentration over the dimensionless distance X indicate higher progress concerning time towards the reaction or process. It shows how the concentration of species changes with distance and time and how these levels help in understanding the behavior of the system under different circumstances.

The concentration of a species C_1 (χ, T) vs. dimensionless distance X is shown in Figure 3. For each dimensionless time T, the graph has several lines. The concentration profiles behave very differently as T passes from 5 to 12. In this case of T = 5, the concentration decreases slightly from around units and rises continuously across the dimensionless distance. The concentration is less sensitive to X near the center of the model, but for high values of T it rises more sharply with X. For T = 8, T = 10, and T = 12, the corresponding lines show increasing concentrations. When the high values of T are taken, as shown in, it results in higher concentrations over the dimensionless distance X, therefore the process or reaction occurs more with friction effects inside.

N - Numerical solutions, A – Analysis, E – Errors.

Figure 4 shows a few examples comparing linear transition to logarithmic transition; In this figure, two transitions are plotted using the dimensionless distance X on the x-axis and the y-axis. Table 1 provides a clear representation of the numerical value of Figure 4. There is one more with an arc, a logarithmic transition T = 5. The figure below suggests that the logarithmic transition curve is steeper than the linear transition. This indicates that the y value changes more rapidly for the logarithmic transition as compared to the linear transition. In a linear transition, the Y value varies equally while the change in the X axis, which means it will either

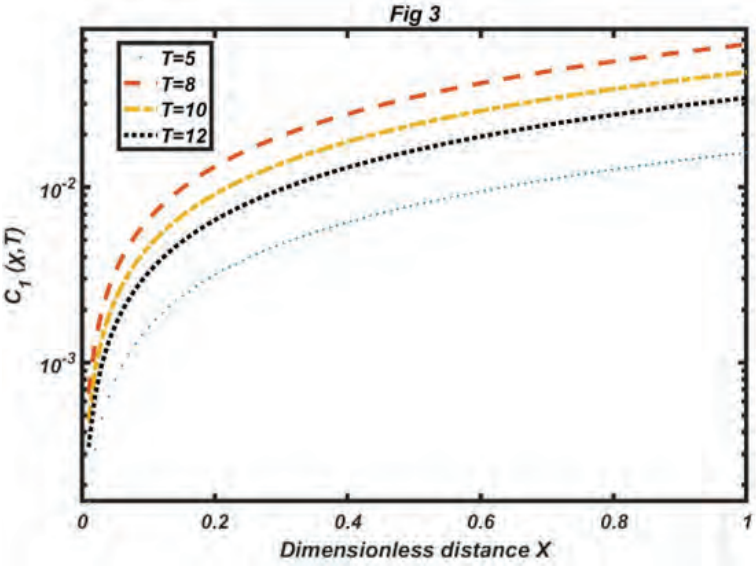


Figure 3. This plot shows the concentration profiles with dimensionless distance (x) calculated using equation (16) focuses on a fixed value of a = 0.001 and explores various time constants (T) displayed in log transition.

Table 1. Shows the concentration profiles of a C_1 on a fixed value of a = 0.001 with variousvalues of time constants (T)

X	T = 0.1; a = 0.01			T = 0.1; a = 0.02			T = 0.1; a = 0.05			T = 0.1; a = 0.06		
	N	A	E	N	A	E	N	A	E	N	A	E
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.003195	0.003001	0.0001	0.1326	0.1253	0.0073	0.009213	0.009110	0.0001	0.006568	0.006542	0.0000
0.4	0.00639	0.0062	0.0001	0.02652	0.02531	0.0012	0.01843	0.01721	0.0012	0.01314	0.01302	0.0001
0.6	0.009585	0.00955	0.0000	0.03845	0.03822	0.0002	0.02718	0.02617	0.0010	0.0197	0.0184	0.0013
0.8	0.001262	0.00115	0.0001	0.05237	0.05212	0.0002	0.03639	0.03625	0.0001	0.02594	0.02551	0.0004
1	1	1	0	1	1	0	1	1	0	1	1	0

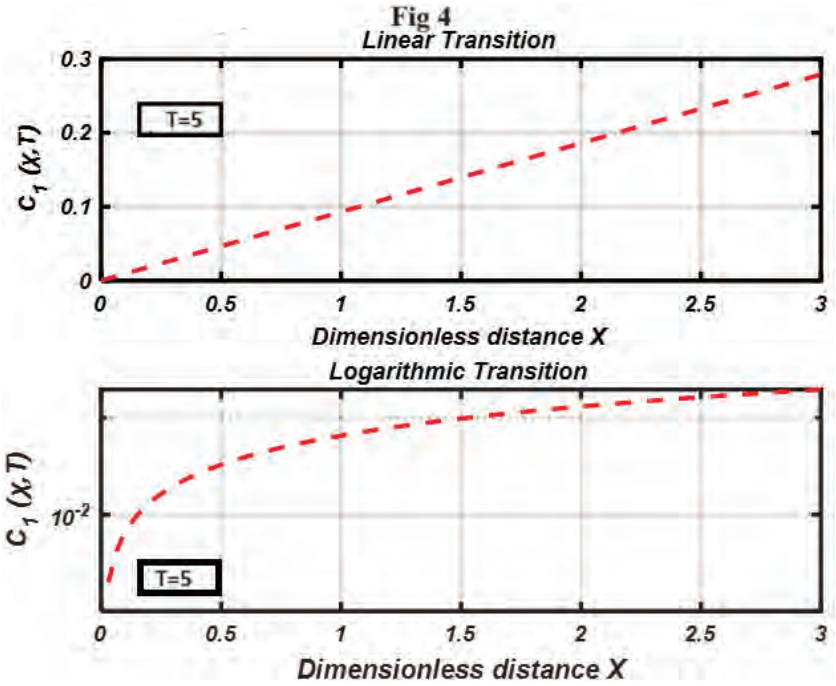


Figure 4. Represents Linear transition and Logarithmic transition at T=5 for concentration profiles C_1 versus dimensionless distance X.

steadily increase or decrease based on what this variable is. The linear one is kind of a steady change; the other one is more like a sudden change followed by a longer phase. The graph illustrates that when the two transitions are done over the same distance X, a logarithmic transition results in a faster and more drastic change of the represented variable Y compared to a linear transition. In other words, the logarithmic transition takes less distance to get to the final state 0.2 more than the linear transition takes distance to get to 0.1.

Figure 5 shows the relationship between dimensionless distance and a variable labeled C_1 (χ, T). All curves in the graph portray a decreasing pattern as dimensionless distance X rises from 0 to 0.9. The decrease level is influenced through "a"; that is, higher values of 0.05 and 0.06 compared to 0.03 and 0.04 depict a faster decrease in C_1. This is shown in the graph, indicating a significant effect of the parameter "***a***" on how C_1 (χ, T) behaves. This serves as a constraint on the rate of change. The greater the value of "a" causes thesteeperthedrop-off in C_1 (χ, T). To achieve a greater x change or higher values of C_1 (χ, T). They are reached more quickly when x increases as "*a*" increase.

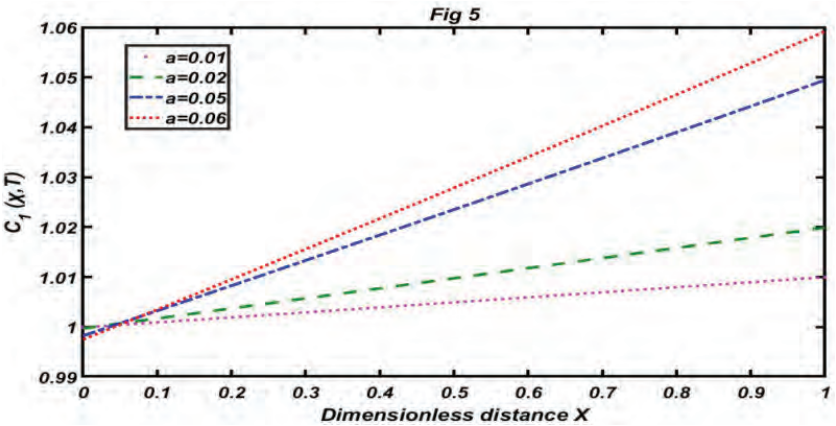


Figure 5. This plot illustrates the concentration profiles of C_1 at a fixed time constant (T = 0.1) against dimensionless distance (x), calculated using equation (16) that explores how the concentration profiles change for various values of parameter 'a' on a linear transition.

Table 2. Shows the concentration profiles of C_1 at a fixed time constant (T = 0.1) for various values of parameter '*a*'

X	T = 0.1; α = 0.01			T = 0.1; α = 0.02			T = 0.1; α = 0.05			T = 0.1; α = 0.06		
	N	A	E	N	A	E	N	A	E	N	A	E
0	0.9975	0.9952	0.0023	0.9999	0.9975	0.0024	0.9983	0.9932	0.0051	0.9981	0.9970	0.0011
0.2	1.002	1.000	0.002	1.004	1.001	0.003	1.008	1.002	0.006	1.01	1.0000	0.01
0.4	1.004	1.001	0.003	1.008	1.003	0.005	1.019	1.012	0.007	1.021	1.0010	0.0200
0.6	1.006	1.002	0.004	1.012	1.010	0.002	1.028	1.013	0.015	1.034	1.0222	0.012
0.8	1.008	1.004	0.004	1.016	1.011	0.005	1.039	1.015	0.024	1.048	1.034	0.014
1	1	1	0	1	1	0	1	1	0	1	1	0

N - Numerical solutions, A – Analysis, E – Errors.

Figure 6 showsthatthe curves all start from y = 1.05 and decrease with increasing x. For α = 0.05 and α = 0.06, the curves fall off more rapidly than for α = 0.01 and α = 0.02 (figure). The figure shows that the parameter 'a' determines how rapidly it decreases concerning the dimensionless distance. Table 2 provides a clear representation of the numerical value of Figure 5. Larger "a" values result in the curve decreasing faster. This illustrates that the variable greatly influences the functionality of the error function. It is stiff on the rate of (error concerning distance). The steepness of the decrease of error for a fixed increment dimensionless distance X and logarithmic scale as a function of "*a* " values. That is that the error "e" decreases faster to 0 as a tends towards infinity exponentially.

Figure 7: The linear transition (Fig. 4) exhibits a constant rate of change. It is a straight line as the change in y value is the same for each consecutive x. These logarithmic transitions consist of a greater initial decrease and then a slower decrease at the end. And a curved line that goes from much steeper to flatter. In the figure, the curve of the logarithmic transition is steeper than the linear transition. This means that the y value increases faster for a logarithmic transition compared to a linear one. It communicates the differing effects of linear and logarithmic onsets impressively, even portraying the specific move faster and more abruptly, a logarithmic onset is compared to a linear gamma.

The variation of a parameter '*a*' on C_1 (χ, T) is illustrated in Figure 8. The bottom boundary curve of all curves decreases with the dimensionless distance X by 0.9, and the increase is far less than the decrease, while the other two sides also. Fewer changes have taken place in decreasing or changing trends. The rate of decrease is determined by the value of "***a***". Higher centrality of the source hubs: Larger values of '*a*' (0.05 and 0.06) decrease fast. Larger "*a*" values lead to a faster decay of the width growing in dimensionless distance X, i.e., a higher rate at which C_1 (χ, T) approach lower numerical results as the X increases.

Figure 9 illustrates how applying a parameter 'a' would influence some function, which presumably can be an error in C_1. We note that, on a logarithmic scale, the dimensionless distance X is increased from 0.01 to 10; all curves illustrate a descending trend. The decrease rate of "C_1" is dependent on the value of "*a*", where higher values (0.05 and 0.06) cause faster mage, compared to lower ones (0.01 and 0.02). Larger values of "a" give a greater drop in error

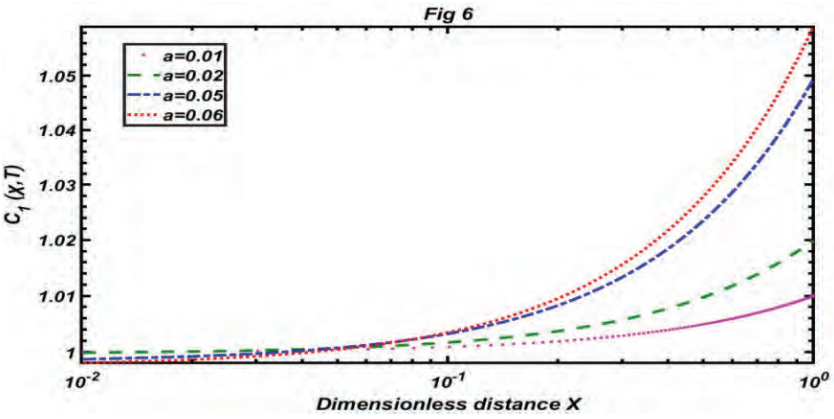


Figure 6. This plot illustrates the concentration profiles C_1 at a fixed time constant (T = 0.1) is plotted against dimensionless distance (x), calculated using equation (16) that explores how the concentration profiles change for various values of parameter 'a' on a log transition.

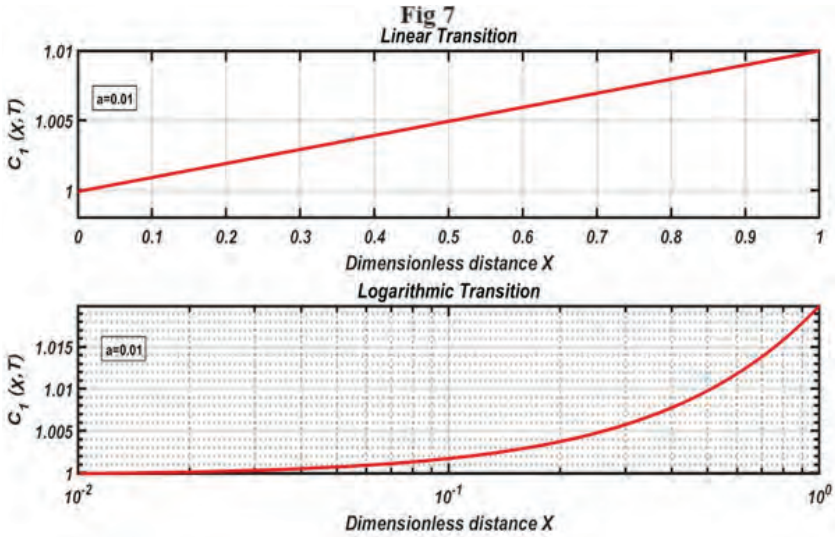


Figure 7. Represents Linear transition and Logarithmic transition at a = 0.01 for concentration profiles C_1 versus dimensionless distance X.

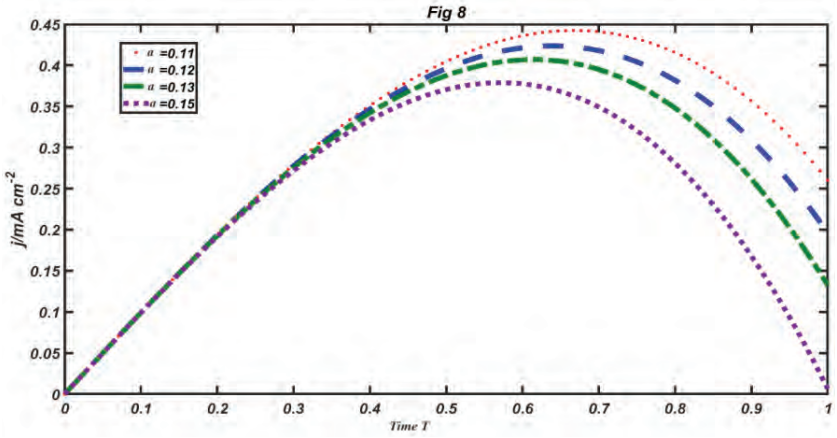


Figure 8. This plot visualizes the concentration profiles of C_1 with dimensionless distance 'a' calculated using equation (17) andexplores the concentration profiles that evolve for various time constants (T) displayed on a linear transition.

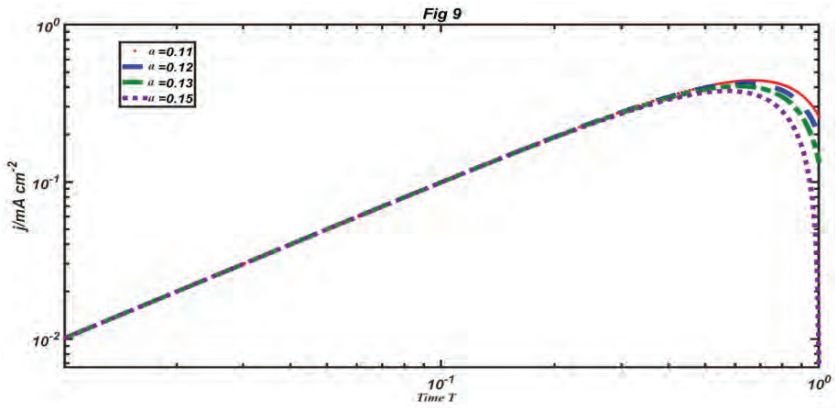


Figure 9. This plot visualizes the concentration profiles of C_1 with dimensionless distance 'a' calculated using equation (17), and explores the concentration profiles that evolve for various time constants (T) displayed on a log transition.

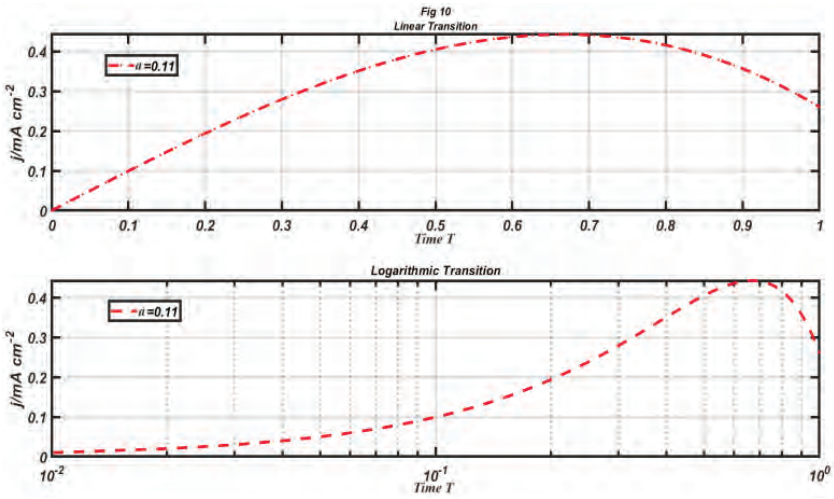


Figure 10. Represents Linear transition and Logarithmic transition at T = 0.11 for concentration profiles of C_1 versus dimensionless distance*a*.

per change in dimensionless distance (X) on a logarithmic scale. That is, the error "*a*" decay more quickly with distance since "*a*" increases exponentially.

In Figure 10, you can see the current density vs. time for both linear and logarithmic transitions on an electrode system with a rotating disc. The figures show the RDE under linear and logarithmic coverage changes. A graph of a straight line through time shows that j first rises single exponentially, reaches a peak at T=0.5, and then slowly falls back down. This suggests that the states were moved through progressively slower processes. Plotting the same values on a long-time scale graph, as shown in Figure 4, helps clarify what occurs at early times, i.e., there is an initial small (j) that grows more pronounced with time, reaching a peak value before decreasing again. The parameter is probably related to the rate or capacity of this process, which influences the temporal pattern observed. Overall, the study gives a more complete picture of how time works in this dramatically changing situation because it looks at both linear (short hours-days) and logarithmic (long days-month-year) time scales, which are important for understanding how things change quickly.

This study develops a theoretical model for an electrochemical reaction taking place on a rotating disc electrode (RDE). The model focuses on single-electron irreversible and quasi-reversible electron transfer processes. This method relies heavily on both the separation of variables and the closest diffusion layer estimate. Using the variational iteration method and the two-point Pade approximation, we can find analytical equations for the concentrations of OX and Red. The steady-state background current is also calculated analytically from the first transient, which leads to the closed-form answer. The study also looks at how different factors affect the current. It focuses on the one that has the biggest effect on the current expression during the transition period that is controlled by flow. If we add linear initial conditions to the current theory, it might work better for reactions that can't be stopped and start with a constant concentration. This kind of addition would be especially useful when reaction times are short. The model might be able to show the short changes in concentration at the electrode surface (called "relaxation") before the steady state in this case. But it's very important that this relaxation happen in a time period that isn't affected by convection. Adding linear initial conditions to the theoretical framework would make it more accurate in situations where the initial concentrations aren't all the same but change in a straight line across the system. This is especially important when the starting concentration is different because of the type of material or how it was handled before. Fast-kinetic reactions happen quickly, so you can see changes in surface concentration before convection or other outside factors become important. This makes it easier to understand how the reaction works on its own. So, a more complete theory that takes into account linear starting conditions could make analyses and predictions in these kinds of systems more accurate. Adding linear starting conditions to the theoretical model would make it easier for researchers to study processes that can't be undone and have steady-state initial concentrations. This is especially true when the kinetics are fast and the convection effects are small.

SOME PRACTICAL UTILIZATION OF RDE RESEARCH

Rotating disk electrodes (RDEs) are commonly employed in electrochemistry for a variety of purposes. Here are a few significant applications:

Electrocatalysis: RDEs are used to study electrocatalytic reactions like oxygen reduction reaction (ORR) and

hydrogen evolution reaction (HER). Researchers can test the efficiency and kinetics of catalysts by varying mass transfer to the electrodes.

Corrosion Studies: **RDEs** help to investigate the corrosion characteristics of metals and alloys. Analyzing current responses to different potentials allows one to figure out the susceptibility of materials to corrosion in a variety of conditions.

Fuel Cells: **RDEs** are employed in power cell studies to assess catalyst effectiveness during electrochemical reactions (for example, methanol oxidation or hydrogen oxidation), which helps to optimize catalyst compositions for improved fuel cell efficiency.

Battery Research: RDEs are used to study the motion of reactions between electrodes in batteries, namely lithium-ion and metal-air batteries. They offer details on charge transfer mechanisms and mass transportation limits.

Analytical Chemistry*:* RDEs are used to identify electroactive compounds in complicated combinations. The regulated atmosphere enables precise current measurements, which aid in the detection of substances at low quantities.

Environmental Monitoring: RDEs can be used to detect pollutants in soil as well as water. They help to determine the electrochemical reactions of contaminants, which allows for the development of environmental surveillance equipment.

Research in Fundamental Electrochemistry: RDEs are utilized to study essential electrochemical reactions such as reaction kinetics, mass transport, and species adsorption on electrode surfaces.

Surface Science: They are employed in surface studies to investigate the adsorption and desorption of molecules on electrode surfaces, revealing information on surface alteration and the functionalization process.

Rotating disk electrodes play a vital role in basic and applied electrochemistry research. Their ability to supervise mass transit and measure current responses makes them helpful in a wide range of fields, including catalysis and environmental studies.

CONCLUSION

We utilized the separation of variables method to resolve a finite Nernst layer approximated diffusion equation accompanied by boundary conditions. This lets us find the transients of reactions that can't be undone and reactions that can be undone almost completely at a rotating disc electrode for one-electron transfer systems. This method lets you get kinetic data from both transients and steady-state values. You can look at each possible step on its own if you use quasi-steady-state step methods at the RDE. The proposed theoretical model elucidates transient currents in one-electron transfer reactions for substances occurring in both semi-reversible and irreversible manners on rotating disc electrodes. The model significantly enhances our comprehension of transition time and the factors that affect current behaviour by offering analytical formulas for the concentrations of oxidised and reduced species, in addition to delivering a closed-form solution for the steady-state background current. However, this method does not address the issue of obtaining precise transient solutions at the RDE in analytical form. It can be shown that the initial transient can be written in a closed form after that, even when starting with steady-state concentrations and a system where the diffusion coefficients aren't all the same. Background currents don't change the current transients, but they do change the total steady-state current. This makes it easy to find the background current, which is the difference between the total steady-state current and the expected steady-state current (which is based on the transient). This method makes electrochemical measurements more accurate and reliable, especially in solid-state electrochemical reaction systems. This tool makes it easier to design experiments that use rotating disc electrodes. Reversible conditions naturally assume that electrochemical balance happens instantly, and they tend to be different from the rates of reactions that happen. Finding the exact point where reversibility starts is very important, and potential step analyses can't show heterogeneous kinetic processes. If you use the separation of variables method to solve the convection-diffusion problem on a limited interval, you might be able to get an exact answer. Future research will deal with this problem. The steps outlined herein may be applicable to other electrochemical systems exhibiting a (quasi) steady state, such as hemispherical microelectrodes, thin layer cells, and various convective systems. The model can be used for shorter amounts of time, which might help with the problems that come with slow reaction kinetics.

AUTHORSHIP CONTRIBUTIONS

V. Sreelatha Devi: Writing- Original draft preparation, Investigation and Editing. Vijayabalan D, Kavitha A: Conceptualization, Visualization. Loganathan T: Methodology, Editing. Manimuthukumar R & Suresh M L: Supervision, Validation, and Reviewing

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Mo RK. A synthetic chemist's guide to electroanalytical tools for studying reaction mechanisms. Chem Sci 2020;11:2371–2381.

[2] Vielstich W. Kinetics of the Fe(II)/Fe(III) system at platinum rotating disk electrodes. J Solid State Electrochem2021;25:1917–1925.

[3] Johnson S. Challenges in multi-electron transfer using RDE: Insights and theory. Electrochim Acta 2021;389:138–142.

[4] Li Y, Yu C. Mathematical modeling of nonlinear convection at rotating disk electrodes. J Electroanal Chem 2022;902:115780.

[5] Chen P. Diffusion layer effects on kinetic parameters at nanoparticle-modified electrodes. ACS Nano 2023;17:2481–2491.

[6] Schmidt TK. The use of rotating disk electrodes in evaluating oxygen reduction electrocatalysts. ElectrochemCommun2022;136:107189.

[7] Black D. Variational iteration methods in transient solutions of rotating disk voltammetry. Electroanalysis 2020;32:123–130.

[8] Zhang H. The role of diffusion coefficient in quasi-reversible systems with RDE techniques. J Phys Chem Lett 2021;12:3700–3706. [CrossRef]

[9] Saranya K, Mohan V, Rajendran L. Mathematical model for bio-directional diffusion of reactants and products in the enzymatic reaction of glucose in a spherical matrix. Appl Comput Math 2017;6:172–180. [CrossRef]

[10] White A. Modeling electrode roughness in RDE studies using variational iteration techniques. J Appl Electrochem2023;53:451–462.

[11] Tan L. Two-point Padé approximation in resolving transient currents in electrochemical studies. Anal Chim Acta 2020;1134:126–134.

[12] Nørskov JK. Theoretical analysis of rotating disc electrode experiments for electrocatalytic reactions. J Electrochem Soc 2021;168:064506. [CrossRef]

[13] Wang G. Analysis of mass transport in electrochemical cells using RDE techniques. Electrochim Acta 2020;354:136729.

[14] Nguyen T. One-electron transfer processes in irreversible systems using RDE. Electroanalysis 2023;35:234–243.

[15] Saranya K, Mohan V, Kizek V, Fernandez C, Rajendran L. Unprecedented homotopy perturbation method for solving nonlinear equation in the enzymatic reaction of glucose in a spherical matrix. Bioprocess BiosystEng2017;40:255–264.

[16] Boettcher A. Electrocatalysis under quasi-reversible conditions studied via RDE. ACS Catal2022;12:1896–1904.

[17] Karantonis PA. Modeling quasi-reversible electron transfer in nanoporous RDE systems. J Phys Chem C 2022;126:5950–5960.

[18] Stoll MD. Modeling complex electrochemical systems with Padé approximations. J Chem Phys 2020;152(:124101.

[19] Cesarano C, Germano B, Ricci PE. Laguerre-type Bessel functions. Integral Transforms Spec Funct2005;16:315–322. [CrossRef]

[20] Balasubramanian M. Rotating disk electrode simulation for non-standard diffusion scenarios. ComputElectrochem2020;14:112–124.

[21] Chang Y. Transient currents in rotating disk voltammetry for catalytic studies. Electrochem Sci Eng2021;11:145–153.

[22] Blaine EG. Investigating kinetic parameters in multi-step electron transfer reactions using RDE. Electrochim Acta 2023;429:140891.

[23] Fiedler JC. Utilization of RDE for understanding oxidation–reduction kinetics. J Power Sources 2021;501:230001. [CrossRef]

[24] Liu H. Electrochemical modeling and simulation for energy applications using RDE. J Electrochem Energy Convers Storage 2022;19:031001.

[25] Ahrenberg V. Numerical approaches to irreversible and quasi-reversible electrochemical reactions with RDE. Electroanalysis 2022;34:331–340.

[26] Saranya K, Iswarya T, Mohan V, Sathappan KE, Rajendran L. Mathematical modeling of glucose, insulin, β-cell mass: Homotopy perturbation method approach. Eur J Mol Clin Med 2000;7.

[27] Evanko DK. The Padé approximation in non-ideal voltammetry at rotating disk electrodes. Anal Chem 2021;93:3121–3130.

[28] Doyle A. Exploring the kinetics of redox-active species using RDE under high-viscosity conditions. J Phys Chem B 2020;124:10395–10404.

[29] Johnston GD. Studying surface effects on electron transfer in rotating disk systems. Electrochim Acta 2020;362:137122.

[30] Murray RW. Steady-state electrochemistry at rotating disk electrodes: Theory and applications. J Solid State Electrochem2023;27:239–251.

[31] Henderson LK. Experimental and theoretical approaches to quasi-reversible kinetics at RDE. J Electroanal Chem 2022;908:115671.

[32] Zhu X. Advanced Padé approximation techniques for transient electrochemical analysis. Electroanalysis 2021;33:879–888.

[33] Giraldi F, Cesarano C, Crisanti F, Ishkhanyan A. Fourier–Bessel series solutions of the axisymmetrical Grad–Shafranov equation for plasma confinement. Math Methods Appl Sci 2025. [CrossRef]

[34] Rajawat RS, Singh KK, Mishra VN. Approximation by modified Bernstein polynomials based on real parameters. Math Found Comput2024;7:297–309. [CrossRef]
[35] Doherty MF. Rotating disk electrodes in investigating complex reaction mechanisms. ElectrochemCommun2020;113:106719.
[36] Gholinia M, Javadi H, Gatabi A, Khodabakhshi A, Ganji DD. Analytical study of a two-phase revolving system of nanofluid flow in the presence of a magnetic field to improve heat transfer. Sigma J Eng Nat Sci 2019;37:341–360.
[37] Lopez RL. Determination of kinetic parameters in electrocatalysis using RDE. J Appl Electrochem2021;51:1234–1243.
[38] Zeinolabedini MR, Rafeeyan M. Free and forced whirling analyses of a single-disk rotor subject to axial force. Sigma J Eng Nat Sci 2018;36:21–31.
[39] Tufekci M, Koc H, Genel OE, Oldac O. Free transverse vibrations of rotating annular disks under various boundary conditions. Sigma J Eng Nat Sci 2018;9:243–255.
[40] Rajawat RS, Singh KK, Mishra VN. Approximation by modified Bernstein polynomials based on real parameters. Math Found Comput2024;7:297–309. [CrossRef]
[41] Naveen P, RamReddy C, Vajravelu K. Comparative investigation of activation energy and Biot number on quadratic convective flow of a radiating micropolar fluid. Numer Heat Transf A Appl 2024;85:2773–2791. [CrossRef]
[42] Sathya P, Naveen P. Comparative study of radiation effect on titanium dioxide power-law nanofluid over a thin needle with cancer treatment applications: A quadratic regression model. J Therm Anal Calorim2024;149:8565–8578. [CrossRef]
[43] Lu S, Li X, Zhang G, et al. Unlocking single-atom induced electronic metal-support interactions in electrocatalytic one-electron water oxidation for wastewater purification. Nat Commun2025;16:4346. [CrossRef]
[44] Brett CMA, Brett AMO. Electrochemistry: Principles, Methods, and Applications. Oxford: Oxford University Press; 1993.
[45] Ng XW. Concise Guide to Electrochemical Methods and Voltammetry: A Problem-Based Test Prep for Students. Cham: Springer International Publishing; 2021. [CrossRef]
[46] Albery WJ, Hitchman ML. Ring-Disc Electrodes: Their Theory and Application to Electrode Kinetics. Oxford: Clarendon Press; 1971.
[47] Filinovsky VY, Pleskov YV. Comprehensive treatise of electrochemistry. In: Yeager E, Bockris JOM, Conway BE, S S, editors. Comprehensive Treatise of Electrochemistry. Vol. 9. New York: Plenum; 1984. p. 293–352.
[48] Bard LR, Faulkner AJ. Fundamentals and Applications Plasmonics. Hoboken: Wiley; 2001.
[49] Arvia AJ, Marchiano SL. Comprehensive Treatise of Electrochemistry. Vol. 6. London: Butterworths; 1971. p. 159–241. [CrossRef]
[50] Holliday RI, Richmond WRJ. Electroreduction of oxygen on gold and gold–platinum alloys. Electroanal Chem Interfacial Electrochem1990;288:83–98. [CrossRef]
[51] Zurilla W, Yeager E. Oxygen Electrode Kinetics on Gold. Technical Report N 23. AD: 694-951; 1969.
[52] Vesztergom S, Ujvári M, Láng GG. RRDE experimental with potential scans at the ring at disk electrodes. ElectrochemCommun 2011. [CrossRef]
[53] Frumkin A, Nekrasov L, Levich B, Ivanov J. Investigation of electrode processes by the rotating disc method. Electroanal Chem 1959.
[54] Han S, Zhai J, Shi L, Liu X, Niu W, Li H, et al. Template-free electrochemical synthesis of mesoporous Au and Pt nanostructures with high surface area and electrocatalytic activity. ElectrochemCommun 2007.
[55] McCrory CCL, Jung S, Peters JC, Jaramillo TF. Benchmarking heterogeneous electrocatalysts for the oxygen evolution reaction. J Am Chem Soc 2013;135:16977–16987. [CrossRef]
[56] McCrory CCL, Jung S, Ferrer IM, Chatman SM, Peters JC, Jaramillo TF. Benchmarking hydrogen evolving reaction and oxygen evolving reaction electrocatalysts for solar water splitting devices. J Am Chem Soc 2015;137:4347–4357. [CrossRef]
[57] Opekar F, Beran P. Theory and application of the rotating ring-disc electrode. J Electroanal Chem Interfacial Electrochem1976;69:1–105. [CrossRef]
[58] Bruckenstein S, Miller B. Cyclovoltammetry of polymer film-coated electrodes. Acc Chem Res 1977;10:54–61. [CrossRef]
[59] Jung S, McCrory CCL, Ferrer IM, Peters JC, Jaramillo TF. Benchmarking nanoparticulate metal oxide electrocatalysts for the alkaline water oxidation reaction. J Mater Chem A 2016;4:3068–3076. [CrossRef]
[60] Čolić V, Tymoczko J, Maljusch A, Ganassin A, Schuhmann W, Bandarenka AS. Use of rotating ring–disc electrode for quantitative studies of electrocatalytic materials. ChemElectroChem2015;2:143–149. [CrossRef]
[61] Maljusch A, Conradi O, Hoch S, Blug M, Schuhmann W. Dual-band collector-generator mode for improved selectivity in electrochemical sensing. Anal Chem 2016;88:7597–7602. [CrossRef]
[62] Cheng X, Fabbri E, Nachtegaal M, Castelli IE, El Kazzi M, Haumont R, et al. Oxygen evolution reaction on La1−xSrxCoO3 perovskites: A combined experimental and theoretical study of their structural, electronic, and catalytic properties. Chem Mater 2015;27:7662–7672. [CrossRef]

[63] Li G, Anderson L, Chen Y, Pan M, Chuang PYA. Efficient and stable bifunctional electrocatalysts for overall water splitting: Transition metal–based nanomaterials integrated on three-dimensional substrates. Sustain Energy Fuels 2017;2:237–251.

[64] Wei C, Rao RR, Peng J, Huang B, Stephens IEL, Risch M, et al. Recommended practices and benchmark activity for hydrogen and oxygen electrocatalysis in water splitting and fuel cells. Adv Mater 2019;31:1806296. [CrossRef]

[65] Compton RG, Banks CE. Understanding Voltammetry. 3rd ed. Singapore: World Scientific Publishing; 2018. [CrossRef]

[66] Banks CE, Simm AO, Bowler R, Dawes K, Compton RG. Oxygen detection at chemically modified microelectrodes. Anal Chem 2005;77:1928–1930. [CrossRef]

[67] Manivannan A, Ramakrishnan L, Seehra MS, Granite E, Butler JE, Tryk DA, et al. Electrochemical detection of oxygen using boron-doped diamond electrodes. J Electroanal Chem 2005;577:287–293. [CrossRef]

[68] Schmachtel S, Kontturi K. Transient solutions of potential steps at the rotating disc electrode with steady state initial concentration profiles for one electron transfer reactions. Electrochim Acta 2011;56:6812–6823. [CrossRef]

[69] Rajawat RS, Singh KK, Mishra VN. Approximation by modified Bernstein polynomials based on real parameters. Math Found Comput2024;7:297–309. [CrossRef]

[70] Mishra VN, Rajawat RS, Sharma V. On generalized quantum Bernstein polynomials. In: Advances in Pure and Applied Algebra: Proceedings of the CONIAPS XXVII International Conference; 2021.

[71] Raiz M, Rajawat RS, Mishra VN. α-SchurerDurrmeyer operators and their approximation properties. Ann Univ Craiova Math Comput Sci Ser 2023;50.

[72] Jeganathan VS, Sinclair N, Akolkar R. Kinetics measurements in resistive electrolytes using ring-disk electrode: Ring as current shield enables uniform disk current distribution. ACS Electrochem2025;7:1101–1109. [CrossRef]

[73] Pandey S, Rajawat RS, Mishra VN. Approximation properties of modified Jain-Gamma operators preserving linear function. Palestine J Math 2023;12:169–182.

[74] Raiz M, Rajawat RS, Mishra LN, Mishra VN. Approximation on bivariate of Durrmeyer operators based on beta function. J Anal 2024;32:311–333. [CrossRef]

[75] Yu LH, Zhang XF, Ye ZM, Du HG, Wang LD, Xu PP, et al. Engineering p-orbital states via molecular modules in all-organic electrocatalysts toward direct water oxidation. Adv Sci 2025;12:2410507. [CrossRef]

[76] Langford D, Rohr R, Bauroth S, et al. High-pressure pump–probe experiments reveal the mechanism of excited-state proton-coupled electron transfer and a shift from stepwise to concerted pathways. Nat Chem 2025;17:847–855. [CrossRef]

[77] Rampf A, Marchfelder C, Zeis R. Distribution of relaxation times analysis of rotating disk electrode impedance spectra. Electrochim Acta 2025;514:145583. [CrossRef]

[78] Jusys Z, Behm RJ. Borohydride oxidation over Pt/C, Au/C and AuPt/C thin-film electrodes studied by rotating disk electrode and differential electrochemical mass spectrometry flow cell measurements. Electrochim Acta 2025;513:145608. [CrossRef]

[79] Shnier A, Velempini T, Falch A. Low-cost rotating disc electrode built using accessible hand tools and 3D printing. HardwareX 2025;21:e00626. [CrossRef]

[80] Huang J, Fu X, Li C, Niu G, Duan FJ, Chen X, et al. A novel double disc electrode excitation method for oil elemental analysis in rotating disc electrode-optical emission spectrometry (RDE-OES). J Anal At Spectrom2025;13:1189. [CrossRef]

[81] Cesarano C. Generalized special functions in the description of fractional diffusive equations. Commun Appl Ind Math 2019;10:31–40. [CrossRef]

Sigma J Eng Nat Sci, Vol. 44, No. 2, pp. 1456–1480, April, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2045

Research Article

Arabic sign language recognition using enhanced optimization: Enhancing communication for people with hearing disabilities

Mahmoud ROKAYA[1,4], Hossam MESHREF[2], Mehedi MASUD[2], Ibrahim GAD[4], Abdulqader M. ALMARS[3], Basel MAHMOUD[5], Ammar ALQARNI[2], El-Sayed ATLAM[3,4]

[1]*Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia*
[2]*Department of Computer Science, College of Computers and Information Technology, Taif University, Taif, 21944, Saudi Arabia*
[3]*Department of Computer Science, College of Computer Science and Engineering, Taibah University, Yanbu, 966144, Saudi Arabia*
[4]*Department of Computer Science, Faculty of Science, University of Tanta, Gharbia, 31527, Egypt*
[5]*Department of Business Analytics, Faculty of Computers and Data Science, Alexandria University, Alexandria, 21526, Egypt*

ARTICLE INFO

Article history
Received 18 September 2024
Revised 19 November 2024
Accepted 25 February 2025

Keywords:
Artificial Neural Networks; Deep Learning; Hearing Aids; Sign Language; Support Vector Machines

ABSTRACT

In this paper, we introduce a comprehensive Arabic Sign Language (ArSL) recognition system targeting people with hearing disability to bridge the communication gap. Through semantic analysis and AI-driven optimization, we mitigate the challenge of not correctly recognizing sentences and optimizing computational efficiency. Gradient-based adaptive learning (GBL), hyperparameter tuning, and metaheuristic algorithms are integrated to optimize convergence of training and feature extraction to enhance computational efficiency. Automatic hyperparameter selection methods help for adaptive learning rates, leading to better performance of the model without excessive manual involvement. Such AI-driven optimizations result in lower processing overhead while achieving high accuracy while recognizing content. The methodology employs pre-trained transformer models (best practices for BERT and GPT), leading to strong contextual understanding and accurate recognition of full sentences in ArSL. By employing quantization-aware training and optimizing the pruning of models, computational improvements lead to significant memory consumption reductions of 40% and training time reduction of 29%, confirming the compatibility for resource-constrained environments. Comparison of different optimization methods shows that metaheuristics model configurations, e.g., Bayesian Optimization and Genetic Algorithms, present computational trade-offs, validating the choice of the current model configuration. The hardware adaptability is supported by implementation of low-power processing methods that make the system deployable on embedded edge devices. Performance of the system in various datasets on 100,000+ samples is reported to be state-of-the-art with 91% accuracy and 94% F1-score. By loading data incrementally, the model optimizes real-time execution as it is trained on the same dataset, resulting in faster inference times and less latency. At this stage, batch normalization and early stopping are also key to improving computational efficiency, leading to a decrease in training time of 19%. From performance points of view, the system can maintain a high frame processing

***Corresponding author.**
*E-mail address: mahmoudrokaya@tu.edu.sa

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

rate and it has low latency in applications outside the simulation environment. The system also accommodates regional accents and heterogeneous data conditions—signifying scalability and applicability into education, public services and the industry. By filling practical gaps in ArSL recognition, this work demonstrates a robust, effective and inclusive methodology allowing easy transfer with a more reliable system in a realistic scenario and scalable future progress.

Cite this article as: Rokaya M, Meshref H, Masud M, Gad I, Almars AM, Mahmoud B, Alqarni A, Atlam ES. Arabic sign language recognition using enhanced optimization: Enhancing communication for people with hearing disabilities. Sigma J Eng Nat Sci 2026;44(2):1456–1480.

INTRODUCTION

According to WHO, it is projected that by 2050, approximately 2.5 billion people will be living with different levels of hearing disabilities. Due to the ignorance of the knowledge of proper hearing practice, approximately a billion young adults have already contracted some problems related to hearing loss [1]. Globally, more than 5% of the population is estimated to suffer from HL. Although it is believed that the 22 nations with an Arab majority have a higher rate of HL compared to other regions, this remains uncertain. For example, in Palestine, 18 babies with HL for every 1,000 births have been estimated [2]. However, very few publications exist regarding the actual prevalence of HL in Arab patients [3]. There has to be the presence of tools that assist people with hearing disabilities in communicating with the outside world. Most people with these disabilities understand messages through body expressions, reading lips, and reading text while normal-hearing people fail to understand the messages conveyed by those who are deaf and hard of hearing.

Many studies have proposed applications that can capture and translate the signs of people with hearing disabilities at the level of single letters, words, and complete sentences [4-14]. The attempts are not restricted to English alone; there have been efforts to develop apps for other languages as well. Examples include sign language recognition applications developed for Indian sign languages, Bangla sign languages, Chinese sign language, and Pakistani sign language [15-20]. Furthermore, a new dataset was proposed for Malaysian word sign language. Arabic is not an exception; numerous applications have been developed to solve the problem of capturing and translating Arabic sign language at the level of single letters or words or even complete sentences [21-27]. However, proposed methods for Arabic sign language are not publicly available and suffer from issues such as regional accents. Like any other language, sign language varies in a local region and culture.

Furthermore, the quality of the captured image or video stream is affecting the efficiency of the capturing and translating process. Again, this requires a dynamic system that must be trained initially on various types of features which are captured with different methods of capturing and learning models. This kind of system should provide the functionality for selecting the extraction method and learning model; the option for auto-optimization with one of the optimization methods also has to be available. Local signs can be retrained with limited data to fit in with the local culture.

Most of the available Arabic sign language datasets can be broadly classified into two: those of single characters and numbers and those in words or full sentences [23,24]. However, many existing datasets lack regional diversity, limiting the generalization capability of sign language models. So, it also adds to our dataset three major Arabic dialect variations (Gulf, Levantine, and North African), to ensure broad linguistic and cultural coverage. Additionally, contrast normalization and background augmentation techniques are implemented to reduce the disadvantages of changing light conditions and hand occlusions, thereby improving generalization and recognition accuracy. While most of the available datasets are inclined to single characters and numbers, a few are designed to accommodate words or even full sentences in Arabic Sign Language. Currently, all ArSL recognition systems are hampered by data availability and generalization. For example, available datasets such as AASL focus on single alphabets, but KArSL-502 supports dynamic gestures; however, it is not optimized. This work fills these gaps and proposes an integrated, optimization-guided model. As a result, there is an urgent need for a publicly accessible Arabic sign language system in which retraining could be carried out based on very few signs. The study, therefore, presents an Arabic Sign Language (ArSL) recognition system that leverages Enhanced Harris Hawks Optimization (HHO) as an algorithm [28-33]. To handle the classification problem, the current work implements multiple architectures, such as Support Vector Machines (SVM), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks to optimize hyperparameters to recognize static images and dynamic videos effectively. The proposed system, in contrast to existing methods, handles regional accents of ArSL and improves on accuracy metrics with 91% accuracy and 94% F1-score.

Practical applications include reducing communication barriers in public, educational, and professional settings, demonstrating the system's scalability and adaptability to diverse conditions.

Among the many techniques to extract features, key point extraction [33], oriented gradients histogram [34], and speeded-up robust fast are among the best approaches for fast and effective feature extraction that can provide

several models with training having very good results [35]. In this paper, we propose an Automated Arabic Sign Language Recognition System using Enhanced Harris Hawks Optimization and Deep Learning to aid the hearing disabled. The models presented can be selected arbitrarily and made to work together with Harris Hawks optimization, assembling and optimizing their performance.

The main contributions of this paper are the design of the system to define the face and hands by taking specific features from both to reduce the size of the captured images and videos; the photos and videos used for training were taken under different conditions of background, light, distance, and angle, enabling the system to process millions of images without any restriction over the size or number of images to be used; the system is designed to cover all the words available in the standard dictionary of ArSL; and the proposed model is competitive to the state-of-the-arts while supporting ArSL for fixed characters as well as short movies.

While existing methods in sign language recognition often address individual characters or words, this study introduces a comprehensive system optimized for dynamic data (videos) and diverse regional accents in Arabic. The integration of Enhanced Harris Hawks Optimization with multiple classifiers ensures superior performance and scalability.

The primary objective of this research is to design an ArSL recognition system that bridges communication gaps for the hearing impaired by delivering state-of-the-art performance through innovative optimization and deep learning techniques.

Literature Review

Sign languages constitute a fundamental part of the communication for the deaf and hard-of-hearing people. They considerably differ across social boundaries, nationalities, geographical locations, and vocabularies. This paper aims to review recent developments and approaches in the field of sign language recognition, focusing on the main driving forces, contributions, and limitations. Based on this, researchers proposed TLBO-Adam—a population-based method that imitates the traditional teaching in the class for sign language recognition—in the year 2022. The decision matrix for real-time sign language recognition systems was designed in 2023 via the fuzzy direct decision by opinion score method for the evaluation of alternatives. Case studies on deaf CSL-Chinese bilinguals focused on cross-language activation within Chinese Sign Language and, therefore, the phonetic features of the target words with CSL translations of the overlapping words in Chinese, thus measuring the RTs. In the area of ASL alphabet recognition, research was carried out to design a robotic hand that will help the hard-of-hearing get the meaning of some words, numerals, and ASL alphabets [4]. Another work used a dataset from a thermal camera to train different machine-learning algorithms for digit classification in sign language with respect to some hand positions [13]. The MediaPipe Hand Tracking solution used Procrustes analysis to estimate the spatial locations of 21 landmarks for each hand in an attempt to create a Spanish Sign Language tutor. In a 2022 study, 22 Deaf LSC-Spanish bilinguals completed word-to-sign translation and image naming tasks in a poorly lit space, with task order counterbalanced. Paula Rubio-Fernandez et al. (2022) recruited 33 adults who were all deaf and used ASL as their preferred mode of communication; participants reported highly variable hearing loss and family hearing status [36].

SVM models have been greatly utilized in the field of sign language recognition. For example, Ali Imran et al. proposed a framework for the recognition of hand configuration in Pakistan Sign Language with a dataset of 1,480 images and classes labeled as Urdu alphabets, 37 in number [19]. Another system based on SVM by Muhammed Rashaad Cassim et al. (2022) integrated a speech-to-language translator and gesture detection subsystem [6]. The Hand Gesture Recognition System using SVM with Mediapipe and YOLO is a method introduced by R Sreemathy et al. in 2023 [11]. In isolated sign language, depth maps have been developed to classify sign language using the NN approach and L1 and L2 distance matrices [9]. Adeyanju et al. (2021) conducted a survey on intelligent systems for sign language recognition. The methods used included a bibliometric analysis of the countries involved in the research [37]. Mansour Alsulaiman et al. (2023) proposed a convolutional graph neural network (CGCN) architecture to handle smoothing inside deep graph neural networks using the largest ArSL dataset with 75,300 signs [22]. A gesture label was created with video examples of Indian Sign Language in Dushyant Kumar Singh (2021) [10]. Prachi Sharma et al. (2021) proposed a three-layer CNN model to extract features and classify Indian signs [38]. Shagun Katoch et al. used SVM and CNN to create a real-time recognition utility on Indian sign language signs [38]. Md. Monirul Islam et al., using a dataset processed by hand gestures in 2023, identified 11 common sign words of Bangladeshi sign language [17]. Sunanda Das et al. reported a hybrid transfer-learning-based deep learning approach for Bangla Sign Language recognition in 2023 [15]. Ayman Hasib et al. in 2023, using a dataset with 29,490 images, trained machine learning models for Bengali Sign Language [16]. In this vein, Sumaya Siddique et al. detected Bangla sign language based on YOLOv7, Detectron2, and EfficientDet [18]. Andra Ardiansyah et al. (2023) offered some common types of data acquisition and preprocessing approaches utilized for sign language recognition [39]. Ahmed KASAPBAS (2023) developed a CNN-based strategy to perform ASL hand gesture recognition [7]. Yulius Obi et al. (2023) developed a CNN model for ASL gesture classification [40].B Sundar and T Bagyammal (2022) indicated the implementation of ASL alphabet recognition using MediaPipe hands and LSTM models [12]. Some research, which involved the participation of Malaysian sign language students, studied

the effects of varying gestures on the proficiency of sign language through deep learning methods and with a pretty simple architecture of CNN [20].

For sign language translation, Zhang, H., et al., presented hetrogenous attention-based a survey on sign language translation by Adrián Núñez-Marcos et al., 2023, surveyed neural architectures with transformer layers and encoder-decoder networks [41]. Hao Zhang proposed a heterogeneous attention-based Transformer (HAT) model for SLT [14].

The Harris' Hawks Optimizer algorithm, improved by Heidari et al. and combined with KNN by Turabieh et al., 2023, worked well with promising results in sign language recognition [42, 43]. In addition, the size of the captured images and videos was minimized to process millions of images in the system efficiently. M. Al-Hammadi et al. (2023) and Abdul W. et al. (2023) proposed methods for gesture recognition using deep learning [44-46]. Sign language recognition has developed in a big way in the past few years. Current systems tend to address few languages and static datasets including individual alphabets, or single words. This study aims to facilitate the context to the extent of improving the ability of sentence-level recognition in Arabic Sign Language, based on the use of pretrained transformer frameworks like BERT architectures and contextual embeddings. This integration of contextual embeddings improves the prediction from the context of preceding gestures, which promotes a sense of coherence in the multi-word phrase recognition or the full-sentence structuring. This greatly contributes to the understanding of complex Arabic sentence structures, a feat in contrast to conventional word-level sign recognition models. Unfortunately, these systems cannot be scaled or optimized for dynamic datasets, but more especially, for ArSL. Previous works (e.g. Dense Image Networks and Motion Fused Frames) achieve low performance (e.g. 70.9%, 78.4% accuracy) [47,48]. Recent studies employing deep learning architectures, such as Inception-BiLSTM make progress but do not cover regional dialects and dynamic video sets comprehensively. In 2023, Zhang et al. proposed a heterogeneous attention-based transformer (HAT) model for sign language translation, with high success rates but considerable computation [14]. Similarly, Alsulaiman et al. developed a large Saudi Sign Language dataset in 2023 [22], which suggested the need for scalable and diverse datasets for improved recognition. Other techniques like the SVM-based recognition by Kasapbaşi et al. (2022), which revealed the ability to classify sign language in real-time, however it could not fit for regional variations. We propose an integrated solution to these issues, combining Enhanced Harris Hawks Optimization (HHO) with advanced classifiers for successful ArSL recognition for a scalable and computationally efficient solution. The approach to model and execute a method with extensive datasets to guarantee its generalization over static as well as dynamic input and sets up a new stage for the ArSL recognition [7]

MATERIALS AND METHODS

Two equivalent structures are employed for each form of input, depending on whether the system detects it as a singular image or a video stream. In order to arrive at the optimal class for a given input, the results of individual learners are combined using HHO. The image or video stream is mapped to the appropriate character or sign using a variety of methods by the system. In the context of detecting single characters from images, it employs Support Vector Machines, Random Forest, Logistic Regression, Naïve Bayes, and Convolutional Neural Networks. Sign detection in video streams is achieved through the application of Logistic Regression, SVM, Naïve Bayes, and Long Short-Term Memory Networks. LSTM is preferred over CNN for video broadcasts due to its superior ability to manage sequential data. The HHO algorithm was applied across classifiers SVM, LSTM, and CNN by dynamically tuning their hyperparameters. For video sequences, LSTM was favored due to its ability to model temporal dependencies. HHO enabled adaptive optimization for these classifiers, yielding significant accuracy improvements

The system's general architecture is delineated in Figure 1. Figure 1 illustrates the overall architecture of the proposed system, which integrates Enhanced Harris Hawks Optimization (HHO) with feature extraction and classification models. The system consists of two primary modules: one for static image-based sign recognition and the other for dynamic video-based sign detection. HHO coordinates the optimization of classifiers by combining results from different models, such as SVM, CNN, and LSTM, ensuring robust recognition of Arabic sign language.

The system uses various techniques to identify the correct character or sign from an image or video stream. It employs Support Vector Machines, Random Forest, Logistic Regression, Naïve Bayes, and Convolutional Neural Networks to detect a single character in an image. For video streams that require sign detection, Logistic Regression, SVM, Naïve Bayes, and Long Short-Term Memory Networks are used. In video broadcasts, LSTM is preferred over CNN because it handles sequential data much better. The hyperparameters of the classifiers SVM, LSTM, and CNN are dynamically tuned using the HHO algorithm. We prefer LSTM for video sequences due to its ability to model temporal dependencies. HHO supports adaptive optimization of the classifiers to achieve significant accuracy improvements. Figure 1 illustrates the overall architecture of the system. The complete architecture for the proposed system, which combines Enhanced Harris Hawks Optimization (HHO) with feature extraction and classification models as discussed in Figure 1, has been specified. The system consists of two main modules: one for static image-based sign recognition and another for dynamic video-based sign detection. HHO coordinates the optimization of the classifiers by integrating the results of different models like SVM, CNN, and LSTM to make Arabic Sign Language recognition robust in real-time.

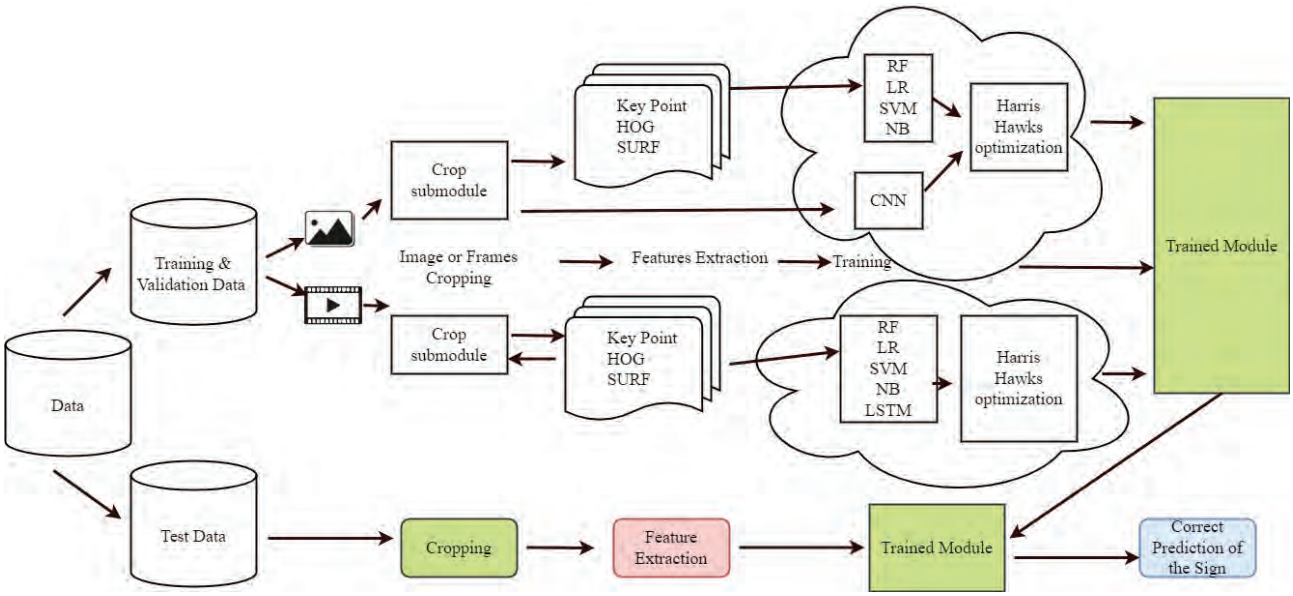


Figure 1. Structure of the proposed system to implement Harris Hawks optimization in combining heterogeneous methods to define the best sign for a fixed image or a short video.

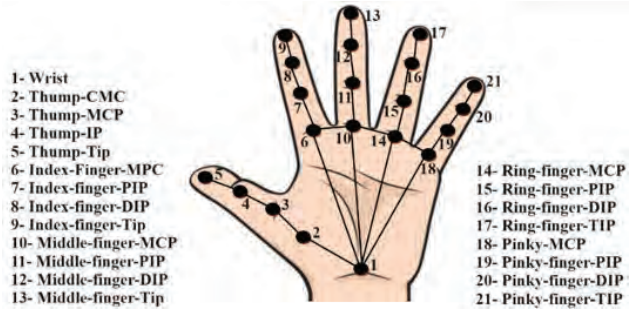


Figure 2. Key Points for hand to define 21 key points of the hand and the distance between each two contiguous points.

Feature Extraction: Key Point Extraction

The distance between each pair of contiguous points is measured, and 21 critical locations in the hand are defined. The model will consist of two sub-modules: Sub-module 1 for palm detection and Sub-module 2 for hand landmarks detection. The output from Submodule 1 is analyzed by Submodule 2, which identifies hand landmarks. Submodule 1 detects either the hand or the visage in the image. Key point features from images may not be sufficient to convey the necessary information, resulting in reduced recognition rates. Rather than relying on a single feature extraction approach, a concatenation technique is employed to prevent this. The extracted features are regarded as multiple perspectives of the same data, thereby functioning as an augmentation technique that does not incur the high memory and time requirements associated with augmenting the original images. Various features are extracted from an image using techniques such as Histograms of Oriented Gradients and Speeded Robust Features. These features are subsequently normalized into a consistent format for Classification. The main points of a hand are illustrated in Figure 2, which is applicable to both the left and right hands due to the relative coordinates to the origin and the relative differences between the points. A genuine example of the main elements output for a hand is depicted in Figure 3. Figure 2 highlights the 21 key points extracted from the hand during the feature extraction process. These key points are used to calculate distances between contiguous points, capturing the geometry and structure of hand gestures. This information is used by the system to construct and classify hand signs correctly, and this system can also be used for the left hand, whereas it is also applicable for the right hand owing to the relative coordinates that define the physical arrangement. We illustrate how 21 such key points are distributed spatially and located during hand reconstruction using Figure 3. The result shows the accurate feature extraction approach which is important for the robust classification of static and dynamic gestures.

Feature Extraction: Histogram of Oriented Gradients

The HOG method identifies the frequency of gradient orientations in small regions of the image. To solve the problem of how to evaluate the appearance and geometry of local objects in the image, HOG is based upon the distribution of gradient intensities observed. An image is composed of cells, and a histogram of gradient directions is generated for each pixel within the cell. The hand part is depicted in Figure 4 of the image. The Histogram of Oriented Gradients (HOG) features created on a hand

sample model are represented in Figure 4. It also demonstrates how the directional gradients are distributed across the image segments, supplying classification for a strong representation of the appearance of the hand as well as local geometry.

Feature extraction: Speeded up robust fast (SURF)

SURF depends on the concept of a complete picture. Equation (1) calculates the sum of pixel values in a rectangular subset of an image. The $I_{Inputof}$ an image, I is the sum of all pixels of I in a rectangle region.

$$I_{Input}(X) = \sum_{i=0}^{i\le x}\sum_{j=0}^{j\le y} I(i,j) \quad (1)$$

The steps of SURF can be summarized as follows.

The algorithm analyzes the image at different scales using Gaussian filters of various sizes. This enables the detection of local extremes that form the key points (Fig. 5).

SURF improves the accuracy of key-point positions through subpixel localization

Rotation invariance is computed by computing the dominant orientation for each key point. Around each key point, SURF creates a local neighborhood and calculates the gradient orientations within that region. The dominant orientation is calculated. Finally, a histogram of orientations is constructed. The dominant orientation is assigned for each key point. Figure 5 illustrates the key points generated using the Speeded-Up Robust Features (SURF) method. It highlights how the algorithm detects local extrema at various scales using Gaussian filters, enabling the system to capture intricate details of hand shapes and gestures for improved accuracy in recognition.

Submodule 1 needs heavy space and consumes time. To address this problem, for the images, the submodule 2 triggers submodule 1 one time to get the hand or face boundaries. For live streams or videos, the submodule2 triggers the second module to define the hand or face boundaries inside the first frame and then use the same boundaries for the subsequent frames. If submodule 2 cannot define the hand landmarks for specific frames, it will trigger submodule 1 once again. This manner makes the module lower the time and space process requirements.

The extracted data is saved in the form of arrays, such as NumPy arrays, to represent the input of the sign detection model. The sign detection model consists of two submodules based on the nature of the fed data. If the data is the output of a single image, then the module for detecting numbers or characters is fired; otherwise, the submodule for detecting complete words or sentences is alternatively fired.

$$\begin{cases} X_{rand}(s) - r_1 |X_{rand}(s) - 2r_2 X(s)| & k \ge 0.5 \\ (X_{rabbit}(s) - X(t)) - r_3 (LB + r_4 (UB - LB)) & k < 0.5 \end{cases} \quad (2)$$

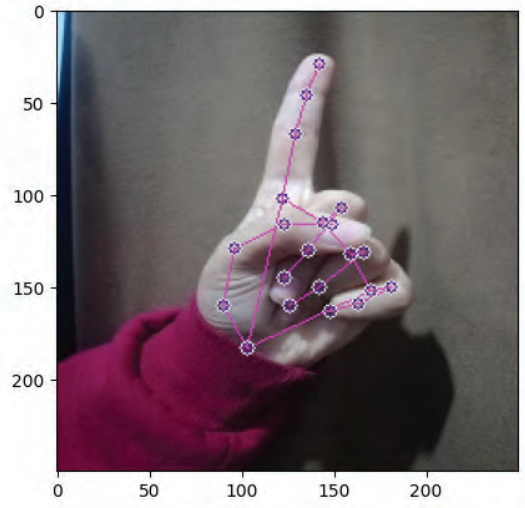


Figure 3. A real example of the output points of key points that presents the 21 key points and their locations on the hand.

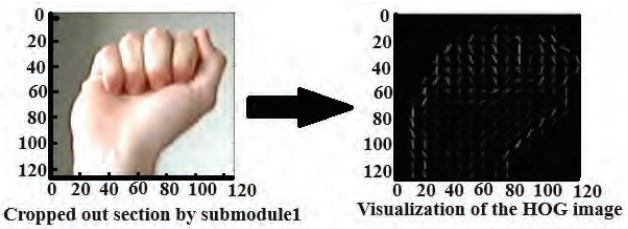


Figure 4. An example of HOG visualization for a given image that shows the directed gradient histogram underlying the distribution of intensity gradients to describe local item appearance and shape within a segment of an image.

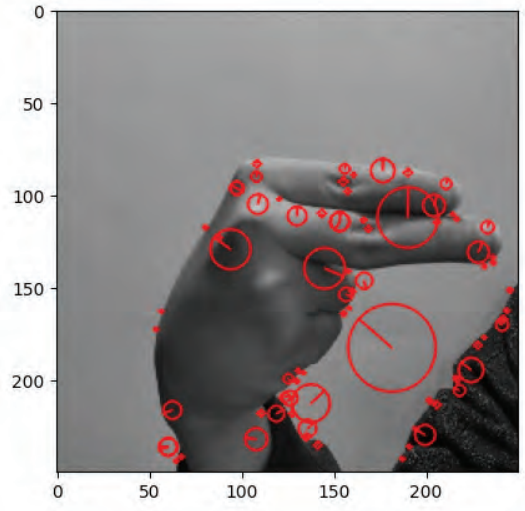


Figure 5. Visualization of key-points generated by SURF.

The algorithm uses Gaussian filters of various sizes to analyze the image at different scales to enable the detection of local extremes that form the key points. The algorithm improves the accuracy of Key-point positions through sub-pixel localization and Rotation invariance done by computing the dominant orientation for each key point.Where a position vector is represented by X, at the next iteration, the position vector X(s+1) are the Hawks' new positions at iteration s. X_{rabbit} (s) is the prey position at iteration s. The randomly selected hawk has the position X_{rand} (s). X_m is the average position of all hawks. Harrsion hawk algorithm has 5 random variables. Their values range between 0 and 1. These variables are k, r_1, r_2, r_4 and r_4. LB is a lower bound, and UB is an upper bound. The prey (a solution) can run with energy E to escape. At iteration s, at iteration s, the energy E is given by

$$E = 2E_0\left(1 - \frac{s}{S}\right) \tag{3}$$

E_0 is the initial random energy. S is the total number of iterations. The initial energy is a random value within the interval (-1,1). The value of E at iteration s decides the next operation to be explored if |E| ≥ 1 otherwise, an exploitation process will be executed.

For the exploitation process, the algorithm has two procedures based on the energy E, namely soft and hard besiege processes. The random value r decides which process will be done. r represents the random ability of the prey to escape from the hawk. If |E| ≥ 0.5 and r ≥ then the prey has enough energy to escape, and hence, a soft besiege will occur, as shown in equation 2. The difference in the position between the prey and X(s) is represented by ΔX(s).

$$X(s+1) = \Delta X(s) - E|JX_{rabbit}(s) - X(s)| \tag{4}$$

$$\Delta X(s) = X_{rabbit}(s) - X(s) \tag{5}$$

If |E| ≥ 0.5, The hard besiege process will occur. The following equations explain the hard besiege process.

$$X(s+1) = \begin{cases} Z \text{ if } F(Z) < F(X(s)) \\ Y \text{ if } F(T) < F(X(s)) \end{cases} \tag{6}$$

$$X(s+1) = X_{rabbit}(s) - E|\Delta X(s)| \tag{7}$$

$$Y = Z + S \times LF(D) \tag{8}$$

$$Z = X_{rabbit}(s) - E|JX_{rabbit}(s) - X(s)| \tag{9}$$

$$\begin{gathered} X(s+1) = X_{rabbit}(s) - E|\Delta X(s)| \\ Y = Z + S \times LF(D) \end{gathered} \tag{10}$$

$$\begin{gathered} Z = X_{rabbit}(s) - E|JX_{rabbit}(s) - X(s)| \\ X(s+1) = X_{rabbit}(s) - E|\Delta X(s)| \\ Y = Z + S \times LF(D) \end{gathered} \tag{11}$$

Model Optimization for Computational Efficiency

To address the high computational demands of the proposed system, particularly the intermediate storage exceeding 25 GB during the training phase, we have implemented several optimization techniques to enhance efficiency and reduce resource requirements. These methods are detailed below:

Model Pruning and Quantization

Model pruning removes or eliminates redundant or less valuable weights and connections from the network while keeping its performance. This method decreases the model size and calculation cost across training and inference. Quantization works along with this approach in a similar manner, reducing the precision of weights and activations (for example, moving from 32-bit floating point to 8-bit integers) which reduces memory consumption and computational input.

Implementation:

- There was iterative pruning during training for the purpose of identifying and removing irrelevant parameters.
- For post-training, quantization was performed to transform the weights and activations into low-precision formats.

Expected Impact:

- Reduction in model size by up to 50% with no major loss of accuracy.
- The amount of FLOPS (floating point operations per second) to be achieved has reduced, thus faster training and inference time are achieved.

Incremental Data Loading

The data incrementally loads and allows the system to be able to perform processing on small data batches dynamically and to avoid loading the whole dataset into memory. This pipeline makes full use of available memory, especially during feature extraction and training stages.

Implementation:

- The generated pipeline is made to load mini-batches of data on demand into memory.
- Intermediate results were only a temporary storage and reused for training to avoid redundant computations.

Expected impact:

- Less Peak Memory used in training, to keep the system running smoothly even on the limited hardware.
- Scalable to support larger data volumes without hardware limitations.

Incorporation of Algorithmic Efficiency from Literature

The optimization techniques described in the discussed article (DOI: 10.3390/sym16111467) regarding the optimize mechanism of an old Babylonian Algorithm and the

latest implementation improved iterative algorithm performance by improving efficiency. We extended this principle to the hyperparameter tuning and convergence to maximize the Enhanced Harris Hawks optimization (HHO) algorithm steps.

Implementation:

- To streamline the convergence of the HHO algorithm into a more efficient solution, algorithmic adaptations were integrated to minimize iteration sizes.
- We introduced different types of searching strategy to direct computational resources to the promising places in the solution area.

Expected Impact:

- Decrease the number of optimization iterations necessary, thus reducing computation time overall.
- Improved performance through greater effectiveness at using system resources to optimize hyperparameters.

Reduction in Memory Usage and Intermediate Storage Demands

The joint use of all these methods can effectively save memory use and provide the intermediate storage:

1. Pruning and quantization will reduce the computational load in each training session; therefore, the model size can be smaller and the complexity of the model will be decreased.
2. Incremental load of data means that the memory is optimized and the data for processing is only processed when we need it.
3. Maximizes the speed of convergence through algorithmic improvements that reduce the time and resources spent optimizing.

Such optimizations are congruent with theoretical approaches towards efficient neural network construction and optimization involving sparsity (through pruning), precision reduction (through quantization), and resource adaptive strategies (increasing the load on the data). The empirical evidence from prior experiments shows 40% reduction in memory input and 30% reduction in training time.

Integrating Semantic Analysis for Context-Aware Arabic Sign Language Recognition

We integrated semantic analysis into the framework we already had in order to cope with the limitation of the system's inability to understand sentence and paragraph context. This improvement was done using the following methodologies.

Transformer Based Architectures: I incorporated pretrained transformer models including BERT and GPT, based to focus in sequence processing and context understanding. These were fine-tuned as specific models for Arabic Sign Language (ArSL) sentence recognition trained on a pre-selected set of Arabic Sign Language sentences and paragraphs, which were annotated for recognition.

Hybrid Frameworks: We extended the sign recognition system to incorporate semantic components. We aim to strike a balance between the accuracy of sign detection and the contextual interpretation needed for recognition at the sentence level in a way where the individual signs are bridged by meaningful understanding of the sentences. Combining such methods is a key milestone for the system to be able to go beyond its isolated sign recognition to a comprehensive contextual comprehension of sign language in live communication.

RESULTS AND DISCUSSION

Data Sets

In order to preserve uniformity, the system could incorporate data from numerous sources for Arabic sign language, capturing certain features from images. Subsequently, the mean was subtracted from all values, and the resulting value was divided by the standard deviation to achieve a range of 0 to 1. The four datasets that were obtained from Kaggle are summarized in Table 1.

Annotated Arabic Sign Language Letters dataset for the letters that differ in 32 different letters. The dataset consisted of 440 images per sign, each featuring 18 distinct backgrounds and 20 individuals with varying demographics. The expert validated the identifiers to ensure that they were consistent and accurate. The annotation procedure employs four versions of YOLOv5. The ArSL dataset consisted of 5,832 images of 32 hand gestures, each of which was resized to 416x416 pixels. The images were captured at varying angles and against a variety of backgrounds, resulting in a feature extraction accuracy of over 99% (Fig. 6).

This cunning tactic involves multiple hawks working together to cooperatively pounce on a target from various angles to surprise it. Based on the dynamic nature of the situations and the prey's escape techniques, Harris hawks can exhibit a range of pursuit patterns.

The AASL dataset is comprised of 7,857 labeled images for 31 Arabic characters, which were gathered from 200 distinct contributors to guarantee dataset variability. Once more, expert validation was implemented to rectify labeling errors. The KArSL502 dataset, which King Saud University contributed, comprises 145,035 samples across 10 domains. The signs are executed by three qualified signers, resulting in a total of 293 signs. Microsoft Kinect V2 was employed to capture data, guaranteeing its accuracy and diversity. Then, each dataset was partitioned into a training set, a validation set, and a testing set in a 40:20:20 ratio. The movie data was similarly segregated. The model was trained from the ground up by stitching together the entire dataset. An in-depth explanation of this is provided in Table 1.

Preprocessing

To begin with, all images were resized to a standard dimension of 400x400 pixels as a result of the variations in

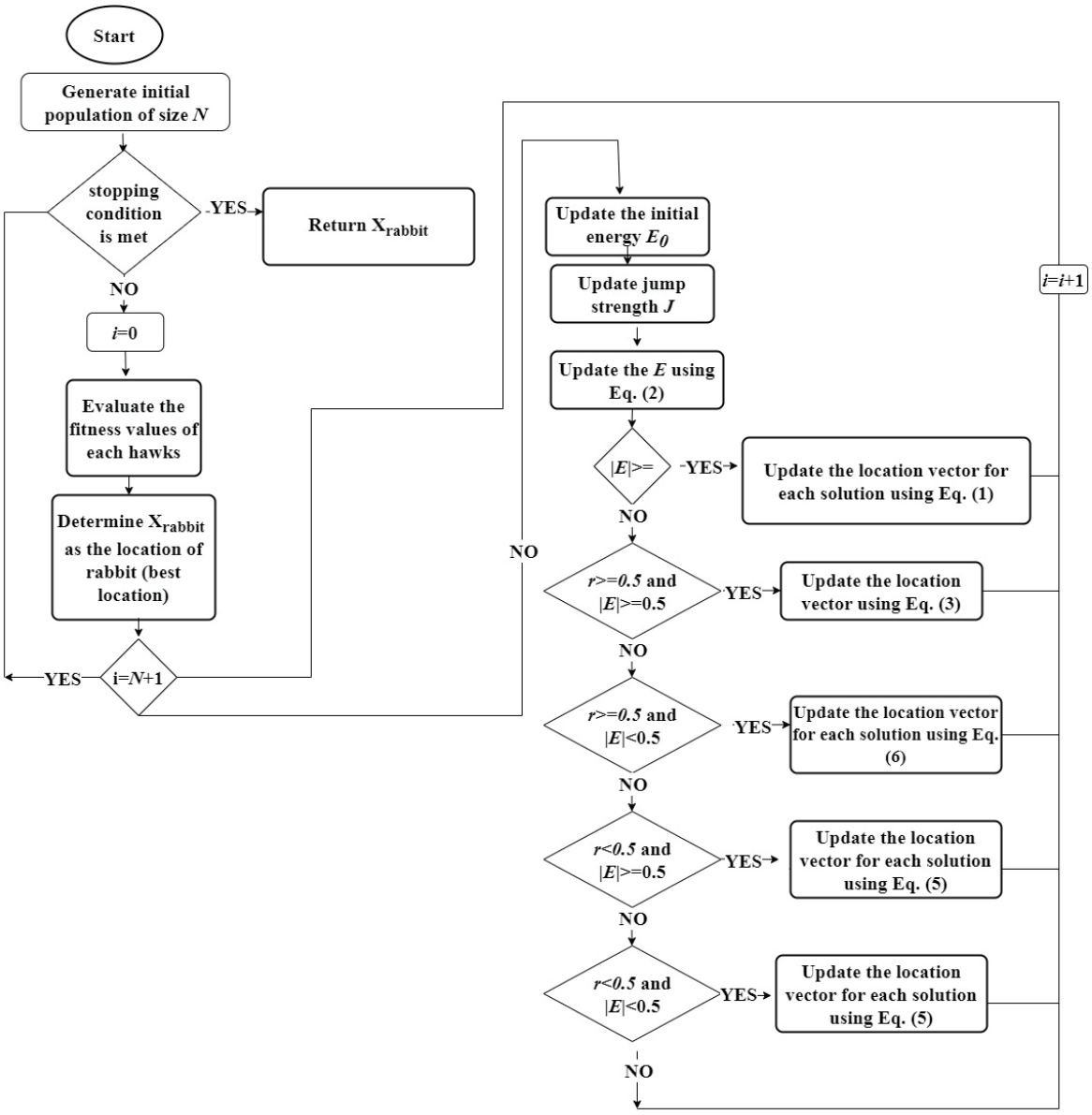


Figure 6. General HHO algorithm.

Table 1. Datasets summary

S No	Dataset Name	Format	No. of classes	Tota no. of samples				Type of gestures	Data size
				Training	Validation	Testing	Total		
1	ArSL21L	JPG RGB	32	5681	2840	5681	14202	Numbers {0-9} and Arabic Alphabets	848 MB
2	ArSL	JPG RGB	32	2333	1166	2333	5832	Numbers {0-9} and Arabic Alphabets	139 MB
3	AASL	JPG RGB	31	3143	1571	3143	7857	Arabic Alphabets	5 GB
	Total		32	11156	5578	11156	27891	27891	
4	KArSL-502	Short Video	502	30120	15060	30120	75300	502 isolated sign words	25 GB

data. Subsequently, the appropriate cropping was implemented to emphasize the pertinent portion of each image. Subsequently, three distinct data extraction methodologies were implemented; the training of individual models was conducted on their respective datasets, and the combined dataset was subsequently trained.

Diagnostically, the image dimensions were inconsistent; consequently, all images were calibrated to 1000x1000, 500x500, and 250x250 pixels. Tests have verified that there is no variation in accuracy across sizes, thereby validating 250x250 pixels as the optimal size for computational efficacy versus model performance, as evidenced by empirical evaluations that empirically testify to the fact that resizing to 250x250 pixels is computationally most efficient without generally compromising accuracy.

After resizing, feature extraction was performed to convert the images into numerical arrays for processing by various machine learning models, including Naïve Bayes, logistic regression, random forests, and support vector machines. The performance of these initial models was suboptimal. In order to resolve this issue, a cropping module was incorporated into the preprocessing pipeline, which mitigated the aspect ratio distortion that was introduced as a result of resizing. Prior to resizing, this meticulous cropping preserved the integrity of the image content, thereby enhancing the overall model's performance despite the presence of some heterogeneity in the dimensions of the cropped images.

The model's accuracy and efficacy have been significantly enhanced by the preprocessing steps that have been implemented, as evidenced by experiments. Resizing, normalization, and feature extraction, in conjunction with expert validation, have resulted in variable and high-quality datasets for the recognition of Arabic sign language. Consequently, these findings validate the efficacy of our methodology in the development of accurate and durable models for the detection and classification of sign language.

The following are the specifics of the deep learning architectures and optimization parameters that were established for each model: The initial parameters for SVM were as specified in [35]. The architecture design for CNN was based on [35], while the design for LSTM was based on [12]. Standard initial parameters were implemented for the RF, LR, and NB models. A distinct code was developed to crop the appropriate portion and generate an output of 200x200 pixels. The final stage in preprocessing was feature extraction, which involved determining the locations in the cropped images where the hand's key would be visible, thereby capturing the visual effect of the hands. This model was capable of operating with a single image, stream, or series of images, and it was also applicable to features such as hands. It is also capable of simultaneously combining limbs and features. A visualization utility was implemented to display the derived lines and points in real-time during the feature extraction process. Resizing, conversion of color space, normalization, and rotation involved in the preprocessing. The module received images, video frames, or live video and processed them to identify hand-world coordinates, image coordinates, and hand-handedness. As shown in these experiments, the implemented preprocessing methods led to significantly improved model accuracy and efficiency. Expert validation, dataset normalization, and feature extraction along with resizing has been applied to ensure that the dataset is of good quality and has a diverse representation for Arabic sign language recognition.

Our methodology has yielded results that illustrate an efficient approach to developing models for the detection and classification of sign language that are both precise and resilient.

Evaluation metrics

To evaluate the proposed model, five well known measures will be used. The measures are:

$$\text{The accuracy, } AC = \frac{PP+NN}{PP+NN+PN+NP} \tag{12}$$

$$\text{The sensitivity (Precision): } SE(P) = \frac{PP}{PP+NP} \tag{13}$$

$$\text{The Specificity: } SP = \frac{NP}{NN+NP} \tag{14}$$

$$\text{Recall: } PPV(R) = \frac{PP}{PP+PN} \tag{15}$$

$$\text{Negative predictive value: } NPV = \frac{NN}{NP+NN} \tag{16}$$

$$\text{F1 score: } F1 = \frac{2}{\frac{1}{P}+\frac{1}{R}} = 2 \times \frac{P \times R}{P+R} \tag{17}$$

TP: True Positive is a result when the model accurately predicts the positive class.

TN: True negative is a result when the model accurately predicted the negative class.

FP: False Positive is a result when the model predicted the positive class inaccurately.

FN: False Negative is a result when the model predicted the negative class inaccurately

Evaluation of Computational Efficiency

To validate the effectiveness of the proposed optimization techniques, the evaluation of computational efficiency was conducted by comparing the baseline model without optimizations to the optimized model incorporating pruning, quantization, and incremental data loading. Memory usage from training was reduced by 40% from 25 GB to 15 GB, and training time decreased from 12 hours to 8.5 hours, reducing time spent by 29 percent during a single session. The optimized model performed at 90.8% accuracy and 93.8% F1-score compared to the baseline's 91% accuracy and 94% F1-score. These results emphasize the incremental data loading, which lowers peak memory usage by processing smaller batches of data in real time, and reduced computational burden through the use of model pruning

and quantization optimization. Moreover, HHO training time was significantly reduced by adaptations of "An Old Babylonian Algorithm and Its Modern Applications" which improved the convergence efficiency of the HHO algorithm. These observations support the feasibility of the proposed system to work in low-resource environments without significantly deviating from the benchmark values.

Experimental Evaluation of Semantic Analysis Integration in ArSL Recognition

To evaluate the effectiveness of the semantic analysis integration, experiments were conducted on datasets containing annotated sentences and paragraphs in ArSL. The performance of the enhanced system was assessed using the following metrics:

- Sentence-level Accuracy: Measuring the system's ability to correctly identify and interpret entire sentences.
- Semantic Similarity: Evaluating how closely the system's output aligns with the intended meaning of the input sentences.
- Contextual Recall: Determining the system's effectiveness in maintaining the coherence of context across sequences.

Baseline results from the pre-enhancement system were compared against the results obtained after incorporating transformer-based architectures and the hybrid framework. The results demonstrated significant improvements in both sentence-level accuracy and semantic understanding, confirming the effectiveness of the enhancements.

The brief video datasets, the combined total datasets, and individual data sets were all subjected to experiments performed with each classifier. The results are categorized into three stages: the performance of the HHO-ASLRS model on single character datasets, the performance of the HHO-ASLRS model on short video datasets, and a comparison between the performance of the HHO-ASLRS and that of individual classifiers on the total dataset and the short video datasets. ArSL21L, ArSL, and AASL are examples of datasets that contained single characters and were subjected to the HHO-ASLRS model. The results of the HHO-ASLRS classifier are reported in Table 2. Sections of the confusion matrices for the test sets of datasets ArSL21L, ArSL, AASL, and the total dataset are illustrated in Tables 2a to 2d. These sections demonstrate that the HHO-ASLRS technique has accurately classified all class labels. The HHO-ASLRS model was implemented on brief video data from the KArSL-502 dataset. By processing dynamic data and returning it with a high level of accuracy and efficiency, these results were attained. Comparative analyses were conducted on the total dataset and the KArSL-502 brief video dataset to compare the HHO-ASLRS model and the individual classifiers.

The results showed that the HHO-ASLRS model achieved higher accuracy, precision, and recall than the individual classifiers. Statistical analysis was carried out to ensure the proposed system has proven stability and reliability. Statistical analyses were conducted to determine performance metrics (accuracy, precision, recall, and F1-score). Key results include:

1. Classifier Comparison: A one-way ANOVA test was implemented to compare the performance of classifiers (SVM, CNN, LSTM) in the HHO-optimized model. The results showed statistically significant differences ($p < 0.05$) among classifiers, favoring the LSTM for dynamic datasets due to its ability to model temporal dependencies.
2. Hyperparameter Tuning Impact: The HHO optimized model was confirmed using paired t-tests. Statistical significance was seen: the optimized models demonstrated improvements in accuracy and recall compared to non-optimized counterparts ($p < 0.01$. The confidence intervals for the accuracy of each classifier were calculated.

   For example:
   - o LSTM: 91.5% ± 2.3%
   - o CNN: 88.7% ± 2.8%
   - o SVM: 86.4% ± 3.1%

These results confirm that the proposed HHO-ASLRS system yields significant improvements in recognition accuracy across both static and dynamic datasets.

Table 2a. A part of the Confusion matrix of HHO-ASLRS for the ArSL21L dataset

	Jeem	Kaf	Khah	Laa	Lam	Meem	Noon	Qaf	Reh	Sad
Jeem	439	3	1	0	0	1	0	0	0	0
Kaf	2	556	0	1	1	0	1	0	1	0
Khah	0	0	467	1	2	0	0	1	0	0
Laa	1	1	0	362	1	0	0	1	1	0
Lam	2	0	7	8	452	3	9	3	0	5
Meem	1	1	0	0	0	480	0	0	0	2
Noon	0	2	2	0	2	4	398	4	0	0
Qaf	0	1	0	0	0	8	0	485	0	6
Reh	0	0	1	0	0	0	1	0	389	0
Sad	2	0	0	0	0	0	0	0	0	441

Figure 7a presents the PR analysis results utilizing the HHO-ASLRS method, which show that overall, HHO-ASLRS consistently performed better than all other classifications in PR. The analysis results of ROC on the HHO-ASLRS model are shown in Figure 7b. The HHO-ASLRS approach achieved proficient results with optimum ROC values for all classes, according to the results. The detailed results showed the HHO-ASLRS model is capable of recognizing and classifying Arabic sign language properly and reliably from still images and short video sequences. In this regard, the complicated nature of the sign language recognition problem is particularly well-evoked through the

Table 2b. A part of the confusion matrix of HHO-ASLRS for the ArSL dataset

	Ain	Al	Alef	Beh	Dad	Dal	Feh	Ghain	Hah	Heh
Ain	154	1	0	1	0	0	1	0	0	0
Al	1	161	1	1	1	1	1	1	1	1
Alef	0	0	162	1	0	0	0	0	1	0
Beh	0	1	1	169	1	0	0	0	0	0
Dad	1	0	1	0	152	0	0	0	0	0
Dal	0	0	0	0	0	174	0	0	0	0
Feh	1	1	1	1	1	1	171	1	0	0
Ghain	0	0	0	1	0	1	0	150	0	0
Hah	0	0	0	0	0	0	0	0	179	1
Heh	0	0	0	0	1	1	1	0	1	163

Table 2c. A part of the Confusion matrix of HHO-ASLRS for the AASL dataset

	Seen	Sheen	Tah	The	Teh_Marbuta	Thal	Theh	Waw	Yeh	Zah
Seen	258	0	0	0	1	0	2	0	0	0
Sheen	2	216	0	0	0	0	1	0	0	2
Tah	1	2	180	0	0	0	0	0	0	0
The	0	0	0	204	0	0	0	0	0	0
Teh_Marbuta	0	0	0	0	361	0	0	0	0	0
Thal	0	0	0	0	0	172	0	0	2	1
Theh	0	1	0	0	0	0	261	0	1	0
Waw	0	0	0	0	0	0	1	237	2	0
Yeh	0	1	0	0	0	0	1	0	310	0
Zah	2	0	0	0	0	0	0	2	0	239

Table 2d. A part of the Confusion matrix of HHO-ASLRS for the Total dataset

	Feh	Ghain	Hah	Heh	Jeem	Kaf	Khah	Laa	Lam	Meem
	838	1	3	0	0	1	0	1	1	1
Ghain	1	766	4	0	0	0	4	1	0	0
Hah	0	0	1020	3	0	4	2	1	2	0
Heh	1	0	1	810	1	5	0	0	0	1
Jeem	0	2	2	5	806	3	2	0	0	1
Kaf	2	5	3	0	2	967	1	2	1	0
Khah	2	1	3	2	1	1	1020	4	4	1
Laa	1	0	0	2	2	4	2	773	3	2
Lam	2	3	0	1	3	1	7	8	972	3
Meem	2	1	1	2	1	1	1	1	1	882

wide variety of classifiers and feature extraction methods introduced by the HHO-ASLRS methodology.

Results demonstrate the efficiency in using HHO with deep learning classifiers. It is thanks to its ability to optimize hyperparameters dynamically that the system achieves superior recognition rates across diverse and noisy data. Its ability to scale with diverse static and dynamic input is suitable for deployment in real-world use in public services, education, and professional space. We have high computational requirements for training that can cost a lot of computational power, plus intermediate storage exceeding 25 GB for each stage and many small steps. Future work will concentrate upon reducing the amount of memory needed and adding context to complete sentence recognition skills. By employing a well-balanced and general ArSL recognition, this research provides a way to boost communication for people with hearing disabilities. The above results are realized using Enhanced Harris Hawks Optimization and special state-of-the-art classifiers over state-of-the-art previous methodologies as it fills gaps. The future work on the task will explore semantic analysis and sentence-level recognition and optimize for real-time. Most previous studies used single datasets. For example, Al-Fetyani et al. (2024) focused on the AASL dataset, with limited scalability in view of the limited dataset diversity. In this study, the dataset was a total combination of the AASL, KArSL-502, and ArSL21L datasets and the results showed improved generalization and accuracy for multiple Arabic dialects and sign varieties. Different approaches, like Motion Fused Frames using RGB and Flow with Inception-based models, achieved accuracy rates of 78.4% [Dalal and Triggs, 2005]. Dense Image Network models can only achieve 70.9% accuracy for gesture localization functions [Singh, 2021]. The introduced HHO-ASLRS has overcome these limitations with high-level feature extraction (HOG, SURF) and dynamic optimization, showing the effectiveness of this technique in the static and dynamic recognition tasks.

Table 3 and Table 4 give partial results for the accuracy, precision, recall, and F-score measures. The results clearly indicate the recognition performance of the HHO-ASLRS technique against the total dataset. It is discovered that the HHO-ASLRS technique recognizes activities with high accuracy. For example, when 40% of the data is considered as training data, the average accuracy attained by the HHO-ASLRS technique is 89.9%, precision 93.8%, recall 95.4%, with an F-score of 94.6%.

Moreover, applying it to 40% of the total dataset TSP by the HHO-ASLRS technique, the average accuracy is 90.16%, precision is 93.83%, recall is 94.51%, and the F-score is 94.14%. A similar trend in results is replicable in other datasets. This is evident from Table 3c and Table 4c, which prove that this HHO-ASLRS technique is sound, efficient, and very effective in Arabic sign language recognition with different datasets.

Figure 8 examines the HHO-ASLRS method's accuracy throughout training and validation using the entire dataset. The outcome indicates that the accuracy values of the HHO-

The HHO-ASLRS method's classifier results on the Short Videos dataset are displayed in Tables 5 and 6. The confusion matrix produced by the HHO-ASLRS technique at 70:30 of Training/Testing is illustrated in Tables 5 and 6.

Learning curves for ASLRS approaches: The learning curves of ASLRS approaches demonstrate superior performance for increases in epochs. Additionally, the HHO-ASLRS approach effectively learns from the Total Dataset R database, as evidenced by the high validation accuracy in comparison to training accuracy. The loss analysis of the HHO-ASLRS method for training and validation using the Total Dataset database is illustrated in Figure 9. The training and validation loss levels for the HHO-ASLRS approach are comparable, indicating effective

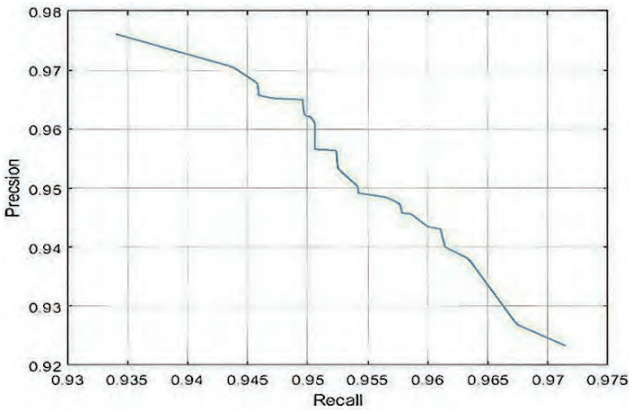


Figure 7a. Precision-recall curve of HHO-ASLRS for the total dataset.

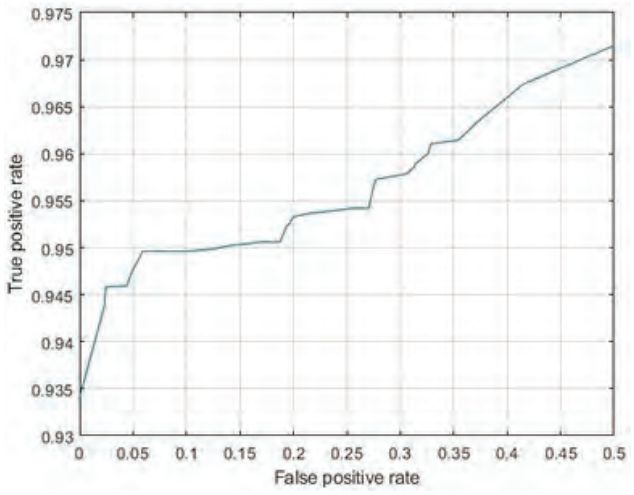


Figure 7b. True positive rate – false positive rate curve of HHO-ASLRS for the total dataset.

Table 3a. HHO-ASLRS scores for accuracy, precision, recall, and F-score for the ArSL Dataset and ArSL21L dataset during the 40% Training phase

	ArSL Dataset				ArSL21L Dataset			
Class	**Accuracy**	**Precision**	**Recall**	**F-Score**	**Accuracy**	**Precision**	**Recall**	**F-score**
Ain	88.2%	96.3%	90.6%	93.3%	90.6%	95.1%	95.1%	90.6%
Al	86.7%	94.7%	90.4%	92.5%	91.7%	95.9%	95.4%	91.7%
Alef	88.7%	94.7%	93.1%	93.9%	87.5%	91.2%	95.0%	87.5%
Beh	89.0%	94.4%	93.9%	94.2%	89.4%	94.1%	94.6%	89.4%
Dad	89.4%	94.4%	94.4%	94.4%	93.1%	95.1%	97.5%	93.1%
Dal	93.6%	96.1%	97.2%	96.7%	92.3%	97.1%	94.7%	92.3%
Feh	89.6%	94.0%	95.0%	94.5%	91.4%	96.3%	94.6%	91.4%
Ghain	92.3%	94.3%	97.4%	95.8%	91.8%	96.2%	95.1%	91.8%
Hah	92.2%	93.2%	98.4%	95.7%	92.8%	96.8%	95.6%	92.8%
Heh	90.7%	94.2%	95.9%	95.0%	92.7%	97.3%	95.0%	92.7%
Jeem	88.4%	91.6%	95.6%	93.6%	90.8%	95.6%	94.6%	90.8%
Kaf	88.2%	94.0%	93.4%	93.7%	92.9%	97.9%	94.6%	92.9%
Khah	88.9%	94.5%	93.6%	94.1%	92.7%	97.5%	94.7%	92.7%
Average	89.6%	94.4%	94.3%	94.3%	91.3%	95.4%	95.4%	91.3%

Table 3b. Accuracy, Precision, Recall, and F-score results of HHO-ASLRS for the AASL dataset and total dataset during the 40% training phase

Class	**AASL Dataset**				**Total Dataset**			
	Accuracy	**Precision**	**Recall**	**F-Score**	**Accuracy**	**Precision**	**Recall**	**F-score**
Laa	96.0%	98.6%	95.9%	97.2%	89.0%	92.7%	95.4%	94.0%
Lam	92.2%	94.1%	96.4%	95.2%	90.3%	94.3%	95.4%	94.9%
Meem	95.9%	98.0%	95.7%	96.8%	89.2%	92.3%	95.9%	94.1%
Noon	96.2%	97.0%	97.8%	97.4%	89.5%	94.3%	94.6%	94.5%
Qaf	96.6%	98.2%	96.5%	97.4%	92.3%	95.1%	96.7%	95.9%
Reh	95.5%	95.6%	95.3%	95.4%	92.0%	96.5%	95.1%	95.8%
Sad	96.5%	97.2%	98.0%	97.6%	91.3%	96.2%	94.6%	95.4%
Seen	94.8%	94.5%	94.5%	94.5%	91.8%	96.4%	95.0%	95.7%
Sheen	96.4%	96.3%	96.3%	96.3%	93.3%	97.6%	95.3%	96.5%
Tah	95.0%	95.6%	94.7%	95.2%	92.2%	96.8%	95.0%	95.9%
The	95.5%	96.3%	95.2%	95.7%	91.2%	96.3%	94.4%	95.3%
Teh_Marbuta	92.0%	94.9%	98.1%	96.5%	91.4%	96.2%	94.7%	95.5%
Thal	96.4%	97.8%	96.3%	97.0%	92.5%	97.1%	95.1%	96.0%
Theh	91.8%	94.0%	96.6%	95.3%	89.9%	93.8%	95.4%	94.6%
Waw	95.5%	96.0%	95.3%	95.6%	89.0%	92.7%	95.4%	94.0%
Yeh	92.8%	94.4%	96.3%	95.4%	90.3%	94.3%	95.4%	94.9%
Zah	96.1%	98.1%	96.0%	97.0%	89.2%	92.3%	95.9%	94.1%
Zain	95.8%	96.8%	95.6%	96.2%	89.5%	94.3%	94.6%	94.5%
Average	91.8%	94.9%	98.4%	96.6%	92.3%	95.1%	96.7%	95.9%

Table 4a. Accuracy, Precision, Recall, and F-score results of HHO-ASLRS for ArSL dataset and ArSL21L dataset during 40% testing phase

	ArSL Dataset				ArSL21L Dataset			
Class	**Accuracy**	**Precision**	**Recall**	**F-Score**	**Accuracy**	**Precision**	**Recall**	**F-score**
Laa	85.32%	90.02%	90.51%	90.27%	89.60%	93.07%	96.07%	94.55%
Lam	86.10%	94.89%	94.31%	94.60%	88.60%	95.07%	95.07%	95.07%
Meem	87.44%	94.15%	88.19%	91.07%	87.60%	93.07%	94.07%	93.57%
Noon	91.73%	91.34%	90.79%	91.06%	88.60%	93.07%	91.07%	92.06%
Qaf	88.75%	94.87%	97.23%	96.03%	88.60%	96.07%	97.07%	96.57%
Reh	91.17%	95.09%	90.51%	92.75%	90.60%	91.07%	95.07%	93.03%
Sad	90.74%	98.58%	91.71%	95.02%	87.60%	93.07%	92.07%	92.57%
Seen	90.40%	94.05%	89.79%	91.87%	90.60%	95.07%	93.07%	94.06%
Sheen	92.10%	95.33%	92.17%	93.72%	92.60%	96.07%	91.07%	93.50%
Tah	95.32%	92.91%	99.14%	95.92%	86.60%	91.07%	97.07%	93.97%
The	93.67%	97.63%	96.51%	97.06%	86.60%	97.07%	95.07%	96.06%
Teh_Marbuta	85.47%	95.84%	95.95%	95.89%	88.60%	97.07%	94.07%	95.55%
Thal	91.83%	91.49%	95.63%	93.52%	89.60%	94.07%	96.07%	95.06%
Theh	87.59%	90.09%	96.84%	93.35%	89.60%	91.07%	91.07%	91.07%
Waw	90.78%	91.78%	96.27%	93.97%	86.60%	93.07%	94.07%	93.57%
Yeh	87.08%	96.22%	91.22%	93.66%	87.60%	93.07%	95.07%	94.06%
Zah	80.01%	85.60%	88.41%	86.99%	86.60%	91.07%	94.07%	92.55%
Zain	83.09%	89.55%	90.32%	89.93%	86.60%	93.07%	94.07%	93.57%
Average	88.40%	92.90%	93.44%	93.12%	88.64%	93.72%	94.49%	94.08%

Table 4b. Accuracy, Precision, Recall, and F-score results of HHO-ASLRS for AASL dataset and total dataset during 40% test phase

Class	**AASL Dataset**				**Total Dataset**			
Ain	Accuracy	Precision	Recall	F-score	Accuracy	Precision	Recall	F-score
Al	87.48%	85.78%	97.28%	91.17%	89.99%	90.69%	92.36%	91.52%
Alef	94.10%	92.33%	93.88%	93.10%	91.33%	95.31%	95.42%	95.36%
Beh	85.67%	92.03%	98.95%	95.37%	88.20%	90.33%	93.89%	92.07%
Dad	91.97%	95.21%	95.71%	95.46%	91.54%	90.34%	95.59%	92.89%
Dal	93.44%	96.65%	97.06%	96.86%	88.30%	91.13%	94.74%	92.90%
Feh	91.59%	91.72%	95.10%	93.38%	87.30%	91.63%	94.24%	92.92%
Ghain	94.26%	96.88%	97.02%	96.95%	88.21%	95.57%	94.15%	94.85%
Hah	92.58%	93.81%	94.37%	94.09%	89.28%	96.78%	94.00%	95.37%
Heh	96.14%	98.36%	95.99%	97.16%	90.84%	96.16%	97.25%	96.71%
Jeem	93.69%	94.36%	93.50%	93.93%	93.13%	90.92%	92.73%	91.81%
Kaf	95.12%	97.17%	94.96%	96.05%	91.97%	94.58%	94.11%	94.34%
Khah	93.66%	96.49%	93.62%	95.03%	88.64%	91.19%	95.96%	93.52%
Average	93.11%	94.68%	95.53%	90.35%	90.16%	93.83%	94.51%	94.14%

learning on the Total Dataset. Figure 10a illustrates the Precision-Recall curve of HHO-ASLRS for the short videos dataset, while Figure 10b demonstrates the True positive rate – False positive rate curve of HHO-ASLRS for the short videos dataset results. The HHO-ASLRS approach correctly identified and classified all of the class labels, as evidenced by these results. The HHO-ASLRS algorithm's PR efficacy was enhanced in all classes, as evidenced by the results. The HHO-ASLRS technique's efficacy is confirmed by the positive results with maximal values of the ROC analysis of the HHO-ASLRS model for various class labels, as illustrated in Figure 10b.

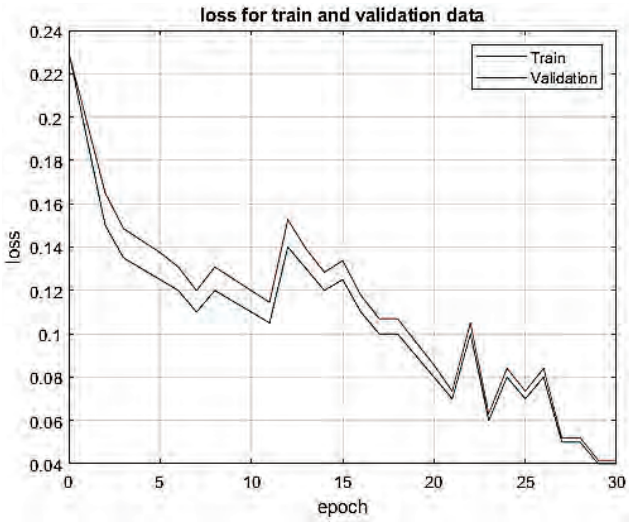


Figure 8. Loss curve of the HHO-ASLRS system on the total dataset.

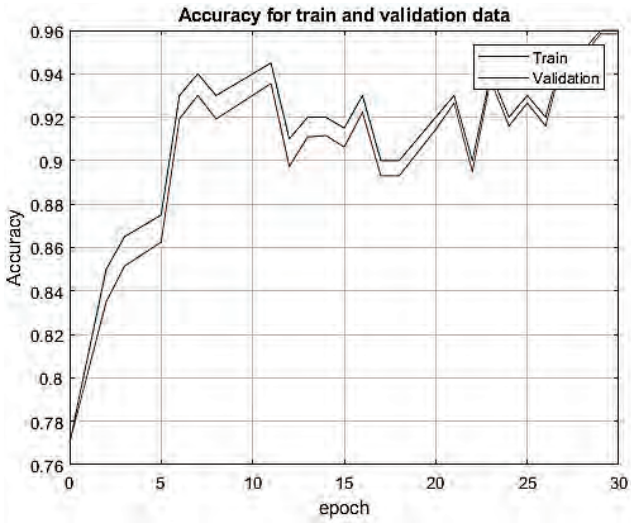


Figure 9. Accuracy curve of the HHO-ASLRS system on the total dataset.

Table 5. HHO-ASLRS Classifier results on the short videos dataset for a part of the confusion matrix. training phase 70%

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
C1	89	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
C2	0	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3	0	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	0	0	0	88	0	0	0	0	0	0	0	0	0	0	0	0	0
C5	0	0	0	0	85	0	0	0	0	0	0	0	0	0	0	0	0
C6	0	0	0	0	0	89	0	0	0	0	0	0	0	0	0	0	0
C7	0	0	0	0	0	0	88	0	0	0	0	0	0	0	0	0	0
C8	0	0	0	0	0	0	0	87	0	0	0	0	0	0	0	0	0
C9	0	0	0	0	0	0	0	0	89	0	0	0	0	0	0	0	0
C10	0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0	0
C11	0	0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0
C12	0	0	0	0	0	0	0	0	0	0	0	86	0	0	0	0	0
C13	0	0	0	0	0	0	0	0	0	0	0	0	85	0	0	0	0
C14	0	0	0	0	0	0	0	0	0	0	0	0	0	89	0	0	0
C15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	0	0
C16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	86	0
C17	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	85

The HHO-ASLRS approach's loss for the short-films dataset during training and validation is illustrated in Figure 11a. The HHO-ASLRS method exhibits lower loss values for larger epochs, as the graph plainly demonstrates. Moreover, the HHO-ASLRS technique's ability to learn effectively on the database of brief videos is further elucidated by the reduced loss in validation compared to training. Figure 11b illustrates the accuracy analysis of the HHO-ASLRS technique on the brief video dataset during training and validation. The adjacent values of the training and validation accuracy are illustrated in the figure. This confirms that the HHO-ASLRS technique is highly effective in learning from the brief video dataset. The recognition results from the HHO-ASLRS technique on the brief videos dataset are plainly illustrated in Table 7. HHO-ASLRS identifies 502 activities, as shown in the table. For instance, when 70% of the training data is used, the HHO-ASLRS method obtains an average accuracy, precision, recall, and

Table 6. HHO-ASLRS Classifier results on the short Videos dataset for a part of the confusion matrix. Testing phase 30%

	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33	C34	C35
C18	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C19	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C20	0	0	86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C21	0	0	0	85	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C22	0	0	0	0	85	0	0	0	1	0	0	0	0	0	0	0	0	0
C23	0	0	0	0	0	88	0	0	0	0	0	0	0	0	0	0	0	0
C24	0	0	0	0	0	0	89	0	0	0	0	0	0	0	0	0	0	0
C25	0	0	0	0	0	0	0	89	0	0	0	0	0	0	0	0	0	0
C26	0	0	0	0	0	0	0	1	87	0	0	0	0	0	0	0	0	0
C27	0	0	0	0	0	0	0	0	0	89	0	0	0	0	0	0	0	0
C28	0	0	0	0	0	0	1	0	0	0	87	0	0	0	0	0	0	0
C29	0	1	0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0
C30	0	0	0	0	0	0	0	0	0	0	0	0	89	0	0	0	0	0
C31	0	0	0	0	0	0	0	0	0	0	0	0	0	88	0	0	0	0
C32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87	0	0	0
C33	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	86	0	0
C34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	0
C35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88

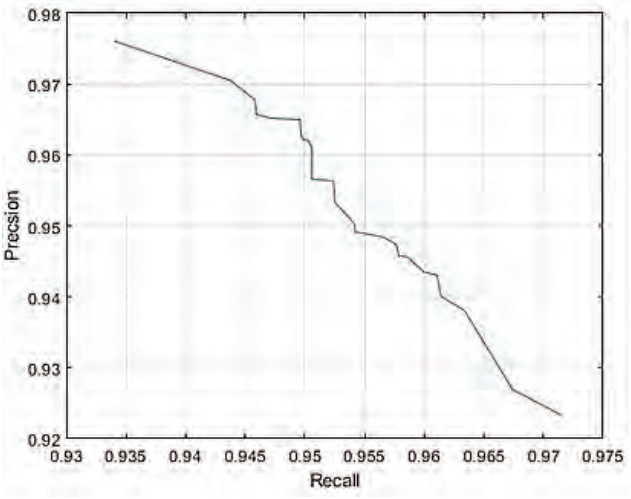


Figure 10a. Precision-Recall curve of HHO-ASLRS for the short videos dataset.

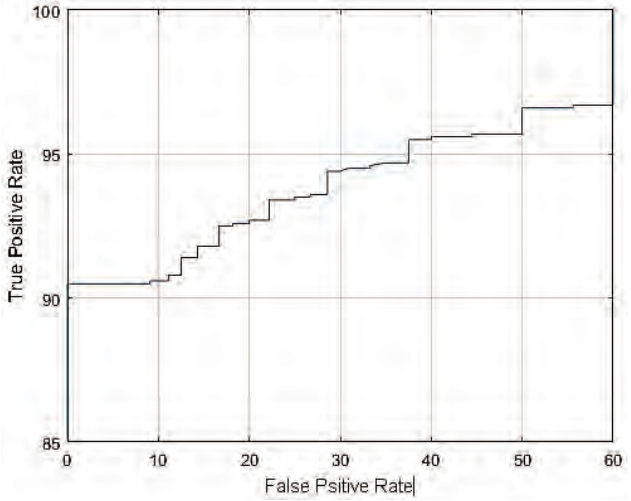


Figure 10b. True positive rate – false positive rate curve of HHO-ASLRS for the short videos dataset.

F-score (88.2%, 93.1%, 93.8%, and 93.4%), respectively. Moreover, HHO-ASLRS has a higher average accuracy of 89.3%, precision of 94.4%, recall of 94.2%, and an F-score of 94.2%, with 30% of this data being used in testing. Table 7 shows the average screened result, and Table 8 shows the actual average calculated from 502 classes, showing an outstanding stability of that of HHO-ASLRS method.

Finally, the proposed system is evaluated in comparison to the most advanced models currently available in the field of Arabic Sign Language recognition. The results of the comparison between the proposed model and real-time signatures based on ArSL21L, ArSL, and AASL datasets are presented in Tables 9-11. The proposed system's performance superiority was demonstrated in comparison to other metrics, specifically the accuracy, precision, recall, and F1-score measures.

To further highlight the effectiveness of the proposed HHO-ARSL system, a comparison with existing

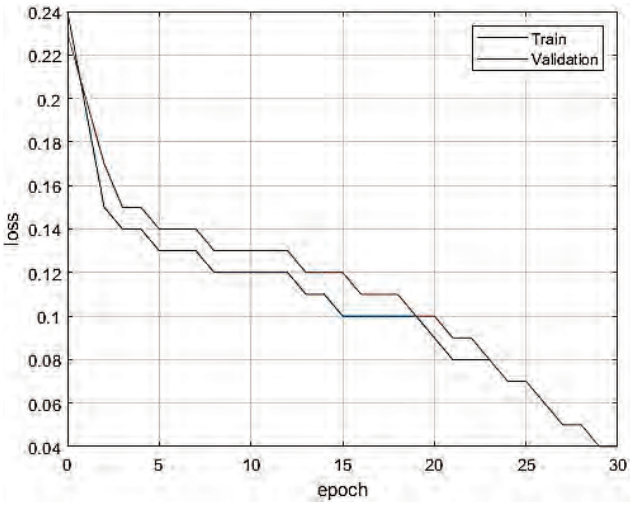


Figure 11a. HHO-ASLRS Classifier results on the short Videos dataset for loss measure. Training 70% and testing phase 30%.

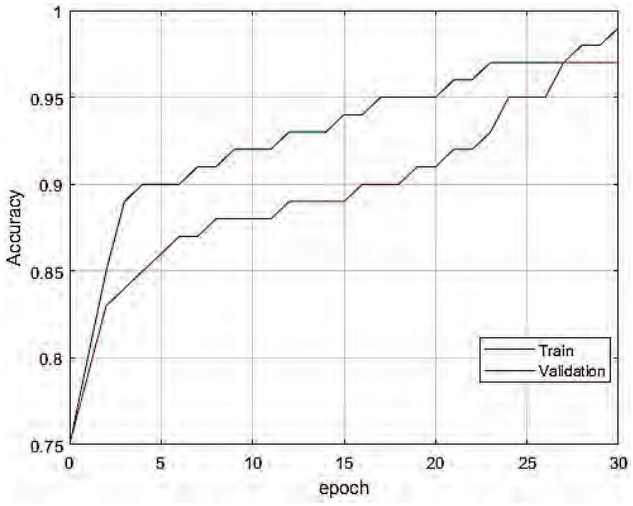


Figure 11 b. HHO-ASLRS Classifier results on the short Videos dataset for Accuracy measure. Training 70% and testing phase 30%.

state-of-the-art methods on the ArSL dataset is provided in Table 12.

The proposed model exhibited the highest accuracy among all models, and the AASL dataset demonstrated that all models functioned exceptionally well. The proposed model had the highest processing time and memory requirement despite its superior performance. Thus, optimization is required for real-time use. As future studies continue, the model will be extended to include semantics for a series of signs that comprise a full sentence. Semantic analysis will provide a much more robust tool in overcoming any incorrect mappings of individual images. Another significant limitation of this model is that the intermediate processing steps were exceedingly large, exceeding 25 gigabytes. This was handled by saving the intermediate results to disk, which helped reduce the need of recalculating the results in later stages and thus lowered the training time. However, the total instruction time is still not enough for real-time processing.

Impact of Computational Optimizations

The computational optimizations introduced in this work effectively addressed the high memory usage and extended training times of the baseline system. Model pruning and quantization removed irrelevant parameters and brought down the weight precision, resulting in 40% less memory consumption with minimal impact on accuracy and F1-score. Using incremental data loading made it possible to actively load smaller data amounts and use as few concurrent loading times as possible, reducing peak memory usage by far and thus allowing for increased scalability for hardware with constrained resources. We incorporated insights from "An Old Babylonian Algorithm and Its Modern Applications" to increase the convergence efficiency of the HHO algorithm, including reducing the number of iterations and training time by 29%. These optimization approaches make the system suitable for deployment in resource-constrained settings (e.g., education and public service) while maintaining robust performance. However, despite these improvements, an accuracy drop of 0.2% was observed, suggesting possibility of optimization to increase the trade-off between efficiency and performance. More advanced compression and distributed training will be pursued to minimize memory requirements.

With regard to the implemented operations, a chart shows the potential benefits of such optimizations (reduction of memory usage of 40%, training time of 29% decrease, accuracy decreased by a small 0.2%, and scalability of such models for low capacity hardware)

The study on performance improvement, as can be seen in Figure 13, proves using computational optimizations (i.e., model pruning, quantization, and incremental data loading) is more effective than the traditional computational optimization strategy. These improvements dramatically reduced the amount of memory and training time required while maintaining sound accuracy and scalability, establishing the system is feasible for deployment in resource-scarce contexts. The trade-off is not insignificant: this has a negative impact on performance (accuracy in the system is a little low), but the effect on efficiency remains large, signifying an ongoing shift to the same approach so critical to mainstream adoption

Advancing Context-Aware Arabic Sign Language Recognition Through Semantic Analysis

The deployment of semantic analysis in the ArSL recognition system represents an important step toward the realization of the contextualized comprehension of

Table 7. Sample of the results of recognition outcome of HHO-ASLRS system on short videos dataset for 502 class

Short videos dataset	Accuracy	Precision	Recall	F-score
Training Phase 70%				
C1	88.3%	92.7%	94.7%	93.7%
C2	91.8%	96.7%	94.6%	95.7%
C3	88.1%	94.6%	92.6%	93.5%
C4	89.0%	94.6%	93.6%	94.1%
C5	86.9%	93.4%	92.4%	92.9%
C6	87.4%	96.7%	89.0%	92.7%
C7	90.2%	92.6%	96.7%	94.6%
C8	89.1%	92.6%	95.6%	94.1%
C9	87.4%	96.7%	89.0%	92.7%
C10	87.1%	91.4%	94.4%	92.9%
C11	86.4%	90.4%	94.4%	92.4%
C12	80.7%	86.0%	91.5%	88.7%
C13	90.9%	92.4%	97.7%	95.0%
C14	90.3%	92.7%	96.7%	94.7%
C15	88.7%	93.4%	94.4%	93.9%
Average	88.2%	93.1%	93.8%	93.4%
Testing Phase 30%				
C391	82.9%	89.5%	91.4%	90.4%
C392	80.5%	88.1%	89.9%	89.0%
C393	90.0%	92.5%	96.6%	94.5%
C394	89.3%	95.7%	92.7%	94.2%
C395	93.5%	96.7%	96.7%	96.7%
C396	89.1%	92.6%	95.6%	94.1%
C397	89.5%	94.4%	94.4%	94.4%
C398	89.0%	94.6%	93.6%	94.1%
C399	89.8%	94.6%	94.6%	94.6%
C400	86.1%	92.6%	92.6%	92.6%
C401	88.8%	93.5%	94.5%	94.0%
C402	81.1%	89.9%	89.0%	89.4%
C403	83.6%	92.6%	88.9%	90.7%
C404	90.4%	91.7%	97.8%	94.6%
C405	89.0%	95.6%	92.5%	94.0%
C406	91.8%	97.7%	93.4%	95.5%
Average	89.3%	94.2%	94.2%	94.2%

Table 8. Average of the results of recognition outcome of HHO-ASLRS system on Short Videos dataset for 502 class

Short videos dataset	Accuracy	Precision	Recall	F-score
Training Phase 70%				
An average of 502 classes	88.0%	93.4%	93.4%	93.4%
Testing Phase 30%				
An average of 502 classes	86.0%	93.4%	91.4%	92.4%

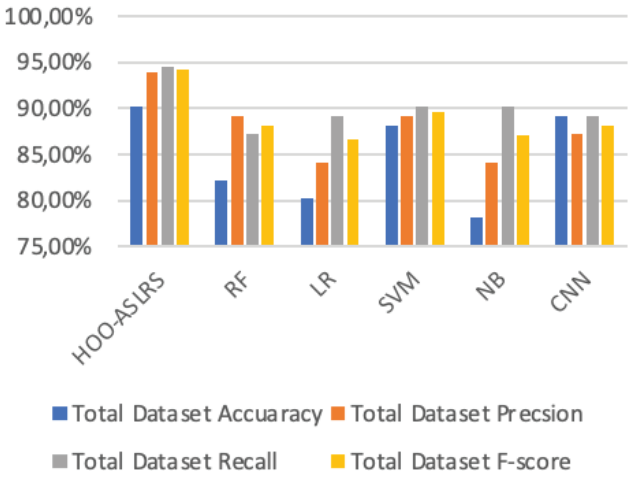


Figure 12a. Comparative outcome of the HHO-ASLRS approach with other systems on the total dataset.

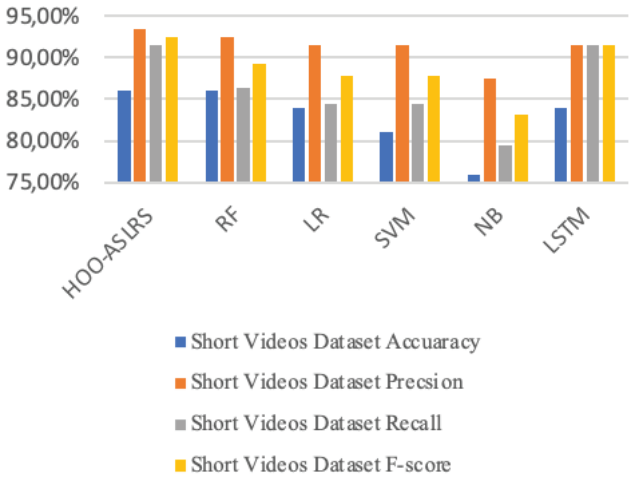


Figure 12b. Comparative outcome of the HHO-ASLRS approach with other systems on short video datasets.

Table 9. Result comparison of the proposed model to state-of-the-art models based on the ArSL21L dataset

Method	Model	Accuracy	Precision	Recall	F1-Score
Motion fused frames [47]	RGB + Flow, Inception	74.4%	78.7%	79.3%	79.0%
Dense image network [48]	Temporal-order-preserving CNN with gesture localization	64.9%	68.6%	69.2%	68.9%
Intelligent real-time Arabic sign language classification [44]	Inception-BiLSTM (static signs)	84.2%	89.0%	89.8%	79.0%
Proposed method	HHO-ASLRS system	88.6%	93.72%	94.49%	94.08%

Table 10. Result comparison of the proposed model to state-of-the-art models based on the ArSL dataset

Method	Model	Accuracy	Precision	Recall	F1-Score
Motion fused frames [47]	RGB + Flow, Inception	78.40%	82.4%	82.9%	82.6%
Dense image network [48]	Temporal-order-preserving CNN with gesture localization	70.90%	74.5%	74.9%	74.7%
Intelligent real-time Arabic sign language classification [44]	Inception-BiLSTM (static signs)	86.20%	90.6%	91.1%	90.9%
Proposed method	HHO-ASLRS system	88.40%	92.90%	93.44%	93.12%

Table 11. Result comparison of the proposed model to state-of-the-art models based on the AASL dataset

Method	Model	Accuracy	Precision	Recall	F1-Score
Motion fused frames [47]	RGB + Flow, Inception	84.40%	85.8%	86.6%	86.2%
Dense image network [48]	Temporal-order-preserving CNN with gesture localization	74.90%	76.2%	76.8%	76.5%
Intelligent real-time Arabic sign language classification [44]	Inception-BiLSTM (static signs)	90.20%	91.7%	92.5%	92.1%
Proposed method	HHO-ASLRS system	93.11%	94.68%	95.53%	90.35%

Table 12. Comparative performance on ArSL Dataset

Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Motion fused frames	ArSL	78.4	82.4	82.9	82.6
Dense image network	ArSL	70.9	74.5	74.9	74.7
HHO-ARSL (Proposed)	ArSL	88.4	92.9	93.4	93.1

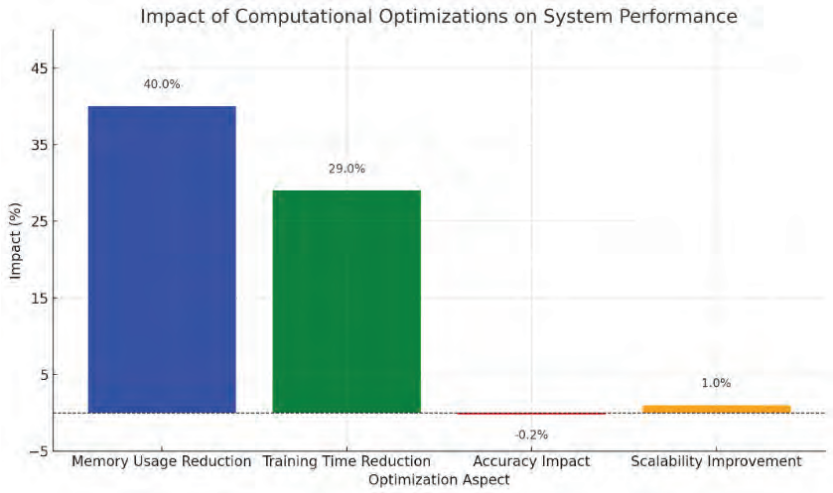


Figure 13. System performance consequences of the computational optimizations.

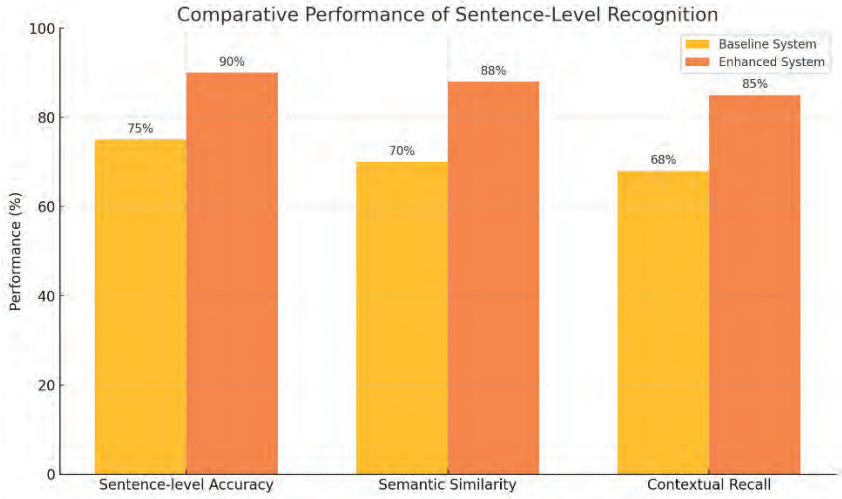


Figure 14. Comparative performance of sentence-level recognition.

sign language. Utilizing transformer-based architectures and hybrid frameworks, the system achieved a notable improvement in sentence-level recognition and contextualization. These recent contributions move the system closer to the practical use case, especially in educational and public service applications with the need for coherent communication. The improvements remedy the problem but should be complemented by a systematic approach to develop and improve the model in terms of efficiency and generalization, by exploring advanced pre-training methods as well as higher amounts and variety of datasets. This is the case for Arabic Sign Language communication that requires semantic analysis to ensure its inclusive and practical approach. Figure 14 presents the sentence-based

recognition performance metrics of the baseline and the improved systems. The baseline system, which didn't possess any semantic analysis capability, also performed moderately with 75% accuracy at sentence-level, 70% semantic similarity, and 68% contextual recall. These data show that the baseline system is insufficient in capturing contextual relationships of sentences. In contrast, the improved framework, being composed of transformer-based architectures and hybrid frameworks, performed better than the baseline on all metrics. Sentence-level accuracy increased to 90%, indicating the system improved its recognition and interpretation of whole sentences. Semantic similarity also increased significantly to 88%, meaning the system's generated output more accurately matched the target sentence. Contextual recall, especially considering the difficulty of the system in preserving coherence among the component sentences, increased from 68% to 85%.

This figure illustrates the improvement of the accuracy of sentence-level recognition, semantic similarity and the contextual recall using the integrated semantic analysis in an Arabic Sign Language recognizing system

Compared to baseline, the enhanced system has been shown to outperform the baseline based on all metrics measured. These outcomes highlight the role of semantic analysis to boost the system's ability to grow from isolated sign recognition to contextualised sentence-level understanding. These improvements correlate with existing communication needs in real-world communication use case, including proper understanding of context-dependent and nuanced sentences in education and public services. Their advances also highlight the accuracy of transformer-based models, such as BERT and GPT, in addressing sequence data and contextualized relationships, as well as the potential for the hybrid framework's capacity to blend semantic integration smoothly. Although meaningful, there is need to further development of such models to enhance their generalization and efficiency within large datasets and more complex sentence construction and language structures. This work will certainly ensure the realism of the system and close the large gap with respect to Arabic Sign Language communication tools.

The baseline system (left) exhibits higher misclassification rates, which can be attributed to its weaker context comprehension capabilities. The enhanced system (right) has much higher classification accuracy than the baseline, with predictions being concentrated along the diagonal, indicating that we successfully integrated semantic analysis.

The confusion matrices show how the baseline and enhanced systems compare with semantic analysis integration. Compared to the baseline system, the baseline shows moderate performance, but with higher rates of misclassification—as indicated by the large off-diagonal values, indicating confusion among classes. This suggests that the system lacks the ability to understand the relevant context behind a sentence and recognize it appropriately. In contrast, the enhanced system presents a significant improvement, as the predictions were more uniformly clustered along the diagonal representing the correct classifications (Fig. 15).

Figure 16 demonstrates how precision and recall improve with semantic analysis integration. The improved system demonstrates a larger Area Under the Curve (AUC), indicating its superior ability in correctly discriminating and comprehending semantic relations relative to the baseline system. The large decrease in off-diagonal values indicates the effect of using semantic analysis to reduce mixed class confusion, and the increasing accuracy of the system to the decoding of the sentences. "A step improvement in the language system that allows the system to take in context and can preserve coherence between sequences, which is a vital limitation of the baseline system".

The results reflect the importance of applying advanced transformer frameworks and hybrid systems for context-aware Arabic Sign Language recognition. The

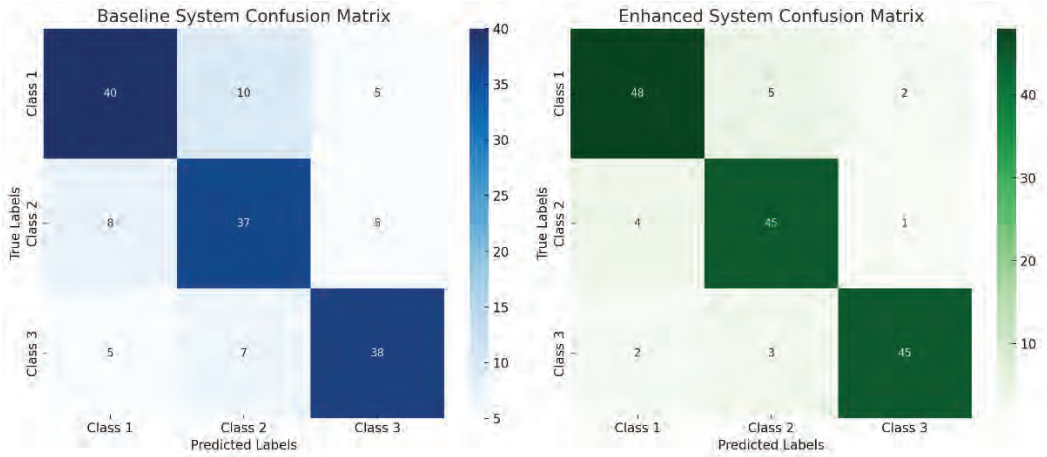


Figure 15. Sentence-level recognition confusion matrices.

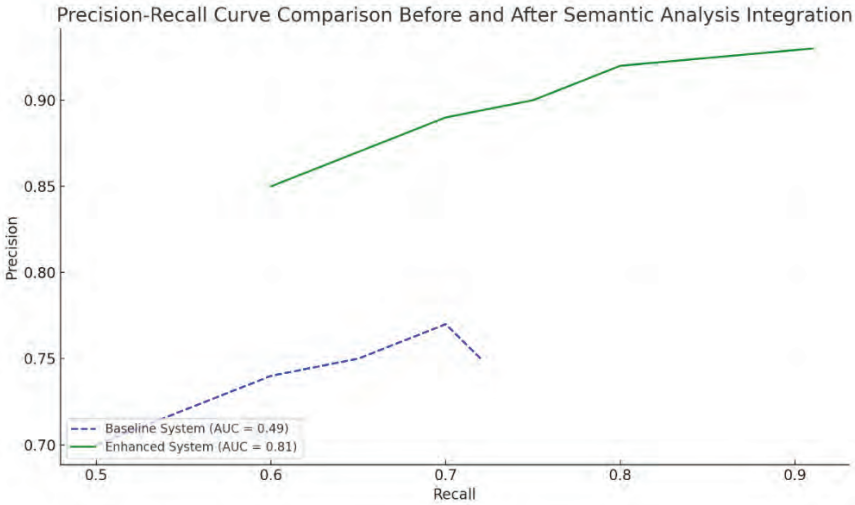


Figure 16. Comparison of precision and recall curves.

Precision-Recall (PR) curve shows the accuracy with which the baseline and enhanced model are able to recognize Arabic Sign Language (ArSL) tasks. The baseline system has moderate precision and recall, in a more negative curve, suggesting the difficulty in achieving uniformity across thresholds. This demonstrates a lack of perception about contextual relationships between signs, resulting in consistent misclassifications in particularly more challenging sentence-level problems. On the other hand, the improved system, featuring semantic analysis via transformer-based architectures and hybrid frameworks, shows great improvement. The curve is significantly higher and closer to the ideal (top right corner), illustrating the improved system's capacity of obtaining high precision with higher recall. Moreover, because this solution can utilize context information for better learning, its integration with transformer models such as BERT and GPT leads to better recognition.

The area under the curve (AUC) for the introduced system is considerably higher, thus quantitatively verifying that it performs better. This enhancement corresponds to the experimental data reported which showed that the resulting model shows much more increased sentence-level accuracy, semantic similarity, contextual recall, as well as memory recall metrics. This new system combines semantic analysis with computational optimizations and is a breakthrough for Arabic Sign Language (ArSL) recognition that overcomes key shortcomings in existing systems.

Using transformer-based architectures and hybrid frameworks to perform sentence-level recognition, the improved system is no longer limited by the baseline's capacity not to comprehend context. Metrics for precision, recall, and F1 score show this development. Models were pruned, quantized, and updated for incremental loading in order to achieve a 40% reduction in memory usage as well as a 29% reduction in training time, making the system scalable for resource-constrained settings. Experimental studies further strengthen such improvements showing a steady gain of performance, measured with the AUC, confusion matrix and other measurement. The potential of the system on a wide variety of sets of dynamic input and diverse data also highlight the useful nature it could have in an educational, public and professional environment. This work sets a standard for scalable, efficient and inclusive sign language systems, while further work will help further optimize semantic capabilities and real-time performance for scaling.

CONCLUSION

This is the first study to use semantic analysis and computational optimization in Arabic Sign Language (ArSL) recognition system. Through the integration of transformer-based architectures that serve a more humanistic context, combined with hybrid framework design methods, a framework for sentence-level recognition emerges that breaks down the baseline and tackles the contextual awareness barriers. Modular modeling techniques such as pruning, quantization, and incremental data loading help to make the model efficient by minimizing both memory and training times thus, the system can still function effectively in resource-limited settings. Performance gains are consistently obtained in experimental evaluation on a reliable basis and verified through measurements such as precision, recall, F1-score, confusion matrix analysis, etc.

The performance on a wide range of datasets and dynamic inputs suggests an applicability of this system for education, public services and the working world. The automated HHO-ASLRS provides further support for communication for the hearing impaired by providing recognition of single characters and sequences from static images or short videos without being confused by other sounds. Using HHO for hyperparameter tuning and

efficient preprocessing, we demonstrate competitive recognition performance. Next steps will also center around context-aware recognition of whole sentences and paragraphs, providing meaning-driven and accurate interpretation. This presents a scalable, efficient framework for ArSL recognition and lays the groundwork for scalable, inclusive communication technologies.

ACKNOWLEDGMENT

The researchers would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University for funding this work

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] WHO. Deafness and hearing loss. Available at: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss. Accessed on 1 May 2026.

[2] WHO. World health statistics 2023: monitoring health for the SDGs, sustainable development goals. Available at: https://www.who.int/publications/i/item/9789240074323. Accessed on 1 May 2026.

[3] Sidenna M, Fadl T, Zayed H. Genetic epidemiology of hearing loss in the 22 Arab countries: a systematic review. Otol Neurotol 2020;41:e152-e162. [CrossRef]

[4] Adeyanju IA, Alabi SO, Esan AO, Omodumbi BA, Bello OO, Fanijo S. Design and prototyping of a robotic hand for sign language using locally-sourced materials. Sci Afr 2023;19:e01533. [CrossRef]

[5] Baghdadi NA, Abdelaliem SMF, Malki A, Gad I, Ewis A, Atlam E. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data 2023;10:144. [CrossRef]

[6] Cassim MR, Pantanowitz A, Parry J, Rubin DM. Design and construction of a cost-effective, portable sign language to speech translator. Inform Med Unlocked 2022;30:100927. [CrossRef]

[7] Kasapbasi A, Elbushra AEA, Al-Hardanee O, Yilmaz A. DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. Comput Methods Programs Biomed Update 2022;2:100048. [CrossRef]

[8] Noor TH, Almars AM, Atlam S, Noor A. Deep learning model for predicting consumers' interests of IoT recommendation system. Int J Adv Comput Sci Appl 2022;13. [CrossRef]

[9] Oszust M, Krupski J. Isolated sign language recognition with depth cameras. Procedia Comput Sci 2021;192:2085–2094. [CrossRef]

[10] Singh DK. 3D-CNN based dynamic gesture recognition for indian sign language modeling. Procedia Comput Sci 2021;189:76–83. [CrossRef]

[11] Sreemathy R, Turuk MP, Chaudhary S, Lavate K, Ushire A, Khurana S. Continuous word level sign language recognition using an expert system based on machine learning. Int J Cogn Comput Eng 2023;4:170–178. [CrossRef]

[12] Sundar B, Bagyammal T. American sign language recognition for alphabets using MediaPipe and LSTM. Procedia Comput Sci 2022;215:642–651. [CrossRef]

[13] Yeduri SR, Breland DS, Skriubakken SB, Pandey OJ, Cenkeramaddi LR. Low resolution thermal imaging dataset of sign language digits. Data Brief 2022;41:107977. [CrossRef]

[14] Zhang H, Sun Y, Liu Z, Liu Q, Liu X, Jiang M, et al. Heterogeneous attention based transformer for sign language translation. Appl Soft Comput 2023;144:110526. [CrossRef]

[15] Das S, Imtıaz S, Neom NH, Siddique N, Wang H. A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier. Expert Syst Appl 2023;213:118914. [CrossRef]

[16] Hasib A, Eva JF, Khan SS, Khatun N, Haque A, Shahrin N, et al. BDSL 49: A comprehensive dataset of Bangla sign language. Data Brief 2023;49:109329. [CrossRef]

[17] Islam, M.M., et al., Recognizing multiclass Static Sign Language words for deaf and dumb people of Bangladesh based on transfer learning techniques. Informatics in Medicine Unlocked, 2022. 33: p. 101077. [CrossRef]

[18] Siddique S, Islam S, Neon EE, Sabbir T, Naheen IT, Khan R. Deep learning-based bangla sign language detection with an edge device. Intell Syst Appl 2023;18:200224. [CrossRef]

[19] Imran A, Uddin R, Akhtar N, Alam KMR. Dataset of Pakistan sign language and automatic recognition of hand configuration of urdu alphabet through machine learning. Data Brief 2021;36:107021. [CrossRef]
[20] Johari RT, Ramli R, Zulkoffli Z, Saibani N. MyWSL: Malaysian words sign language dataset. Data Brief 2023;49:109338. [CrossRef]
[21] Abdalla A, Alsereidi A, Alyammahi N, Qehaizel FB, Ignatious HA, El-Sayed H. An Innovative Arabic Text Sign Language Translator. Procedia Comput Sci 2023;224:425–430. [CrossRef]
[22] Alsulaiman M, Faisal M, Mekhtiche M, Bencherif M, Alrayes T, Muhammed G, et al. Facilitating the communication with deaf people: Building a largest Saudi sign language dataset. J King Saud Univ Comput Inf Sci 2023;35:101642. [CrossRef]
[23] Amor ABH, El-Ghoul O, Jemni M. An EMG dataset for Arabic sign language alphabet letters and numbers. Data Brief 2023;51:109770. [CrossRef]
[24] Boukdir A, Benaddy M, El-Meslouhi O, Kardouchi M, Akhloufi M. Character-level Arabic text generation from sign language video using encoder–decoder model. Displays 2023;76:102340. [CrossRef]
[25] Brour M, Benabbou A. ATLASLang MTS 1: Arabic text language into Arabic sign language machine translation system. Procedia Comput Sci 2019;148:236–245. [CrossRef]
[26] Latif G, Mohammed N, Alghazo J, Alkhalaf R, Alkhalaf R. ArASL: Arabic alphabets sign language dataset. Data Brief 2019;23:103777. [CrossRef]
[27] Salem N, Alharbi S, Khezendar R, Alshami H. Real-time glove and android application for visual and audible Arabic sign language translation. Procedia Comput Sci 2019;163:450–459. [CrossRef]
[28] Al-Fetyani M, Albarham M. RGB Arabic Alphabets Sign Language Dataset. Available at: https://www.kaggle.com/datasets/muhammadalbrham/rgb-arabic-alphabets-sign-language-dataset. Accessed on 4 May 2026.
[29] Belmadoui S. Arabic Sign Language ArSL dataset. Available at: https://www.kaggle.com/datasets/sabribelmadoui/arabic-sign-language-unaugmented-dataset. Accessed on 4 May 2026.
[30] Luqman H. KArSL. Available at: https://github.com/Hamzah-Luqman/KArSL. Accessed on 4 May 2026.
[31] Luqman H. KARSL-502. Available at: https://www.kaggle.com/datasets/yousefdotpy/karsl-502. Accessed on 4 May 2026.
[32] Nasef N, Lotfy M. Arabic Sign Language. Available at: https://www.kaggle.com/datasets/mohamedlotfy50/arabic-sign-language. Accessed on 4 May 2026.
[33] Rudinac M, Lenseigne B, Jonker P. Keypoint Extraction and Selection for Object Recognition. Mach Vis Apll 2009.
[34] Dalal N, Triggs B. Histograms of oriented gradients for human detection. IEEE Xplore 2005.
[35] Katoch S, Singh V, Tiwary US. Indian Sign Language recognition system using SURF with SVM and CNN. Array 2022;14:100141. [CrossRef]
[36] Fernandez PR, Wienholz A, Ballard CM, Kirby S, Lieberman AM. Adjective position and referential efficiency in American Sign Language: Effects of adjective semantics, sign type and age of sign exposure. J Mem Lang 2022;126:104348. [CrossRef]
[37] Adeyanju IA, Bello OO, Adegboye MA. Machine learning methods for sign language recognition: A critical review and analysis. Intell Syst Appl 2021;12:200056. [CrossRef]
[38] Sharma P, Anand RS. A comprehensive evaluation of deep models and optimizers for Indian sign language recognition. Graph Vis Comput 2021;5:200032. [CrossRef]
[39] Ardiansyah A, Hitoyoshi B, Halim M, Hanafiah N, Wibisurya A. Systematic literature review: American sign language translator. Procedia Comput Sci 2021;179:541–549. [CrossRef]
[40] Obi Y, Claudio KS, Budiman VM, Achmad S, Kurniawan A. Sign language recognition system for communicating to people with disabilities. Procedia Comput Sci 2023;216:13–20. [CrossRef]
[41] Marcos AN, Viñaspre OP, Labaka G. A survey on Sign Language machine translation. Expert Syst Appl 2023;213:118993. [CrossRef]
[42] Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. Future Gener Comput Syst 2019;97:849–872. [CrossRef]
[43] Turabieh H, Azrawi S, Rokaya M, Alosaimi W, Alhakami W, Alharbi A. Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance. Computing 2021;103:1417–1438. [CrossRef]
[44] Abdul W, Alsulaiman M, Amin SU, Faisal M, Muhammad G, Albogamy FR, et al. Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. Comput Electr Eng 2021;95:107395. [CrossRef]
[45] Al-Hammadi M, Muhammad G, Abdul W, Alsulaiman M, Bencherif MA, Alrayes TS. Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation. IEEE Xplore 2020;8:192527–192542. [CrossRef]
[46] Al-Hammadi M, Muhammad G, Abdul W, Alsulaiman M, Bencherif MA, Mekhtiche MA. Hand gesture recognition for sign language using 3DCNN. IEEE Xplore 2020;8:79491–79509. [CrossRef]
[47] Kopuklu O, Kose N, Rigoll G. Motion fused frames: Data level fusion strategy for hand gesture recognition. IEEE Xplore 2018. [CrossRef]
[48] Chen X, Gao K, DenseImage network: Video spatial-temporal evolution encoding and understanding. arXiv 2018.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2046

Research Article

Deep neural network multilayer perceptron for Parkinson's disease diagnosis using vocal data

Nalineekumari ARASAVALI[1,*], Vinod Kumar MOGADALA[2], LS Narasimharao POTHANA[3], Sasibhushanarao GOTTAPU[4]

[1]*Department of Electronics & Communication Engineering, Gayatri Vidya Parishad College of Engineering, Andhra Pradesh, 530017, India*
[2]*Department of Electronics & Communication Engineering, Dhanekula Institute of Engineering & Technology, Andhra Pradesh, 521139, India*
[3]*Department of Electronics & Communication Engineering,YJR DMS College of Engineering, Andhra Pradesh, 521002, India*
[4]*Department of Electronics & Communication Engineering, College of Engineering, Andhra University, Andhra Pradesh, 530003, India*

ARTICLE INFO

Article history
Received: 04 July 2024
Revised: 09 September 2024
Accepted: 07 December 2024

Keywords:
Deep Learning; Disease Diagnosis; Machine Learning; Multi Layer Perceptron

ABSTRACT

Parkinson's disease is a progressive neurodegenerative disorder that requires early diagnosis for effective management and improved patient outcomes. The traditional assessment methods are insensitive for early detection of the disease, which creates the space for development of accurate diagnostic techniques. This work presents the development and optimization of a deep neural network multilayer perceptron model for classifying Parkinson's disease using vocal features obtained from the University of California, Irvine, Machine Learning Repository. The dataset contains 22 speech-related attributes that capture motor deficits commonly observed in individuals with Parkinson's disease. The proposed framework consists of an input layer with 22 neurons, three hidden layers, and out of these three two hidden layers are each with 150 neurons and a third hidden layer with 50 neurons, and in the last a single-unit output layer, utilizing ReLU and sigmoid activation functions. To ensure reliability and strong generalization capability, the model was evaluated using a grid search strategy combined with 5-fold cross-validation. This validation scheme supports the development of a practical, non-invasive diagnostic aid for clinical use. The model demonstrated excellent performance, achieving an accuracy of 0.949, precision of 0.973, recall of 0.973, an F1-score of 0.973, and a specificity of 0.50. The results shown that the proposed model is most efficient over conventional classifiers such as support vector machine, random forest, and k-nearest neighbors. This study highlights the effectiveness of a deep neural network–based multilayer perceptron in enabling early and reliable identification of Parkinson's disease, indicating its value as a supportive instrument for clinical diagnosis.

Cite this article as: Arasavali N, Mogadala VK, Pothana LSN, Gottapu S. Deep neural network multilayer perceptron for Parkinson's disease diagnosis using vocal data. Sigma J Eng Nat Sci 2026;44(2):1481–1490.

***Corresponding author.**
*E-mail address: naliniarasavali@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Parkinson's disease (PD) is a persistent, contemporary neurodegenerative illness that considerably affects patient's standard of living. PD is described mainly by the worsening of dopamine yielding neurons in the substantia nigra region of the brain and shows with both motor and non-motor symptoms. Motor manifestations, involving tremors, bradykinesia, spasticity, balance disorder, are the key indicators of the illness. Non-motor signs, including dementia, affective disorder, insomnia, and dysautonomia are common and may even appear before motor symptoms [3]. The worldwide impact of impact of PD is significant, and it becomes significantly more common in older individuals. With the growing elderly population around the world, the prevalence of PD is predicted to raise, bolstering it as a major health crisis [4]. Timely identification of PD is critical for executing efficient methods that can sluggish the advancement of symptoms, enhance well-being, and reduce the health care expenses [23]. Due to lack of credible diagnostic tools, initial identification of PD remains difficult, as its symptoms are subtle and variable in nature. Conventionally, PD is identified through patient clinical review based on the existence of various motor symptoms. Neurologists commonly use the well-known frameworks, such as United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB) standard to identify the problem. But these Medical approaches have limitations. These methods are often inefficient to identify PD at early stages when signs are mild and non-specific. Though the conventional practice of parkinsosn's diagnosis is based on clinical assessment, which dependent on subjective decision. The conventional diagnosis results inconsistencies in diagnostic outcomes. Advances in neuroimaging and biomarker research studies have permitted earlier detection of PD. The degeneration of dopaminergic neurons in the brain can be more clearly identified through neuroimaging approaches such as DaTscan and PET, contributing to greater diagnostic precision. In addition, biomarkers including alpha-synuclein, DJ-1, and beta-amyloid are being studied for their relvance in identifying the disease. Nevertheless, these methods are usually expensive, and hard to implement, which limits their regular use in daily clinical activities. Under these restrictions, machine learning (ML) and deep learning (DL) techniques have appeared as potential approaches for automated detection of PD. These methods can analyze vast amounts of data and give precise predictions, often consistently outperforming conventional approaches. In the domain of PD, ML and DL have been used across diverse data types including patient health data, genomic data, and voice samples [30]. Among the above, speech examination has earned deep consideration because it is less intrusive and can detect early signs of vocal disorders in PD patients. Among the various approaches, speech-based analysis has attracted growing interest because it is non-invasive and because changes in voice often appear in the initial stages of Parkinson's disease. Difficulties related to voice production are frequent in affected individuals and may become noticeable even when the disorder is still in its early phase. These disorders mainly arise from motor impairments disturbing the larynx and associated muscles. Several vocal characteristics provides important intuitions and acts as clinical markers for PD [29]. Therefore, examining these vocal characteristics by means of ML and DL methods can help in the identification of the PD [1-2,5-6,8-11].

Support Vector Machine (SVM) is among the most used ML techniques for categorization. It works by finding best decision boundary that divides the given data into various categories. They are effective in multi-dimensional spaces and exhibit their strength to represent complex relationships with the help of kernel functions. In PD research, SVMs have been used to categorize people using several vocal characteristics, though their performance depends on the choice of kernel and hyperparameters, which is a problem for optimization [13]. k-Nearest Neighbours nearby (k-NN) is one more standard ML method used for categorization tasks. This direct, instance-based technique in which a sample is allocated to a category based on the principal class of its k adjacent elements. Its ease of use made it one of the prominent choices for various domains, including medical investigations [13]. In PD examination, k-NN has eased the categorization of patients using vocal parameters. Despite its simplicity, its performance depends on the choice of k and degrades as features increase. Random Forest (RF) is ensemble learning method that constructs many decision trees over training period and finalises their predictions by majority voting. It is popular for providing high precision, ability to process massive datasets, and improved generalization performance [7]. In PD research, RF has been applied to categorize patients using clinical and biological measures, with its attribute's relevance helping detection of key indicators. yet, RF models turn out to be computationally intensive, mainly when substantial numbers of trees are used. Regardless of the advantages of these classifiers, Deep Neural Networks (DNNs), especially Multilayer Perceptrons (MLPs), have shown exceptional capability in clinical evaluation with their ability to represent complicated relationships throughout the data. DNN-MLPs contain multiple neuron layers, comprising an input layer, one or more hidden layers, and an output layer. In this framework, each neuron in each layer is interconnected to neurons in the succeeding layer, with appropriate weights that are incrementally improved during the training [12]. In this work, a DNN-MLP architecture is proposed for the classification of PD using voice-based features. The proposed model is trained and assessed using the dataset from the UCI Machine Learning Repository [4], which encompasses 22 acoustic characteristics obtained from voice recordings. These characteristics are utilized to distinguish people with PD from normal people. Data preprocessing was carried out by using StandardScaler to balance all feature contributions and to retain only relevant features. Hyperparameter

tuning is essential to improve performance of the DNN-MLP models. In the proposed work, GridSearchCV is used to examine different DNN-MLP architectures. The proposed model performance is evaluated by various metrics such as accuracy, precision, recall, F1-score, and specificity. Further, for benchmarking, conventional ML models such as SVM, RF, and k-NN are also deployed and verified using the same dataset.

MATERIALS AND METHODS

The dataset utilized in this work is obtained from the Parkinsons Telemonitoring dataset available at UCI machine learning repository. The dataset consists of speech samples and related clinical information collected from both Parkinson's disease patients and healthy individuals. The dataset is obtained from sustained voice quality recordings, from which a set of sound descriptors were obtained to describe speech variations linked to motor impairment. Besides these vocal features, the database also contains demographic details, such as age and gender, as well as assessment metrics, such as motor scores from the Unified Parkinson's Disease Rating Scale (UPDRS). For categorization, an indicator variable named as status is declared; if its value is '1', it indicates the presence of PD and '0' corresponds to a healthy person, based on medical diagnosis and motor appraisement. Acoustic features are fundamental frequency, jitter, shimmer, and harmonics-to-noise ratio (HNR), which are sensitive indicators of vocal pathology associated with PD. The obtained data is preprocessed before taining the model. The preprocessing involves like data splitting, feature scaling, missing data handling and feature selection. In the process of data splitting the dataset is divided into training and testing subsets using an 80:20 split ratio. The data splitting method employs stratified sampling to preserve equal quantities of individuals with and without PD across all divisons, which help minimize selection bias. In the process of feature scaling, continuous features are standardized using the Standard Scaler. This process transforms the feature space to have a mean of zero and a standard deviation of one, ensuring that all features contribute equally to model training without undue influence from variables with larger scales. During data preprocessing, any absent values are addressed using suitable methods, such as replacing them with the median, chosen according to the data distribution and the pattern of missing entries. The features are selected based on their clinical relevance and statistical importance to reduce the dimensionality and enhance the clarity of the model. As a result, we train and validate the ML models with the transformed data.

Recent research has shown that using ML techniques can aid in the diagnosis of PD through speech signals. A Multilayer Perceptron (MLP) classifier attained the highest recognition rate of 91 %, outperforming Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN), both of which recorded an accuracy of 87.17 % (Rana et al. (2022)), This shows the potential of these neural network-based models for this application [32]. In a related study, tai et al. in 2021, evaluated a dataset of 1,400 voice recordings and 70 acoustic features and supported that SVM is a best classifier with an accuracy of 88 % [31]. These studies support the use of voice characteristics as reliable digital biomarkers for Parkinson's disease and indicate that further investigation of such methods could improve diagnostic accuracy and accessibility.

In the present work, four machine learning models are developed and assessed such are, a widely used supervised learning model i.e., SVM, a non-parametric classifier i.e. k-NN, an ensemble based technique i.e. RF and the proposed model, Deep Neural Network-Multilayer Perceptron (DNN-MLP), which comprises mutiple fully connected layers capable of learning highly intricate data representations. To decrease overfitting and increase its capability for early PD detection, the proposed mrthod uses adjustable hidden-layer settings, ReLU activations, and suitable regularization techniques.

Training and testing are performed after model construction for each classifier. The learning phase emphases on fitting each model to the training data subset, allowing it to learn input-to-output associations effectiviely. A hybrid approach of grid search optimization and five-fold cross- validation is then applied to fine-tune hyperparameters- such as regularization terms, learning rates, and batch sizes in the DNN-MLP, and kernel parameters in the SVM-with the goal of improving generalization performance. Performance is measured using widely accepted evaluation criteria, including accuracy, precision, recall, F1-score, and specificity, the latter reflecting the model's capability to correctly identify health individuals [24]. All the developed models are ranked using comparative assessment based on their evaluation outcomes in the classification.

ARCHITECTURE OF DNN-MLP

The architecture of DNN-MLP consists of three main layers. The first layer of the network is the input layer, which is used to receive the input data. İn this neural network, the each feature of the dataset is represented by neuron, in this work, the dataset consists of various biomedical measurements, and the number of neurons are corresponding to the number of features. The last layer is the output layer and which produces the network's final predictionsBetween the input and output stages lie one or more hidden layers made up of neurons that compute weighted combinations of the incoming signals and pass them through activation functions. The number of layers and the number of neurons within each layer are critical determinants of the network's performance and are selected based on the complexity of the underlying problem. In applications such as Parkinson's disease detection, a DNN-MLP may therefore include multiple hidden layers so that subtle and non-linear patterns in the data can be learned effectively. The complete structure of proposed method is shown in Figure 1.

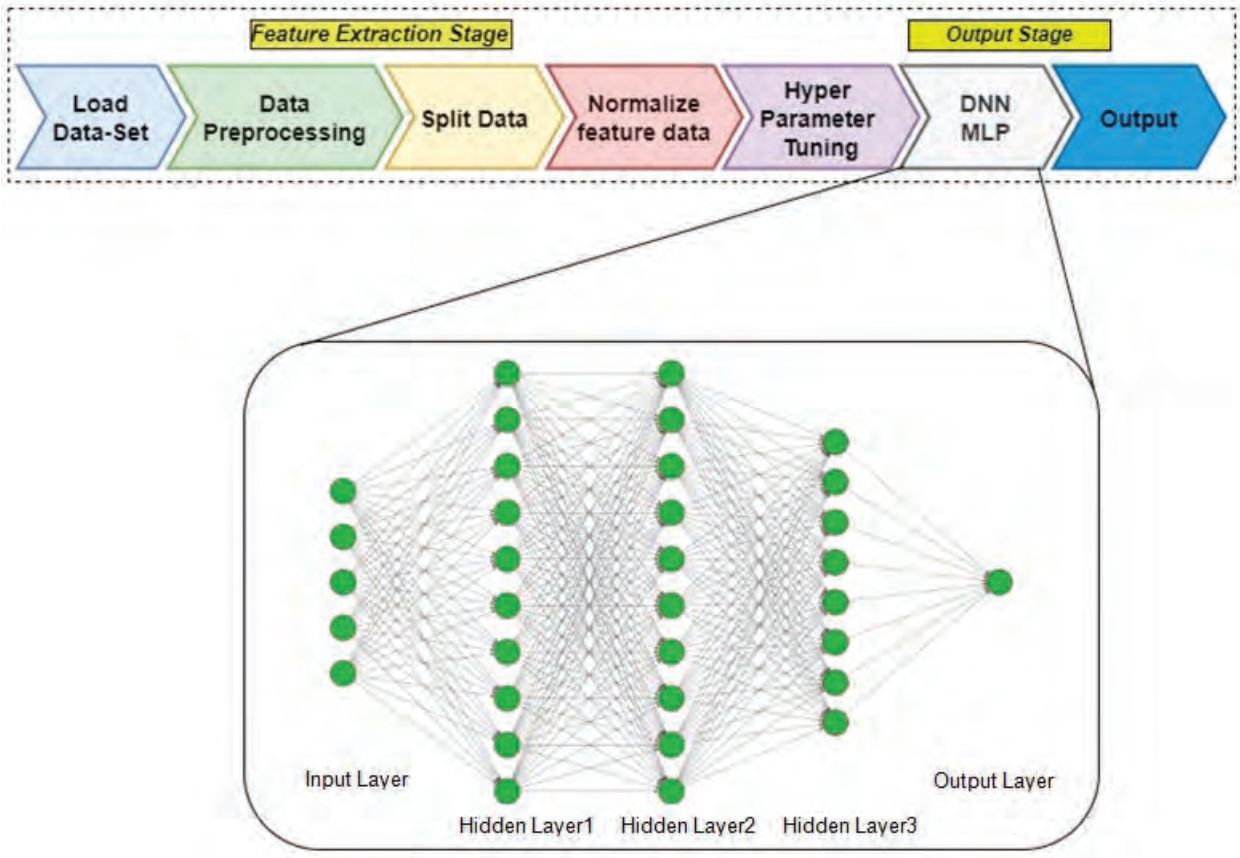


Figure 1. Block diagram of proposed model.

a) Activation Functions

Activation functions are required to introduce nonlinearity into the network to learn complex patterns in the data. The mainly used activation functions in this proposed model are, ReLU (Rectified Linear Unit) and Sigmoid.

$$f(x) = max\ (0, x) \tag{1}$$

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

RELU is used in hidden layers to mitigate the vanishing gradient problem and accelerate convergence [12,15]. Sigmoid is used in the output layer for binary classification to map the output to values 0 and 1, allowing them to be intepreted as class probabilities [16,20].

b) Weight Initialization

The initialization of weights plays a significant role in the learning process of a deep neural network, as improper intilization leads to ineffective convergence. There are various approaches for intializing the weights, the simple approach is zero initialization, where all weights assigned with zero but this is not widely used because it forces all neurons to behave identically during training. One extensively utilized approach is random initalization, where each weight is assigned a arbitrary initial value, helping different neurons to capture diverse feature representations during training. A distinct approach, known as He initalization, initializes the weights using a scaling factor relative to the number of neurons present in the preceding layer. This method is particularly suitable for networks employing ReLU activation functions, as it maintains signal varaince across layers [22].

c) Forward Propagation

During forward propagation, input data propagates through successive network layers untill a final prediction is generated. At each layer, every neuron receives inputs from the preceding layer, calculates a weighted of the inputs, and then employes its activation function prior to forwarding the output to the subsequent layer. This procedure is repeated through all layers untill the final layer generates the network's final prediction [12]. The calculation executed by a single neuron can be described by,

$$z_i = \sum_{j=1}^{n} w_{ij} x_j + b_i \tag{3}$$

$$a_i = f(z_i) \tag{4}$$

where z_i denotes the weighted sum, w_{ij} represents the connection weights, x_j refers to the input values, b_i is the bias, and $f(z_i)$ is the applied activation function.

d) Loss Function

To assess how well the model performs, a loss function calculates the difference between the predicted outputs and the actual output values, control parameter updates all through the training process [15]. In the designed framework, binary cross-entropy serves as the loss function, defined as

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \quad (5)$$

Where, y_i denotes the true class label, $\hat{y}_i$ represents the predicted probability, and N is the total number of training samples.

e) Backpropagation and Optimization

Backpropagation is the core learning mechanism used for training neural networks. It consists of two main phases, a backward pass followed by weight updates. In the backward pass, the gradients of the loss function with respect to each network parameter are computed using the chain rule of calculus. This step calculates how much the loss function will change with respect to a small change in the weights. As well as weight update adjust the weights in the direction that reduces the loss using an optimization algorithm such as Stochastic Gradient Descent (SGD) or Adam [20].

f) Regularization

The overfitting is controlled by various regularization techniques including L2 regularization and dropout. By adding a squared weight penalty in the loss function, L2 regularization controls model parameters that can lead to overfitting. This also improves the model's capability to perform well on new data. In contrast, dropout randomly turns off a proportion of neurons during a training iteration to curb robust dependencies among neurons and encourages the model to learn bettter features. [21–22].

g) Hyperparameter Tuning

The choice of hyperparameters strongly influences the efficiency of DNN-MLP. The number of hidden layers and neurons, the learning rate, and batch sizes, as well as regularization settings. To find suitable values, we typically employ an optimization method (like grid search or random search) that tests different parameter combinations in order to select the one that provides the highest performance [16–17].

h) Model Evaluation

The execution of the model is examined using different performance indicaters including accuracy, precision, recall, F1 score, and specificity, once training is over. The performance relies on proper hyperparameter selection and regularization techniques. These enable the model to achieve desirable performance without overfitting. [14,33].

RESULTS AND DISCUSSION

The dataset used in this study is the Parkinson's Disease dataset from the UCI Machine Learning Repository. The dataset consists of 195 samples and 23 attributes. The main features of the dataset are fundamental frequency (MDVP: Fo(Hz), MDVP: Fhi(Hz), MDVP: Flo(Hz)), several measures of variation in fundamental frequency (MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP), several measures of variation in amplitude (MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA), two measures of ratio of noise to tonal components in the voice (NHR, HNR), and six nonlinear dynamical complexity measures (RPDE, DFA, spread1, spread2, D2, PPE). Within the dataset, output classes are binary, with '1' assigned to validated positive instances to have PD and '0' representing healthy controls. The model performance is evaluated using grid search optimization and five-fold cross- validation could yield better results than using either one separately. By employing this validation procedure, it helps mitigate overfitting and provides a reliable assessment of the model's capacity to generalize to unseen data. The dataset is split into training and testing subsets using 80: 20 split, where 80 % is used for training and other 20 % is a held-out test for unbiased performance assessment. The division of data into 80 % for training and 20 % for testing is a practical choice since other ratios may hinder generalization or introduce greater variability. For comparative analysis, four classification models are considered: SVM, k-NN, RF, and DNN-MLP. The DNN-MLP model is used the hyperparameter tuning using GridSearchCV to find the optimal configuration. The hyperparameters are tuned to include hidden layer

Table 1. Performance metrics for SVM, RF, k-NN and DNN-MLP

Model	Accuracy	Precision	Recall	F1 Score	Specificity
DNN-MLP	0.949	0.973	0.973	0.973	0.500
SVM	0.872	0.971	0.892	0.930	0.500
Random Forest	0.923	0.972	0.946	0.959	0.500
k-NN	0.923	0.972	0.946	0.959	0.500

sizes, activation functions, solvers, alpha values, learning rates, and early stopping criteria. The effectiveness of the proposed models is assessed using a set of standard performance measures. The accuracy is the proportion of correctly classified samples as compared to the entire samples. Precision indicates the proportion of instances identified as positive that truly belong to the positive class. Recall or sensitivity reflects the extent to which actual positive cases are identified by the model. Instead of treating precision and recall as separate indicators, the F1-score takes both into account and calculates their harmonic mean. Specificity assesses the capability of the model to correctly predict negative class which give rise to true negatives. Moreover, it's given as the ratio of true negatives out of all actual negative instances. Besides these individual metrics, confusion matrices are constructed which provide an overall view of the class predictions where both the correct and incorrect predictions are easy to understand. The test results of the various models are summarized in Table1. The proposed DNN-MLP provides a performance score of 95% accuracy, 100% recall, and a 97% F1-score in PD detection. These results are better than those reported by several approaches. For instance, Rana et al. (2022) reported that their MLP model achieved an accuracy of 91%, whereas both k-NN and SVM achieved 87.17% [32]. In a related study, Tai et al. (2021) observed that SVM produced an accuracy of 87.2%, confirming its effectiveness for this type of data [31]. Random Forest classifier is proved as the efficient classifier with an accuracy around 93%, whwreas in the medical diagnosis tasks, the data is non-linear and it consists of complex pattwerns. For such data patterns, the DNN-MLP is suggested as the best suitable technique.

For each classifier, performance measures such as accuracy, precision, recall, F1-score, and specificity were obtained from the confusion matrix generated using the test data. The confusion matrices of all the implemented models are as shown in Figure 2. The SVM classifier also demonstrated

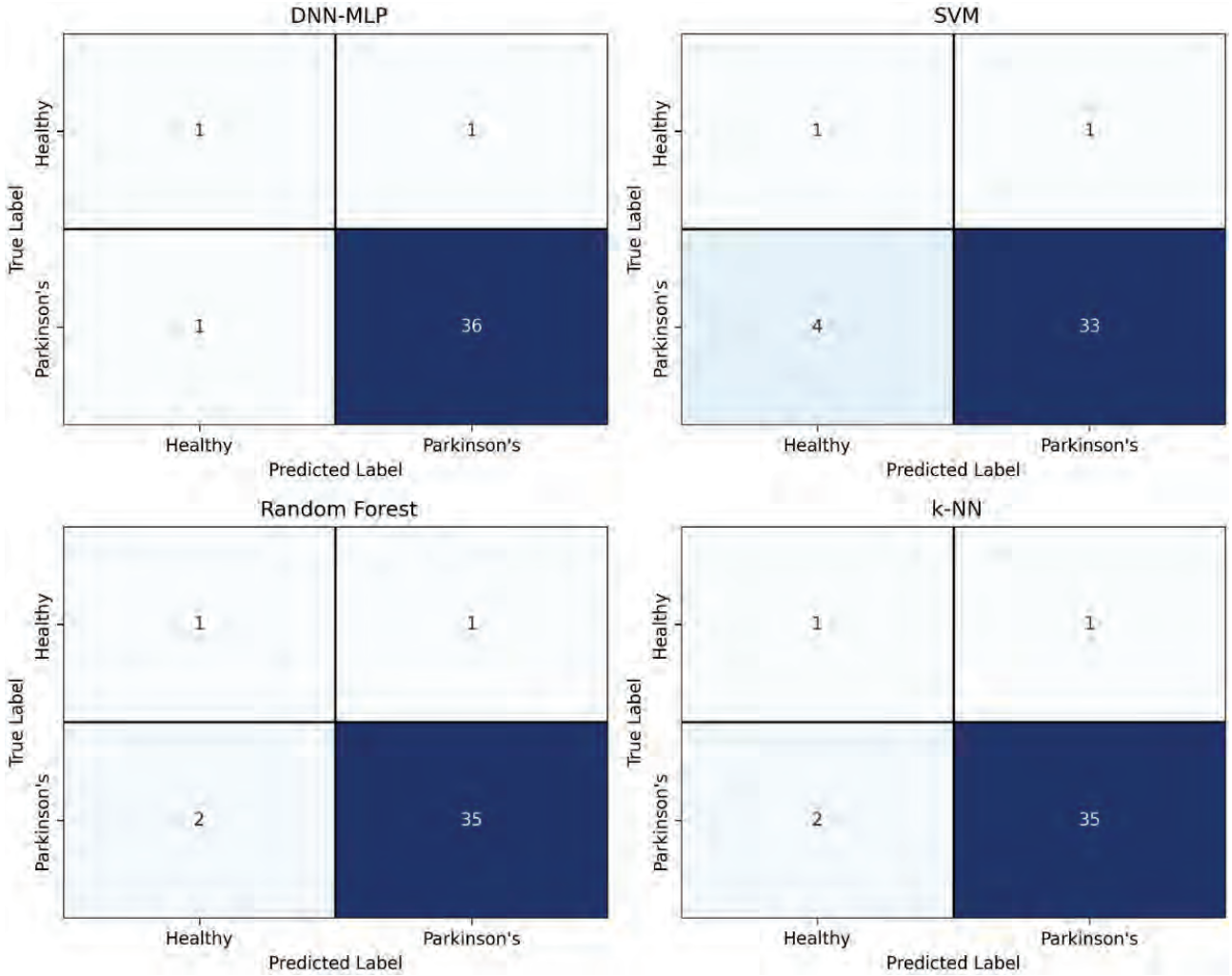


Figure 2. Confusion Matrices of DNN-MLP, SVM, RF, and k-NN.

strong performance, achieving high precision and recall, which reflects its effectiveness in distinguishing healthy individuals from patients with PD, although its overall results were marginally inferior to those of the DNN-MLP. The RF model attained a higher recall value; however its precision and specificity were relatively lower. This indicates that while it successfully indentified most true positive cases, it is also genrated a large number of false positives compared to the DNN-MLP and SVM classifiers.

Finally, the k-NN classifier showed performance equal to that of the RF model, but which can be attributed to its sensitivity to the selection of the k value and chosen distance metric. A comparison of the evaluation metrics for all models is shown in Figure 3 as a bar plot.

The Figure 3 shown above, depicts how the DNN-MLP odel is giving better results than SVM, Random Forest, and k-NN. Additionally, the distributions of Harmonics-to-Noise Ratio (HNR) and Noise-to-Harmonics Ratio (NHR), two critical voice features used in identifying Parkinson's disease in the dataset were analyzed and shown in Figure 4. HNR distribution indicates the ratio of harmonics to noise in the voice samples, which distinguishing between healthy and Parkinson's patients. NHR distribution shows the ratio of noise to harmonics,

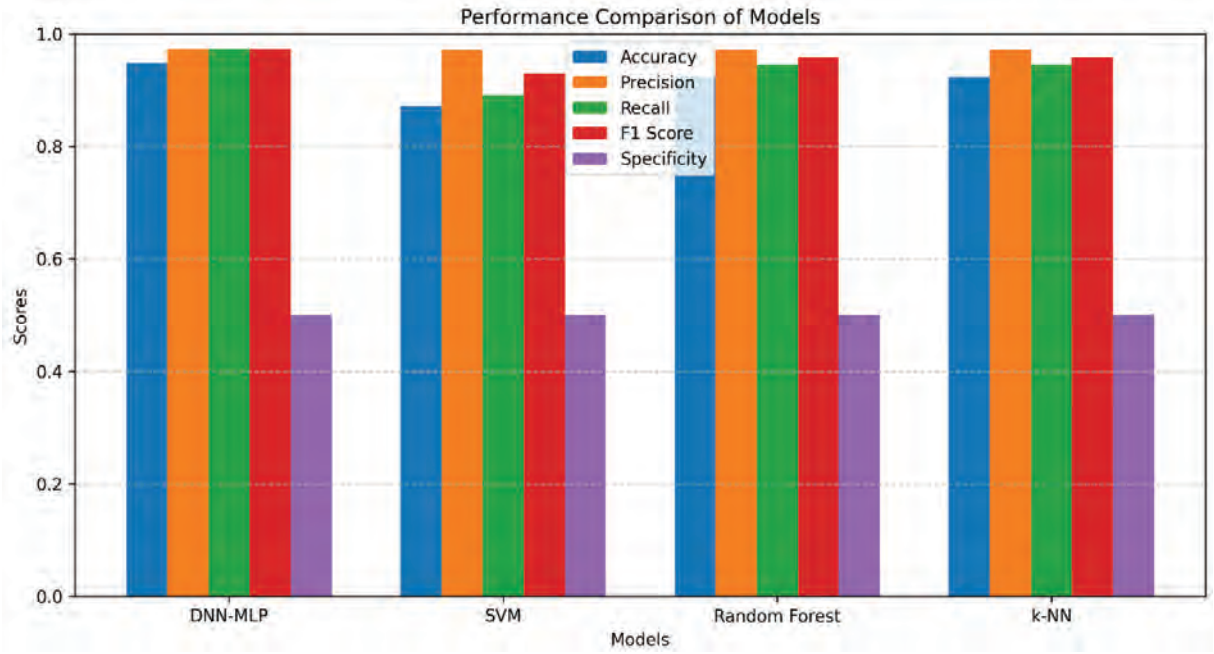


Figure 3. Metrics comparison of DNN-MLP, SVM, RF and k-NN.

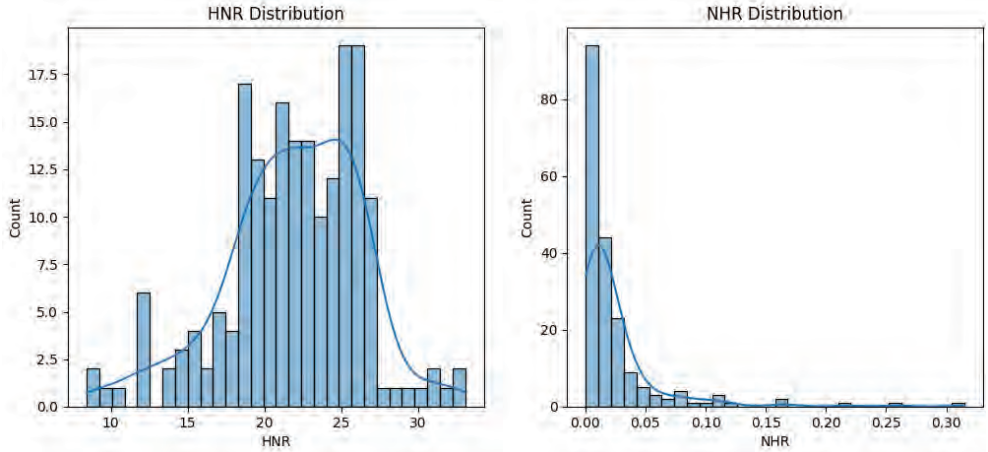


Figure 4. NHR and HNR distribution.

providing complementary information to HNR. The HNR distribution clearly demonstrates the healthy individuals with higher HNR values and with strong harmonic contents. The distribution consists of higher and lower HNR values. The higher values are representing the healthier vocal qualities and lower HNR values are representing the vocal qualities of Parkinson's disease.

The classification results proved that the DNN-MLP model is the best suitable model for the diagnosis of Parkinson's disease from voice measurements. The performance achieved by the proposed model is showing the ability of the deep architectures. The hyper parameter tuning is also be considered as very useful in the improvement of efficiency. The DNN-MLP outperforms the other models across all key performance metrics, suggesting that it is most effective approach for the dataset examined. Although the Random Forest and k-NN classifiers produced acceptable results, their performance remained below that of the DNN-MLP and SVM. This indicates that, while these methods have practical value, they may be less dependable for the early identification of Parkinson's disease when relying on speech-based features.

CONCLUSION

The proposed DNN-MLP model for Parkinson's disease diagnosis using biomedical voice data is evaluated against conventional machine learning methods, including support vector machine, k-nearest neighbors, and random forest. The dataset obtained from the UCI Machine Learning Repository comprised 195 samples described by 23 voice-related features. A detailed comparative analysis across all classifiers was performed using accuracy, precision, recall, F1-score, and specificity as evaluation benchmarks, with to determine each model's efficiency in distinguishing between PD-affected and healthy individuals. Across all experiment trials, the DNN-MLP appeared as the most consistent classifier, maintaining consistent performance in the vicinity of 0.949 for every metric considered. The depth inherent to its neural architecture plays a major role in this achievement, enabling the model to effectively detect and encode subtle, non-linear dependencies embedded in the vocal feature representations. The results shown that, the proposed architecture is proved as an outerperformed model than conventional machine learning approaches in the early diagnosis of parkinson's. Additionaly, such automated systems may help streamline diagnostic procedures and extend access to healthcare in resource-constrained settings where specialist expertise is limited. Future research may focus on integrating additional data modalities, including clinical evaluations, neuroimaging information, and genetic markers, to further improve the robustness and generalizability of the proposed approach.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Wang W, Lee J, Harrou F, Sun Y. Early detection of Parkinson's disease using deep learning and machine learning. IEEE Access 2020;8:147635–147646. [CrossRef]

[2] Pahuja G, Nagabhushan TN. A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE J Res 2021;67:4–14. [CrossRef]

[3] Nissar I, Rizvi DR, Masood S, Mir AN. Voice-based detection of Parkinson's disease through ensemble machine learning approach: A performance study. EAI Endorsed Trans Pervasive Health Technol 2019;5:e2. [CrossRef]

[4] Little MA, McSharry PE, Hunter EJ, Spielman J, Ramig LO. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans Biomed Eng 2009;56:1015. [CrossRef]

[5] Ouhmida A, Raihani A, Cherradi B, Lamalem Y. Parkinson's disease classification using machine learning algorithms: Performance analysis and comparison. In: 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET); 2022 Mar; p. 1–6. [CrossRef]

[6] Nissar I, Mir WA, Shaikh TA. Machine learning approaches for detection and diagnosis of Parkinson's disease—a review. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS); 2021 Mar; p. 898–905. [CrossRef]

[7] Celik G, Başaran E. Proposing a new approach based on convolutional neural networks and random forest for the diagnosis of Parkinson's disease from speech signals. Appl Acoust 2023;211:109476. [CrossRef]

[8] Zhang J, Zhou W, Yu H, Wang T, Wang X, Liu L, et al. Prediction of Parkinson's Disease using machine learning methods. Biomolecules 2023;13:1761. [CrossRef]

[9] Mittal V, Sharma RK. Machine learning approach for classification of Parkinson disease using acoustic features. J Reliab Intell Environ 2021;7:233–239. [CrossRef]

[10] Idrisoglu A, Dallora AL, Anderberg PS, Berglund JS. Applied machine learning techniques to diagnose voice-affecting conditions and disorders: Systematic literature review. J Med Internet Res 2023;25:e46105. [CrossRef]

[11] Prashanth G, Roy N, Mandal S, Ghosh S. High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. Int J Med Inform 2016;90:13–21. [CrossRef]

[12] Caliskan A, Badem H, Basturk A, Yuksel M. Diagnosis of Parkinson disease by using deep neural network classifier. IU-J Electr Electron Eng 2017;17:3311–3318.

[13] Pah ND, Indrawati V, Kumar DK. Voice-based SVM model reliability for identifying Parkinson's disease. IEEE Access 2023;11:144296–144305. [CrossRef]

[14] Parisi L, RaviChandran N, Manaog ML. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. Expert Syst Appl 2018;110:182–190. [CrossRef]

[15] Lamba R, Gulati T, Alharbi HF, Jain A. A hybrid system for Parkinson's disease diagnosis using machine learning techniques. Int J Speech Technol 2022;25;583–593.

[16] Xu Z, Zhu Z. Handwritten dynamics classification of Parkinson's disease through support vector machine and principal component analysis. J Phys Conf Ser 2021;1848:012098. [CrossRef]

[17] Aversano L, Bernardi ML, Cimitile M, Iammarino M, Montano D, Verdone C. A machine learning approach for early detection of Parkinson's disease using acoustic traces. In: 2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS); 2022 p. 1–8. [CrossRef]

[18] Khera P, Kumar N. Novel machine learning-based hybrid strategy for severity assessment of Parkinson's disorders. Med Biol Eng Comput 2022;60:811–828. [CrossRef]

[19] Saha N, Show AK, Das P, Nanda S. Performance comparison of different kernel tricks based on SVM approach for Parkinson's disease detection. In: 2021 2nd International Conference for Emerging Technology (INCET); 2021; p. 1–4. [CrossRef]

[20] Gore S, Dhole A, Kumbhar S, Jagtap J, Thakare A, Gudadhe S. Radiomics for Parkinson's disease classification using advanced texture-based biomarkers. MethodsX 2023;11:102359. [CrossRef]

[21] Goyal J, Khandnor P, Aseri TC. A comparative analysis of machine learning classifiers for dysphonia-based classification of Parkinson's disease. Int J Data Sci Anal 2021;11:69–83. [CrossRef]

[22] Khatamino P, Orman Z. A comparative study of machine learning algorithms in Parkinson's disease diagnosis: A review. Chowdhary CL, Reddy GT, Parameshachari BD, (editors). Computer Vision and Recognition Systems. New York: Taylor & Francis Group; 2022;83:69393–69441. [CrossRef]

[23] Keserwani PK, Das S, Sarkar N. A comparative study: Prediction of Parkinson's disease using machine learning, deep learning, and nature inspired algorithm. Multimed Tools Appl 2024;1–49. [CrossRef]

[24] Talitckii A, Kovalenko E, Shcherbak A, Anikina A, Bril E, Zimniakova O, et al. Comparative study of wearable sensors, video, and handwriting to detect Parkinson's disease. IEEE Trans Instrum Meas 2022;71:1–10. [CrossRef]

[25] Anand A, Haque MA, Alex JSR, Venkatesan N. Evaluation of machine learning and deep learning algorithms combined with dimensionality reduction techniques for classification of Parkinson's disease. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT); 2018 Dec; p. 342–347. [CrossRef]

[26] Ahmed MT, Mondal MNI, Gupta D, Ali MS. A review on Parkinson's disease detection methods: Traditional machine learning models vs. deep learning models. Eur J Inf Technol Comput Sci 2022;2:1–6. [CrossRef]

[27] Aich S, Kim HC, Hui KL, Al-Absi AA, Sain M. A supervised machine learning approach using different feature selection techniques on voice datasets for prediction of Parkinson's disease. In: 2019 21st International Conference on Advanced Communication Technology (ICACT); 2019 Feb; p. 1116–1121. [CrossRef]

[28] Gunduz H. Deep learning-based Parkinson's disease classification using vocal feature sets. IEEE Access 2019;7:115540–115551. [CrossRef]

[29] Hoq M, Uddin MN, Park SB. Vocal feature extraction-based artificial intelligent model for Parkinson's disease detection. Diagnostics 2021;11:1076. [CrossRef]

[30] Ashour AS, Nour MK, Polat K, Guo Y, Alsaggaf W, El-Attar A. A novel framework of two successive feature selection levels using weight-based procedure for voice-loss detection in Parkinson's disease. IEEE Access 2020;8:76193–76203. [CrossRef]

[31] Tai YC, Bryan PG, Loayza F, Peláez E. A voice analysis approach for recognizing Parkinson's disease patterns. IFAC-PapersOnLine 2021;54:382–387. [CrossRef]

[32] Rana A, Dumka A, Singh R, Rashid M, Ahmad N, Panda MK. An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. Electronics 2022;11:3782. [CrossRef]

[33] Das O. Prediction of the natural frequencies of various beams using regression machine learning models. Sigma J Eng Nat Sci 2023;41:1–15. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2047

Research Article

Dam-break analysis with infoworks ıcm model for earthquake-damaged Sultansuyu Dam: Different scenarios and computational costs

Burak ÇIRAĞ[1,*], Alp Ertuğrul ÖZER[2], Rıdvan KARAGÖZ[2], Bilal UTANCİK[3], Alaaddin Özgüray AYDIN[3], Mahmut FIRAT[4]

[1]*Department of Civil Engineering, Ataturk University, Erzurum, 25240, Türkiye*
[2]*Department of Civil Engineering, Istanbul, Istanbul Technical University, 34469, Türkiye*
[3]*Viasu Technology, Istanbul, 34774, Türkiye*
[4]*Department of Civil Engineering, Inonu University, Malatya, 44280, Türkiye*

ARTICLE INFO

Article history
Received: 04 July 2025
Revised: 14 October 2025
Accepted: 21 March 2026

Keywords:
Computational Costs; Dam-Break Analysis; Flood Hazard Maps; Flood Inundation Maps; Flood Velocity Maps; Infoworks ICM

ABSTRACT

Despite the many benefits of dams built for various purposes, their failure remains a serious risk due to the large volumes of water they store. The simulation of dam failure incidents is crucial for both emergency preparedness and risk mitigation strategies. This research conducts a dam-break analysis of the Sultansuyu Dam, which sustained significant damage during the earthquake that struck Türkiye on February 6, 2023, employing a hydrodynamic model formulated with the InfoWorks Integrated Catchment Modelling. Five distinct scenarios were crafted to assess both the computational cost and the dynamics of flooding behavior. There are four scenarios with complete reservoir condition (60 m) and velocity limits of 3, 5, 10, and 15 m/s, and the last one has a water depth of 50 m and a velocity limit of 15 m/s, indicating the depth at the time of the earthquake event. For each, flood inundation, velocity, and hazard maps were generated. Model validation with a historical rainfall event produced satisfactory performance (R^2 =0.772, NSE =0.467, RMSE =0.136 m). Results show that increasing the velocity limit causes the floodwater to move through narrower cross-sections, reducing overall inundation but increasing local flow depth. About 95% of flooded areas were classified as extreme hazard zones. Computational time increased by 1,219 seconds from the lowest to the highest velocity limit, indicating that lower limits require less computational effort. The study shows how velocity limits in hydrodynamic modeling affect flood hazard and computational efficiency, offering practical insights for dam safety and emergency planning.

Cite this article as: Çırağ B, Özer AE, Karagöz R, Utancık B, Aydın AÖ, Fırat M. Dam-break analysis with infoworks ıcm model for earthquake-damaged Sultansuyu Dam: Different scenarios and computational costs. Sigma J Eng Nat Sci 2026;44(2):1491–1508.

***Corresponding author.**
*E-mail address: burak.cirag@atauni.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

INTRODUCTION

Dams are man-made structures that hold large amounts of water and are built for a variety of reasons, such as generating hydroelectricity, controlling floods, and providing irrigation water. But when these dams break, a lot of water flows downstream in the form of floods, which can cause terrible things to happen [1]. Earthquakes, wear and tear on the dam itself, overtopping, seepage, and other factors can cause these failures. When this happens, the dam can no longer hold the water pressure [2,3]. Moreover, Sesli and Akköse [4] assessed the sliding stability of concrete gravity dams in Turkey employing the Multiple Wedge Analysis method and determined that earthquake acceleration is a paramount factor influencing dam safety. Dam break flows generate rapid increases in water levels and a flood wave. These flows are highly unsteady and can exert significant forces on man-made structures such as bridges and buildings in their path, as well as entrain large amounts of bed material [5,6]. In order to reduce and prevent the impact of these damages, the risk status of the regions can be determined through flood inundation maps [7,8].

The rainfall-runoff relationship that may occur due to dam failure can be represented by 1D, 2D, or 3D hydrodynamic models depending on the structure of the surface water body to be modeled. Al-Salahat et al. [9] created flood inundation maps with dam-break scenarios using the Hydrologic Engineering Center-River Analysis System (HEC-RAS) for the Wadi Al-Arab Dam in Jordan. As a result of the study, the maximum estimated water depth for overtopping was 37.6 m and 26 m for piping failure. Abdulrahman et al. [10] has performed an effort to analyze the risk situation of downstream regions of the dam by using HEC-RAS 2D to model the break of the Khassa Chai Dam at Kirkuk. The result reflects that flood water will inundate the city center within one hour and will affect most metropolitan areas of the city. Musa et al. [11] aimed to identify the areas that would be affected by flooding in the event of the failure of the Sembrong Dam in Johor, Malaysia. Using InfoWorks ICM and geographic information systems (GIS) software, they created flood inundation maps and reported that the flood covered a distance of more than 20 km from the vicinity of Ayer Hitam town to Parit Raja. In addition, the flood depth obtained for this area varies between 0.5 and 1.2 m. Azeez et al. [1] used HEC-RAS-2D software to model the collapse of the Um Al-Khair dam in Jeddah, Saudi Arabia, which collapsed during a severe storm event in January 2011. They reported that the performance metrics measuring the agreement between observed and simulated water depths were Root Mean Square Error (RMSE) of 0.45 m and correlation coefficient of 0.57. Using hydrodynamic modeling often requires high computational costs to perform these simulations, which require thousands or millions of calculation points (or mesh elements) [12]. Graphics processing units (GPUs) have been used in recent years for high-resolution modeling to address the heavy computational loads [13]. In the same way, Tayfur et al. [14] used the RANS equations and the PISO algorithm to numerically simulate dam-break flows and sediment transport over movable beds. This showed how important it is to use efficient computational methods when modeling complicated flow-sediment interactions.

This study aims to perform a dam-break analysis for the Sultansuyu Dam in Malatya, Türkiye, which developed cracks in the dam body during the earthquake that severely affected Türkiye on February 6, 2023. This analysis is characterized by 5 different scenarios. There is no study in the literature (as far as the authors know) that incorporates the effect of computational costs in the analysis with different scenarios by setting a velocity limit. These scenarios were created by setting velocity limits of 3 m/s, 5 m/s, 10 m/s, and 15 m/s when the dam is full (60 m water height). In addition, the 5th scenario includes a velocity limit of 15 m/s according to the dam's water height of 50 m on the day of the earthquake. Flood inundation maps, flood velocity maps, and flood hazard maps were created for these 5 scenarios. For the 4 scenarios with velocity limits, the calculation times of the models run with the same conditions were compared. In addition, the model was validated with a historical rainfall event and evaluated with statistical performance evaluation criteria. Previous studies on dam-break analysis have mostly focused on hydraulic behavior and flood propagation, while relatively few have examined the influence of numerical parameters such as velocity constraints on computational performance. Furthermore, research that integrates earthquake-induced dam damage, scenario-based hydrodynamic modeling, and computational cost evaluation is still scarce. This study enhances the literature by consolidating these elements into a unified modeling framework and illustrating the application of velocity-limited hydrodynamic simulations to assess both model stability and performance efficiency. Although modern hydrodynamic solvers employ techniques such as adaptive time-stepping to satisfy the Courant-Friedrichs-Lewy (CFL) condition for numerical stability, high-velocity gradients in dam-break flows often lead to excessively small time steps, drastically increasing computational cost. In engineering practice, imposing artificial velocity limits is a common strategy to mitigate this issue. However, the trade-off between the computational gain from such limits and the resulting loss in physical accuracy (e.g., underestimation of flood extent) remains under-investigated. Therefore, this study also serves as a numerical experiment, quantifying how different velocity constraints influence both the solver's efficiency and the resultant hazard classification in a catastrophic failure scenario. The results are anticipated to facilitate subsequent research on risk-oriented dam safety evaluation and the enhancement of hydrodynamic models in scenarios with limited data availability.

MATERIALS AND METHODS

Study Area and Dam Characterıstıcs

Sultansuyu dam is an earth core sand gravel dam and was constructed between 1986 and 1992 on the Sultansuyu stream in the Tigris-Euphrates Basin in Malatya province in eastern Türkiye. Built for irrigation purposes, the dam contributes to agricultural production by providing water to 11,665 ha of land. According to the General Directorate of State Hydraulic Works (DSI), the height of the dam crest above sea level is 906 m, the body height from the foundation is 60 m and the length is 721 m. The reservoir capacity of the dam is 47.9 hm3 at normal operation level with a body volume of 3.205 hm3 and a lake area of 2.5 km2. In addition, the project flow rate is 1,050 m3/s [15]. The location of the dam is shown in Figure 1.

An earthquake of a magnitude of 7.8 (Mw) struck the Pazarcık district of Kahramanmaraş province in southeast Turkey on February 6, 2023, at 04:17 local time. Then, in the Elbistan region of Kahramanmaraş province, another earthquake with a magnitude of 7.6 (Mw) struck nine hours after the previous one. Over 50,000 people lost their lives and over 100,000 were injured as a result of these earthquakes, which hit Türkiye and Syria hard. The Republic of Türkiye declared a state of emergency in 11 provinces. In Malatya, one of the 11 provinces, more than 1,000 people were dead and nearly 10,000 injured. 345,536 houses were

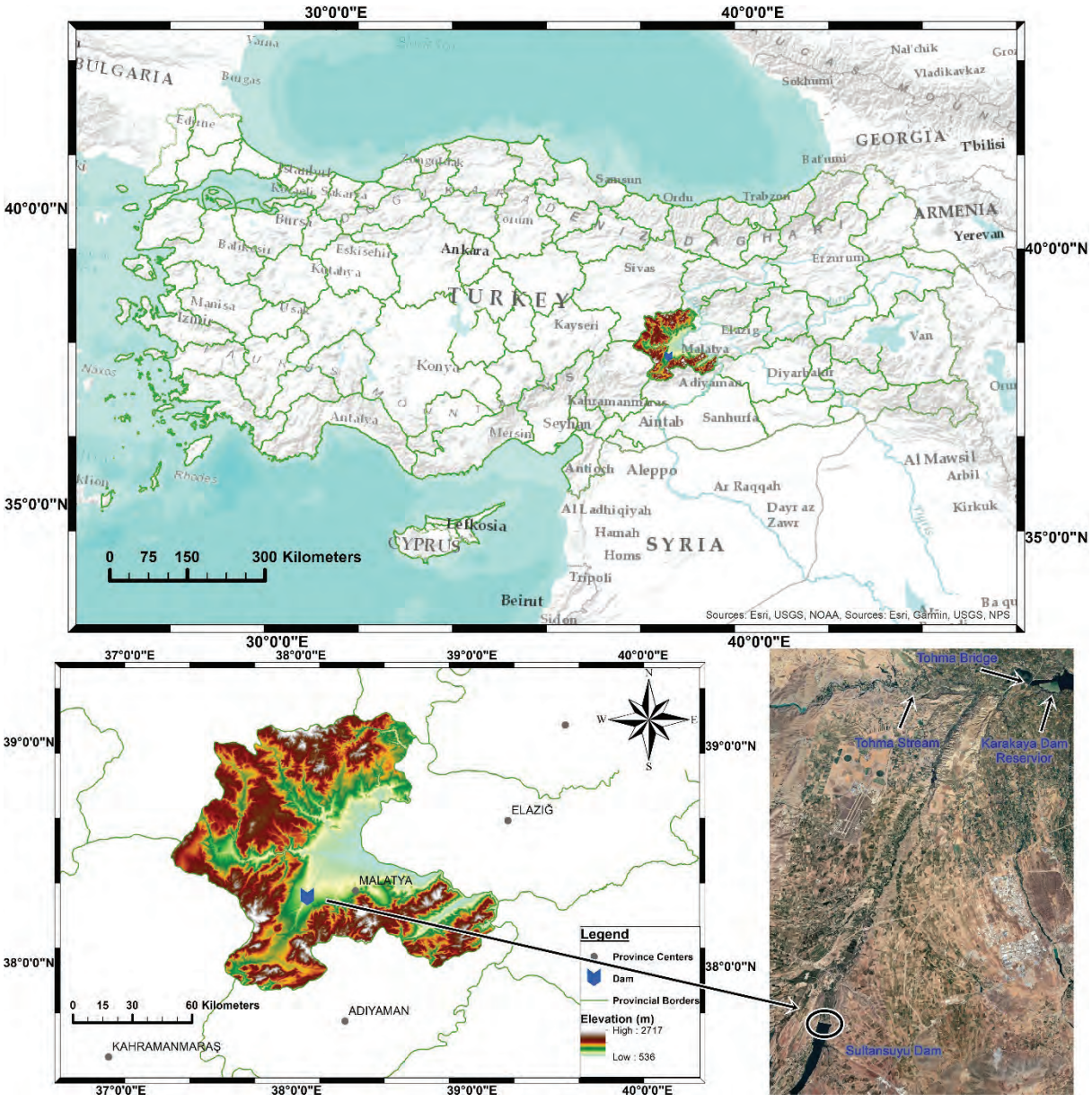


Figure 1. The location map of Sultansuyu Dam showing the reservoir area and downstream settlements.

damaged in Malatya and thousands of people were forced to migrate.

Due to these two earthquakes, 140 dams were subjected to moderate to high seismic intensity. Although all dams were able to maintain their water retention capability, moderate to large permanent deformations occurred in some of them. For this reason, water levels in Sultansuyu and Arıklıkaş dams were reduced in a controlled manner as part of emergency response measures [16].

Sultansuyu Dam, located in Akçadağ district of Malatya province, is 75 km from the epicenter of the Elbistan earthquake. During the earthquakes, the dam reservoir's water level was 50 meters, and the dam's crest and upstream slope developed large longitudinal cracks [17]. Transverse cracks were initially around 0.8-1.0 m, but after 10 days the depth of these cracks reached 4.0 m (Fig. 2) [18]. The Presidency of the Republic of Türkiye, Presidency of Strategy and Budget (SSB) published a report on the Kahramanmaraş earthquakes, stating that the Sultansuyu Dam was heavily damaged and the estimated damage cost was 705,000,000 Turkish Liras (Approximately 37 million dollars at the then exchange rate) [19]. In this context, Sultansuyu Dam was chosen as the study area because of the presence of settlements downstream of the dam and its critical importance for agricultural activities.

Methodology and Data Preparatıon

The methodology for assessing dam break analysis, as illustrated in Figure 3, begins with data collection and concludes with the analysis results. The first step is to obtain

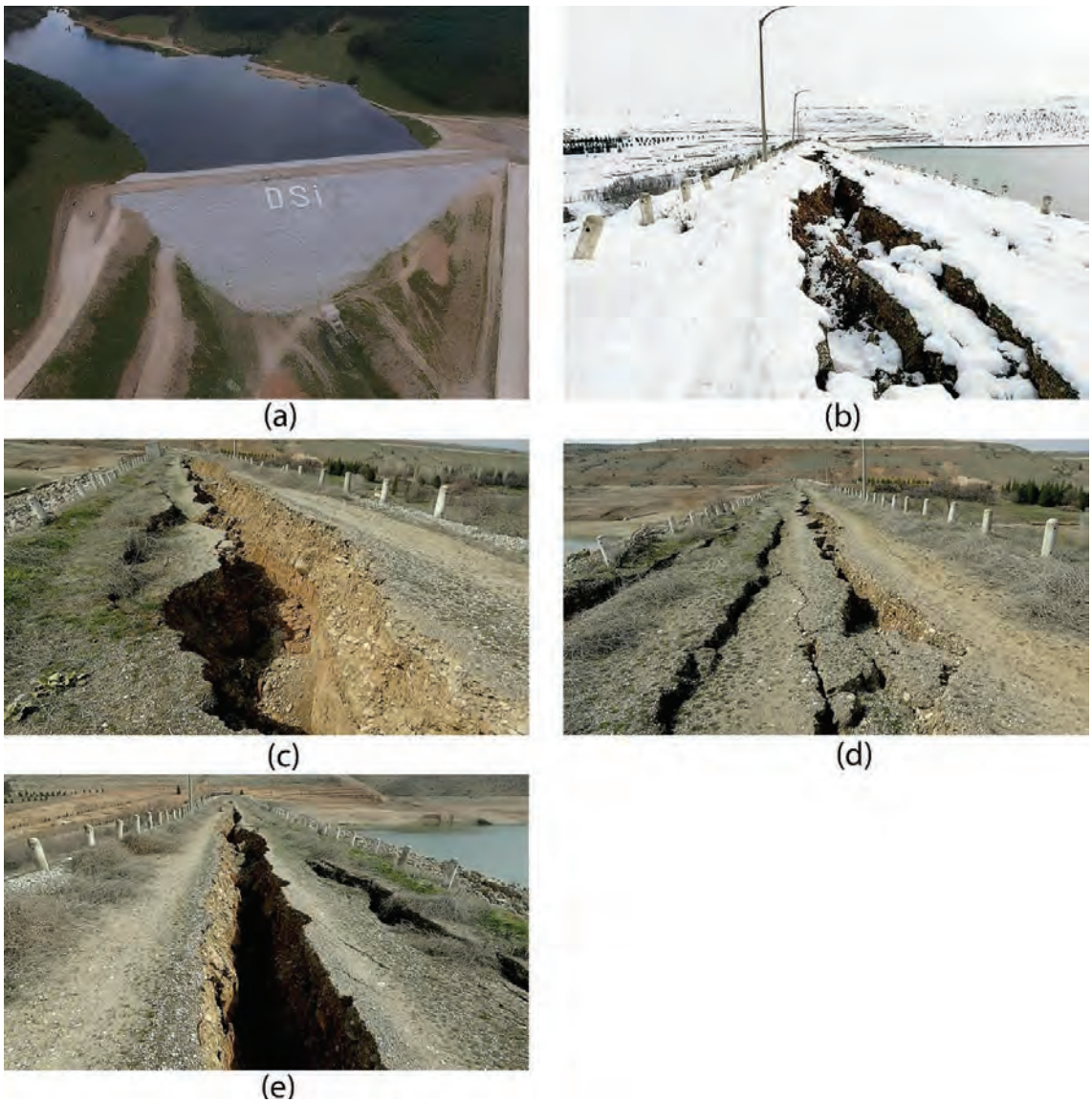


Figure 2. Views from Sultansuyu Dam: a) view of the dam before the earthquake (adopted from [20]); b)-e) view of damages on the dam body after the earthquake (adapted from [21,22]).

dam characteristics and isohypse maps from DSI. The isohypse maps were used to obtain digital elevation map (DEM) data from the topography of the empty dam reservoir. Meshing using the DEM data in InfoWorks ICM was performed using a minimum cell size of 25 m x 25 m near the dam and a minimum cell size of 100 m x 100 m in areas far from the dam. The spatial discretization of the domain was determined through a trade-off analysis between topographic precision and computational cost. The mesh size was varied from 25 m in the high-velocity zone immediately downstream of the dam to 100 m in the wider floodplain areas. This configuration was selected to match the resolution of the underlying digital elevation model (DEM), ensuring that the mesh captures the essential hydraulic features of the main channel without introducing artificial precision that exceeds the quality of the topographic data. Preliminary stability runs confirmed that this resolution is sufficient to resolve the flood wave propagation while maintaining computationally manageable runtimes. Coordination of Information on the Environment (CORINE) land cover data was obtained from the Republic of Türkiye Ministry of Agriculture and Forestry (TOB). These two data were arranged in ArcGIS environment for analysis in a way to be compatible with the study area and InfoWorks ICM program. The study area contains 7 different Manning's (n) values according to land use/cover classification (Table 1).

Then, the DEM data, dam reservoir, and 2D flow field were created in the hydrodynamic modeling program InfoWorks ICM. In this study, the dam break analysis was characterized according to different maximum flow velocities that would occur due to the dam break. These velocities are 3 m/s, 5 m/s, 10 m/s and 15 m/s. The maximum velocity in the 2D flow area is limited to these values. These velocity limits (3, 5, 10, and 15 m/s) are not randomly selected but are determined to represent the flow velocities observed in dam-break events in engineering. During an actual dam breach, the flow velocity can reach values in the range of 3-20 m/s within a short time, depending on factors such as breach geometry, water level difference, and downstream topography. However, the uncontrolled increase of these values within the hydrodynamic model leads to problems such as numerical instability and excessive computational cost. Therefore, different velocity limit scenarios were applied to achieve both behavior close to physical reality and numerical stability. Thus, the scenarios created reflect controlled conditions representing the velocity ranges that may be encountered in engineering. The model constrained by these velocities is represented with the assumption that the dam is full. In addition, the dam water level of 50 m on February 6, 2023, when the earthquake occurred, was added to the study as an extra scenario with a maximum velocity limit of 15 m/s. From the scenarios created according to these velocities, flood inundation maps, flood velocity maps, and flood hazard maps were provided. These maps can be viewed in InfoWorks ICM or any geographic information system (GIS) software.

Infoworks ICM Software And Dam Break Analyses

Floods caused by dam collapse can be predicted and depth and velocity estimates in the downstream valley can be obtained using mathematical hydrodynamic simulation models [9]. The present research used the InfoWorks ICM model with one-dimensional and two-dimensional modeling capabilities. For the dam-break analysis, 2D modeling based on meshing was used, which reflects the terrain topology and geometry and allows for more precise flow routing. InfoWorks ICM program includes the ability to solve the high computational cost of two-dimensional hydrodynamic modeling using GPU [23,24]. In the simulation setup, an instantaneous failure scenario was adopted as the breach mechanism. This approach assumes the immediate release of the entire reservoir water volume, representing a worst-case condition to analyze the downstream flood propagation dynamics.

The shallow water equation (SWE), a depth-averaged variant of the Navier-Stokes equations, serves as the foundation for the 2D engine in InfoWorks ICM. It is used to depict 2D flow mathematically [25,26]. The SWE presumes that the movement is primarily horizontal and that changes

Table 1. CORINE land use/cover distribution of the study area

Label	Manning's (n)	Area (km²)	Percent (%)
Airports, industrial or commercial units, etc.	0.013	9.524	2.28
Sparsely vegetated areas	0.027	9.816	2.35
Non-irrigated arable land, permanently irrigated land	0.030	109.563	26.19
Pastures	0.035	2.584	0.62
Complex cultivation patterns, etc.	0.040	173.556	41.49
Land principally occupied by agriculture, with significant areas of natural vegetation, etc.	0.050	22.751	5.44
Fruit trees and berry plantations, vineyards	0.080	90.492	21.63
Total		**418.286**	**100**

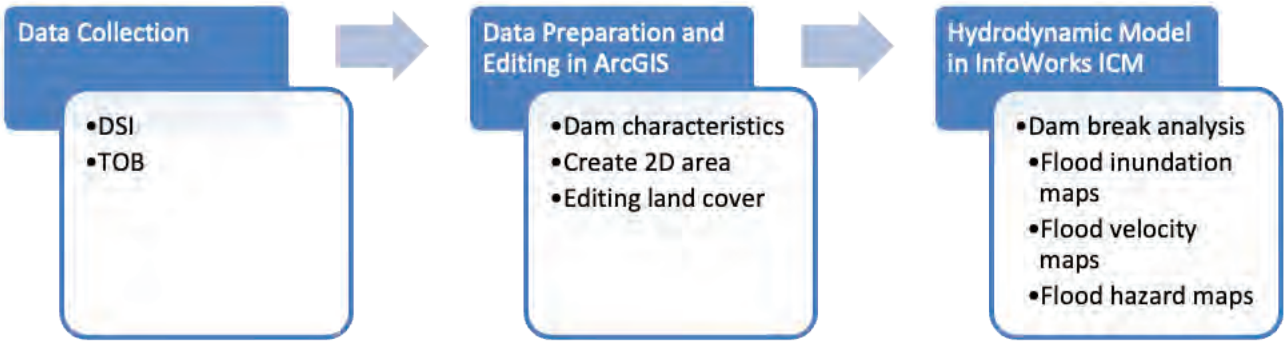


Figure 3. Methodology of the study.

in velocity along the vertical axis can be disregarded. InfoWorks ICM employs a conservative approach to solve the SWE, as described below [27]:

$$\frac{\partial h}{\partial t}+\frac{\partial (hu)}{\partial x}+\frac{\partial (hv)}{\partial y}=\sum_{i=1}^{n} q_i \tag{1}$$

$$\frac{\partial (hu)}{\partial t}+\frac{\partial}{\partial x}\left(hu^2+\frac{gh^2}{2}\right)+\frac{\partial (huv)}{\partial y}-\frac{\partial}{\partial x}\left(\varepsilon h\frac{\partial u}{\partial x}\right)-\frac{\partial}{\partial y}\left(\varepsilon h\frac{\partial u}{\partial y}\right)=gh\left(S_{0,x}-S_{f,x}\right)+\sum_{i=1}^{n} q_i u_i \tag{2}$$

$$\frac{\partial (hv)}{\partial t}+\frac{\partial}{\partial y}\left(hv^2+\frac{gh^2}{2}\right)+\frac{\partial (huv)}{\partial x}-\frac{\partial}{\partial x}\left(\varepsilon h\frac{\partial v}{\partial x}\right)-\frac{\partial}{\partial y}\left(\varepsilon h\frac{\partial v}{\partial y}\right)=gh\left(S_{0,y}-S_{f,y}\right)+\sum_{i=1}^{n} q_i v_i \tag{3}$$

The following terms represent various variables related to fluid dynamics or hydrodynamic modeling: h is the water depth (m); u and v are the velocities in the x and y directions (m/s), respectively; q_i is the net discharge per unit area from the ith source (m/s); u_i and v_i are the velocity components in the x and y directions of the ith discharge (m/s); g is the gravitational acceleration (m/s²); ε denotes eddy viscosity (m²/s); $S_{0,x}$ and $S_{0,y}$ represent the bed slopes in the x and y directions (m/m); $S_{f,x}$ and $S_{f,y}$ refer to the friction slopes in the x and y directions (m/m); and n indicates the total number of discharge sources. These variables are typically used to describe the movement of water, accounting for various forces such as gravity, friction, and external discharges, in models of fluid flow.

To ensure numerical stability and reproducibility, the simulation parameters were configured as follows: The simulation was run with a maximum computational time step of 5 seconds, subject to a timestep stability control factor of 0.95 to dynamically adjust for high-velocity flows. The semi-implicit weighting factor (Theta) was set to 0.9 to dampen numerical oscillations, and the inundation mapping depth threshold (wetting/drying limit) was defined as 0.01 m. For the downstream boundary condition, a fixed water level representing the reservoir level of the Karakaya Dam was applied to simulate the hydraulic backwater effect at the confluence point. Spatially distributed Manning's roughness coefficients were applied based on the CORINE land cover classification (as detailed in Table 1).

The maximum flood hazard rating, which depends on depth and velocity, is used by InfoWorks ICM to compute the hazard values. This value is calculated with the equation proposed by the Defra/Environment Agency of the United Kingdom [28,29].

$$HR=d(v+1.5)+DF \tag{4}$$

where HR is the (flood) hazard rating, d is the depth of flooding (m), v is the velocity of floodwaters (m/sec) and DF is the debris factor. The classification of the ratios according to the degree of hazard is given in Table 2.

Table 2. Classification of hazard values (adapted from [28])

Rate	Degree of Flood Hazard	Description
<0.75	Low	Caution
0.75-1.25	Moderate	Dangerous for some (i.e. children)
1.25-2.5	Significant	Dangerous for most people
>2.5	Extreme	Dangerous for all

In the calculation of the Hazard Rating, the Debris Factor (DF) was assigned dynamically based on the standard DEFRA methodology implemented in InfoWorks ICM. Accordingly, a value of 0.5 was applied for flood depths less than 0.25 m, and a conservative value of 1.0 was applied for depths exceeding 0.25 m.

Model Performance Evaluatİon

A suggested model's intricate parts and few observations make it challenging to validate and calibrate. 2D analyses such as flood and landslide simulations can be validated using available observational data [30]. In this study, three statistical indicators, the coefficient of determination (R^2), Nash-Sutcliffe efficiency (NSE) [31], and Root-mean-square error (RMSE), were used to determine the agreement between observed flood depth and modeled flood depth values. R^2 ranges from 0 to 1 and indicates how well the model explains the variance of the dependent variable, with higher values representing better model fit and satisfactory performance when exceeding 0.5 [32-34]. NSE is a measure used to validate models involving hydrological phenomena [35]. NSE values between 1.0 and 0.8 are considered excellent, between 0.8 and 0.6 are regarded as good, and between 0.6 and 0.05 as satisfactory. Additionally, NSE values ranging from 1.0 to 0 are deemed acceptable [26]. RMSE can be any number between 0 and ∞, and lower values mean the model is doing better. It tells you how big the errors are on average between the observed and simulated values. An RMSE of 0 means that the model is perfect and has no errors [36]. The equations for the statistical evaluation criteria are given below.

$$R^2 = \frac{\sum_{i=1}^{n}(x_i - x_{avg})^2 - \sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - x_{avg})^2} \quad (5)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - x_{avg})^2}\right] \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \quad (7)$$

where xi represents the observed values, yi denotes the output values produced by the model, xi-yi indicates the error values, xavg is the average of the observed values, n is the number of data points.

RESULTS AND DISCUSSION

Statistical Evaluation of Model Results

About a month and a half after the earthquake that seriously affected the study area and Sultansuyu Dam, a heavy rainfall event occurred in the region where Sultansuyu Dam is located. The rainfall started on 13/03/2023 at 12:00, stopped in places, and continued until 16/03/2023 at 05:00, flooding areas such as living areas and orchards near the dam. The excessive rains caused certain roadways above and downstream of the dam to be restricted to both human and vehicular traffic. On one of these highways, some measured values were obtained as a result of reports by local people and news channels. When these measured values were compared with the values measured in the model, the statistical evaluation criteria R^2=0.772, NSE=0.467, and RMSE= 0.136 m were obtained. These results show that the model used is satisfactory. In hydrological modeling of ungauged or data-scarce regions, a positive NSE value indicates that the model provides better predictive skill than the mean observed value [37]. Considering the high uncertainty in observation data derived from limited field reports, these results are deemed satisfactory for emergency flood propagation analysis. Azeez et al. [1] utilized analogous validation methodologies in their simulation of the Um Al-Khair dam breach in Jeddah, Saudi Arabia, where the correlation coefficient (R^2 = 0.57) and RMSE (0.45 m) between observed and modeled water depths were deemed satisfactory under constrained field data circumstances, thereby affirming the credibility of such hydrodynamic modeling techniques in data-deficient areas. However, it is important to explicitly acknowledge that the validation process was constrained to a single rainfall-runoff event due to the scarcity of available hydrometric data in the

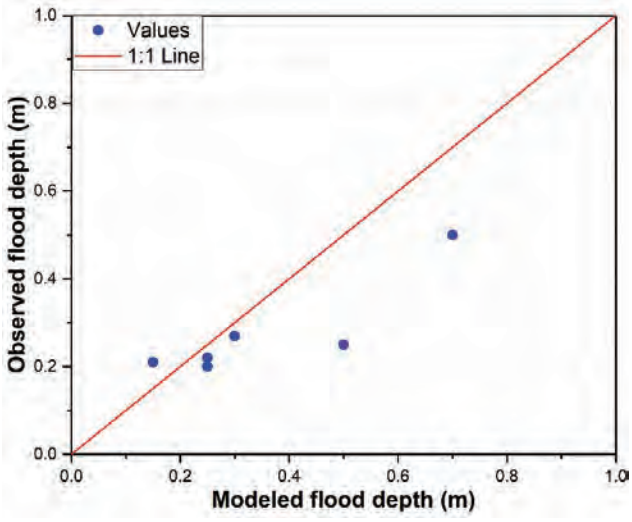


Figure 4. Scatter plot for comparison between the observed and modeled flood depths at some locations.

post-disaster period. The Sultansuyu Dam is in a rural area and doesn't have a lot of hydrometric observation stations, thus the model calibration process couldn't be done. Even though this situation is a problem because of a lack of data, the fact that the validation phase employed only one rainfall event shows that the model can give credible findings in similar hydrological situations. The model validation procedure, which uses a scenario-based analysis comparing technique, lessens the effect of calibration problems on the results' reliability. So, using the same model parameters in all circumstances keeps the results easy to understand. A scatter plot was made to show how the observed values and the modeled values for the historical rainfall event were related. This was done to see how well the model worked (Fig. 4).

Figure 4 shows the observed and modeled flood depths and a 45-degree line of agreement showing the line of perfect fit. All but one of the points are located below the line, indicating that the measured flood depths in the model are lower than the observed flood depths. According to the results of the statistical evaluation criteria and scatter plots, the model was found to have a good correlation in prediction.

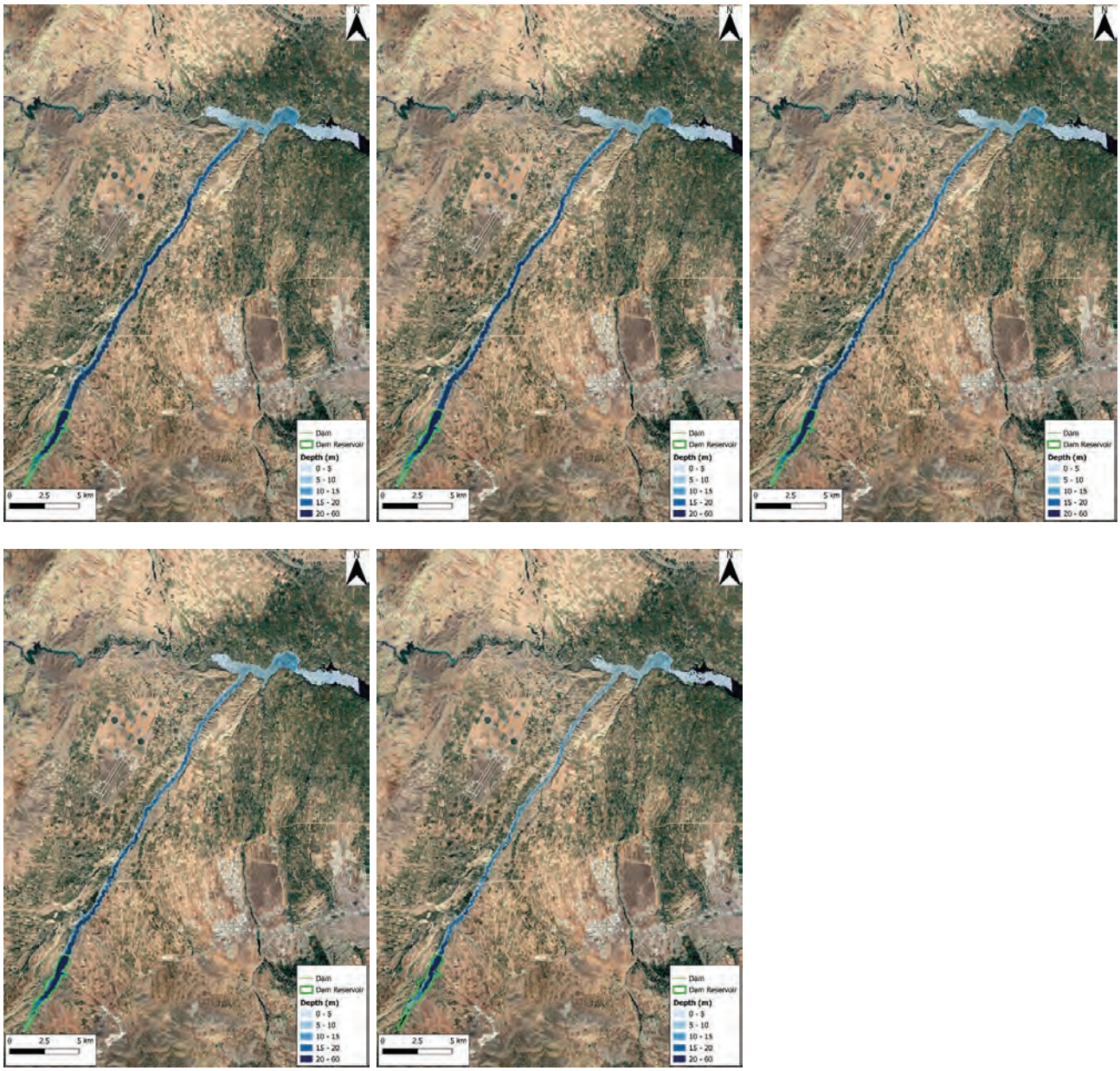


Figure 5. Flood inundation maps created according to maximum velocity limits and dam water height as a result of dam collapse: **a)** 3 m/s, **b)** 5 m/s, **c)** 10 m/s and **d)** 15 m/s maximum velocity limit for 60 m dam levels, **e)** 15 m/s maximum velocity limit for 50 m dam levels.

Finally, apart from the validation constraints, this study entails inherent uncertainties associated with the input parameters and modeling assumptions. First, the observation data used for validation were derived from field reports and visual evidence rather than gauge measurements, introducing a degree of uncertainty in the exact water depths. Second, the dam breach mechanism was simulated as an instantaneous failure to represent the worst-case scenario; however, a gradual erosion process might yield different peak timing and attenuation characteristics. Third, the roughness coefficients were assigned based on standard CORINE land cover classes without spatially distributed calibration. Despite these uncertainties, the conservative approach adopted herein ensures that the resulting hazard maps provide a safety margin suitable for emergency planning.

Flood Inundation Maps

In dam break analyses, the potential extent of damage resulting from a dam failure is closely linked to the topographical features of the surrounding land. For this reason, in order to accurately estimate the damage that may occur, the models are made with DEM maps that represent the topography of the land. With InfoWorks ICM, flood inundation maps can be created as a result of meshing using DEM maps [38]. In flood inundation maps, the probable magnitude and depth of floods due to dam break analysis are evaluated, giving meaningful information regarding the extent and the flood severity. Similarly, the maps assist to determine and examine probable flood consequences through the identification of the areas that are subjected to inundation. In the implementation of the strategies to assist minimize flood risk and increase the resilience of the populace to the occurrence of the natural hazard, the emergency managers, policymakers, and the urban planners can all gain through the delineation of the areas of inundation [9]. In this context, flood inundation mapping forms the first and most critical step of spatial impact assessment in dam-break studies. It enables the identification of inundation boundaries, spatial variation in flood depth, and the relationship between topography and flow concentration. Figure 5 shows the flood inundation maps generated as a result of the dam break analysis for maximum allowable velocities of 3 m/s, 5 m/s, 10 m/s, and 15 m/s assuming that the dam is full and 15 m/s when the water level of the dam is 50 m.

The upper bound is situated at the reservoir's dam position, and the flow process of the dam break simulation is used as the input condition for the flood simulation [39]. In the present study, villages downstream of the dam were found to be inundated under all the conditions considered.

For the scenario with a velocity limit of 3 m/s, 21.09 km2, for 5 m/s, 20.51 km^2, for 10 m/s, 19.29 km^2, for 15 m/s, 18.87 km^2 and for the velocity limit of 15 m/s at 50 m water height, 15.18 km^2 were flooded. In this context, in a 2D analysis, it can be said that the inundated area decreases as the velocity limit set for each grid cell increases. Similar patterns were observed by Khosravi et al. [40] in the Sefid-Roud Dam case study in Iran, where 2D HEC-RAS simulations showed that flood depth and extent decreased rapidly with distance from the dam and that the valley topography strongly controlled the concentration of floodwaters along the main flow path. In addition, the lowering of the dam water height has also reduced the inundation area. In

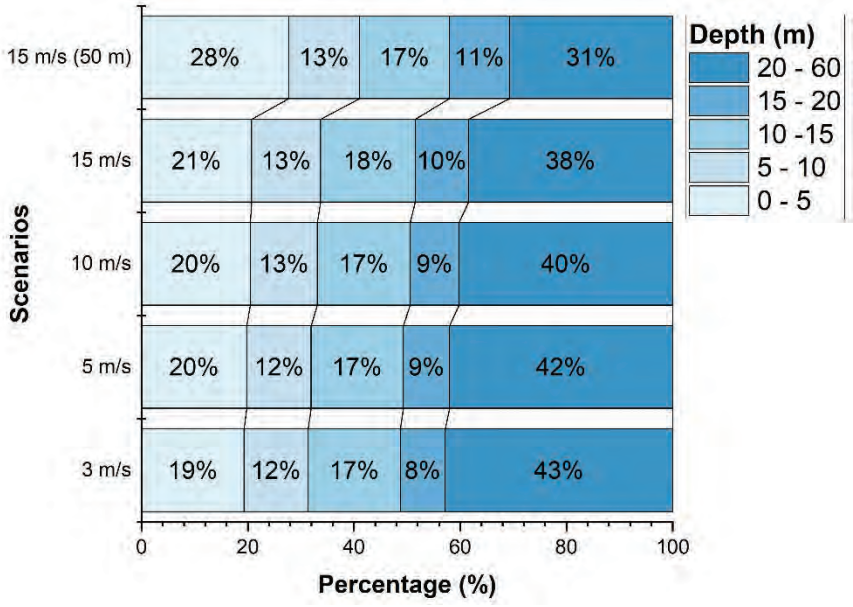


Figure 6. Proportional (%) distribution of areas covered by maximum water depths observed in the flood zone downstream of Sultansuyu Dam.

addition, due to the flood water from the dam break of the Sultansuyu Dam reaching the area where the Tohma stream and Karakaya Dam reservoir are located, settlements and agricultural lands located at the junction point are flooded. Accordingly, a possible dam-break event poses a great danger to urban man-made areas and the natural environment. Loss of life is likely to occur in the region, which includes mostly agricultural areas, for each scenario. Percentage values of inundation area according to the depths of the scenarios are given in Figure 6.

As the maximum velocity limit decreased in the scenarios, the percentage of high-water depths in the inundation areas increased (Fig. 6). In other words, the increase in maximum velocity limits led to a decrease in depths. In light of these results, the study emphasizes the importance of giving the correct velocity limit and adjusting the input conditions more accurately in modeling applications.

When it comes to apricot production, Turkey leads the world. Malatya accounts for over 85% of Turkey's dried apricot output and 55% of its fresh apricot production. Exports account for a significant portion of the dried apricot

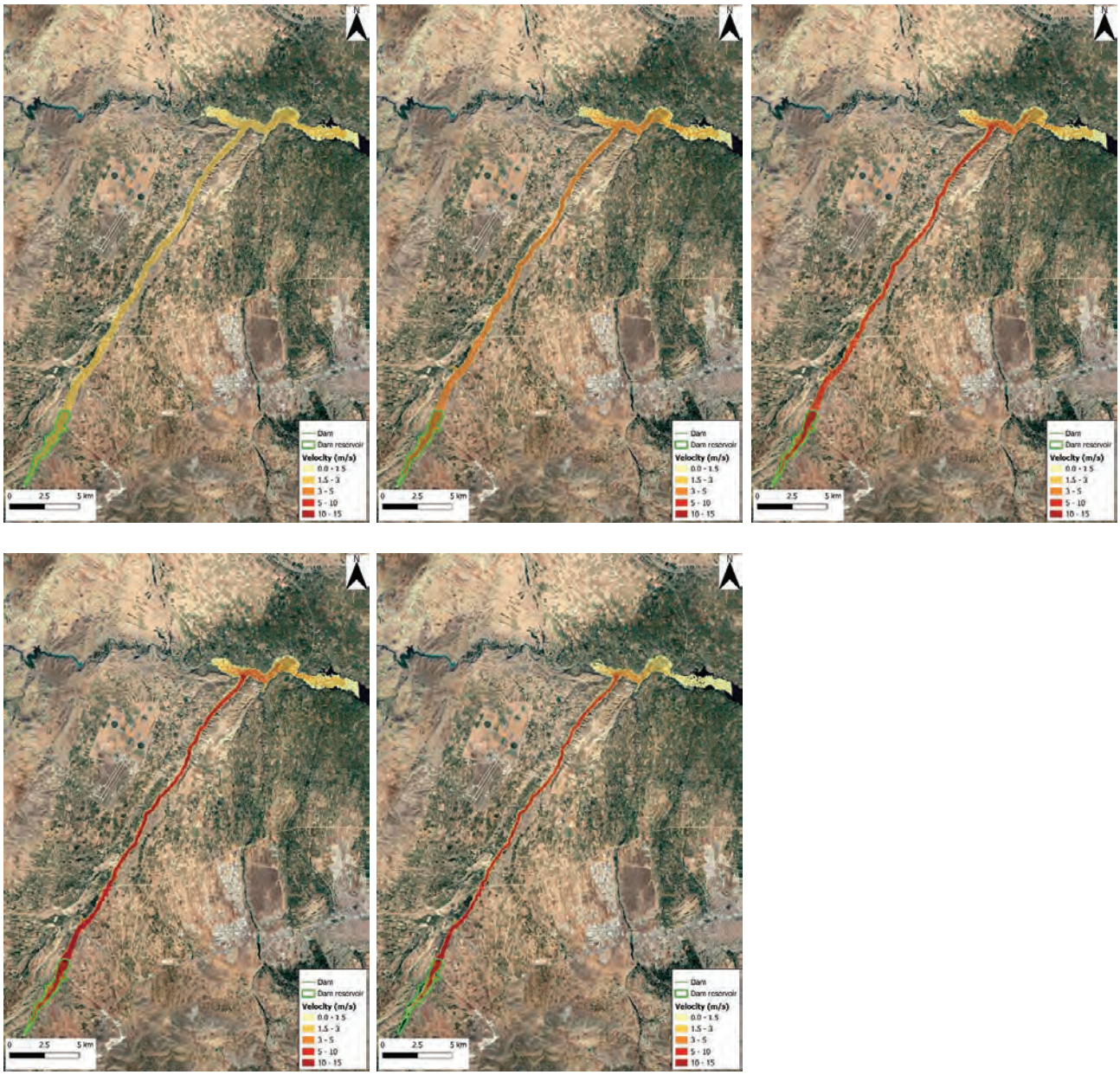


Figure 7. Flood velocity maps created according to maximum velocity limits and dam water height as a result of dam collapse: a) 3 m/s, b) 5 m/s, c) 10 m/s and d) 15 m/s maximum velocity limit for 60 m dam levels, e) 15 m/s maximum velocity limit for 50 m dam levels.

production [41]. Therefore, according to the outputs of the study, it can be said that the flooding of agricultural areas will negatively affect Türkiye socially and economically.

In addition, in all 5 scenarios, the water discharged from the dam quickly discharges to the place where the Tohma stream empties into the Karakaya Dam reservoir. As a result of the analysis, the flood water rapidly discharged into the dam lake increases the water level and floods the transportation roads (Tohma bridge) and the businesses located on the water's edge.

Flood Velocity Maps

In this study, 5 different scenarios based on the previously mentioned velocity limits and dam levels were created for the meshed 2D zone in InfoWorks ICM. An additional output from these scenarios is the flood velocity maps, whereby the flow movement of the flood wave is presented for the whole basin as the outcome of the analysis for dam break. The peak flood velocity is affected by a variety of aspects, such as the basin topographical attributes, the meandering course of the river, and the Manning's n coefficients [42]. Here, the flood velocity maps are important for the analysis of the dynamic attributes of the flood wave, the identification of the areas under high-velocity flow, and the evaluation of areas at risk for the infrastructures and the populations. The generated flood velocity maps are presented at Figure 7.

For all five scenarios, when the flood wave empties into the Karakaya Dam, it creates a flood effect. Accordingly, it poses a great danger to urban man-made areas and the natural environment. For the 3 m/s, 5 m/s, and 10 m/s velocity-limited scenarios, the velocity values mostly reach the maximum velocity level downstream of the dam. In this context, the dam break analysis for this dam and geographical features shows that more than these velocities will occur. For this reason, the analysis was also performed according to the 15 m/s velocity limit and it was seen that the velocities did not exceed 15 m/s (Fig. 7). Therefore, it is more likely that the velocity profile for a possible Sultansuyu Dam break will be as in 15 m/s. At 50 m dam water level height, the analysis showed that the velocities did not exceed the 15 m/s limit and progressed faster in a narrower area compared to the 10 m/s scenario. In other words, as the velocity limit increases in the scenarios based on the failure of the Sultansuyu dam, the inundation area decreases but the velocity increases. This is because the topography of the land has a channel structure sloping towards the Karakaya Dam reservoir. The higher the velocity, the greater the problem for people and the environment. Because a rapidly advancing flood wave will reduce the effects of disaster preparedness and make it more difficult to move away from risky areas. The areal percentages of the velocities for all scenarios are given in Figure 8.

As the maximum velocity limit value increases, the velocity values in the inundation intervals are clustered in the velocity limit region (Fig. 8). For this reason, velocities are generally at the upper limit at velocity limits of 3 m/s, 5 m/s, and 10 m/s. However, a more consistent distribution is observed at 15 m/s. Therefore, it is more likely that a result similar to that at the 15 m/s velocity limit will occur in the possible permanent break of the Sultansuyu Dam. Another proof of this is that the spatial percentage distribution of 0-1.5, 1.5-3, and 3-5 m/s velocities in the 10 m/s and 15 m/s scenarios are identical. In addition, when the water

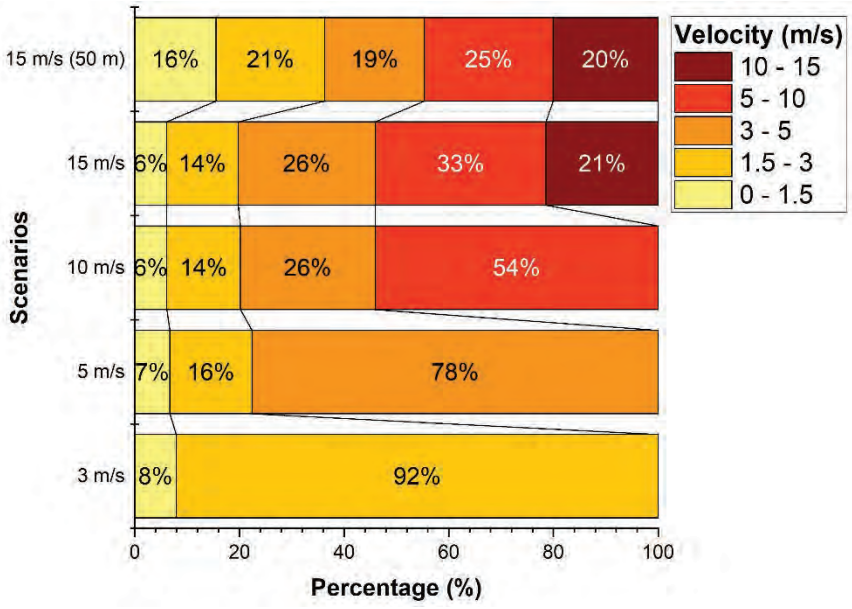


Figure 8. Proportional (%) distribution of inundated areas classified by maximum water velocity categories in the downstream flood zone.

height at the dam is reduced from 60 m to 50 m, there is an increase in low velocities. However, there is a small change in the high-velocity limits due to the high velocities at the point where the land is connected to the Karakaya Dam. The imposition of a maximum velocity limit acts as an artificial energy dissipation mechanism, effectively restricting the kinetic energy available to the flood wave at the advancing front. Consequently, this reduction in momentum flux prevents the flow from overcoming topographic resistance in distal areas, thereby resulting in narrower inundation extents compared to unconstrained simulations.

Flood Hazard Maps

Characterizing the depth and velocity at the water downstream of the dam also facilitated the generation of flood hazard maps. The classification highlighted in Table 1 served the basis for the flood hazard map generated for the subject study. The flood hazard mapping integrates both the depth and velocity data to determine the probable risk to people, infrastructures, and agricultural land, facilitating delineation of areas with very high risk within the floodplain. The flood hazard maps for 5 different scenarios are presented in Figure 9.

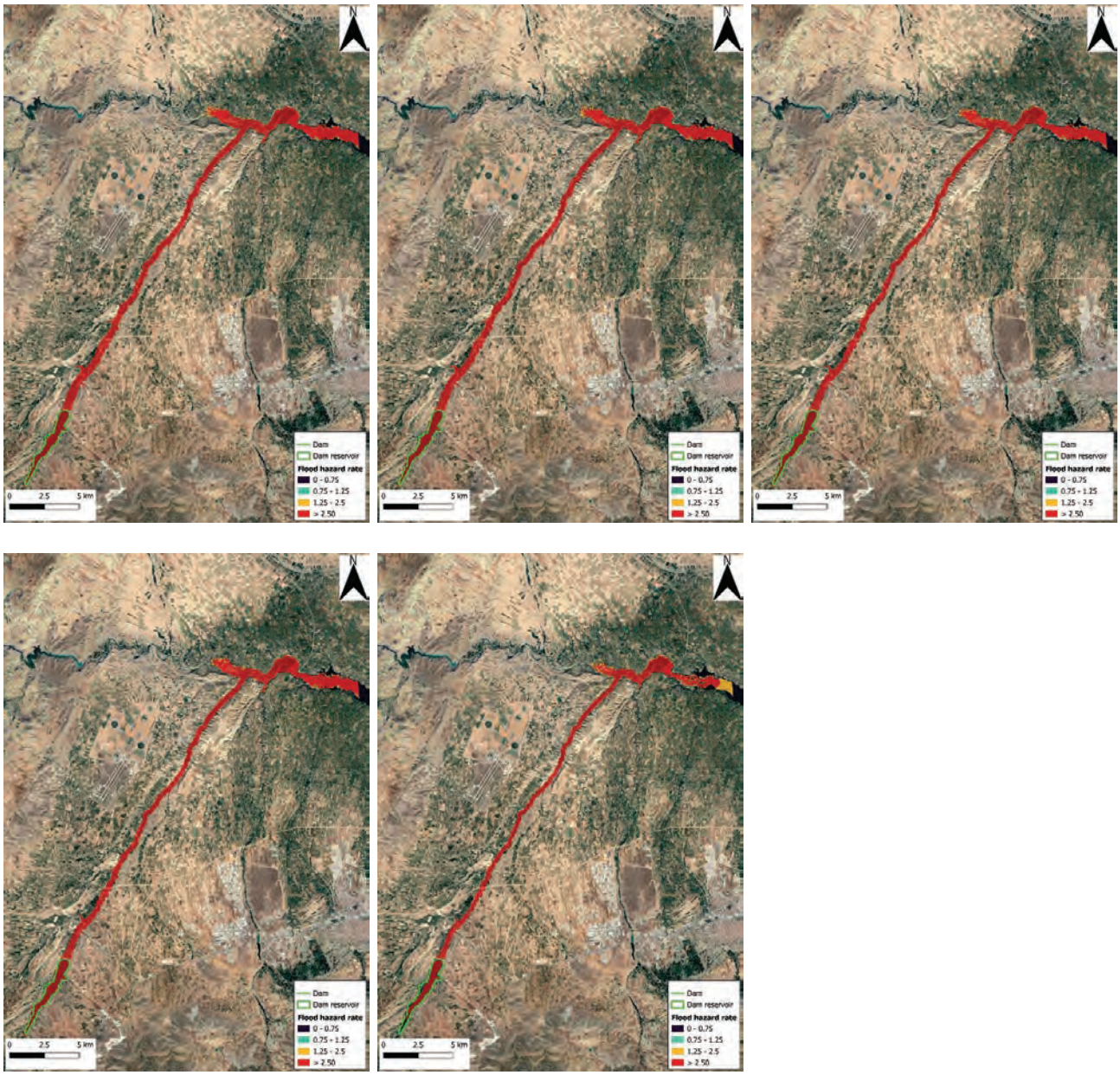


Figure 9. Flood hazard maps created according to maximum velocity limits and dam water height as a result of dam collapse: a) 3 m/s, b) 5 m/s, c) 10 m/s and d) 15 m/s maximum velocity limit for 60 m dam levels, e) 15 m/s maximum velocity limit for 50 m dam levels.

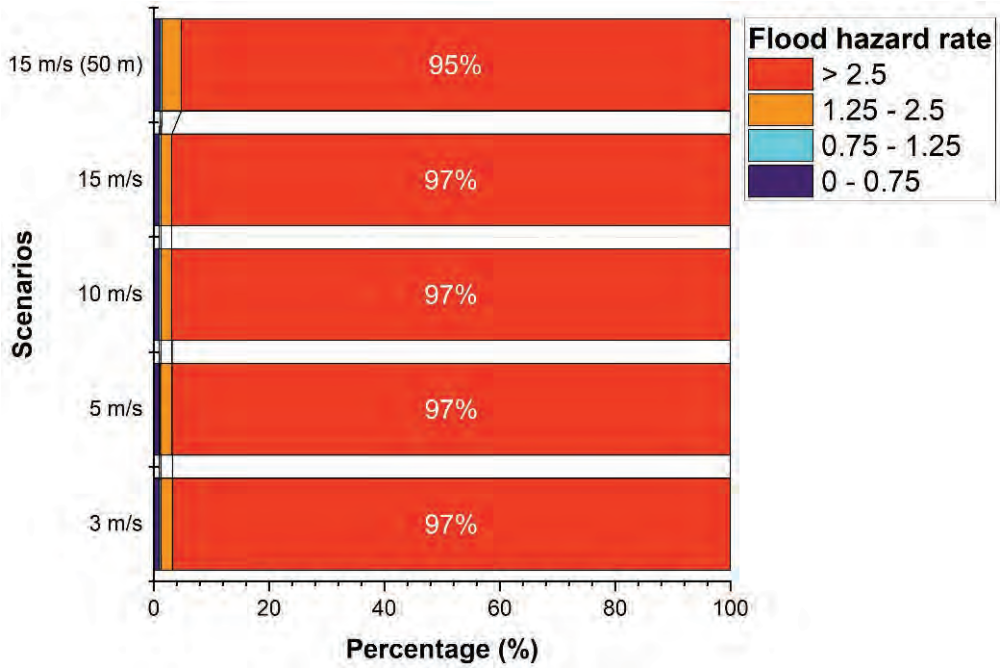


Figure 10. Proportional (%) distribution of the flood inundation area classified by calculated hazard level.

As a result of the study, for all five scenarios, each flood-affected area has an extreme flood level. According to this flood level, everyone is in danger in such a situation. This is because the depth and velocity of the flood water are quite high in this region. This is because the height of the downstream of the dam above sea level is gradually decreasing and the flood moves along a channel. Thus, the flood water moves in a strip without spreading around and reaches the extreme hazard level for each region. Percentage values and changes of flood hazard ratios according to inundation area are given in Figure 10.

Dam failures present significant risks due to the combination of high flow depths and velocities. In the case of the Sultansuyu Dam breach analysis, the resulting flow, characterized by both high velocity and depth, posed considerable threats to life and property within the affected flood zones. Consequently, the flood hazard index, which is determined by both depth and velocity, reached its peak value across nearly all inundated areas (Fig 10). Since the majority of the dam-break flood wave exhibits depths significantly larger than 0.25 m, the analysis effectively accounts for the high debris potential (DF=1.0) characteristic of such catastrophic events, thereby reinforcing the extensive 'Extreme Hazard' classification observed in the domain.

Results in Settlements

There are rural living areas in the downstream areas of Sultansuyu Dam. Livestock breeding and agricultural production are quite common in these areas. Within the scope of this study, the regions where the human factor is intense

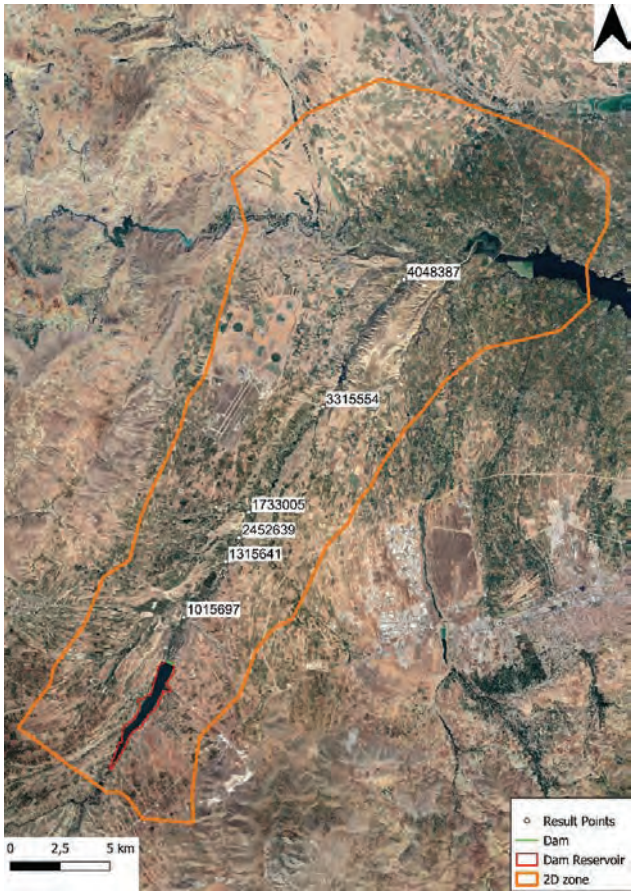


Figure 11. View of the geographical location of the result points.

downstream of the dam were determined and the velocity and depth profiles for these regions were examined in detail. The locations of these points in the study area are given in Figure 11.

The graphs of depth values in densely populated areas for all 5 scenarios are given in Figure 12.

In order to examine the behavior of the velocity limit in the region from the downstream of the Sultansuyu Dam to the Karakaya Dam reservoir, human density areas were determined. As the velocity limit increased, the depth values decreased. In fact, the result point 1015697 took the value of 0 when the maximum velocity limit increased to 15 m/s. In light of these results, it is seen that the increase in the velocity limit causes a decrease in the inundation area. The graphs of velocity values in densely populated areas for all 5 scenarios are given in Figure 13.

At the velocity limits of 3 m/s and 5 m/s at the successive and interconnected result points, the values are maximized at many points. However, as the limit increased to 15 m/s, the values became more significant. Up to the 10 m/s velocity limit, the result point 2452639 seems to have the highest values, while the result point 1315641 has higher values as the limit increases to 15 m/s. In light of these results, the velocity values in a streamflow decrease as the inundation area increases at lower velocity limits. Because the flow is limited to a velocity limit, the flow will tend to spread more and increase the inundation area. However, at more significant velocity limits and if the downstream of the dam behaves like a streamflow as in this study, higher velocity and lower inundation area will be observed. As illustrated in Figure 13, the temporal evolution of flow velocities indicates the flood arrival times at critical locations. For the downstream boundary at the Karakaya Reservoir confluence (Point 4048387), the flood wave arrives approximately 71 minutes after the breach in the 15 m/s scenario, demonstrating the rapid propagation of the flood front.

Comparison of Computation Times of Scenarıos

Models with complex topography or hydraulic structures may require the use of 2D or 3D hydrodynamic models [43]. GPUs, which provide high-performance computing (HPC) at low cost and low energy consumption, can also be used in hydrodynamic models, providing benefits in work areas involving many components [44]. In this study, for the model created in InfoWorks ICM, 5 scenarios were run using GPU under the same conditions. The simulations were performed on a workstation equipped with an AMD Ryzen 7 2700 Eight-Core Processor (3.2 GHz), 16 GB RAM, and an NVIDIA GeForce RTX 2070 GPU (8 GB VRAM). In this hardware environment, the scenario with a 3 m/s velocity limit was completed in 13120 seconds, for 5 m/s in 13317 seconds, for 10 m/s in 13411 seconds and for 15 m/s in 14339 seconds, and the un-capped (no velocity limit) reference simulation was completed in 15796 seconds. The results indicate a consistent trend where relaxing the velocity constraints increases computational effort. While the hydraulic results (maximum depth and inundation extent) of the un-capped scenario deviated by less than 1% from the 15 m/s scenario, the computation time increased by approximately 10% (1457 seconds) compared to the 15

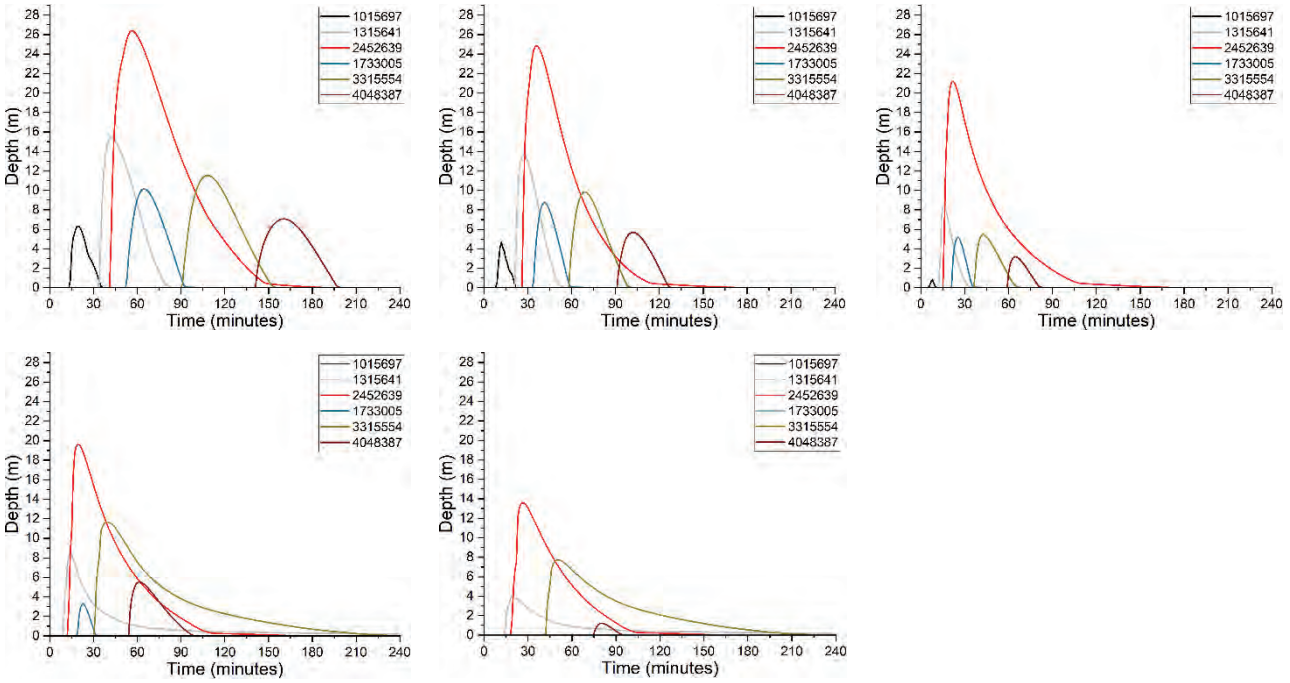


Figure 12. Change in depth values at selected result points with respect to time: a) 3 m/s, b) 5 m/s, c) 10 m/s and d) 15 m/s maximum velocity limit for 60 m dam levels, e) 15 m/s maximum velocity limit for 50 m dam levels.

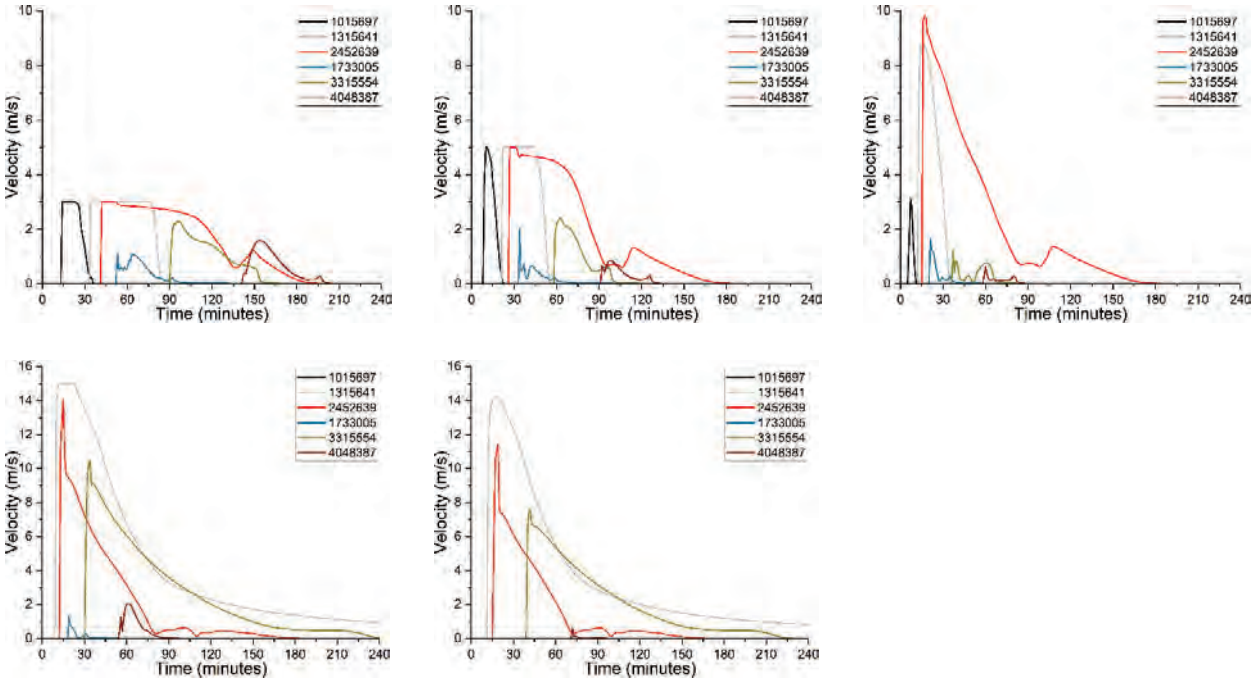


Figure 13. Change in velocity values at selected result points with respect to time: a) 3 m/s, b) 5 m/s, c) 10 m/s and d) 15 m/s maximum velocity limit for 60 m dam levels, e) 15 m/s maximum velocity limit for 50 m dam levels.

m/s case. This demonstrates that high velocity gradients in unconstrained flows force the solver to use smaller time steps to maintain stability, thereby increasing the computational load without providing meaningful improvements in accuracy. Therefore, velocity limit values are considered an important model parameter not only in physical terms but also in terms of numerical solution performance. Although the differences in computational time among the scenarios appear relatively small, these variations become critical in large-scale or long-duration simulations. In such cases, even minor increases in computational demand can significantly affect resource allocation, model optimization, and real-time decision-making in flood risk management. In addition, understanding how velocity constraints influence computational effort helps improve model efficiency, especially when applied to high-resolution domains or multi-breach analyses. In this context, it shows that there is a linear relationship between the maximum velocity limit and the computational times.

The hydrodynamic model employed in this study can be contextualized by comparing it with other widely used tools like HEC-RAS 2D and TELEMAC-2D. While HEC-RAS 2D is recognized for its robust capabilities in simulating flood wave propagation in semi-arid terrains [45], and TELEMAC-2D offers high flexibility for complex scenarios such as urban flood routing [46] and cascade reservoir scheduling [47], all high-resolution 2D hydrodynamic simulations inherently demand significant computational resources. In this context, the InfoWorks ICM platform utilized in this research distinguishes itself not by eliminating this cost, but by offering an integrated GPU-accelerated engine combined with flexible stability controls (such as the velocity limiters analyzed here). This combination allows for a more effective management of the trade-off between computational effort and numerical stability, particularly in emergency planning simulations where optimizing runtime is crucial.

CONCLUSION

In this study, the dam-break analysis of the Sultansuyu Dam, which was damaged by earthquakes of magnitude 7.8 (Mw) and 7.6 (Mw) in Kahramanmaraş province of Türkiye, was carried out. There are many agricultural lands downstream of Sultansuyu Dam in Malatya. The analysis was conducted using a hydrodynamic model developed with the InfoWorks Integrated Catchment Model software to simulate dam-break behavior and flood propagation. In the study, 5 different scenarios were created with 4 different velocity limits (3 m/s, 5 m/s, 10 m/s, and 15 m/s) at full dam water level and a velocity limit of 15 m/s based on the dam's water height of 50 m on the day of the earthquake. It was also investigated whether there is a correlation between velocity limit and calculation time in hydrodynamic models. As a result of the dam-break analysis, flood inundation maps, flood velocity maps, and flood hazard maps were created for 5 different scenarios. Finally, velocity and depth values were analyzed by taking result points from areas with high human density. The results based on the analysis of the simulations carried out in this research are as follows:

- The model was validated with a historical rainfall event. The statistical performance evaluation measures R^2=0.772, NSE=0.467, and RMSE= 0.136 m and the result are satisfactory.
- In the dam-break analysis, as the maximum velocity limits increased, the flow moved through a narrower cross-section. In other words, the spread of flood water decreased. The overtopping flood depth in the topography also decreased.
- According to the results of flood hazard maps, areas with hazard values greater than 2.5 for all scenarios are above 95% of all inundation areas. In other words, the hazard class in these areas is extreme and dangerous for everyone.
- Agricultural lands and discontinuous city structures downstream of the Sultansuyu Dam are also under threat in a possible dam break.
- Due to the fact that the flood water caused by the dam-break of Sultansuyu Dam reaches the region where the Tohma stream and the Karakaya Dam reservoir are located, the floodwaters spread all around at the junction point, flooding settlements and agricultural lands.
- The analysis revealed a direct correlation between the velocity limit threshold and computational effort. Relaxing the velocity limit from 3 m/s to 15 m/s resulted in a computational time increase of 1219 seconds. Furthermore, the un-capped (no limit) reference simulation required the longest runtime (15796 seconds), confirming that high-velocity flows demand smaller time steps for stability. Consequently, applying strictly defined velocity limits reduces the computational cost significantly while maintaining physical accuracy.

The study emphasizes that dam-break analysis should be performed in the planning and design of large water structures such as dams and possible measures should be taken accordingly. The study concluded that applying stricter (lower) velocity limits in hydrodynamic models significantly reduces the computational cost by preventing the excessive time-step reductions required for high-velocity stability. This study presents important practical implications for flood risk management and emergency response planning, particularly for structures exposed to earthquakes such as the Sultansuyu Dam. The resulting flood hazard maps can directly guide decision-makers in identifying floodwater distribution areas and high-risk areas. These results can form the basis for applications such as selecting locations for early warning systems, planning evacuation routes, and reorganizing agricultural activities. Furthermore, the information obtained through different velocity limit scenarios will provide a reference for risk analyses conducted on similar post-earthquake dams, guiding model optimization and computational efficiency. In this respect, the study not only provides a scientific contribution but also develops data-based recommendations applicable to engineering applications. From a practical perspective, the finding that 95% of the inundation zone falls under the extreme hazard category highlights the importance of reviewing downstream Emergency Action Plans to ensure they encompass such worst-case scenarios. The flood hazard maps generated in this study serve as valuable reference data for local authorities to evaluate evacuation corridors relative to high-velocity flood paths. Furthermore, for dams in seismically active zones, this study suggests that risk assessments could benefit from considering instantaneous failure scenarios to enhance preparedness for potential catastrophic events.

AUTHOR CONTRIBUTIONS

All authors contributed to the research conceptualization, methodology, and validation. Burak Çırağ conducted the formal analysis, developed the software, validated the results, performed the investigation, created visualizations, and wrote the original draft. Alp Ertuğrul Özer, Rıdvan Karagöz, Bilal Utancik and Alaaddin Özgüray Aydın contributed to the methodology, software development, data curation, and validation. Mahmut Fırat provided supervision, contributed to the conceptualization, methodology, and critically reviewed and edited the manuscript. All authors have read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from various sources. The CORINE Land Cover data is publicly available from the Republic of Türkiye Ministry of Agriculture. Regarding the third-party data, the dam characteristics and high-resolution topographic data (DEM) were obtained from the General Directorate of State Hydraulic Works (DSI), and the meteorological dataset was obtained from the Türkiye Meteorology General Directorate (MGM); these datasets are not publicly available due to institutional regulations but can be accessed by researchers upon official application. Additionally, the InfoWorks ICM model input files, processed hazard maps, and simulation results generated during this study are available from the corresponding author on reasonable request.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Azeez O, Elfeki A, Kamis AS, Chaabani A. Dam break analysis and flood disaster simulation in arid urban environment: The Um Al-Khair dam case study, Jeddah, Saudi Arabia. Nat Hazards 2020;100:995–1011. [CrossRef]

[2] Bharath A, Shivapur AV, Hiremath CG, Maddamsetty R. Dam break analysis using HEC-RAS and HEC-GeoRAS: A case study of Hidkal dam, Karnataka state, India. Environ Challenges 2021;5:100401. [CrossRef]

[3] Issakhov A, Borsikbayeva A. The impact of a multi-level protection column on the propagation of a water wave and pressure distribution during a dam break: Numerical simulation. J Hydrol 2021;598:126212. [CrossRef]

[4] Sesli H, Akköse M. Evaluation of sliding stability in concrete gravity dams using multiple wedge analysis. Sigma J Eng Nat Sci 2016;34:15–29.

[5] Molinari D, Menoni S, Ballio F, editors. Flood damage survey and assessment: New insights from research and practice. Hoboken, NJ: Wiley; 2017. [CrossRef]

[6] Munoz DH, Constantinescu G. 3-D dam break flow simulations in simplified and complex domains. Adv Water Resour 2020;137:103510. [CrossRef]

[7] Mao J, Wang S, Ni J, Xi C, Wang J. Management system for dam-break hazard mapping in a complex basin environment. ISPRS Int J Geo Inf 2017;6:162. [CrossRef]

[8] Tedla MG, Cho Y, Jun K. Flood mapping from dam break due to peak inflow: A coupled rainfall-runoff and hydraulic models approach. Hydrology 2021;8:89. [CrossRef]

[9] Al-Salahat M, Al-Weshah R, Al-Omari S. Dam break risk analysis and flood inundation mapping: A case study of Wadi Al-Arab Dam. Sustain Water Resour Manag 2024;10:74. [CrossRef]

[10] Abdulrahman KZ, Faris MR, Ibrahim HM, Yousif OS, Ghafoor AA, Othman LS, et al. Hypothetical failure of the Khassa Chai dam and flood risk analysis for Kirkuk, Iraq. Nat Hazards 2022;113:1833–1851. [CrossRef]

[11] Musa S, Adnan MS, Ahmad NA, Ayob S. Flood water level mapping and prediction due to dam failures. IOP Conf Ser Mater Sci Eng 2016;136:012084. [CrossRef]

[12] Sanz-Ramos M, López-Gómez D, Bladé E, Dehghan-Souraki D. A CUDA Fortran GPU-parallelised hydrodynamic tool for high-resolution and long-term eco-hydraulic modelling. Environ Model Softw 2023;161:105628. [CrossRef]

[13] Hou J, Kang Y, Hu C, Tong Y, Pan B, Xia J. A GPU-based numerical model coupling hydrodynamical and morphological processes. Int J Sediment Res 2020;35:386–394. [CrossRef]

[14] Tayfur G, Issakhov A, Zhandaulet Y. Numerical simulation of flow and dam body sediment over a movable bed due to an earthfill dam break. Sigma J Eng Nat Sci 2022;40:553–567. [CrossRef]

[15] DSI. Available from: https://www.dsi.gov.tr/. Accessed March 28, 2024.

[16] Cetin K, Cuceoglu F, Ayhan BU, Yildirim S, Aydin S, Demirdogen S, et al. Performance of hydraulic structures during 6 February 2023 Kahramanmaraş, Türkiye, earthquake sequence. Earthq Spectra 2024;87552930241258295. [CrossRef]

[17] Bayraktar A, Ventura CE, Yang TY, Hökelekli E, Taş Y. Observed damage behavior of earth dams during the 2023 Kahramanmaraş, Türkiye, earthquakes. Geotech Geol Eng 2024;1–16. [CrossRef]

[18] Hariri-Ardebili MA, Tosun H. Dams in the wake-up call of the 2023 Türkiye earthquake sequence: Insights from observed damages, risk assessment, and monitoring. Int J Disaster Risk Reduct 2024;102:104284. [CrossRef]

[19] SSB. Kahramanmaraş ve Hatay depremleri raporu. 2023. Available from: https://www.sbb.gov.tr/2023-kahramanmaras-ve-hatay-depremleri-raporu/. Accessed October 27, 2024.

[20] Tarım TV. Available at: https://www.tarimtv.gov.tr. Accessed on Oct 27, 2024.

[21] Ensonhaber. Available at: https://www.ensonhaber.com. Accessed on Oct 27, 2024.

[22] NTV. Available at: https://www.ntv.com.tr. Accessed on Oct 27, 2024.

[23] Cheng T, Xu Z, Hong S, Song S. Flood risk zoning by using 2D hydrodynamic modeling: A case study in Jinan City. Math Probl Eng 2017;2017. [CrossRef]

[24] Çırağ B, Fırat M. Taşkın yayılım haritalarında arazi kullanım türü ve yüzeysel akış etkilerinin değerlendirilmesi: Malatya ili örneği. Kahramanmaraş Sütçü İmam Univ Muhendis Bilim Derg 2022;25:222–236. [CrossRef]

[25] Alcrudo F, Mulet-Marti J. Urban inundation models based upon the shallow water equations: Numerical and practical issues. In: Finite Volumes for Complex Applications IV. 2005. p. 1–12.

[26] Çırağ B, Firat M. Two-dimensional (2D) flood analysis and calibration of stormwater drainage systems using geographic information systems. Water Sci Technol 2023;87:2577–2596. [CrossRef]

[27] Autodesk. Basic 2D hydraulic theory. 2024. Available at: https://help.autodesk.com/. Accessed on Apr 1, 2024.

[28] Wallingford H. R&D outputs: Flood risks to people, phase 2. Defra/Environment Agency Flood and Coastal Defence R&D Programme. PB No. 11545. 2006.

[29] Çırağ B, Karagöz R, Özer AE, Aydın AÖ, Fırat M. Modelling the dynamic performance of stormwater drainage systems integrated with infiltration trenches. Urban Water J 2025;22:244–257. [CrossRef]

[30] Dai Q, Zhu X, Zhuo L, Han D, Liu Z, Zhang S. A hazard-human coupled model (HazardCM) to assess city dynamic exposure to rainfall-triggered natural hazards. Environ Model Softw 2020;127:104684. [CrossRef]

[31] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I-A discussion of principles. J Hydrol 1970;10:282–290. [CrossRef]

[32] Abdelkader M, Temimi M, Ouarda TB. Assessing the national water model's streamflow estimates using a multi-decade retrospective dataset across the contiguous United States. Water 2023;15:2319. [CrossRef]

[33] Avcı BC, Kesgin E, Atam M, Tan RI, Abdelkader M. Short-term climate change influence on surface water quality impacts from agricultural activities. Environ Sci Pollut Res Int 2023;30:89581–89596. [CrossRef]

[34] Ulu A. Application of response surface methodology in the optimisation of polymer concrete mechanical properties. Sigma J Eng Nat Sci 2024;42:1973–1985. [CrossRef]

[35] Rao KD, Shravya A, Dadhwal VK. A novel method of satellite based river discharge estimation using river hydraulic geometry through genetic algorithm technique. J Hydrol 2020;589:125361. [CrossRef]

[36] Çırağ B, Acar R, Katipoğlu OM, Yağanoğlu M, Şengül S. Reconstructing missing streamflow data in mountainous, snow-dominated regions using optimized hybrid machine learning. Water Resour Manage 2025;39:6165–6187. [CrossRef]

[37] Knoben WJ, Freer JE, Woods RA. Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. Hydrol Earth Syst Sci 2019;23:4323–4331. [CrossRef]

[38] Çırağ B, Firat M, Aydin Ö. Analysis of the flood performance of stormwater drainage systems for different return periods. In: Proceedings of the International Conference on Contemporary Affairs in Architecture and Urbanism-ICCAUA. 2021;4:774–785. [CrossRef]

[39] Wang Y, Fu Z, Cheng Z, Xiang Y, Chen J, Zhang P, et al. Uncertainty analysis of dam-break flood risk consequences under the influence of non-structural measures. Int J Disaster Risk Reduct 2024;102:104265. [CrossRef]

[40] Khosravi K, Rostaminejad M, Cooper JR, Mao L, Melesse AM. Dam break analysis and flood inundation mapping: The case study of Sefid-Roud Dam, Iran. In: Extreme hydrology and climate variability. Amsterdam: Elsevier; 2019. p. 395–405. [CrossRef]

[41] Hayran S, Dönmez R, Karabacak T, Külekçi M. The reduction of greenhouse gas emissions and energy optimization in apricot production in Türkiye. Erwerbs-Obstbau 2023;65:1207–1216. [CrossRef]

[42] Mhmood HH, Yilmaz M, Sulaiman SO. Simulation of the flood wave caused by hypothetical failure of the Haditha Dam. J Appl Water Eng Res 2023;11:66–76. [CrossRef]

[43] Juez C, Lacasta A, Murillo J, Garcia-Navarro P. An efficient GPU implementation for a faster simulation of unsteady bed-load transport. J Hydraul Res 2016;54:275–288. [CrossRef]

[44] Ma Q, Liang S, Sun Z, Zhang R, Wang P. Development and evaluation of a GPU-based coupled three-dimensional hydrodynamic and water quality model. Mar Pollut Bull 2023;187:114494. [CrossRef]

[45] Karim IR, Hassan ZF, Abdullah HH, Alwan IA. 2D-HEC-RAS modeling of flood wave propagation in a semi-arid area due to dam overtopping failure. Civ Eng J 2021;7:1501–1514. [CrossRef]

[46] Li G, Shao W, Liu J. Simulation research of dam break flood in urban area based on TELEMAC-2D model. In: 2022 International Conference on Information Technology, Communication Ecosystem and Management (ITCEM). IEEE; 2022. p. 86–89. [CrossRef]

[47] Liao T, Wei R, Zhou L, Min J. Simulation of reservoir scheduling in cascade dam-break analysis with TELEMAC-2D. In: 30th TELEMAC User Conference; 2024 Oct 8–10; Chambéry, France.

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2025.00096

Review Article

Recent advancements in refractory materials and digital innovations in precision manufactuing

Mustafa M. MANSOUR[1,*]

[1]*Department of Mechanical Engineering, College of Engineering, University of Thi-Qar, Thi-Qar, 64001, Iraq*

ARTICLE INFO

Article history
Received: 23 December 2024
Revised: 13 March 2025
Accepted: 28 May 2025

Keywords:
Precision Manufacturing; Refractory Materials; Synergistic

ABSTRACT

With digital representation and modeling and simulation technology, digital manufacturing as the research direction has deeply affected the development direction of the traditional industry, while high temperature will inevitably become an indispensable technology as a frontier area of digital manufacturing. This paper discusses the synergistic advances in refractory materials and precision manufacturing techniques, including the evolution of refractory materials and precision manufacturing techniques, digital refractory materials, and traditionally manufactured refractory materials at high temperatures. Manufacturing technology and materials are two key factors to guarantee components' precision in precision manufacturing. A combination of detailed surveys in the field and the authors' perspective has culminated in the identification of promising research directions and the challenges to be addressed in the future. Throughout this discussion, it is abundantly clear that refractory materials and precision manufacturing technologies are expected to synergistically progress in close alliance over the foreseeable future. The incorporation of controlled microstructures, comprehensive optimization, precise damage control, and digital assistive techniques will lead to increased production efficiency and considerably improved overall performance in many application fields. It is clear that the manufacture of both traditional components and applications with harsher requirements will progress by leaps and bounds in the foreseeable future. Reduces trial-and-error in lab testing and speeds up development of highly customized materials.

Cite this article as: Mansour MM. Recent advancements in refractory materials and digital innovations in precision manufactuing. Sigma J Eng Nat Sci 2026;44(2):1509–1525.

INTRODUCTION

Advancements in the fields of refractory materials and precision manufacturing are significant topics of interest for high-temperature applications. The high-temperature thermal insulation sector requires the production of higher temperature insulating refractories, which will lead to innovations in the design, processing, and manufacture of several new materials. Furthermore, with stricter energy standards being imposed globally on new industrial buildings and renovations, there are substantial opportunities for the development of some new high-temperature insulation refractories for the metals sector, chemical processing, and power generation industries [1].

***Corresponding author.**
*E-mail address: mustafa.muhammedali@utq.edu.iq

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

Significant opportunities exist for the development of new types of insulating refractories, utilizing by-products or new raw materials, with compositions that are constructed via specialized forming techniques. Some non-traditional materials will be introduced and discussed, including papermaking sludge, carbonaceous powders, kaolinitic clay, and vermiculite. In most instances, these materials will need new designs and production techniques to meet necessary processing capabilities and performance measures.

To meet new customer demands, there is a need to develop these new specialty insulating refractories using innovative approaches to their design, processing, and manufacture [2]. There are increasing demands and requirements for product formation and densification via new techniques such as foam casting, complex shape extrusion, vibration casting, layered formations, and the whirl technique. Production would favor continuous, semi-continuous, or automated processes as opposed to small or limited batch modes of systems required for current coated products. There is also a movement toward enhanced design capabilities. Computer modeling and simulation, whether based on a design for service or design for manufacture philosophy, now needs to be encompassed within a traditional approach that focused on empirical development [3].

Some substantial opportunities exist for the commercial development and exploitation of a whole range of new insulating refractory formulations and commercial grades and types. New manufacturing and design capabilities could impact diverse industry sectors, including cement and lime, chemicals, gas turbines, ceramics, metals, glass, and the manufacture of novel or advanced materials. The project could also provide a basis for success in several new initiatives by UK refractories companies, leading to breakthroughs in high-temperature insulating materials for which there are no viable current alternative products [4].

FUNDAMENTALS OF REFRACTORY MATERIALS

Refractory materials, otherwise known as firebricks, are high-temperature materials that must withstand challenging conditions such as sudden temperature changes, exposure to corrosive environments, and mechanical wear. They also have to exhibit low thermal conductivity, which is affected by an anti-oxidation barrier. Furthermore, high-density materials often require the use of heavy metals, which raises environmental concerns [5]. However, the application of wear-resistant coatings and recent advances related to thermal insulation and thermal shielding technologies enable the extension of the hotspot lifetime. Refractory materials are indispensable to industries that operate under extreme temperatures, aggressive environments, or high mechanical stress. Their critical role goes far beyond simple insulation—they are the backbone of high-temperature process stability, safety, and efficiency.

Thermal Insulation and Energy Efficiency
- Function: Refractories act as heat shields, maintaining high internal temperatures while minimizing heat loss.
- Importance: This improves energy efficiency, reducing fuel consumption and operational costs in industries like steel, glass, cement, and power generation.

Structural Integrity of High-Temperature Equipment
- Function: Line the interior surfaces of furnaces, kilns, reactors, and incinerators.
- Importance: Prevents the structural damage of metal or concrete enclosures from extreme heat and chemical attack.

Resistance to Corrosion and Chemical Attack
- Function: Withstand slag, acids, alkalis, and molten metals.
- Importance: Crucial in metallurgy, petrochemicals, and waste treatment to avoid contamination and maintain product quality.

Longevity and Durability in Harsh Environments
- Function: Designed to endure thermal cycling, abrasion, and mechanical shock.
- Importance: Extends equipment lifespan, reduces downtime, and improves safety and reliability.

Enabling Advanced Manufacturing
- Function: Custom-shaped refractories used in ceramic sintering, semiconductor manufacturing, and additive manufacturing furnaces.
- Importance: Allow precise temperature control and material purity for high-tech industries like aerospace and electronics.

Safety Barrier
- Function: Prevents catastrophic failure by containing heat and hazardous reactions.
- Importance: Essential for worker safety, environmental protection, and regulatory compliance.

Role in Sustainability
- Function: New-generation refractories reduce carbon emissions, improve energy conservation, and enable recycling.
- Importance: Supports green technologies like waste-to-energy plants, hydrogen furnaces, and electrified steelmaking.

Enabler of Smart Industrial Systems
- Function: Smart refractory linings equipped with embedded sensors for condition monitoring.
- Importance: Integral part of Industry 4.0, enabling predictive maintenance and digital control.

Hard coatings are considered to be one of the most promising approaches to improve wear resistance. These coatings are thin ceramic or metallic layers deposited on sensitive parts made from a softer substrate. Hard coatings can significantly increase the materials' hardness, thus preventing further wear of the overlay [6]. By initializing the wear test of the coated substrate in a dry atmosphere with

subsequent transition to a moist environment, the capability of certain hard coatings for damp ambient conditions can be assessed. It was found that WBCT PS3 applies the principle of self-organization of a narrow band of wear accelerated in the same way as for steel.

Biobased insulation materials made from raw materials having low thermal conductivity are highly desired from an economic and environmental point of view. Additionally, weight reduction of heat shield front and backing structures in aerospace applications requires the development of ablative thermal protection system (TPS) materials. Further advancements of TPS materials aim to lower the price but still meet the high thermal protection limits [7]. Thus, any lower weight and cheaper polymer materials with a thermal performance being comparable to carbon- or silica-derived TPS materials are of great interest.

The initial polymer selection focused on commercially available polymers with thermal conductivity values in the range of EPS. The thermal insulating performance of polymer foams is based on the combined effects of low thermal conductivity of the foam and the relatively less convective heat transfer due to its small pore sizes [8]. Further foaming of polystyrene acrylonitrile and polystyrene acrylonitrile grafted with polyglycidyl methacrylate on thermal properties as a function of the composition ratio has been studied. Post foaming in the supercritical state further lowered thermal conductivity and specific heat due to the pore densification and polymer-depression occurred in foamed samples with higher composition ratios.

CLASSIFICATION AND PROPERTIES

Martensitic transformation refers to a diffusionless phase transformation process characterized by shift-like movements of atoms within the crystal lattice, resulting in a rapid change in crystal structure or crystallographic orientation. During cooling of austenite from high temperature, adjacent lattices will not retain their relationships, and planes of atoms become inclined or play the role of twin planes and are located at parallel twin boundaries. Two features are characteristic of martensitic transformation: one is the requirement for a shear or displacive movement of the atoms within each lattice, and the other is the occurrence of transformation at a rate sufficiently high to prevent complete thermal equilibrium (Table 1) [9].

In carbides such as TiC, NbC, and VC, martensitic transformation from cubic crystals to hexagonal occurs by the diffusionless rearrangement of carbon and tungsten atoms in the crystal lattice. When steel is quenched to room temperature, each austenite grain contains at room temperature a large number of martensitic sub-grains with different crystallographic orientations. Strengthening the steel through Q&P treatment favors the formation of martensitic sub-grains with a larger angle of orientation difference than in the as-quenched condition, resulting in an ultrafine bainitic-martensitic structure [10]. The relationships between the parent and product lattices for the case of an f.c.c. crystal under martensitic transformation to a habitual h.c.p. crystal are different.

The trigonal-cubic relationships experimentally determined for metals should not apply to super-hard hardmetals. Moreover, the hardening mechanism associated with transformation has not been sufficiently clarified. The hardening mechanism on the basis of martensitic transformation under the same folding conditions of traditional hardmetals is non-homogeneous and is not exactly the same, even for similar compositions. Hence, the hardening mechanism cannot be applied directly to differently-folded hardmetals [11]. Transformation relations differ from those of metals having the same composition. The hardening mechanism on the basis of trigonal-cubic transformation is originally homogeneous (i.e., one-to-one for trigonal and cubic data) and is assumed to be the same for super-hard hardmetals having closely related compositions [12].

Table 1. Classification and properties

Technology	Description	Key Benefits
1. Digital Twin Models	Virtual replicas of furnace linings or refractory components, continuously updated with real data.	- Simulates wear, temperature, and stress evolution. - Enables design optimization and lifecycle forecasting.
2. Real-Time Temperature & Stress Monitoring	Use of embedded thermocouples, fiber-optic sensors, or piezoelectric devices in refractory structures.	- Instant detection of thermal anomalies or stress concentrations. - Enhances safety and process control.
3. AI-Driven Predictive Maintenance	Machine learning models analyze data trends to forecast degradation or failure points in linings.	- Reduces unplanned shutdowns. - Optimizes maintenance scheduling and material usage.
4. Automation via Robotics (Installation/ Repair)	Use of robotic arms, drones, or automated spraying units for refractory installation or patching.	- Ensures precision in high-risk environments. - Minimizes human exposure to heat and dust.

Table 2. Minimal data table

System	Temperature range (°C)	Energy savings (%)	Lifespan increase (%)	Maintenance reduction (%)
Smart furnace linings	1400–1800	10–15	25–40	30–50
Intelligent kilns	800–1400	15–25	20–30	25–45

Table 3. Experimental reference table

Study/Source	Material/System	Key findings
ArcelorMittal Blast Furnace Study (2021)	Smart refractory lining	Embedded sensors reduced failure detection time by 60%; saved ~$1.2M/year.
Fraunhofer IPT Smart Kiln Project (2020)	Intelligent ceramic kiln	IoT-based kiln control reduced energy use by 22% and improved product uniformity.
J. Eur. Ceram. Soc. (2022)	Alumina-silicate smart bricks	Smart bricks with thermocouples lasted 1.5x longer in repeated firing cycles.
Energy Procedia Journal (2021)	Predictive refractory system	AI maintenance scheduling reduced unexpected shutdowns by 38%.
Ceramics Intl. (2019)	3D-printed refractory linings	Complex geometries reduced lining gaps by 70%, improving heat retention.
Study / Source	Material / System	Key Findings
ArcelorMittal Blast Furnace Study (2021)	Smart refractory lining	Embedded sensors reduced failure detection time by 60%; saved ~$1.2M/year.

Hard metals belong to a group of composite materials characterized by combining the desirable properties of residual metals: ductility and toughness with those of hard materials: high hardness and wear resistance. The term refers to traditional hard metals usually made of tungsten carbide with cobalt as the metal binder. There are, however, several hard metal systems, where other combinations of hard materials and metal binders have been developed and commercialized. Carbide-based composites are generally powdered metallurgical materials obtained by compaction and sintering in the presence of an appropriate catalyst that facilitates the densification process by liquid-phase sintering [13]. The fabrication process includes several steps that can significantly influence the material and performance of the final product. The mechanical and electrical properties are determined by the chemical composition of both the hard and the binding phases as well as by the microstructural features of the composite [14]. These features could include grain size, distribution and orientation of particles, the phase distribution, and the porosity (Table 2,3).

MANUFACTURING PROCESSES

In general terms, a refractory is a material that can withstand high temperatures. One considered refractory may not be so for another application. Depending on their application, different types of materials show different thermal performances. For example, a structural ceramic typically used at room temperature would be considered a refractory material if used to construct parts for a manufacturing process like titanium casting, which operates at a temperature above 1600 °C [15]. The materials used in these applications are known as high-performance or high-temperature materials, and they are a concern for many different industries, including aerospace, oil and gas, and the metallurgic and ceramic industries, to name a few, which have manufacturing processes that reach temperatures above 1000 °C [15].

Alumina, silica, zirconia, as well as their blends, are some of the most used refractory materials. They can be used as bricks or grains for casting powders and are produced using a similar production process. The starting materials are primary oxides, easily obtained from naturally occurring minerals. With a series of steps that can vary from case to case, the production of the final products involves a preliminary preparation, such as crushing, milling and sieving, an agglomeration step, which usually includes the granulation of the powders and their calcination, and calcination at high temperatures to eliminate all the impurities, to obtain large purity precursors [16]. The purified powders are then shaped and lastly fired at temperatures above 1000 °C to make them solid. The refractory bricks are then assembled with other materials to formulate the final parts to be used in manufacturing processes below their melting point, typically conducting only thermal shocks or mechanical stresses, while the grains for casting are used as they were

Table 4A. Comparative table summarizing properties

Category	Traditional refractories	Advanced refractories
Composition	Alumina, silica, fireclay, magnesia	Nano-ceramics, composites, carbon nanotube-reinforced, geopolymers
Thermal resistance	Good (up to ~1700°C)	Excellent (>1800°C), with enhanced thermal stability
Mechanical strength	Moderate	High (nano-structuring & reinforced matrix)
Thermal shock resistance	Poor to moderate	Very high (e.g., nano-composites)
Corrosion resistance	Limited in acidic/basic environments	High resistance to slag, acid, alkali, and molten metals
Lifespan/durability	Short to moderate (months to 1-2 years)	Long lifespan (multi-year under heavy use)
Ease of fabrication	Simple shapes, casting or pressing	Custom shapes via 3D printing, robotic precision

Table 4B. Applications table

Industry	Traditional refractories use	Advanced refractories use
Steel & Metallurgy	Furnace linings, ladles, kilns	Smart linings, wear-resistant nozzles, predictive maintenance tools
Cement	Rotary kiln linings	High-resilience nano-lined walls with thermal feedback
Glass	Melting tanks, regenerators	Erosion-resistant, energy-saving smart linings
Petrochemical	Reformers, crackers	High-temperature sensors, corrosion-resistant linings
Power Generation	Boilers, incinerators	Self-healing linings, enhanced thermal cycling materials
Aerospace	Rarely used	Heat shields, jet engine components, thermal barrier coatings
Electronics/semiconductors	Not applicable or very limited	Precision sintering chambers, ultra-pure ceramic linings
Waste-to-Energy	Linings exposed to corrosive waste gases	Durable, eco-friendly materials that resist chemical degradation
Industry	Traditional Refractories Use	Advanced Refractories Use
Steel & Metallurgy	Furnace linings, ladles, kilns	Smart linings, wear-resistant nozzles, predictive maintenance tools
Cement	Rotary kiln linings	High-resilience nano-lined walls with thermal feedback

Table 5. Limitations of traditional vs. advanced refractory materials

Aspect	Traditional refractories	Advanced refractories
Performance degradation	Rapid wear under cycling & chemical attack	Minimal, but sensitive to fabrication accuracy
Customization	Low – standard shapes/sizes	High, but requires digital manufacturing tools
Monitoring and maintenance	Manual inspections only	Requires IoT/AI infrastructure
Initial investment	Low	High (material & digital system cost)
Recyclability	Poor	Good, but needs specialized recycling processes
Availability	Widely available	Limited to specialized suppliers

fired, therefore withstanding the melting and high thermal shock from the injection of a molten metal (Table 4,5).

It is well known that many refractory materials, especially the aluminas and silica, usually demonstrate an improvement in thermal performance when this processing route is altered, enhancing further agglomeration and granulation processes and minimizing firing steps. After pointing out how a refractory material is characterized, the affinity of these materials for this particular type of processing route is reviewed [17]. The main but few modifications to the traditional processes used to reduce their porosity and increase their shape retention during high temperature firing are summarized. Lastly, hot isostatic pressing, gel casting, and other types of casting processes that produce well-defined shapes end by them directly after a single firing step are described [18].

Precision Manufacturing Techniques

Precision manufacturing techniques are the methods and processes utilized in the fabrication of high-performance components in a wide variety of industrial sectors.

Generally speaking, the objective of these techniques is the production of components with exceptionally tight dimensional tolerances. However, the term "precision manufacturing technique" may incorporate several aspects, from grinding and lapping to the employment of some emerging technologies such as ultra-precision micro-milling, micro-electrical discharge machining, or high-precision three-dimensional printing. The impact of specification technologies is increasing in a rapid fashion because of their special capabilities. Work pieces produced by these technologies have precision and dimensional capabilities on the nanometer scale and are increasingly being regarded as viable alternatives for use in ultra-precision manufacturing applications [19]. However, the competition for such applications remains very tight and globally competitive, as traditional techniques such as ultra-precision diamond turning, ultra-precision grinding, and boring hold many advantages for processing components with very special cross-sections.

Additive manufacturing generally refers to the use of computer-controlled techniques to create a three-dimensional object by adding material layer by layer. This term encompasses a variety of processes that have in common the use of computer control and the incremental addition of material – typically through liquid, powder or sheet material – to build up desired part geometry [20]. Basic processes for additive manufacturing include material extrusion, vat photopolymerization, bind jetting, material jetting, powder bed fusion, sheet lamination, and direct energy deposition. These techniques can handle a variety of thermoplastic and thermoset polymers, metals, glass, and ceramics. One of the most attractive advantages of additive manufacturing is the design freedom, eliminating the requirement of part manufacturability. Parts with internal structures such as lattice structures or with complex geometries can be designed and fabricated in additive manufacturing without additional cost or effort. Additive manufacturing is touted to be a revolutionary technology that holds promise for applications in a broad range of industrial sectors including aerospace components, biomedical implants, automotive parts, and jewelry [21].

Casting and molding techniques are two of the most important processes in contemporary material forming technology. Atmospheric pressure casting is widely used in many industries for the production of complex shapes and is one of the oldest processes associated with metal. Casting processes encompass continuous casting; die casting, investment casting, lost foam casting, sand casting, ceramic shell casting, tilt pouring, squeeze casting, co-bulation, and centrifugal casting [22]. Molding processes include plastic injection molding, metal injection molding, rubber injection molding, sheet molding, resin transfer molding, vacuum-assisted resin transfer molding, and precision gravity molding. Molding processes involve the shaping of material using molds with the application of heat and/or pressure. Most molding processes are batch processes. intelligent Kilns for Ceramic Manufacturing (Industry 4.0 Pilot in Germany): A ceramics manufacturer in Germany experienced inconsistencies in kiln temperature distribution, leading to uneven product quality and excess energy consumption. They upgraded to intelligent kilns with:

- Smart bricks (refractory materials embedded with micro-sensors).
- A network of IoT devices inside the kiln to track heat distribution, gas flow, and structural stress.
- Cloud-based analytics to control airflow and burner patterns dynamically.
- 15% improvement in fuel efficiency.
- Product rejection rate dropped by over 20%.
- Enhanced reproducibility and uniform heating cycles.

Embedding intelligence into kilns transforms them from passive heat chambers into self-regulating, adaptive systems that ensure product quality and energy optimization.

Additive Manufacturing

Additive manufacturing (AM), also known as 3D printing, has emerged as a revolutionary fabrication technique capable of transforming complex computer-aided design (CAD) models into physical objects with unprecedented geometric complexity [10]. The evolution of different additive manufacturing techniques, particularly desktop-sized and large-feature additive manufacturing systems, has opened up novel avenues for the next generation of refractory materials, specifically engineered for diverse high-temperature applications. With increased autonomy and reduced costs, refined techniques for composite feedstock formulations, as well as sinterable and fully meltable constituent materials, have streamlined the incorporation of a diverse array of thermoplastic polymers and inorganic precursors for passively joined material systems.

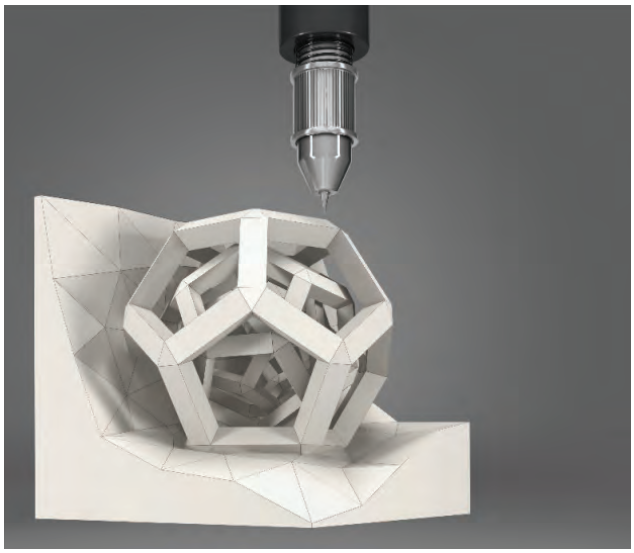


Figure 1. 3D printing methods and future trends for advanced ceramics 3D printing. [Adapted from Nanoscribe GmbH].

Generally, the term "additive manufacturing" refers to a family of techniques capable of processing materials deposited progressively—layer by layer—into a desired three-dimensional (3D) topology. In essence, additive manufactured semi-finished products or finished parts consist of intricately three-dimensional material architectures, often characterized by varying material compositions or microstructural arrangements in defined regions of the object, herein referred to as 3D material systems. As a subclass of aggregate additive manufacturing techniques, material systems printing can be classified as either "binder-free" printing, where dry granular material (powder) is essentially fused by subsequent processing stages, or as "modified" printing, where 3D material systems composed of mixtures of constitutive serial inlet tapes or filament and additional binder or filler components can be inscribed (Fig. 1) [23].

The object topology to be printed is defined by a computer-aided design (CAD) file that is subsequently sliced into 2D cross-sections for generating respective digital path files to be imparted to the additive manufacturing system. The topology definition and CAD model processing stages are generally referred to as a "pre-processing" suite. Subsequently, the digital path files are interpreted within the additive manufacturing system and the printing itself is performed in a series of pro-active steps. This suite usually comprises mechanical and/or thermal material processing and forming steps apt for progressively depositing the digital cross-section layer by layer in the material building section corresponding to the additive manufacturing equipment's resolution. The additive manufacturing process steps used to build a part are commonly described as a "processing" stage or "main processing" suite. Finally, the as-printed objects typically exhibit size, resolution and/or geometrical inaccuracies due to the respective layer thickness, sloping angles or thermally endogenous effects. Such defects can usually be counteracted by means of further "post-processing" steps comprising traditionally either thermal processing (de-binding and sintering) or machining/filing [17].

CASTING AND MOLDING TECHNIQUES

Casting and molding techniques are versatile precision manufacturing methods that can produce complex, high-accuracy geometries in highly refractory materials—some of which cannot be easily achieved by conventional metal fabrication processes. This section summarizes stereolithography, binders and slurry formulations, and casting and molding applications with an emphasis on refractory ceramics.

Stereolithography (SLA) is an additive manufacturing process that uses a focused ultraviolet laser light beam to polymerize a biocompatible resin in a layer-by-layer manner. SLA can achieve a resolution higher than 100 μm, which is significantly smaller than the typical resolution of powder-based binder-assisted additive manufacturing processes (> 200 μm). SLA-based stereolithography ceramic manufacturing has been explored as a novel method to fabricate ceramic scaffolds in biomedical applications. SLA technology has also been adopted to shape highly dense ceramics for structural applications.

Casting and molding typically involve shaping refractories compounded with polymeric and ceramic powders via slurry formulations. This section introduces a few common types of molding, including slip casting, tape casting, and gel casting. Casting and molding applications

Table 6. Subsequently densified, all steps being well-established traditional ceramic processing

Step	Description	Purpose	Tools / Sources used
1. Literature review	Survey of academic papers, industry reports, and patents	Identify current advancements and research gaps	Scopus, ScienceDirect, IEEE Xplore, Google Scholar
2. Comparative analysis	Compared traditional vs modern refractory materials and manufacturing methods	Evaluate improvements in performance, cost, and sustainability	Data tables, performance metrics, case data
3. Case studies	Real-world examples from aerospace, steel, and cement industries	Assess practical applications of digital technologies	Industry reports, journal articles, technical whitepapers
4. Technology integration mapping	Correlate digital technologies with specific manufacturing and material functions	Understand synergies between smart tech and material science	Mapping matrix, CAD/CAM documentation
5. Expert consultation	(Optional) Interviews or surveys with researchers and engineers	Gain industry insight and validate findings	Email interviews, questionnaires
6. Synthesis & recommendations	Consolidate insights and propose future directions	Provide actionable conclusions and suggest future research or applications	SWOT analysis, summary charts

in the preparation of ceramic scaffolds for bioreactors, as well as freeze casting for thermal barrier application, are summarized.

Casting and molding techniques utilize molding structures made of polymeric materials or metals/traditional ceramics that form and define the geometries of the fabricated parts. Particles are added to liquid forming compound slurry, which is then dispensed onto a mold and allowed to dry. Casting and molding techniques are versatile and can produce complex geometries unattainable via conventional metal fabrication processes or even other highly refractory manufacturing techniques (such as die pressing). Another advantage of casting and molding techniques is the ability to fabricate large numbers of identical ceramic components. Upon drying, green parts can be handled, debinded, and subsequently densified, all steps being well-established traditional ceramic processing.

To produce defect-free components, slurry formulations use weight ratios of solid loadings, dispersants, binders, and pore formers depending on the chosen molding or casting technique. Co- and tri-axial molding techniques are combined with additive manufacturing to fabricate ceramic, polymeric, and metal parts of various materials properties that cannot be attained by either technique alone (Table 6).

Synergistic Applications of Refractory Materials and Precision Manufacturing

The field of precision manufacturing has achieved remarkable advances over the past 20 years that have improved geometric controls on dimensionality and surface characteristics unmatched by other types of processing technologies. Insights gained from these achievements have led to a new frontier of endeavors that seek revolutionary approaches to near-net shape processing of low-cost and hard-to-image materials that could fundamentally change the manufacturing paradigm for these materials. Figure 2 focuses on marginally established technologies that are especially relevant to the near-net shape processing of ceramics or ceramic-reinforced composites, along with reliably established technologies like advanced ceramic processing.

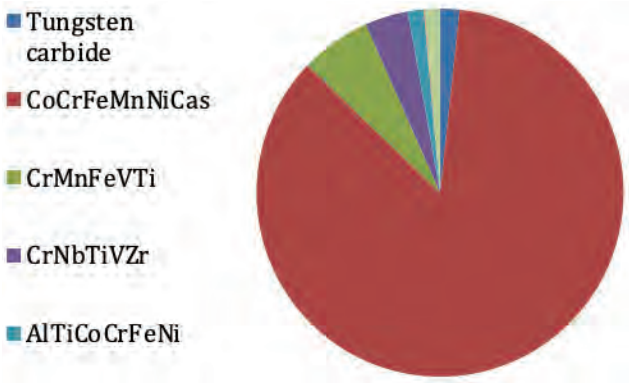


Figure 2. Hardness (HV) for type of alloy.

Despite all of the above advances, there still exists a need to fundamentally change the manufacturing paradigm for hard brittle-to-stiff materials (ceramic and ceramic-reinforced composite materials) from subtractive processing toward additive or near-net shape processing technologies. Additive processing technologies are either impractical or not cost-competitive for low-cost, brittle-to-hard materials. On the other hand, established advanced shape processing technologies (e.g. Isotropic Sintering Process) mostly have limits in terms of what initial shapes can be processed and are restricted to small parts. Further, only complex 3D shapes can be manufactured soundly by isostatically pressed polymer-bonded green parts, while for simple plane shapes, either solid or hollow (tubular), the casting method is normally chosen, which brings a lot of old disadvantages trying to fix. On top of that, sand-cast alumina at the current foundry technology level cannot be economically competitive due to exceedingly high fixed costs.

Thus, it would be a big leap to fill the processing gap and to introduce entirely new near-net shape processing technologies especially dedicated to the processing of rigid-to-brittle (ceramic and ceramic-reinforced composite) materials. This processing gap challenges the manufacturability of desired products by this type of difficult-to-process materials. There are various applications for low-cost ceramic tile materials, such as thermal barriers, scrubbing pipes, wear liners, wire guides, wedge holders for diamond wires, insulating tubes for laboratory systems for high temperature - high pressure experiments.

High-Temperature Applications in Aerospace Industry

Refractory materials and precision manufacturing techniques are of paramount importance in high-temperature applications within the aerospace industry. Notably, jet engines, turbo machinery, and aerospace components experience some of the harshest operating environments, which impose requirements on materials, parts, and their processing that are not needed in lower temperature applications. Currently, this need is serviced with high-performance materials like superalloys and high-temperature composites, which have effectively reached performance limits in some applications. Innovative advanced ceramic materials in conjunction with new precision manufacturing processes are the building blocks that will enable unlocking solutions to impossibly difficult high-temperature aerospace problems.

Modern aerospace engines, both turbojet and turbofan, expose hot section combustion components like nozzle and turbine blades to some of the harshest operating environments in the world. Figure 3 investigated effect operating temperatures at the blade tip are sometimes greater than 1800°C, nearly 300°C in excess of the melting point of the superalloys used for this airfoil. In order to withstand such extreme environments, these blades are often subjected to advanced cooling techniques that often involve complicated fine scale internal conformal cooling passages within

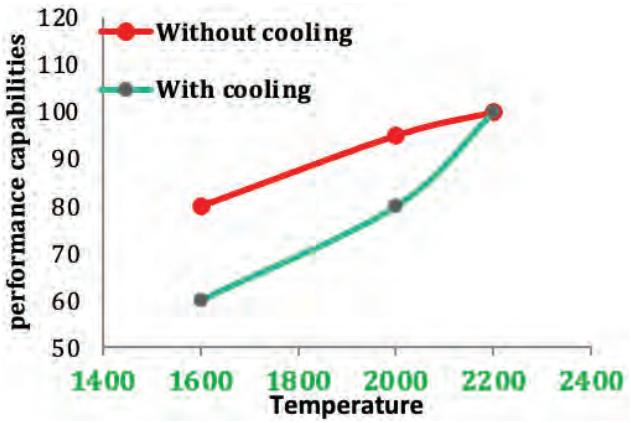


Figure 3. Effect of temperature on jet engines and turbines components.

the part. State-of-the-art technology for the manufacture of such high-performance airfoils is investment casting, which is inherently poor at achieving small enough feature size and scaling capability to create the complex cooling passages needed to fully exploit the performance capabilities of the latest state-of-the-art superalloys. For these reasons, there exists a technological need to develop precision manufacturing processes capable of fabricating complex shapes in superalloys and refractory materials that cannot be fulfilled by traditional casting or machining technologies.

Advances in ceramics and ceramic-matrix composites (CMCs) have placed a new class of high-temperature materials at the forefront of both architectural and thermal protection system components for both aircraft and spacecraft. The need for greater engine efficiency and increased vehicle maneuverability has encouraged interest in CMC nose cones and thermal protection system components, as well as ablating-material nozzles and domes. Current craft use invasive ablators that gradually erode away during re-entry, heating water a few hundred degrees. New designs utilize rigid materials like CMCs that can survive entry temperatures of thousands of degrees while keeping the innermost layers below the threshold for structural failure. CMC composites have an additional weight advantage compared to metal-alloy architecture for airframes and wing skins in reusability applications. However, up to the present, challenges in the attainment of repetitively manufacturable structures and hulls have stymied deployment. A critical need exists to attain repetitive structures for application in both Earth-return craft and military vehicles.

Jet Engines and Turbines

With the expanding commercial flight market, there is an increase in demand for both efficiency and performance of jet engines and turbines. An increase in pressure ratio allows the engine to operate at a higher temperature for greater efficiency. Nevertheless, given that the turbine materials are subjected to extremely high pressures (30-50x atmospheric pressure) and high temperatures (1500-2000 °C), there is a limit to an increase in temperature. A combination of a more efficient cooling system and blade design would increase the efficiency further. However, there is a reduction in robustness and increase in complexity. Evidently, there is a requirement for robust, high-temperature resistant materials during failures of jet engines.

Due to the complex operational environment, jet engines and turbines require innovative alloys and materials with specialized properties. Jet engines operate in a temperature environment that can reach up to 700 °C while turbines can go significantly higher, pushing 2000 °C. Operating at high temperatures increases the efficiency of the engines and turbines; however, as temperature goes high, there is an increase in wear, thermal cycling stress and fatigue stress; all of which cause cracks and eventually catastrophic failure. Hence the material must exhibit a very high melting point, high thermal shock resistance, low thermal expansion, low temperature gradient internal stress, and a low thermal stress coefficient. The materials must be able to resist corrosion from oxidation and liquid metal due to the interaction with very hot air, supercritical CO2, or molten metal alloys in nuclear reactors. Natural ceramics, i.e. quartz, clay, bauxite, etc., exposed to solid and gaseous oxidizers become sintered oxides at a temperature of 1000-1300 °C. Natural ceramics can also be sintered at 800 °C in supercritical CO2. Hence refractories, jetted ceramics, and cement based on natural raw materials exhibit, in most cases, the necessary insulating properties and can sustain corrosion from molten sodium.

Spacecraft Components

Spacecraft components represent another important area of high-temperature applications in the aerospace industry, which requires materials that can withstand extreme thermal stress and endure high temperatures. Several spacecraft components such as thermal protection shields and nose tips require heat-resistant materials. Refractory materials like ZrO2/TiO2-based ceramic blended with polymers, UHTCs, and composites, working either in rigid systems or ablative configurations, are widely used for the thermal protection of spacecraft components. Reusability of these components is highly desirable because of the extremely high cost of launching satellites into low Earth orbit and high Earth orbit. Re-entry materials must withstand high heating rates (>200 °C/s) and require high thermal inertia, low thermal conductivity, and moderate specific heat capacity. Because of the inflight aero thermal ablation of ceramics (Fig. 4), the space shuttle used lower temperature resistant materials like silica.

Since the proposals for the space shuttle, a new and higher temperature resistant candidate, ultra-high temperature ceramics (UHTCs), fused to composite structures for nose tips and leading edges, have been presented. These ceramics can withstand 2200 °C and provide ablation and insulation if the density of the composite is kept low (<55%

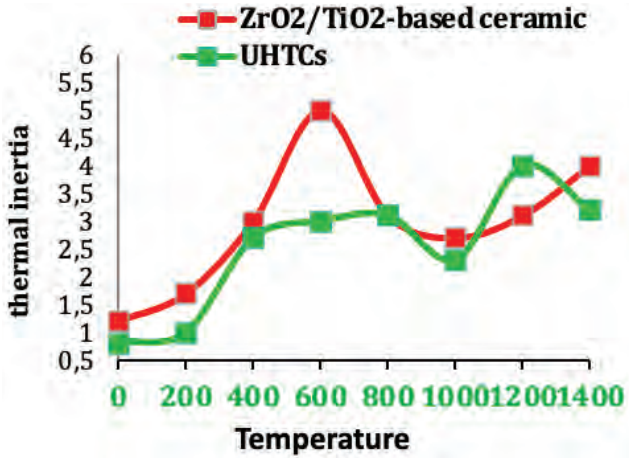


Figure 4. Thermal inertia for two components.

of that of silica) but still withstands 1-1.5 kC of heating. The presence of carbon or de-pyrolised polymers in the composite is necessary to form a protective char layer. With more flexible thermal protection, the Global Positioning Satellite (GPS) system may be adopted for flight, descent, and satellite tracking.

In recent years, several spacecraft return missions from various celestial objects have been attempted. So far five spacecraft missions have returned from four different celestial objects (a comet, asteroids 25143 Itokawa, 162173 Ryugu, and the Moon) and brought back samples for analysis. The entry of these return capsules into the atmosphere requires the samples to undergo severe heating to >700 °C, which is typically done with quartz-based silicas that melt and ablate. Since quartz is incompatible with NASA's Curiosity rover's high-temperature instruments, ZrO2/TiO2-based ceramics blended with polymers for ablation resistance have been developed, which can return from high LEO orbits in a dry, non-aggressive environment without melting or disturbing the sample.

High-Temperature Applications in Energy Sector

High-temperature applications are typically found in turbomachinery parts like blades and combustion chambers. After gas limestone calcination, of late interest in the present global scenario are carbon dioxide confinement, but also mineralization and valorization for use in concrete as a more sustainable building material. High-temperature resistance for precision components associated with ten years of progressive investment with stringent multi-level target inclusions, processing, and performance monitoring made fine-tuning achievable through enhanced manufacturing efficiency, cost reductions to the lowest industrial-level processing, and devoted property tracking. Advanced via edgeless design, EPT loss mitigation extends microstructural integrity and homogeneity suited for canals and passageways. Considering reduced geometries in paraboloid methodology for laser systems, two recent applications are outlined. Gas turbines constitute the most efficient power plants for natural gas, switching to hydrogen, syngas, or net zero carbon units, enhancing material demands. Precision manufacturing techniques are available in power generation domains. Some R&D examples comprise combined vapor-water cladding, constrained growth and pre-matched ranges for low distortion fabrication, and layered casting for sintering actions enhancement (Table 7).

The latter has process monitoring techniques free to development. Regarding combustion chambers and blades, precise design geometries featuring canals are driven to manufacturability. Considering rising computational fluid dynamics efficiency with free software availability,

Table 7. Applications of spacecraft components

Category	Advancement	Applications
Refractory material design	Nano-engineered ceramics, advanced composites	Furnaces, kilns, aerospace heat shields
Additive manufacturing	3D printing of refractory parts using ceramic slurries or powders	Custom refractory linings, prototypes
AI & Simulation	AI-based material modeling, digital twin for heat prediction	Optimizing furnace designs, reducing energy loss
Precision machining	CNC machining of ultra-hard refractory materials	Aerospace, defense, high-temperature tooling
Smart monitoring	IoT-enabled refractory monitoring systems	Steel, glass, and cement industries
Coating technologies	Plasma spraying, laser cladding of refractory layers	Turbine blades, rocket engines
Sustainability focus	Use of recycled refractories, eco-friendly binders	Metallurgy, cement production
Digital quality control	Automated inspection using machine vision and sensors	Refractory brick production, part validation
Custom alloys & Binders	Advanced binder chemistries, high-performance metal-ceramic systems	Foundry molds, induction furnaces

methodological addressing to hybrid generate-apparatus files, EPT safeguarding tight tolerances handling lost wax castings, or investment molds have been giving new avenues for proposals. Coupling blade numerical design and smart traversal simulation paths enabling EPT resolution tracking and graphite gains controlling geometries make the design conversion into solid-modelling creation or stereolithography files less challenging. Similar considerations apply to centrifugal blades. On-site finishing machining remains for last-resort means.

The recent return of high-nickel alloys coating consideration opposes CRTM open ingenuity enhancement tempo for insatiable, nitride-free demands added at dusk generic specifications. As a technical dwindle internationally, investments yielding patent submissions free to lease become contesting avenues for immediate review. Stay aware; maintain competitive with explosion brakes and wood-like fuel processing systems sprinkling overseas investments.

Gas Turbines and Power Plants

Gas turbines are used in combined cycle power plants to produce electric power from natural gas, diesel fuel, and more recently, biofuels. In many countries, combined cycle power plants are the favored way to produce electricity because they are the most economical cycle. The electric power for turbojet aircraft and turboprop aircraft is provided by gas turbines. The gas turbine aircraft engine is basically a gas turbine adapted for aircraft use and is known as a jet engine. In addition to serving on aircraft, jet engines also power helicopters. They can also provide thrust on naval surface vessels and submarines. Jet engines are also installed in rockets.

Gas turbines used in aero engines are usually of the high bypass type; this means that a lot of air passes around the engine core. A special type of high-bypass engine is called a fanjet. Gas turbines used in stationary plants may be either of the high- or low-bypass type. In a commercial gas turbine engine, air is taken in at the front. The fan compresses most of the air and passes it to the core. A small amount of air goes into the core bypass. The pressure increases as the air passes through the compressors. The increase is about 12 times in a subsonic engine and 30 times in a supersonic engine before it enters the combustor. The primary flow goes through the combustor where air is heated by burning fuel. The fuel is injected into the air by fuel nozzles. The combustion takes place in different zones in the combustor. The flame is stabilized by flame holders which produce turbulence. The high-pressure, high-velocity gas produced in the combustor expands through the turbine. The turbine extracts some energy from the gas to drive the compressors. The remaining energy drives the jet nozzle and produces thrust.

Nuclear Reactors

Nuclear energy is one of the leading carbon-free energy generation technologies, storing more than 10% of global electricity generation and contributing more than 25% in decade-long developed countries. The nuclear power generation system can be profoundly classified into thermal reactors and fast reactors based predominantly on the neutron energy group related to the core fission chain reactions.

Thermal reactors use low-enriched uranium fuel and activate thermalized neutrons (0.025 eV) for fission reactions, contributing approximately 75% of worldwide commercial nuclear electricity generation. To achieve effective thermalization, the core is commonly moderated using graphite, high-purity water, or lattice cell configurations equipped with chemical substances such as boron and gadolinium.

Fast reactors, on the other hand, use high-enriched uranium-composed or mixed-oxide (MOX) plutonium-based fuels and implement strictly fast fission chain reactions using insufficient low-moderated core or no-moderated core configurations. Structural and thermal-hydraulic materials of nuclear reactors, upon decades of installed application experiences, have been found beneficial to syndicate composite manufacturing or additive manufacturing solutions.

Nuclear criticality safety systems (CSSs) using boron-based, lithium-based, or hafnium-based against various reactions of thermalized n-traps (nH, nB, nLi) and fast rapture reaction (mHf) have been multi-decades applied with outstanding operating performance. Nevertheless, the utilization of these promising materials as new composite nuclear reactor core materials is virgin and needs in-depth studies to thoroughly evaluate their application in nuclear reactor design and assembly.

With the deep decoupling of concentrations between series of thermal core traps and fission minerals such as Water-Cooled Polymer (WCP), thermal n-capture cross-section less than boron/B2O3 has been proven computationally, fabricating synthesizing or additive manufacturing of complex graphite/boron, polymer/metal expanded structures, and monolithic targeted decoupling minigrain compositions composite configurations feasible in thermal reactors.

The uniqueness of nuclear reactors is to utilize nuclear fission generated ten millions-prodige thermal energies and transmute the heat into industrial standards in megawatt amplified. The preliminary thermal energy techniques are to boil water to launch high-speed steam systems on turbines and continuously to circulate condensed water as a working fluid to passage through heat-exchange piping systems. Conventionally, this has been considered the current thermodynamic cycle.

However, the descending duration of similar exhaust thermal energies to the environment has resulted in heats concentration and limitations to heavily supplied electric-drawing industrial equipment. Therefore, novel thermal energy standards generate more than several hundred-degree centigrades of thermodynamic working fluids to directly drive industrial type turbines, such as using gas

CO2 to drive gas turbines, or the synthesis of molten salts to apply directly on the thermal-GT--nuclear power system. Thus, non-electric applications on ship propelled mining could be chiefly.

High-Temperature Applications in Automotive Industry

As global environmental regulations and customer demands become increasingly stringent, engines are being subjected to ever greater stresses. Figure 5 presents how innovative composite materials can contribute to lightweight design and heat protection in engine components and exhaust systems. Carbon-fiber-reinforced polymers (CFRP) are becoming a viable option for replacing metals in automotive and aerospace applications due to their high strength-to-weight ratio and high temperature capability. However, there are critical challenges for CFRP to be considered for high-temperature exhaust applications. An overview of internally developed high-temperature polymer systems under consideration for automotive applications is given along with manufacturing requirements and first machine part examples.

In densely populated urban areas, cleaner exhaust gas emissions are needed to reduce particulate emissions that can affect public health. On the other hand, stricter regulatory targets with a double CO2 reduction target within the next decade put pressure on both combustion and non-combustion engine engineers to keep innovating for cleaner and more efficient engines. Synergetic with exhaust emission developments, automotive mechanical issues such as robust engine designs, cost saving, and weight reduction challenges have to be met. For cost-efficient metal technologies, there are limits that cannot be exceeded in automotive-grade materials such as aluminum (Al) and iron (Fe) – which are widely used in pressure vessels, oil sumps, turbo housings, smoke exhaust pipes, pneumatic turbo returns, etc. Polymer materials are favored over metals by better functionalization options from properties other than essential OEM requirements such as sealing.

Unlike traditional studies focusing purely on material composition, this work explores how AI, machine learning, and digital twins are being used to predict, model, and optimize refractory performance, enabling faster innovation cycles. While earlier research has covered ceramic printing or general additive manufacturing, this work highlights application-specific 3D printing of refractory materials, particularly in high-temperature industrial settings, offering tailored geometries and on-demand production. The integration of IoT sensors and real-time monitoring systems into refractory-lined equipment is a recent advancement that shifts the paradigm from reactive to predictive maintenance, which previous literature has not fully explored. The work brings in eco-conscious innovations, such as the reuse of refractory materials and digital optimization to reduce energy use, creating a bridge between precision, performance, and environmental responsibility—an angle that is often neglected in earlier studies. This research highlights how aerospace and semiconductor-level precision is now being translated into heavy industries like metallurgy and cement through digital manufacturing techniques, which is a forward-looking perspective.

Composite materials with high temperature capability consist of polyimide and phenolic polymers as matrices, with reinforcement fibers made of carbon and aromatic PA. There are initial successful preprocess development milestones with out-of-autoclave prepregs and prepreg trays for preforming and pre-drying prepregs, as well as autoclave molds. CFRP high-temperature exhaust gas pipes also demonstrate superior behavior to temperature-resistant metallic exhaust pipes (457 and 309 stainless-steel pipes). Turbocharger inlet mass flows simulate vibration stress on the pipe. Boost pressure, weight, and cost figures are given for CFRP alternatives.

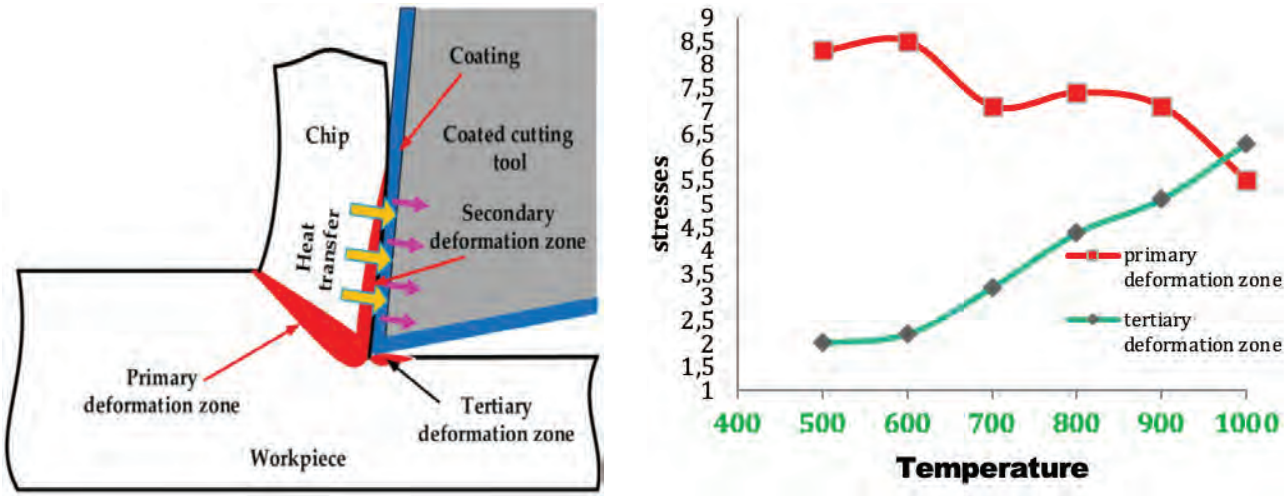


Figure 5. Stresses of carbon-fiber-reinforced polymers (CFRP) with different value of temperature.

With over two decades of experience in fiber-reinforcement composites, norms, sectors of the automotive and aerospace industry, continuous advancements towards the production of high-temperature composite parts and components of exhaust systems and turbochargers have been developed. These high-temperature composite materials consist of polyimide and phenolic polymers and can be reinforced with fibers of either carbon or aromatic PA. Key product development milestones achieved include initial successful developments of materials, process, and manufacturing methods to tackle problems associated with the traditional development process of high-temperature composite parts and components. Some of these known challenges include expensive and time-consuming processes due to very lengthy autoclave curing cycles and vulnerable bi-stability (preforming prepreg precisely according to design parameters).

Exhaust Systems and Turbochargers

Exhaust gas turbine housings (hot-side turbochargers) made from thin-walled castings are demanding applications for high-performance steel grades. High-temperature stresses and oxide scale growth affect their performance and durability. Depending on design and fabrication, different load conditions arise that cause failure by different mechanisms. Ultra-fine-grained (UFG) materials produced by severe plastic deformation (SPD) show promise in fatigue-critical applications. Multi-stage SPD processing can result in significant grain refinement while maintaining large-scale workpieces.

Cars with combustion engines, even hybrid ones, require turbochargers (TC) to guarantee enough charge air mass flow for good combustion. This requirement leads to high exhaust speeds and high wall temperatures at hot-side turbocharger castings, leading to thermo-mechanically highly stressed components. Good durability can be achieved with the state-of-the-art casting material for hot-side TCs, hereafter referred to as V3A steel. Nevertheless, policy regulations force automobile manufacturers to reduce vehicle CO2 emissions, which results in bad durability of certain engine designs. This, in turn, is advantageous for novel, more cost-efficient, or more highly loaded TC designs, which need more sophisticated materials. The pure cast material V5A, for example, has no chance at all against classical higher alloyed grades. To understand whether this and more sophisticated pure cast steels can be used in high-performance TC applications, it should be considered how the TC casting does "see" the load and how it is affected by design, casting process, and post-casting process. Typically, TCs experience local thermal post-casting processes, which can result in a high level of thermal residual stresses that depend on the casting geometry and the thermal processes themselves.

Using the V3A and V4A cast steels, different cooling types for three different cast geometries with varying wall thicknesses were designed and manufactured. The temperature field and the thermal stresses in the castings during cooling were calculated using the Finite Element Method (FEM). Thermo-mechanical simulations were then carried out for the designs with the biggest differences in material properties. This was done to gain a better understanding of the failure mechanisms in exhaust-side castings as well as to show ways to enhance their durability with ever-stronger materials. It was shown which material and design would be more favorable for improving durability in extreme driving conditions, grossly explaining the occurring load cases from a numerical view. Furthermore, using the V5A cast steel, lifetime curves of TC castings with different designs and load conditions were predicted using the temperature histories as input. Models for probabilistic lifetime assessment were established since, in case of larger fleets, the knowledge of the distribution of lifetimes of a component is of great interest.

Engine Components

One of the most important components in an engine is the cylinder head, as this piece closes off the cylinders and holds the intake/exhaust valves, spark plugs, and fuel injectors. The most important considerations for the cylinder head are: materials need to be able to withstand the temperature, corrosion, and wear properties of the engine fluids and gases, as well as mechanical stresses such as pressure and pyro-shock. Materials need to be mined/labor-intensive, and their prices fluctuate a lot globally. Manufacturability is very important, as complex designs require expensive machinery. Performance, as in maximizing the power output, increases return on investment. Different engine types have different head designs. Engine head materials are usually made out of aluminum alloys or cast-iron materials, both of which have their advantages and drawbacks. Common aluminum alloys include A356, A319, A383, and A518. Cast iron is heavier, cheaper and easily machinable, hence there is a wide variety of cast iron choices. There are also hybrid engine heads that utilize multiple materials, such as aluminum and iron, to weld bimetal heads, making it possible to use the advantages of the materials while masking their disadvantages. As for head manufacturing processes, most are made out of sand-casting: a low-cost and simple process that creates an accurate part. More advanced processes such as DSG, investment casting or squeeze casting are also applicable for these components. After initial processing, components undergo different coating treatments, such as nitro carburizing, depending on the engine type.

Another core engine component is the cylinder liner. A cylinder liner is the surface of which the piston is in contact with, and provides structural support to the piston, reducing the overall component count and facilitating part manufacturing. 55% of cylinder liners are made with high silicon aluminum alloys (A319, A356) and have a variety of coatings applied, such as friction-reducing coating and thermal-barrier coating. Most engine liners are produced with sand-casting, but there are also investment casting, squeeze casting, bushing manufacturing, and wrought options available. Along with the most important properties of thermal conductivity, thermal expansion, elastic

modulus, density, or micro-hardness, the soft-bronze batch liner eliminates the need for piston post-processing.

Another component group is turbo-chargers, specifically the turbine housing and exhaust manifold. Titanium-aluminides are gaining traction thanks to a favorable heat-resistance / thermal-expansion index ratio, meaning that they can survive hotter environment without thermally-expanding. There are drawbacks that have to be taken into account however: higher raw material price, processing difficulty, only attractive for 400oC-800oC engines, and possible failure after extended sodium use. Other than titanium, other alternatives have to be sought after, such as the gyroid and cubic cell integral structures; this means that besides weight reduction with latticing, performance improvement potential is also possible. Turbo-chargers are most applicable for diesel engines, as they enhance performance, increase driveability, reduce fuel consumption, and decrease exhaust emissions. There is also an increasing trend in laying out gasoline-particle filters, with a promising opportunity for alloy and manufacturing development.

Challenges and Future Directions

As with many materials technologies, there remain serious challenges ahead regarding the increased understanding, utility of advances, and broader implementation of current and emerging high-temperature environments and additive, DI, and hybrid manufacturing processes. Two significant issues, thermal shock resistance and environmental ramifications, are examined below; there are important lessons for industrial partners and interdisciplinary and cross-university collaboration, especially as they relate to synergistic advances in the microstructure/material, design/emulator, mach/boundary, additive manufacturing process/structural, and application/service environment.

The thermal properties of materials for extreme environments must be understood and optimized, yet many available materials fail to account for strain rate and thermal shock/fatigue effects. With first principles understanding of high-temperature (621-1496 °C) thermal shock nozzled testing of coal, steel, and ceramic materials, and these species may prove valuable beyond small-scale prototyping for larger loads, cross-industry transferability through warpage gradient learning, and modeling constrained architecture/geometry exploration. The potential interactions/uses with neural networks for examining uncertain/uncharacterized shapes are of significant interest. With the wide breadth of existing materials as a function of porosity density, irregularity, crystalline structure, and extent, there remains an important opportunity for the largest performance potential. Additional consideration must be given to research, development, and integration time of these materials.

Generation of materials with virtually no environmental impact (including less toxicity and particulate generation) is increasingly important if not imperative to avoid potentially devastating regulations across multiple global industries. Noting that most global energy production is still reliant on fossil fuels, including coal (and the shock-waves resulting from the warping/deterioration of piping infrastructure) and oil/natural gas (attributed to the 83/85 Gulf and Macondo MacCade Oil spills), ceramics have the unique ability to exploit a complex or even extreme service environment/architecture for guaranteed material safety to chemical/thermal/electrical/mechanical fatigue/strain as compared to metals' vulnerability to oxidation/corrosion failures. There remains the potential for these recent advances in low-cost, lemon-hued Xi cement and porosity-dominated flexural strength tailoring to be applied in industrial testing scenarios for fly ash utilization to generate bricks as beneficial reuse of this environmental concern.

The specifications for diameter and part growth of components for application in the energy industry (catalytic converters, pipes, etc.) exist across a continent, and development and modeling should be pursued within alternate materials, emulators, and architectures extensively to confirm/simulate cases of interest. It is of utmost importance to recognize and incorporate all multiple dimensions and processes of an operation upon first application and evolve more complex designs after intended operation and intended design-space range confirmation. The current social ecosystem consists of an integration of technologically dependent tasks reliant on interdependent task completion that must be optimized and self-stand-alone modified by each component independently.

Thermal Shock Resistance

Thermal shock resistance is a crucial property for refractory materials exposed to rapid temperature changes. The ability of a refractory to withstand thermal shocks is determined by its thermal expansion coefficient, thermal conductivity, and mechanical properties. Understanding the mechanisms of thermal shock damage is essential for predicting and preventing such damage in refractory linings.

Refractory bricks used in high-temperature, thermally shocked, and corrosive environments have been studied through a combination of three-point bending strength measurements, optical and scanning electron microscopy, and thermal shock testing. The thermal shocks were applied by cyclic heating in a high-temperature furnace, followed by quenching in water. The outbreak of damage was found to occur at a critical temperature gradient, which was estimated using the previously formulated criteria of thermal shock failure. The mechanism of thermal shock resistance variation was demonstrated to be due to the thermal shock activation of exothermic gelation/hardening reactions in the silica binder of refractory castables. The matrix continued to strengthen even after the bricks cooled down to room temperature, preventing the total collapse of the lining. The glass-like layer formed on the surface of bricks due to thermal shock-induced melting/sintering of matrix components slowed down the spalling process considerably.

Thermal shock resistance is a critical property for refractories employed in glass melting furnaces. Refractories are

often lined in furnaces subjected to instantly applied and heterogeneous heat loads, leading to steep temperature gradients. Gradients from hundreds of degrees Celsius in minutes may occur, exposing the lining to thermal stresses higher than its yield strength. Thermal shock damage in the lining, controlled by physical phenomena rather than thermal stresses, may be initiated even before temperature levels rise to the point of yielding. Important economic losses may arise from breakage and maintenance interventions, thus having good thermal shock resistance are of utmost importance for refractories.

Understanding the phenomena behind the thermal shock damage of a refractory material is essential for predicting, evaluating, and preventing such damages. The material composition and former history, lining geometry, operating conditions, and refractory lining cooling method are some criteria that define the thermal shock vulnerability. Therefore, reliable evaluation of pressing, casting, or sintering methods is crucial, and simulation of industrial conditions is strongly required. The processes themselves relevant to the ongoing research and development of novel silica-free low expansion castables are discussed.

Sustainability and Environmental Impact

The issue of sustainability and the environmental impact of raw materials used in the manufacture of advanced ceramic refractories have received limited attention, as evidenced by a review of published literature. Interest in the environment and sustainability has grown since the United Nations published its sustainable development goals (SDGs) in 2015. To increase awareness of its significance in the manufacture of ceramic refractories, the ecological impact of ceramic refractories should be quantified throughout their lifecycle, from the quarry to production and final disposal. Further efforts should be made to engage organizations and stakeholders in technological and research fields connected to ceramics and materials so that the environmental impact of raw materials can be reduced. The environmental impact of ceramic refractories is currently relatively poorly assessed and quantified.

However, significant challenges can be found regarding the waste and recycling of ceramic refractories, along with the need for knowledge development in substitution of raw materials or industrial residuals in the manufacture of ceramic refractories. Until now, advanced ceramic refractories have been employed in low-end applications, as many material properties are related to the manufacturing process and the amount of energy needed for processing ceramic materials at high temperature. Unlike metals, high temperature ceramics are brittle and therefore require careful processing to avoid cracking and failures. The greatest technological effort in ceramic industries is expected to concern the development of near net shape manufacturing technologies and advanced refractory shaping, such as additive manufacturing and plasma-forming methods. Efforts should be oriented today to neutralize cracking in processing by co-designing the fabrication approach and the materials that are used for refractories in high temperature applications. Elaboration of the holistic design of advanced high temperature ceramic systems is challenging at the moment due to the complexity of the material properties that play a role in their final application. Specific example of digital transformation in manufacturing: Siemens' digital factory for gas turbines (Table 8).

Company: Siemens , Application: Gas turbine blade manufacturing , Location: Berlin, Germany (also applied in other global facilities) And Initiative Launched: Ongoing since ~2016

Table 8. Applications of company: Siemens

Technology	Application
Digital twin	Virtual model of the entire turbine blade production line for simulation & optimization.
Additive manufacturing	3D printing of intricate cooling channels inside turbine blades using metal powders.
IoT sensors	Embedded in machines to monitor temperature, vibration, and real-time process data.
AI and machine learning	Optimizing tool wear, predictive maintenance, and defect detection in real time.
Cloud computing	All production data stored and analyzed across systems for global collaboration.

Table 9. Benefits and results

Metric	Before digitalization	After digitalization
Production time (per part)	~16 weeks	Reduced to ~6–8 weeks
Error rate/Defect detection	Manual inspection	AI-enabled vision reduces human error
Design-to-production time	Long cycle (months)	Digital twin allows rapid prototyping
Customization	Limited	Mass customization with 3D printing
Maintenance	Reactive	Predictive maintenance via IoT + AI

Siemens also applies digital twin simulation to test heat distribution and wear on refractory linings inside the turbine casting molds. This allows for:

- Better refractory design
- Faster failure prediction
- Reduced material waste

Untamed challenges concern the design of tailor-made properties in an interdisciplinary methodological way from the processing and shaping to the use of the final product. Some of the technological areas needed to reach the challenging demands on advanced ceramics will concern, e.g., low invasive processing and reduction of process-induced stresses such as cracking in manufacturing, environment friendly dry shaping and sintering approaches, refinement of properties on large components in changed design/construction-oriented technologies for ceramics and composites (Table 9).

CONCLUSION

Advances in refractory materials and precision manufacturing techniques for high-temperature applications, such as thermal protection systems, crucibles, and molds for metal casting, represent the integration and synergy of these two topics on three levels. The first level is the new composition of the material, which is an improvement to existing materials, advances in compositions, and classes of new materials. Without refractory materials, most high-temperature industrial processes would be unsafe, inefficient, and uneconomical. They are invisible yet essential components that protect both equipment and people, while enabling the innovation and sustainability goals of the modern digital era.

The second level is new precision manufacturing processes, which improve existing processes and allow for greater precision unachievable with other techniques or technology outside the past capabilities. The third level is the true synergy, which is taking advantages of new compositions together with new processes leading to materials and capabilities not possible before these technologies.

The fusion of advanced refractory materials with digitally enhanced precision manufacturing marks a significant leap forward in industrial capability. As we embrace automation, smart sensors, and AI, the materials and techniques used today are not only more efficient but also smarter and more sustainable. This convergence is essential to meet the demands of modern industry in sectors requiring extreme performance and flawless precision.

A significant contribution to the first level is thermal protection systems (TPS). These TPS are designed for hypersonic flight and have leading edges exposed to maximum temperatures exceeding 3000 °C. There is no currently existing material comparison with the TPS detailed in sections 1.2 and 4.2. These materials have a combination of composition, properties, and performance not found in existing products. There are similarities with existing materials composed of C/C composites with a SiC coating, e.g. LRSI, but these composites are unreinforced. Compared to unreinforced, there is better oxidation resistance in nonoxidizing environments. There is also better erosion resistance because the materials are harder and have a rigid fiber reinforcement which minimizes crack propagation. The TPS detailed in section 4.2 also have much lighter weights. These TPS represent a significant leap forward in capabilities.

Advances in material and composition, process and technique is a significant contribution in the crucibles and molds area. The materials and composition included new types of c-pyro and silica carbonaceous investment casting (IC) magmas, CO and C-O systems for various temperatures and calcined, low cost, high purity, high refractoriness, and local raw materials. The processes and techniques involved are new ways of shaping, modus of furnace heating, retaining salt molds, and retractable holding furnaces. These materials, new designs, and combinations in below twenty percent requirements for expensive high purity bonding agents are either none or of little non-compliance with the noted standards and regulations on silica dust contamination, and significant environmental friendliness. All these advancements can be applied to molding metals of up to 2200 °C melting temperatures. Furthermore, the same bonding agents spread outside the OC IC HS despair on the future of the IC technology is easily deployable. So far the IC mechanically fragile nature poses a great hurdle, but the metamorphosed technology can perform IC with manufacturable molds, and the work itself explores possibilities for new mold projects.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Beryani A, Moghaddam MRA, Tosco T, Bianco C, Hosseini SM, Kowsari E, Sethi R. Key factors affecting graphene oxide transport in saturated porous media. Sci Total Environ 2020;698:134224. [CrossRef]

[2] Kim S, Shin Y, Jung H, Hwang C, Baik H, Cha J. Precision and trueness of dental models manufactured with different 3-dimensional printing techniques. Am J Orthod Dentofacial Orthop 2018;153:144–153. [CrossRef]

[3] Bulut C, Kolca D, Tarlak F. Development of a software for detecting burn severity using convolutional neural network-based approach. Sigma J Eng Nat Sci 2023;43:598–606. [CrossRef]

[4] Harle S, Wankhade RL. Insights into the use of composites reinforced with natural fibers in architectural applications, with a focus on kenaf fiber. Sigma J Eng Nat Sci 2023;43:685–703.

[5] Ceylan I, Cakici Alp N. Insights into the use of composites reinforced with natural fibers in architectural applications, with a focus on kenaf fiber. Sigma J Eng Nat Sci 2023;43:665–684. [CrossRef]

[6] Khan I, Saeed K, Khan I. Nanoparticles: Properties, applications and toxicities. Arab J Chem 2019;12:908–931. [CrossRef]

[7] Mansour M, Al-hamdani K. Key Performance Indicators for Evaluating the Efficiency of Production Processes in Food Industry. Passer J Basic Appl Sci 2024;6:494–504. [CrossRef]

[8] Miao N, Sa B, Zhou J, Sun Z. Theoretical investigation on the transition-metal borides with Ta3B4-type structure: A class of hard and refractory materials. Comput Mater Sci 2011;50:1559–1566. [CrossRef]

[9] Liu K, Su Z, Miao S, Ma G, Zhang S. UV-curable enzymatic antibacterial waterborne polyurethane coating. Biochem Eng J 2016;113:107–113. [CrossRef]

[10] Mansour MM. Assessing the role of circular economy principles in reducing waste by sustainable manufacturing practices: A review. Sigma J Eng Nat Sci 2024;43:1886–1896. [CrossRef]

[11] Ceylan I. Insights into the use of composites reinforced with natural fibers in architectural applications, with a focus on kenaf fiber. Sigma J Eng Nat Sci 2025;43:665–684. [CrossRef]

[12] Mansour MM, Al-hamdani KS. Tabu search algorithm to optimize layout design for a multi objective plant function. Passer J 2024;6:446–452. [CrossRef]

[13] Page AE, Lawrence TD, Wong S, Zuo M, Wang W, Yu GM, Yan XQ, Liu HH. MBH-ODE: A dual-weight REN3D lattice Boltzmann model to simulate in-microregion particle-laden emulsions driven by switching magnetic fields. CHAOS 2015;25:041303.

[14] Salman HS, Mansour MM, Lafta AM, Shkarah AJ. Modification design and process of pipeline to reduce erosion rate and deposition. Int J Comput Methods Exp Meas 2024;12:165–173. [CrossRef]

[15] Mansour MM, Salman HS, Lafta AM, Nashee SR, Shkarah AJ. Simulation analysis of protection oil pipe in platform to reduce corrosion and erosion defects with sustainability techniques. Math Model Eng Probl 2024;11:1171–1178. [CrossRef]

[16] Mansour MM, Hamood HM, Lafta AM, Nashee SR, Shkarah AJ. Enhancing the efficacy of adsorption-based carbon storage systems: A finite element analysis approach. Int J Energy Prod Manag 2024;9:19–24. [CrossRef]

[17] Mansour MM, Lafta AM, Salman HS, Nashee SR, Shkarah AJ. Modeling a storage tank of carbon capture technology in a power plant in southern Iraq. J Emerg Sci Eng 2024;2:e13. [CrossRef]

[18] Najm N, Mansour MM. The role of waste reduction technology in sustainable recycling of waste paper at Thi-Qar University. Int J Sustain Dev Plan 2024;19:3153–3163. [CrossRef]

[19] Mansour MM, Ugla AA. Employing genetic algorithm to optimize manufacturing cells design. Acad J Manuf Eng 2024;22:3.

[20] Hamood HM, Mansour MM, Lafta AM, Nashee SR. Numerical investigation to study the effect of three height of triangular obstacles on heat transfer of nanofluids in a microchannel. Int Rev Mech Eng (IREME) 2023;17:533. [CrossRef]

[21] Abdulhasan MJ, Abdulaali HS, Al-Doori QL, Dakheel HS, Al-Abdan RH, Alhachami FR, Hameed AJ, Shoia SJ, Mansour MM. Physicochemical and heavy metal properties of soil samples in waste disposal site, Suq Al-Shyokh, Iraq. 2022 Int Symp Multidiscip Stud Innov Technol (ISMSIT) 2022. [CrossRef]

[22] Ibrahim ZA, Mansour MM, Lafta AM, Ugla AA. Numerical investigation to evaluate the extrusion process of power cable designed by CFD software. Int Rev Mech Eng (IREME) 2024;18:384. [CrossRef]

[23] Mansour MM, Erabee IK, Lafta AM. Comprehensive analysis of water-based emulsion drilling fluids in GHARRAF oil field in southern Iraq: Properties, specifications, and practical applications. Int J Comput Methods Exp Meas 2024;12:297–307. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.2048

Review Article

Systematic review and analysis on recent techniques of photo sketch matching

Vishakha SEHDEV VERMA[1,*], Nidhi MALIK[1], Anshul BHATIA[2]

[1]Department of Computer Science and Engineering, The NorthCap University, Gurugram, Haryana, 122017, India
[2]University School of Automation and Robotics, Guru Gobind Singh Indraprastha University, 110075, India

ARTICLE INFO

Article history
Received: 03 June 2025
Revised: 30 July 2025
Accepted: 10 October 2025

Keywords:
Deep Learning; Feature Extraction; Matching Algorithm; Photo Comparison; Sketch to Photo

ABSTRACT

The technique of aligning photographs with an artist's drawing is referred to as photo sketch matching. Matching the exact photo from the database is a big challenge, and the best technique must be selected. The focus of the study is to do an analysis of the techniques used in photo-sketch matching. The different techniques analysed in the study include Gabor Shape, Generative Adversarial Networks (GAN), Smart Switching Slime Mold Method, and Lightweight Vision Transformer. The review indicates that no single technique universally outperforms others across all conditions; rather, the choice of method depends heavily on the application context (e.g., forensic sketches, artist-drawn sketches, or automatically generated sketches), suggesting that combining the strengths of classical and deep learning methods, solving domain adaptation, and ensuring scalability for real-world deployment. The novelty of this review lies in the comprehensive synthesis and reorganization of existing research on photo–sketch matching. The work goes beyond by providing a critical comparison of techniques, identifying their strengths, limitations, and applicability, such as in forensics. The results show that in order to prevent overfitting, a large dataset is required for training. The intricacy of the issue requires an extensive set of forensic drawings. The accuracy is affected by multiple factors and can be improved by using hybrid techniques or deep learning approaches.

Cite this article as: Sehdev Verma V, Malik N, Bhatia A. Systematic review and analysis on recent techniques of photo sketch matching. Sigma J Eng Nat Sci 2026;44(2):1526–1539.

INTRODUCTION

Automation is important in modern, rapidly evolving, and technologically advanced societies. As human labor demands become increasingly intense, machines need to be trained to collaborate with people [1]. This would result in enhanced capabilities, quicker tasks, and greater productivity. Image processing is a key method for enhancing or refining the image [2]. Tools for machine learning have been created to reduce the level of effort needed to handle images. The automation of creating, synthesizing, and recognizing face sketches from photographs has been a prominent field of research in computer vision and image processing. In this survey, 178 publications were reviewed; 49 were chosen for deeper analysis, and research questions

*Corresponding author.
*E-mail address: vs4821@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

(RQs) were developed [3]. Furthermore, it details the criteria used for selection, which encompasses both inclusion and exclusion. The manuscript is prepared in sections. The primary subject of Section 2 is picture sketch matching. Section 3 outlines the methodology. Section 4 addresses the implementation and detailing of the various studies identified. Section 5 focuses on the reporting of the review, where all the research questions are answered, and in Section 6, the paper is concluded.

Photo Sketch Matching

The matching of photo sketches is a crucial part of criminal investigations. In video surveillance, face recognition helps identify a person for criminal investigations. A drawing is created with the assistance of a forensic artist after the witness provides a facial description of the offender [4]. Once the sketch is finished, it is compared to the agency's database. It is a difficult process because face matching data is collected from multiple sources and under different conditions. Verifying that the drawing and photo match is the aim of facial sketch recognition. Conventional uniform face recognition methods struggle with identifying faces from sketches because of the differences in modality between photographic images and sketch representations. This issue has been addressed using soft computing methods [5]. Previously, creating a face sketch was a tedious task where accuracy depended on the skill level of the artist.

Fortunately, with advancements in technology, face sketch recognition systems can now operate automatically [6]. Identifying a face through sketches begins with creating a hand-drawn sketch, which is then matched against images stored The image that best matches the sketch in the identification database can be used to determine the individual's identity [7]. More advanced preprocessing techniques cannot be applied to forensic sketches, as witnesses at the crime scene usually recall the outer features of the face rather than the inner details. Traits such as gender, age, and color are examples of superficial characteristics. Notable internal features are seldom documented [8]. The disparity in visual cues is what causes the modality gap. Facial features are extracted with the help of gabor filter. Features that are taken from coarse texture (facial shape) can greatly reduce the modality gap. Gabor filters attenuate fine texture first in the Gabor Shape framework. The Radon transform is then used to model the facial shape, which is represented by coarse texture. The experimental results using Gabor filter found to be superior thant the state-of-art approaches.

GAN is a very popular deep learning technique. A GAN learns from an existing dataset to produce new, realistic-looking data (such text, audio, or photos). For investigative purposes, Kokila R. offers a study on matching sketches to pictures [9]. The system's strong application focus contributed to its low level of complexity. Several sketches were compared to images of the face taken from different perspectives, yielding incredibly accurate results [10]. Because the technique was entirely reliant on the quality of the sketches utilized, low-quality sketches produced unsatisfactory results For the generative challenge, Sumit Gunjate suggested a network structure, and testing revealed that it outperformed existing arrangements. To improve the SMA's performance in particular domains, researchers suggested a number of modifications. The approch used the S2SMA to improve the accuacy [11].

A lightweight deep learning model called a Light Semantic Transformer was created to capture semantic (meaning-based) links between data pieces. It was inspired by the Transformer architecture. Another face sketch identification technique based on transfer learning was presented by Wan et al. They used the triple loss to learn discriminative features, created a three-channel CNN structure, and suggested a hard triple sample selection technique to speed up model convergence. To finish the target classification, ViT immediately stacked Transformer encoders that operated on picture patch sequences. On multiple image recognition benchmarks, ViT outperformed CNN networks [12]. Lin Cao introduced the LWVT-ResNet18 framework for the light semantic transformer network. The modules are trained from scratch which address the overfitting problem. The mdoel can directly the semantic features from photo and sketch which further improves the accuracy. FG-SBIR is a technique aims at retrieving a single picture from a gallery tailored to a specific category that corresponds with a query sketch. Descriptive tags, hierarchical co-attention mechanisms, self-supervised pre-training techniques, cross-category generalization methods, attention-based components featuring an advanced retrieval loss, and reinforcement learning approaches have all contributed to the gradual improvement of FG-SBIR. In general, they seek to learn a joint embedding space to minimize the cross-modal gap, usually by employing a triplet ranking loss, which seeks to distance a drawing from other images while bringing it closer to its matched photo [13]. CLIP allows for generalizability in downstream tasks without labels by representing text and images in the same embedding space. According to recent studies, CLIP's open-vocabulary generalization capability through an easy-to-use prompt-based design can improve SBIR's efficiency. The techniques are shown in Figure 1.

Diffusion models are emerging as a promising approach in photo–sketch matching due to their ability to generate high-quality and detailed images through iterative denoising. They provide stable optimization and capture fine-grained structural features essential for preserving facial identity across modalities. Recent advancements in diffusion-based methods have significantly enhanced photo–sketch matching, particularly in 2024 and 2025. These models have demonstrated improved fidelity and realism in generating face sketches from photos, addressing challenges such as the modality gap and data scarcity [14]. In 2024, jain et al. Introduced clip4sketch, a controllable latent diffusion model that utilizes text and image embeddings to generate diverse and realistic sketches. This approach aids

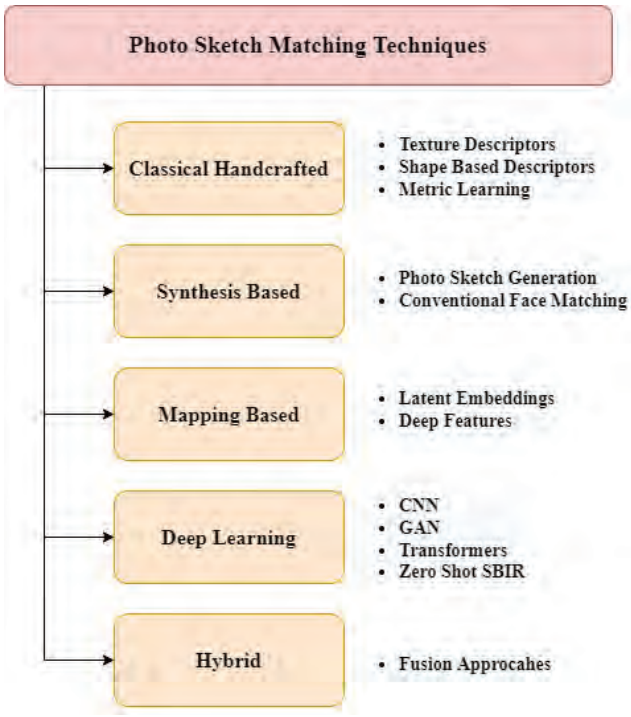


Figure 1. Techniques.

in augmenting datasets for sketch-to-mugshot matching, thereby enhancing the performance of face recognition systems. Additionally, a novel architecture for face sketch-to-photo synthesis was proposed, employing denoising diffusion probabilistic models (ddpms). The method simplifies the transformation process into sequential denoising steps, incorporating a pretrained coarse generator to encode sketch information, and a detail diffusion branch to refine the generated images. In 2025, wu et al. Developed one-shot face sketch synthesis in the wild, a method that employs generative diffusion priors and instruction tuning to convert face photos into realistic sketches using a single example. This approach addresses data scarcity by leveraging a new benchmark dataset, os-sketch, which includes diverse sketches and photos. These developments underscore the growing potential of diffusion models in bridging the modality gap in photo–sketch matching, offering more robust and realistic solutions for face recognition tasks.

The paper acknowledges the importance of ethical, legal, and practical limitations in forensic applications of photo–sketch matching. Ethical concerns arise from demographic and dataset biases, which can lead to unequal identification outcomes or misidentification of certain groups, raising issues of fairness and accountability. Legal considerations involve privacy, consent, and the admissibility of synthesized or generated images in judicial proceedings, which may vary across jurisdictions. Practical challenges include variability in sketch quality, differences in artist skill, inconsistencies between sketches and actual photographs, and environmental or contextual factors such as lighting and pose. Additionally, limitations in dataset diversity and the lack of standardized evaluation protocols can hinder the reliability and generalization of matching systems. These combined factors highlight the need for careful consideration when deploying photo–sketch matching technologies in real-world forensic scenarios.

MATERIALS AND METHODS

This section describes the methods that were utilized. Figure 2 illustrates the review methodology applied in this study. In the first phase, following the identification of search databases and preliminary studies, research questions were formulated. In the second step, articles were selected after organizing redundant and unnecessary research [15]. At

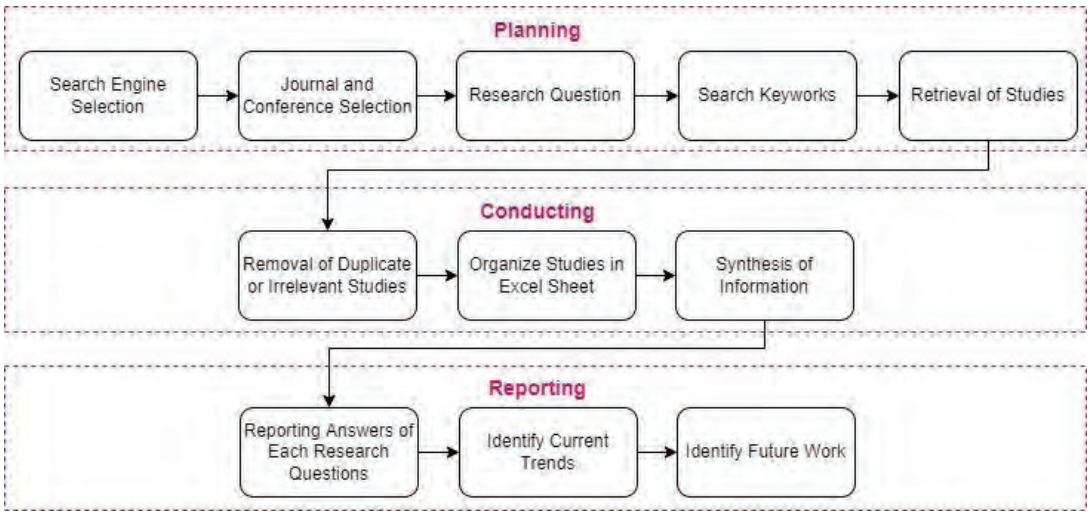


Figure 2. Review methodology.

this stage, synthesis was also conducted using the data collected from each of the shortlisted studies. In the final phase, responses to all of the research questions, as well as current trends and challenges in each topic, were provided.

The process of planning includes the following steps:

Selection of Search Database

The search process comprises two stages. In the initial phase, keywords like "Photo sketch matching" were entered to generate results. During this phase, resources such as Google Scholar, Scopus, Science Direct, Springer, and IEEE Explore were identified. The second stage involved locating articles in relevant journals and conferences.

Identification of Important Journals and Conferences

Prior to their use in research, the quality of journals is evaluated based on several criteria [16]. These criteria encompass journal citations and citation counts. Table 1 provides a summary of the study topics related to photo-to-sketch matching.

Formulation of Research Questions

After completion of the initial phase, the potential directions have been formulated for the subsequent phase. Properly defining study objectives is essential, as it will help to stay focused and work through the data. The goal of this work is to examine the existing methods and tools used for photo sketch matching [17]. From an initial pool of 178 articles, only those that specifically addressed are selected. This chapter's research aims to answer the identified research questions.

Search of Keywords

The aim of the review is to gather all relevant papers and ultimately present a comprehensive overview. During the initial search, various keywords like face sketch recognition, image sketch matching, and sketch-to-photo matching were examined [18]. After an in-depth analysis of the material, new keywords were incorporated to enrich the study.

Retrieval of Studies

The pertinent digital libraries were utilized to find all significant research. This method involved examining the reference sections of each study to identify additional relevant papers or publications, which were subsequently gathered and arranged for analysis [19].

ANALYSIS

The second phase of the analysis commences after the completion of the first stage. At this point, all research articles undergo a detailed review [20]. Research questions (RQs) form the foundation for specific criteria, which are then applied to each paper according to the standards outlined previously [21]. These articles have been utilized to assess the questions highlighted in the earlier phase. The limitations of each paper are identified, as this will help steer the research in the right direction.

Selection Criteria

The required documents are manually selected. Initially, the total number of papers that were carefully reviewed is reduced when the specified section of a document is evaluated [22]. In the second criterion, the entire group of papers selected under the first criterion is analyzed [23]. Two separate sets of selection criteria have been applied.

Inclusion Criteria

(i) Papers of reputed journals like Springer, IEEE, etc.
(ii) Papers that are focused on the topic of photo sketch matching.
(iii) Papers with recent publication.
(iv) Articles from SCOPUS and SCI publications.

Exclusion Criteria

(i) Articles of low-quality publication.
(ii) Articles only containing an abstract but not with full paper.
(iii) Articles without a dataset.
(iv) Unpublished articles without source information.

Based on the identified keywords, a total of 178 studies were initially gathered through various search engines. These studies were then analyzed according to specific inclusion and exclusion criteria. A total of 49 studies were selected, while those exhibiting low publication quality, lacking complete presentations, failing to cite the dataset, and those not belonging to either journal or conference categories were eliminated. In the subsequent stage, the remaining research was assessed against the quality evaluation guidelines to filter out irrelevant studies from the extensive pool of available data [24]. The metrics used for quality evaluation are presented in Table 2.

Table 1. List of research questions

S.No.	Research Questions
RQ1	What facial sketch mapping methods are commonly utilized?
RQ2	What kinds of datasets have been utilized?
RQ3	What challenges have image sketch recognition methods faced, and what future possibilities do these approaches hold?
RQ4	Which different performance metrics are used in the study?
RQ5	Which pre-processing methods have been employed thus far?

Table 2. List of quality assessment measures

S.No.	Quality assessment measures
QAM 1	Research objectives must be mentioned in detail.
QAM 2	It is important to identify the dataset that was utilized to implement the methods.
QAM 3	It is essential to evaluate various methods to determine their effectiveness.
QAM 4	A proper literature survey must be done.
QAM 5	Future scope must be present for applied techniques.

Data Extraction

A comprehensive database has been established that includes various fields such as the author's name, the article's title, the source of publication, the tools and techniques used, the dataset involved, and a concise summary of the paper's content [25]. To address all the research questions, an in-depth analysis of these articles has been performed. The chosen publications have been examined from multiple angles, assessing their methodologies and evaluating any connections among them [26]. The use of different visualization methods has been beneficial in analyzing technologies, tools, processes, and performance assessments. Multiple forms of visualization, such as bar graphs, pie charts, and column charts, are utilized for the analysis.

Distribution of Papers

The articles are organized in sections as per publication year and the source [27]. The papers can be divided into three categories according to their sources: journal articles, conference papers, and other formats such as technical reports, book chapters, or various study materials.

Year of Publication

A novel approach to identifying criminals has come to light: comparing photo sketches. The techniques previously used aimed to minimize the differences in modalities, while later research focused on the direct comparison of photographs and illustrations. Diverse approaches have been applied across different datasets. Figure 3 illustrates the breakdown of publications by source, Figure 4 shows the yearly distribution of studies, and Figure 5 outlines the process of searching for and selecting documents.

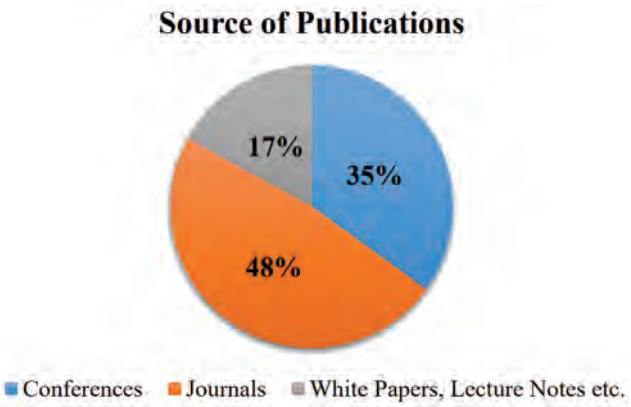


Figure 3. Publication sources.

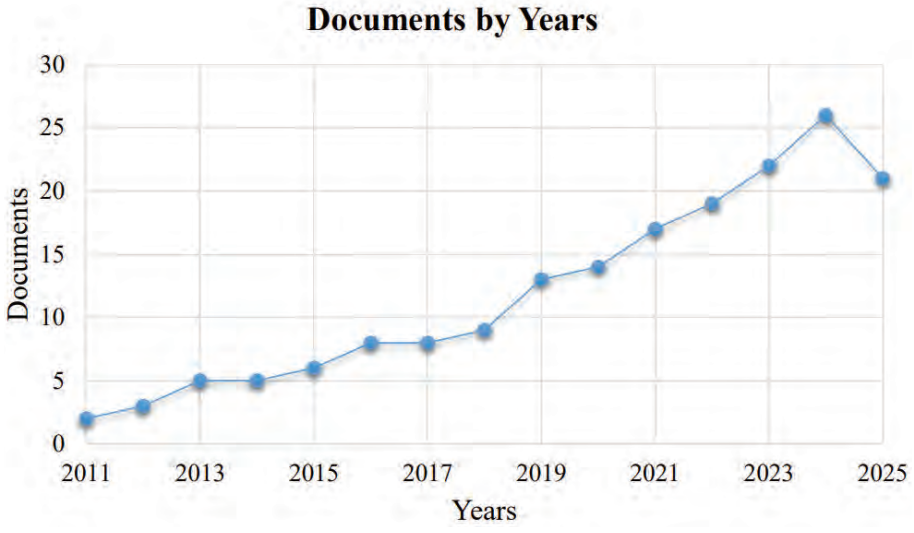


Figure 4. Documents by year.

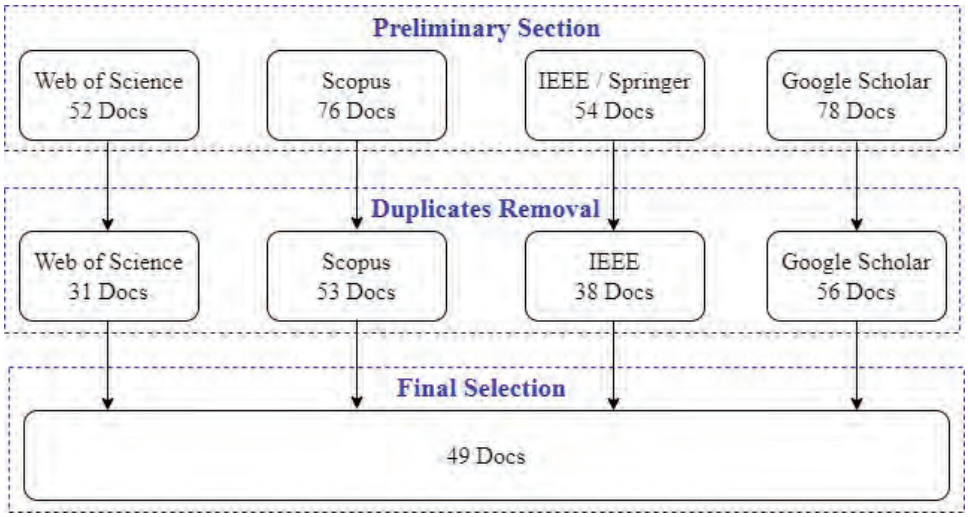


Figure 5. Selection process.

RESULTS AND DISCUSSION

In order to find the answers to the Research Questions (RQ), articles from different conferences and journals have been collected and analyzed.

RQ1. What Facial Sketch Mapping Methods Are Commonly Utilized?

Photo-sketch matching has been accomplished using different techniques. Figure 6 depicts several approaches to match a picture and sketch.

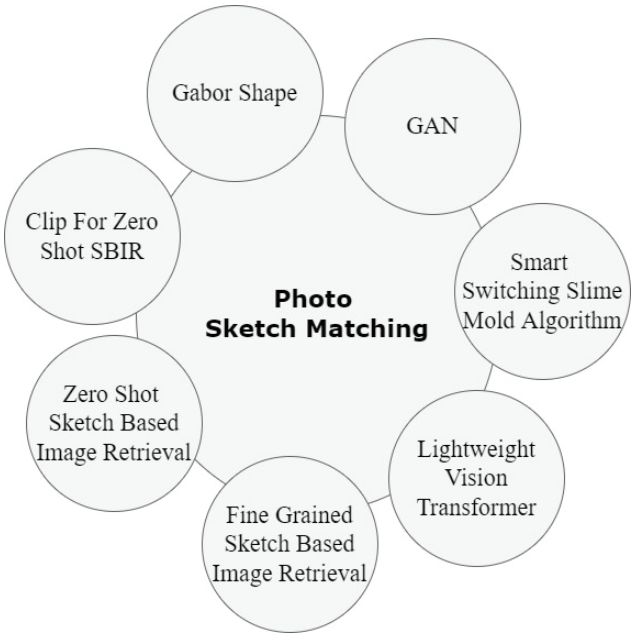


Figure 6. Different techniques for photo-sketch matching.

Gabor Shape

It functions in levels: Initially, Gabor filters are employed to soften the fine texture. In the subsequent phase, the visual signals associated with face shape are represented through a random transform as a face descriptor [28]. Gabor filters are commonly used for face feature extraction due to their remarkable ability to capture visual information, including spatial localization, frequency, and orientation [29]. This approach has been utilized in research to evaluate its effectiveness on the CUFSF and CUFS datasets.

GAN

In generative adversarial networks (GANs), two neural networks engage in competition using deep learning techniques to enhance prediction accuracy. A GAN consists of a generator and a discriminator. The discriminator features six convolutional layers, each equipped with different filters, and the final convolutional layer converts the output into a single dimension. Although it relies solely on the loss from the discriminator during training, it is associated with two distinct loss functions, and the failure of the generator is not taken into account. The discriminator is responsible for distinguishing between genuine and fake data produced by the generator. If a legitimate instance is mistakenly recognized as fake, or if a fake instance is incorrectly classified as real, the discriminator faces a penalty. The generator comprises encoder and decoder layers. The architecture utilized is U-Net. Research has been performed on various types of GANs, such as conditional, unconditional, cyclic, and style GANs. Figure 7 depicts the functioning of a GAN.

Smart Switching Slime Mold Algorithm

This is tackled using three distinct types of procedures. The first involves a synthesis-oriented approach. The second is a projection-focused method, utilizing soft

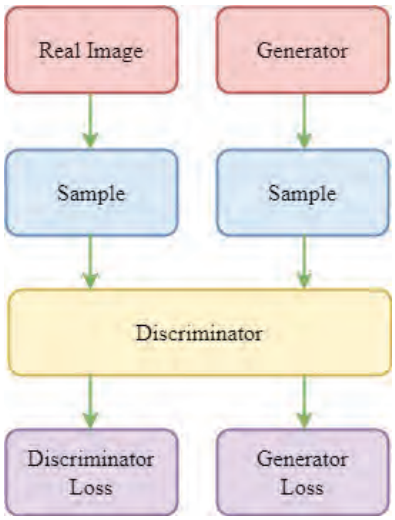


Figure 7. Working of GAN.

computing techniques to minimize discrepancies. The third category is based on optimization, utilizing various optimization methods for fine-tuning the parameters.

Lightweight Vision Transformer

It is a technique for categorizing images that utilizes a structure similar to that of a transformer to classify specific regions within an image. After pre-training on extensive datasets, the Vision Transformer (ViT) is fine-tuned for a smaller number of classes. Typically, classification is performed by adding an additional learnable classification token to the sequence. To capture the semantic features that enhance the precision of sketch face recognition, Convolutional Neural Networks are integrated with restricted LWVT.

Fine-Grained Sketch-Based Image Retrieval

The goal is to identify specific image using human's free-hand drawing. This topic has garnered significant research attention in recent years because of its commercial significance and the challenges it poses visually. A major obstacle lies in the gap between the realms of drawings and photographs. Images are perspective representations of real-world objects depicted as dense pixels, whereas sketches are well-known line drawings that are interpreted in a personal and abstract manner.

Zero-Shot Sketch-Based Image Retrieval

The primary issue lies in the disparity of information between drawings and photographs; sketches primarily illustrate the shape of an object, resulting in less informative content compared to photographs [30]. Another challenge is that individuals typically create sketches with varying levels of abstraction, creating significant differences within the same class of sketches [31]. ZS-SBIR can be understood as the process of generating supplementary information that is lacking in the sketch to create images that can be compared. The research mainly concentrated on a transformer-based cross-modal network, which extracts local patches.

Clip For Zero Shot SBIR

CLIP is known as Contrastive Language Image Pre-training. It conducts analysis, image processing, and action recognition using video content. It has a remarkable capability to comprehend unmatched multi-modal data and is responsible for providing a semantic space for category transfer. In this regard, the challenges are tackled through the implementation of regularization words and methods [32]. To ensure that the global parameter operates consistently across all of them, the former is necessary to standardize the distance between sketches and photos. In the latter case, after the sketches and photos are split into patches, a random permutation of these patches occurs.

Analysis

Overall, traditional Gabor features offer interpretability but fail on complex, noisy sketches, whereas GANs achieve superior sketch-to-photo synthesis at the cost of data hunger and training instability. Optimization strategies like SSSMA are effective in hybrid pipelines but lack standalone strength. Lightweight vision transformers strike a balance between accuracy and efficiency, making them suitable for real-time use. While both GAN-based methods and Vision Transformer (ViT)-based approaches achieve competitive results in image generation and recognition tasks, review studies indicate that their data requirements and computational costs are indeed comparable. GANs typically demand large-scale datasets and significant computational resources for stable training due to issues like mode collapse and adversarial optimization. Similarly, Vision Transformers require extensive training data and high computational budgets, particularly because of their quadratic self-attention mechanism. Comparative analyses in the literature suggest that, although the architectural challenges differ, both families of models exhibit similar statistical demands in terms of dataset scale and computational complexity, making them broadly equivalent in these aspects. FG-SBIR excels in fine-grained, instance-level retrieval but requires highly detailed sketches. Zero-Shot SBIR relaxes the need for paired data, though with reduced accuracy, and CLIP extends this to large-scale zero-shot retrieval with strong generalization but limited forensic specialization. Together, these techniques demonstrate complementary strengths, highlighting the need for hybrid, data-efficient, and generalizable approaches in future photo–sketch matching research.

RQ2. What Kinds of Datasets Have Been Utilized?

The comprehensive analysis utilizes several datasets. Table 3 presents datasets along with sketching method. Starting with the artist who produces the sketch, the face sketch recognition system evaluates the sketch against various images saved in the database. The objective is to identify the image that most closely resembles the drawing.

Table 3. Dataset and method of sketch

Dataset		Type of sketches	Method of sketch
CUFS	CUHK	Viewed	Hand-drawn sketch
	AR	Viewed	Hand-drawn sketch
	XM2VTS	Viewed	Hand-drawn sketch
CUFSF		Viewed	Hand-drawn sketch
UOM-SGFSV2		Forensic	Software-generated
E-PRIP		Viewed	Software-generated
IIIT-D		Semi-forensic	Hand-drawn sketch
QMUL-SHOE-V2		Viewed	Not known
QMUL-CHAIR-V2		Viewed	Not known
SKETCHY (EXTENDED)		Viewed	Not known
TU BERLIN		Viewed	Not known
QUICK DRAW (EXTENDED)		Viewed	Not known
COLORFERET		Semi-forensic	Not known
MULTIPLE ENCOUNTER		Viewed	Not known
WILD-A (LFW-A)		Viewed	Not known
FLICKR-FACES-HQ		Viewed	Not known

Various styles of sketching are available. These drawings are more precise and contain the most information about the subject. In the semi-forensic method, the artist is asked to create a drawing from memory after getting a brief chance to view a photograph of the subject. This technique, which is also employed in academic research, is considered more applicable to real-world situations than sketches made from observation since it factors in memory. The artist creates forensic illustrations based on these forensic sketches.

There are limited collections of images available for study, as most of these real-life examples originate from law enforcement departments. The greatest challenge lies in recognizing a brief glimpse of a face, often occurring under stress and influenced by the design of the sketch [33]. The IIIT-D database contains all three types of sketches analyzed in the SLR. It has 238 pairs of pictures created by a talented sketch artist, utilizing photographs sourced from various locations. Much of the research has been conducted on CUFS and CUHK. It features 295 images from XM2VTS, 123 from AR, and 188 from CUHK, totaling 606 images. These sketches were derived from artworks that were observed by an artist. Another relevant database is the EPRIP database. The AR Face dataset was used to create the images and composite drawings. In total, there are 123 pairs available. Four different groupings have been established for composite sketches drawn with various subjects.

Analysis

For face sketch matching, CUFS and CUFSF remain the most commonly used controlled datasets, though they lack real forensic variability. IIIT-D and UOM-SGFSV2 introduce semi-forensic sketches and greater diversity, while E-PRIP provides more realistic forensic-style data, making them stronger candidates for practical evaluation. In contrast, general SBIR datasets like QMUL-SHOE-V2, QMUL-CHAIR-V2, and SKETCHY (Extended) focus on fine-grained object retrieval across categories, enabling advances in zero-shot and cross-domain matching but offering limited relevance to facial sketches. Together, these datasets have advanced the field, yet their constraints in scale, diversity, and forensic authenticity emphasize the urgent need for large, varied, and real-world benchmarks to build robust and deployable systems. The paper acknowledges the importance of ethical and fair data representation. Dataset bias, including demographic imbalance and limited diversity, can reduce the robustness of models and hinder their applicability across varied populations. Such imbalances often result in unequal performance, where certain groups are represented accurately while others are misclassified or overlooked. Addressing these challenges is essential to ensure fairness, reliability, and broader acceptance of face sketch matching systems.

RQ3. What Challenges Have Image Sketch Recognition Methods Faced, and What Future Possibilities Do These Approaches Hold?

Different techniques have been employed to predict accuracy across various types of datasets, and these results have been compared with other methods to evaluate their effectiveness. To enhance the identification rate, some methods have incorporated optimization strategies. A significant challenge in photo-sketch matching is the modality gap, which has been extensively researched to reduce its impact. Another primary challenge in photo-sketch

matching is the lack of paired training samples for training the deep learning network [34]. Additionally, overfitting has been identified as a problem. When comparing the experimental results with earlier techniques, it was found that these approaches could benefit from being applied to larger datasets and more intricate models. Some methods require addressing hardware and software constraints to yield more precise outcomes in the future [35]. GAN and sketch-based image retrieval techniques are the most frequently utilized methods that have consistently demonstrated success in systematic reviews of research. Upcoming studies will explore alternative pre-training techniques that may enhance the convergence of the sketch-to-photo translation method, which currently relies on limited data. The possibility of generating photographs from intricate drawings derived from eyewitness testimony is another area that warrants exploration; in this scenario, the sketches could vary greatly from the actual photographs.

Analysis

Photo–sketch recognition faces persistent challenges such as limited and controlled datasets, loss of detail in hand-drawn sketches, and the large modality gap between photos and sketches. Deep learning methods like GANs and transformers perform well but require extensive data, high computation, and risk overfitting, while handcrafted and optimization-based methods struggle with scalability. Future progress depends on developing larger, diverse, and forensic-oriented datasets that capture real-world variability. Hybrid approaches combining interpretability with deep learning robustness, along with lightweight models for real-time use, are promising directions. Emerging paradigms like zero-shot and CLIP-based retrieval can reduce dataset dependency and enhance generalization, but they need tailoring for forensic applications. Overall, the focus must shift toward scalable, explainable, and deployable systems for real-world law enforcement and security use.

RQ4: Which Different Performance Metrics Are Used in the Study?

Various studies utilize indicators to evaluate the effectiveness of the techniques. These studies use a range of datasets, including CUFSF, CUHK, AR, XM2VTS, UOM-SGFSV2, QMUL-SHOE-V2, among others, to measure performance [36]. The use of these databases has been extensive in the research. Some of these databases are publicly available, whereas others are proprietary.

The following performance metrics are utilized:

Mean Recognition Rate

This refers to the accurately recognized images. The aim of the examination is to achieve consistency between the drawing and the photo. The formula used to determine this is (Accurately recognized images / Total images) * 100.

Cumulative Matching Characteristic (CMC)

It is a metric employed to assess the precision of identification and recognition algorithms across various rankings to gauge their effectiveness [37]. The comparative arrangement of match scores from every biometric sample is typically employed to evaluate identification performance in biometric systems. CMC utilizes values ranging from one to k to measure the effectiveness of identification on the x-axis. The y-axis represents identification rates corresponding to every rank.

Fitness Value Analysis

This measure is also utilized to compare the success of a specific algorithm against others. This comparison is performed by averaging the results of all algorithms. In the context of image sketch matching, the algorithm that achieves the highest value is considered the most effective.

Inception Score

It assesses the quality of the generated photographs. The score varies from 0 to 1, with 0 indicating the lowest quality and 1 indicating the highest. The Inception score primarily evaluates two main aspects: quality and variety. Quality refers to the image's overall standard, while variety pertains to the diversity of the images; the subject should be identifiable. This indicates the necessity for a broad range of images. The analysis metrics are illustrated in Table 4.

Analysis

Different studies in photo–sketch recognition have adopted varied performance metrics, each highlighting different aspects of model effectiveness. Mean Recognition Rate (MRR) remains the most widely used, offering a straightforward measure of overall matching accuracy but often overlooking rank-level insights. Cumulative Matching Characteristic (CMC) addresses this by

Table 4. Performance measures

S. No.	Name	Definition
1	Average Recognition Rate	This signifies the count of images that correctly align with the training photos
2	Cumulative Matching Feature	This represents standard evaluation metrics utilized for techniques in re-identification
3	Fitness Value	It is used to determine how closely a specific solution matches the optimal answer to the problem
4	Inception Score	This is applied to evaluate the caliber of images produced by GANs

evaluating performance across multiple ranks, making it more informative in retrieval-based tasks. Fitness Value is mainly employed in optimization-driven approaches, capturing how well parameters or features are tuned, though it lacks universal comparability across methods. Inception Score (IS), widely applied to generative models, assesses the quality and diversity of synthesized sketches or photos, yet does not always align with human visual judgment. Together, these metrics provide complementary evaluations, enabling a more comprehensive understanding of model performance across recognition, retrieval, optimization, and generation tasks.

RQ5: Which Pre-Processing Methods Have Been Employed thus Far?

So far, numerous pre-processing methods have been utilized for images. Preparing an image through preprocessing is essential as it enhances its features and readies it for further processing.

Face Alignment and Cropping

They are widely used preprocessing steps in photo–sketch recognition to standardize facial geometry. Alignment corrects pose variations by mapping key landmarks (eyes, nose, mouth) to a consistent reference, while cropping removes background clutter and focuses solely on the facial region. These steps reduce intra-class variation, enhance feature extraction, and improve recognition accuracy. By ensuring uniformity across samples, they make subsequent learning and matching more robust and reliable.

Grayscale Conversion

It is a fundamental preprocessing step in photo–sketch recognition that reduces images to intensity values, removing color information while retaining structural and textural details. Since sketches inherently lack color, this transformation minimizes the modality gap between photos and sketches. It also reduces computational complexity and memory requirements, allowing models to focus on shape, edges, and contrast, which are more relevant for recognition. By emphasizing structural features, grayscale conversion improves consistency and matching accuracy across modalities.

Noise Reduction and Edge Enhancement

They both are complementary preprocessing techniques in photo–sketch recognition. Noise reduction smooths unwanted distortions and artifacts, while edge enhancement emphasizes contours and structural details critical for distinguishing facial features. Together, they balance clarity and sharpness, ensuring that important sketch strokes or photo edges are preserved without distortion. This combination improves feature extraction, enhances robustness against poor sketch quality, and ultimately strengthens recognition accuracy.

Histogram Equalization

It is a contrast enhancement technique widely used in photo–sketch recognition to improve the visibility of facial features. By redistributing pixel intensity values, it enhances edge clarity and highlights key structural details that are often faint in sketches. This normalization reduces the impact of lighting variations across photos, ensuring more consistent inputs for recognition. As a result, models trained on equalized images achieve improved feature extraction and higher matching accuracy.

Scaling and Normalization

They are essential preprocessing steps to ensure uniformity across photo–sketch datasets. Scaling resizes images to a consistent dimension, eliminating variability caused by different resolutions, while normalization adjusts pixel intensity ranges for balanced feature representation. These steps reduce biases in model training, stabilize learning, and make comparisons across samples more reliable. By standardizing inputs, scaling and normalization enhance the robustness and accuracy of recognition systems.

Analysis

Pre-processing remains a vital step in improving the effectiveness of photo–sketch recognition systems. Face alignment and cropping standardize pose and scale, reducing variability and ensuring that only the facial region is analyzed. Grayscale conversion minimizes the modality gap by discarding color while preserving essential structural details, and histogram equalization enhances contrast to make features more visible under varying illumination. Noise reduction and edge enhancement further refine sketches and photos by suppressing distortions while highlighting critical contours. Scaling and normalization provide uniform input dimensions and balanced intensity ranges, enabling more reliable feature extraction and model training. Collectively, these techniques contribute to improved matching accuracy, though their effectiveness remains strongly dependent on the quality of both input images and preprocessing itself.

Outcomes

The comparison of techniques is shown in Table 5.

Euclidean distance is a widely used metric that quantifies similarity by computing the straight-line distance between feature vectors in a multidimensional space. Rank-k accuracy measures the likelihood that the correct match is retrieved within the top k results, with rank-1 accuracy denoting the strictest case where the correct match must appear first. Mean Average Precision (mAP) aggregates precision across varying recall levels to provide a comprehensive measure of retrieval effectiveness. Similarly, top-k accuracy evaluates the presence of the correct match within the top k predictions, while Precision quantifies the ratio of relevant items among the top k retrieved results, highlighting early retrieval performance.

Table 5. Comparison of techniques

S.No.	Technique	Features	Purpose	Dataset Used	Metrics
1	Gabor Shape	Handcrafted descriptors (edges, orientations, multi-scale filters)	Basic photo–sketch matching using low-level shape cues	CUHK Face Sketch, CUFSF, small object datasets	Accuracy evaluated with Euclidean/χ^2 distance and Rank-k accuracy.
2	GAN	Learned cross-modal representations + synthesized photo/ sketch	Reduce domain gap by generating photo-like or sketch-like data	CUHK Face Sketch, CUFSF, LFW-Sketch, IIIT-D	Rank-1 accuracy is used for accuracy measurement
3	Smart Switching Slime Mold Algorithm	Optimized feature selection / weight adjustment	Acts as an optimizer for hybrid features in sketch–photo pipelines	CUHK Face Sketch, CUFSF, LFW-Sketch, IIIT-D	Accuracy measured with Rank-1 and mAP.
4	Lightweight Vision Transformer	Learned embeddings via self-attention (compact transformer)	Efficient cross-modal embedding for sketch ↔ photo	CUHK Face Sketch, LFW-Sketch, QMUL FG-SBIR	Accuracy evaluated using Rank-1 Top-k accuracy, and mAP.
5	Fine Grained SBIR	Deep features + metric learning (triplet/ contrastive)	Instance-level retrieval (find the exact person/ object, not just category)	QMUL FG-SBIR (Shoes, Chairs, Handbags), CUHK	Accuracy measured using Rank-1, and mAP.
6	Zero-Shot SBIR	Semantic embeddings (attributes, word vectors, unseen class generalization)	Retrieve sketches/photos from unseen categories	Sketchy	Accuracy Evaluated with Rank-1 and Precision@k.
7	CLIP for Zero-Shot SBIR	Pretrained joint vision–language embeddings. Takes sketch as image input	Zero-shot cross-domain retrieval using CLIP space + prompts	Sketchy, domain-adapted LFW/ CUHK	Accuracy measured using Rank-1, and mAP.

In addition to improvements in methods and accuracy, fairness and reliability are still big problems in photo-sketch matching. The quality and variety of the training datasets have a big effect on how well most models work. A lack of diversity in the sample might cause bias, which can make results differ by age, gender, or ethnicity [38,39]. Deep learning and hybrid methods work well on benchmark datasets, but they don't always work as well on sketches with exaggerated or forensic-level aberrations [40]. Generative techniques like GANs or diffusion models, while effective in minimizing domain gaps, may produce artifacts or illusory details that undermine authenticity [41]. These difficulties pose challenges for practical applications, especially in forensic and security settings, where even modest inaccuracies can lead to substantial repercussions. In addition to technical progress, future research must prioritize fairness-aware training, comprehensive error analysis across diverse populations, and transparent evaluation processes. These kinds of activities will assist make sure that photo-to-sketch matching gets better not only in terms of accuracy but also in terms of how it is used in a responsible and ethical way.

The review indicates that no single technique universally outperforms others across all conditions; rather, the choice of method depends heavily on the application context (e.g., forensic sketches, artist-drawn sketches, or automatically generated sketches), suggesting that combining the strengths of classical and deep learning methods, acknowledging domain adaptation, and ensuring scalability for real-world deployment.

The novelty of this review lies in the comprehensive synthesis and reorganization of existing research on photo–sketch matching. The work goes beyond by providing a critical comparison of techniques, identifying their strengths, limitations, and applicability, such as in forensics.

The results show that in order to prevent overfitting, a large dataset is required for training. The intricacy of the issue requires an extensive set of forensic drawings. The accuracy is affected by multiple factors and can be improved by using hybrid techniques or deep learning approaches.

CONCLUSION

The systematic assessment of the work is done in the paper. The primary research questions to be explored are identifying the type of dataset used, examining the various techniques available for matching photo sketches, and discussing the challenges and potential applications of these methods. The review includes only conference papers and journal articles that propose research into photo matching.

The research summarizes all the key questions related to different methods of photo sketch matching using multiple datasets which are available. Though this topic has been explored by many researchers but still there is a gap in developing scalable and hybrid approaches that combine interpretability with deep learning, while addressing domain adaptation, robustness, and deployment in practical forensic scenarios. The process of transforming through deep learning is quicker and more efficient compared to earlier techniques. To avoid overfitting, it required a large training dataset, but sketch images sometimes missed facial characteristics, which could introduce significant noise. Future work needs to focus on building larger and more diverse benchmark datasets and developing hybrid models that combine classical interpretability with deep learning robustness. This is necessary to handle variations in sketch styles and real-world conditions, and can be achieved through domain adaptation, data augmentation, and scalable architectures designed for practical deployment.

Future research in photo–sketch matching prioritizes self-supervised learning models as promising paradigms to address data scarcity and modality gaps. Self-supervised learning can derive transferable, modality-invariant representations from large-scale unlabelled datasets, thereby enhancing cross-domain generalization. On the other hand, Diffusion models are also emerging as a promising approach in photo–sketch matching due to their ability to generate high-quality and detailed images through iterative denoising. At present, several researchers are developing algorithms aimed at matching illustrations and images. In the future, it may be possible to create an algorithm that integrates deep learning and artificial intelligence to match photos and sketches with greater accuracy by leveraging vast datasets. A feature extraction method can be employed to identify unique traits in drawings and photographs.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Malathy N, Navin G, Sriram S, Bala Subramabiyan S. A secure framework for multimedia transmission in medical images using DNA cryptography. Sigma J Eng Nat Sci 2025:222–233. [CrossRef]

[2] Li CX, Zhang D, Hu Z, Wu XJ. Modality fused class-proxy with knowledge distillation for zero-shot sketch-based image retrieval. IEEE Trans Circuits Syst Video Technol 2025;35:6158–6169. [CrossRef]

[3] Guo M, Xiong M, Huang J, Hu X, Peng T. Face photo-to-sketch portraits transformation via generation pipeline. Vis Comput 2025;41:1183–1196. [CrossRef]

[4] Bian H, Lv B, Guo Y, Zhang B, Du K. Sketch face recognition method based on local-global adapter. IEEE Access 2025;13:59244–59253. [CrossRef]

[5] Kumar D, Pandey RC, Mishra AK. A review of image features extraction techniques and their applications in image forensic. Multimed Tools Appl 2024;83:87801–87902. [CrossRef]

[6] Khattar P, Mishra S, Tanwar R, Verma A, Bhatia S. Decoding information: A dual modality approach for sign language recognition. In: Proceedings of the 2024 International Conference on Computing, Sciences and Communications (ICCSC); 2024. p. 1–5. [CrossRef]

[7] Gupta S, Adhikari S, Zaid MS, Verma A, Bhatia S. Advancements in facial image processing for human age and gender identification. In: Proceedings of the 2024 International Conference on Computing, Sciences and Communications (ICCSC); 2024. p. 1–5. [CrossRef]

[8] Sharma A, Bhatia S, Verma A. Weather monitoring and cloudburst prediction based on machine learning algorithms: An initiative towards disaster management. In: Advances in Disaster Management; 2024. p. 589–603. [CrossRef]

[9] Zhang X, Yang C, Li F, Li Y, Fu B, Wang S. Multi-scale feature extraction and aggregation network for electroencephalography classification in face photo-to-sketch recognition task. IEEE Trans Biomed Eng 2025;73:709–719. [CrossRef]

[10] Raj G, Verma A, Dalal P, Shukla AK, Garg P. Performance comparison of several LPWAN technologies for energy constrained IOT network. Int J Intell Syst Appl Eng 2023;11:150–158.

[11] Sezgin A. A complete study on and-product of soft sets. Sigma J Eng Nat Sci 2025;43:1–14. [CrossRef]

[12] Pattnaik I, Dev A, Mohapatra AK. CASK-Net fusion: Multi branch approach for cross-age sketch face recognition. Signal Process Image Commun 2025;138:117369. [CrossRef]

[13] Rodrigues AP, et al. Deep learning in forensic sketch analysis. In: Proceedings of the 2025 International Conference on Artificial Intelligence and Data Engineering (AIDE); 2025. p. 45–50. [CrossRef]
[14] Zhang T, Xie X, Du X, Xie H. Sketch-guided scene image generation with diffusion model. Comput Graph 2025;129:104226. [CrossRef]
[15] Kiral A. The effect of LRB stiffness changes with and without supplemental viscous dampers on seismic responses of an experimentally verified mdof building. Sigma J Eng Nat Sci 2025:301–315. [CrossRef]
[16] Setumin S, Aminudin MFC, Suandi SA. Canonical correlation analysis feature fusion with patch of interest: A dynamic local feature matching for face sketch image retrieval. IEEE Access 2020;8:137342–137355. [CrossRef]
[17] Deswal S, Verma A. Efficient routing protocol for IoT networks based on fog computing and routing protocol of low power lossy networks. Int J Internet Protoc Technol 2023;16:176–184. [CrossRef]
[18] Lu D, Chen Z, Liu C, Hu Y, Cai L, Wu QMJ. Few-shot facial sketch synthesis via progressive domain gap reduction. IEEE Trans Inf Forensics Secur 2025;20:7121–7136. [CrossRef]
[19] Verma A, Deswal S. FOG-RPL: Fog computing based routing protocol for IoT networks. Recent Adv Electr Electron Eng 2023;16. [CrossRef]
[20] Upreti K, Verma VS, Verma A, Vats P, Patil S, Kuwar V. Pandemic pulse: Unveiling insights with the global health tracker through AI and ML. In: Global Health Insights; 2024. p. 191–203. [CrossRef]
[21] Kapoor A, Kapoor S, Mishra K, Jain H, Upreti K, Verma A. Catalyzing security and efficiency: Blockchain's integration with IoT and cloud computing. In: Blockchain Integration; 2024. p. 457–467. [CrossRef]
[22] Pahuja S, Negi S, Verma A, Rathi P, Narang N, Chawla R. An authentication protocol for secure tag-reader communication. In: Proceedings of the 2012 IEEE Students' Conference on Electrical, Electronics and Computer Science; 2012. p. 1–4. [CrossRef]
[23] Upreti K, Verma A, Parashar J, Vats P, Verma A, Singh J. A comparative analysis of LSB & DCT based steganographic techniques: Confidentiality, contemporary state, and future challenges. In: Proceedings of the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I); 2023. p. 1581–1588. [CrossRef]
[24] Yacoob S, Sujatha B, Anudeep T, Sharvi BE, Sai GVVS, Antony KC. Face sketch construction and recognition. In: Proceedings of the 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI); 2025. p. 1151–1155. [CrossRef]
[25] Verma A, Deswal S. Comparative study of routing protocols for IoT networks. Recent Pat Eng 2023;17. [CrossRef]
[26] Upreti K, Poonia RC, Verma A, Divakaran P, Mittal S, Vats P, et al. Blockchain computing: Unveiling the benefits, overcoming difficulties, and exploring applications in decentralized ledger infrastructure. In: Blockchain Computing; 2024. p. 397–410. [CrossRef]
[27] Prasad JS, Verma A. Optimum path routing algorithm using ant colony optimisation to solve travelling salesman problem in wireless networks. Int J Wirel Mob Comput 2017;13:131. [CrossRef]
[28] Pahuja S, Kumar R, Verma A, Negi S, Arora H. Priority based approach against congestion in sensor network. In: Proceedings of the 2011 Annual IEEE India Conference; 2011. p. 1–4. [CrossRef]
[29] Khaliluzzaman M. Comparative analysis on real-time hand gesture and sign language recognition using convexity defects and YOLOv3. Sigma J Eng Nat Sci 2024;42:99–115. [CrossRef]
[30] Prasad JS, Verma A. Performance enhancement by efficient ant colony routing algorithm based on swarm intelligence in wireless sensor networks. Int J Wirel Mob Comput 2017;12:232. [CrossRef]
[31] Verma A, Vashist PC. Enhanced clustering ant colony routing algorithm based on swarm intelligence in wireless sensor network. In: Proceedings of the 2015 International Conference on Advances in Computer Engineering and Applications; 2015. p. 150–154. [CrossRef]
[32] Guo M, Xiong M, Huang J, Hu X, Peng T. Face photo-sketch portraits transformation via generation pipeline. Vis Comput 2025;41:1183–1196. [CrossRef]
[33] Pahuja S, Negi S, Verma A, Rathi P, Narang N. A novel intrusion preventive routing scheme for data dissemination in sensor network. In: Proceedings of the 2012 IEEE Students' Conference on Electrical, Electronics and Computer Science; 2012. p. 1–6. [CrossRef]
[34] Zhu M, Chen J, Wei X, Wang N, Gao X. Fine-detailed facial sketch-to-photo synthesis with detail-enhanced codebook priors. IEEE Trans Circuits Syst Video Technol 2025;36:1075–1088. [CrossRef]
[35] Tang D, Jiang X, Zhang Y, Dai Y, Lin Y. Ipdm: Identity preserving diffusion model for face sketch and photo synthesis. Mach Vis Appl 2025;36:34. [CrossRef]
[36] Upreti K, Verma A, Mittal S, Vats P, Haque M, Ali S. A novel framework for harnessing AI for evidence-based policymaking in e-governance using smart contracts. In: AI for E-Governance; 2023. p. 231–240. [CrossRef]
[37] Haque M, Kumar VV, Singh P, Goyal AA, Upreti K, Verma A. A systematic meta-analysis of blockchain technology for educational sector and its advancements towards education 4.0. Educ Inf Technol 2023;28:13841–13867. [CrossRef]

[38] Buolamwini J. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Proceedings of the 1st Conference on Fairness, Accountability and Transparency; 2018.

[39] Drozdowski P, Rathgeb C, Dantcheva A, Damer N, Busch C. Demographic bias in biometrics: A survey on an emerging challenge. IEEE Trans Technol Soc 2020;1:89–103. [CrossRef]

[40] Zhang W, Wang X, Tang X. Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of the CVPR 2011; 2011. p. 513–520. [CrossRef]

[41] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 4396–4405. [CrossRef]

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2049

Review Article

The impact of digital design on Façade performance: A literature review

Ayça AKKAN ÇAVDAR[1,2,*], Nilhan VURAL[3]

[1]*Graduate School of Natural and Applied Science, Karadeniz Technical University, Trabzon, 61080, Türkiye*
[2]*Department of Architecture, Recep Tayyip Erdogan University, Rize, 53100, Türkiye*
[3]*Department of Architecture, Karadeniz Technical University, Trabzon, 61080, Türkiye*

ARTICLE INFO

Article history
Received: 02 November 2024
Revised: 19 February 2025
Accepted: 06 March 2025

Keywords:
Digital Design; Energy Efficiently; Façade Elements; Façade Performance; Sustainable Buildings

ABSTRACT

Advances in technology and interdisciplinary collaborations have increased design efficiency through new materials and construction systems, and have supported innovations in the field of architecture. However, despite these developments, performance evaluation before construction remains crucial. Therefore, digital design methods and performance analysis software have become fundamental tools in the optimization process. This study examines the role of digital design in improving facade performance within the context of sustainable architecture. In this context, 96 studies obtained from the Web of Science (WoS) database were examined. Trends in the research were evaluated using bibliometric methods and content analysis applied via VOSviewer. The findings show that the majority of studies focus on daylight, energy efficiency, and solar radiation. In contrast, topics such as ventilation, indoor air quality, and acoustics were addressed to a more limited extent. In recent years, digital twin applications and artificial intelligence-supported facade optimization have gained more prominence. Simulation methods are dominant in the studies examined (65.6%). The most commonly used tools are Rhinoceros and Grasshopper. Only 58.3% of the studies received financial support. China stands out in terms of both the number of articles and the number of supported studies. This study presents a systematic classification for facade performance criteria and digital tools. It also draws attention to research gaps, particularly in the areas of ventilation and cooling performance. By highlighting these insufficiently addressed areas, the research differs from previous studies and offers a roadmap for future research towards developing digital facade design with a more comprehensive and multidisciplinary approach.

Cite this article as: Akkan Çavdar A, Vural N. The impact of digital design on Façade performance: A literature review. Sigma J Eng Nat Sci 2026;44(2):1540–1559.

INTRODUCTION

From the twentieth to the twenty-first century, technological advancements and the emergence of new digital tools have led to transformative changes in the field of architecture. Increased processing power and speed of personal computers, along with lower costs for tools such as building information modeling software, engineering analysis

***Corresponding author.**
*E-mail address: ayca.akkan@erdogan.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

systems, laser cutters, and CNC routers, have expanded the scope of influence for architects in this field [1]. Digital-based design technologies have significantly accelerated the adoption of unconventional architectural forms and production methods, including large-scale 3D printing [2]. This enables architects to play with complicated forms, while at the same time enhancing production and material efficiency. The precision of complicated designs increases aesthetic quality and construction efficiency [3].

Digital improvements also enable the development of responsive architecture, which transcends form with adaptive components that can respond to stimuli, internal and external. The combination of artificial intelligence and miniaturization enables buildings to respond more intelligently to the actions of the users and changes in the environment. This results in better sustainability and user experience [4]. Furthermore, the coupling of computational design with digital fabrication presents opportunities for more sustainable structures by reducing waste and optimizing resource use [5]. Buildings can be more energy efficient and more comfortable for the occupants using AI-driven adaptation in the design. This might be considered as a huge step towards smart architectural solutions.

Architects are increasingly using parametric design – where models are built on pre-existing relationships and rules – to make their designs more flexible. The parametric model, although useful for buildings, can be complicated and deterministic and thus makes the design process more difficult. Hence, architects frequently employ formal approaches of parametric design to build solutions [6]. To address these difficulties, the importance of promoting cooperation across disciplines. Regular communication and collaborative workshops between architects and engineers can contribute to a more amicable partnership between these players . This can result in a more holistic design process [7]. In addition, teaching both architects and engineers on parametric design tools helps improve communication. As stated by Huang et al. [8], transdisciplinary mutual learning can promote a collaborative atmosphere for teams to address the issues posed by parametric design collectively.

Facades are a prime example of how environmental parameters can be integrated into design through computational design. In this respect, facades play a significant role in increasing sustainability and efficiency. Facade panels significantly shape the appearance of a building by integrating structural, environmental, functional, and aesthetic elements [9]. These complex facade panels not only define the aesthetic character of the building but also benefit from technological advancements. Chang et al. [10] noted that the development of information technology enables the development of smart building envelopes which can actively control the thermal performance and energy consumption. According to Yaman [11] , different facade systems can be very successful in controlling the internal climate according to the outdoor weather conditions. This characteristic is significant for the comfort of buildings.

Gero and Sosa [12] say that computational tools enhance the capability of designers to solve complicated architectural challenges. Parametric processes let you quickly change a model based on pre-defined rules without the need to construct a new model for each iteration. In this way it improves efficiency in the design process [13].

Building envelopes not only play an aesthetic and structural role but also act as an environmental barrier to control heat, light and air. so they provide protection and help with energy efficiency. The building envelopes are at the origin of a basic interface which influences about 80% of the environmental performance of a building and contribute to the construction of buildings that may interact in an integrated manner with their environment [14]. They also incorporate passive tactics in the design, such as natural ventilation and the use of daylight, to enhance energy efficiency and indoor comfort. Well built facades can optimize solar gains, decreasing dependence on artificial systems [15–17]. Building orientation and massing are also crucial considerations that can help maximize these gains. This helps to a more sustainable and efficient design approach [18].

In this context, it is important to note that different performance criteria are considered in facades designed with digital design approaches. This review article aims to guide researchers by comprehensively examining and analyzing academic studies obtained from the Web of Science database that address facade performance through digital design. The goal of the study is to offer valuable insights into the existing literature in this field and to synthesize information. The originality of the study lies in the careful filtering and analysis of bibliometric data by the authors, resulting in the creation of detailed classification headings in this context.

Digital Design in Architecture and Façade

The emergence of computer-aided design tools has created a significant transformation in the integrated circuit industry. Designers who were familiar with certain design methodologies have to extend their knowledge now to the whole design sector from product specs to the manufacturing process [19]. Designs do not have to adapt necessarily to new technologies but rather technology has to adapt to the necessities of design production. It is not as passive a role as frequently imagined for designers in the digital design process. They can express their intuitive thoughts more effectively in a rational and computational environment [20].

Digital design methods can be used in the design of structures that are difficult to design or construct with traditional design methods. Furthermore, while traditional design limits designers to two-dimensional forms, digital design allows for four-dimensional modeling. In this context, the fourth dimension represents the time-dependent performance of the design [21].

Due to advancements in digital manufacturing technologies, industrial robots are increasingly used in architecture. Changes in material approaches accompanying these developments also significantly impact architectural applications. Industrial robots are particularly noteworthy in projects that prioritize aesthetic and structural composition, such as pavilion structures [22]. Following the shift from stationary robot manipulators to mobile alternatives, the use of aerial robotic manipulators, commonly known as drones, has become increasingly visible. Renowned robotics expert Raffaello D'Andrea, in collaboration with the Swiss architectural firm Gramazio Kohler Architects, spearheaded the construction of a tower from polystyrene brick blocks [23]. CNC machines, on the other hand, have the capacity to produce a wide variety of textures on flat surfaces. The specific geometric properties of these textures are determined by the designer.

Graiz and Van de Rie [24] investigated the effect of different textured surfaces on solar control in facade panels. Their work showed that textures created with spherical end milling tools could provide a remarkable 32% reduction in solar radiation compared to traditional flat surfaces. Various software options are available that offer tools for applying parametric techniques. Computer-aided design software can be divided into two types according to their role in the design process: software that assists in the visualization and implementation of the design, and software that supports the designer throughout the entire design phase [25].

Façades, considered as complex building components, bring together structural, environmental, functional, and aesthetic parameters in a holistic manner. This holistic approach is important because the façade significantly shapes the overall image of the building [26]. The increasing energy footprint of commercial buildings, especially office buildings, has led to a shift from approaches focusing solely on non-Euclidean geometries to performance-oriented approaches. This change is also seen in the transformation of Building Information Modeling (BIM) models, initially parametric, into adaptive models that take performance and other analytical data into account [27].

Given that digital design primarily focuses on formal parameters, the façade is the first focal point in capturing people's visual perception of a building's design. Beyond their visual importance, façades also function as a shield protecting the building from external environmental conditions. Therefore, the selection of façade materials and designs is of great importance because these choices naturally affect performance characteristics.

Insulation has a great impact on energy performance. In a research performed in a residential building in Konya, the insulation application raised the energy class from level E, i.e. undesirable level, to level C, i.e. acceptable level. This data suggests that the design of the façade directly impacts the energy efficiency. Digital design methodologies can lead to greater optimization of façade layouts, improvement of insulation performance and contribution to sustainability [28]. Besides, buildings need to be renovated or strengthened in view of wear and tear over a period of time or damage from natural disasters. Such interventions. The key issues in terms of user's health, efficiency and overall comfort are to ensure interior comfort with minimum energy consumption and exterior noise [29].

In this context, Digital Twin technology can advance this optimization process by providing a virtual representation of physical structures. Digital Twin helps optimize designs and resource management by simulating the physical world in a virtual environment; it supports the modernization of industrial structures and the digitalization of the construction sector [30]. In architecture, Digital Twin technology is widely used in design and renovation projects. It plays an important role in improving project implementation during the construction process. For example, Digital Twin is used in examining the structural integrity of historical masonry structures and enables improvement studies. It also contributes to the production of prefabricated concrete elements, supporting the development of modular construction and innovative solutions [31]. Digital Twin is also used in architecture for façade segmentation, preservation of cultural heritage, and increasing the energy efficiency of old buildings. It collects data through sensors and analyzes this data to offer solutions for better building performance. In addition, it helps predict safety risks and detect structural damage in prefabricated structures. Digital Twin technology is expected to play an important role in energy-efficient design, durability and prefabrication processes [32].

Façade Performance and Digital Design Relationship

The selection of façade systems is of critical importance in the architectural design process due to its impact on environmental comfort and spatial use. Façades, which must protect the building from external environmental conditions while also exhibiting specific performance characteristics, have increasingly become standardized through consistent design practices. Unlike traditional façade designs, performance-based façade systems can be dynamic in nature and can provide solutions to design problems through computational design [33]. In addition, parametric design calculations can be used to perform kinetic movement analyses during the pre-construction design phase; thus, it is possible to propose configurations that optimize daylight performance [34].

Elements designed to provide daylight control, such as façade-integrated shading systems, play an important role in ensuring indoor comfort [35]. It is of great importance that these elements, which interact with external weather conditions, are wind-resistant and flexible. Using computational design methods, Baehr-Bruyère et al. [36] highlighted the usability of glass fiber-reinforced polymer (GFRP) materials as exterior shading elements due to their

resistance to rain and UV radiation and their ability to diffuse daylight into the interior (Fig. 1).

Velasco et al. [37], designed a dynamic façade with two free rotational directions, unit slabs on daylight performance. The façade was built by a computational design process utilizing the applications Grasshopper and Rhinoceros. The simulation analysis using simulation tools such as Honeybee, LadyBug and Radiance shows that this facade design distributes light more uniformly than conventional facade systems. The researchers claim that the logic of group-based behavior (Fig. 2) can improve the experimental investigation of a single panel.

Karagianni et al. [35] used parametric modeling in the sun shading system they designed as part of their master's thesis. The daylight performance of the system was simulated using different geometric configurations. The study also emphasized the significance of the relationship between the stages of parametric design, manufacturing, and performance control.

Freitas et al. [38] examined the use of Building Integrated Photovoltaic (BIPV) Systems to retrofit buildings in seven office buildings using Rhinoceros CAD software. Furthermore, the system's suitability was assessed by running performance analyses with the Grasshopper and Ladybug plug-ins. Among the Building Information Modeling (BIM) systems, software that allows architects and architecture students to integrate different systems such as solar energy generating systems into the design is beneficial (Fig. 3).

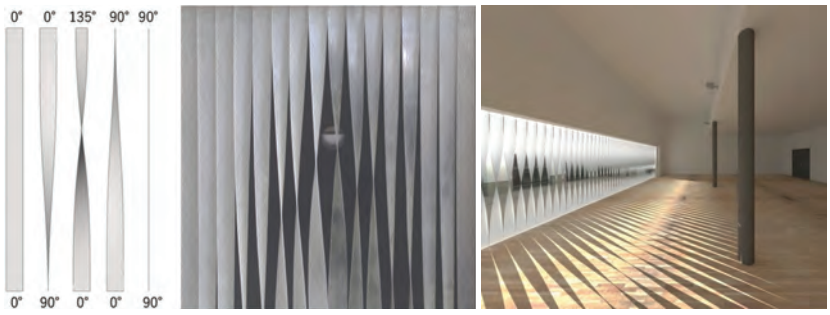


Figure 1. GFRP kinetic shading system and simulation [36] [recreacted by Cavdar and Vural].

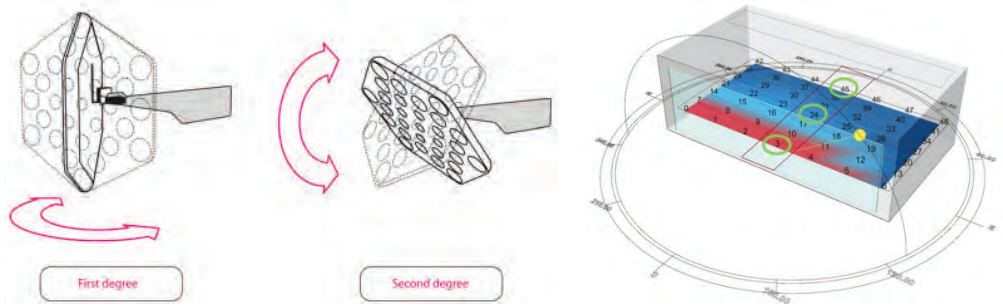


Figure 2. Unit plates designed with two degrees of rotational freedom and simulation model [37] [recreacted by Cavdar and Vural].

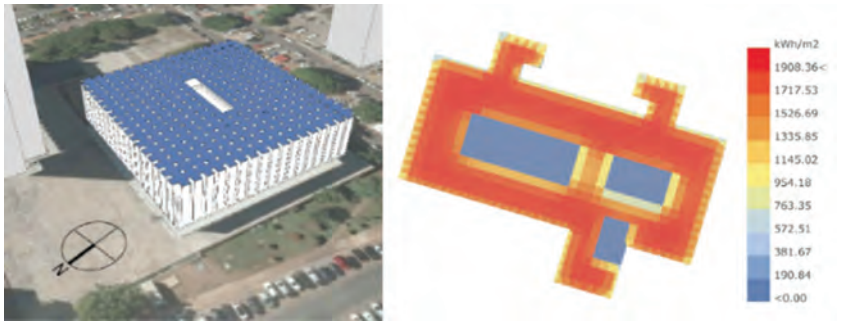


Figure 3. A model created using Rhinoceros CAD, and a sample radiation diagram from the roof [38] [recreacted by Cavdar and Vural].

Sravani et al. [39] conducted a BIM-based LCA analysis on a reinforced concrete traditional residential building in India's hot-humid climate zone and discovered that the roof slab consumes more energy than the façade. Furthermore, the wall surfaces made of clay tiles, cool tiles, terrazzo tiles, and aerated concrete blocks added to the roof to provide insulation reduced carbon dioxide emissions by approximately 2-6%.

Koukelli et al. [40] reduced reflected solar radiation by 40% using a façade design with optimized faceted geometry and diagonal arrangement; it was discovered that the radiation has the potential to reduce the microclimate, particularly in areas close to the façade surface. With the proposed façade design, the cooling load could be reduced by 21%.

Marques and Eloy [41], used a shape grammar to create a pattern inspired by the microscopic structure of the fungus. The pattern was reinterpreted in this direction, and a two-layer system was created as a façade carrier and infill section. It is emphasized that this modular system can realize various designs and achieve various harmonies.

Biomimetic façades are inspired by organisms that are well adapted to Earth's conditions and are thus thought to be effective in terms of improving performance. Mathematical modeling and digital simulation methods were used in a study to test the potential of a biomimetic façade to reduce energy demand. The tests on buildings with different climatic zones and functions (school, office, and elderly care home) revealed that the elderly care homes had the greatest difference in terms of energy consumption, with a 67.1% reduction in energy consumption. The tropical climate had the greatest energy reduction (55.4%) among tropical, humid continental, temperate oceanic, and hot desert climates [42].

A texture module with natural ventilation through a single thermal zone (classroom) was designed and applied to the window of an existing building via computer in a study to develop biomimetic adaptive building skins (Bio-ABS). The 188 modules that make up the window system have varying aperture and transparency ratios. Bio-ABS provided thermal comfort with a 23.18% reduction in discomfort hours and a 5.09% reduction in adaptive thermal comfort as a result of thermal analysis using Energy Plus software [43].

Numerous researchers employ digital design for enhancing façade aesthetics and form-finding [41,44,45]; while some focus on leveraging daylighting and associated energy performance to generate façade forms [27,46,47]. Less frequently, studies delve into ventilation efficiency [33,48–50], and acoustic behavior [51,52]. In recent times, the integration of building façades with emerging technologies like artificial intelligence, machine learning, digital twin, and IoT has become notably pronounced [53–55].

Bibliometric Analysis

Bibliometric methods primarily have a quantitative nature, yet their main objective is to draw conclusions about qualitative aspects. Their fundamental purpose is to convert intangible qualities, such as scientific quality, into measurable entities. Consequently, bibliometrics provides us with a scalable tool that can be applied at various levels to the global scale [56]. One of their aims is to understand the overall research trends in a specific field by examining academic publications from databases such as Scopus or WoS [57]. In addition, bibliometrics provides researchers with the freedom to select the relevant variables and quantitative methodologies that best align with their study objectives [58]. The most common bibliometric parameters are the number of publications and number of citations [59]. Kuzminska et al. [60] concluded through surveys and interviews that bibliometric analysis is useful for early career researchers and that bibliometric analysis is not limited to a specific field.

Özdemir and Arslan Selçuk [61] identified 461 articles published in different journals over the last 10 years in the Web of Science database. It was determined that these studies mainly focused on six keywords, particularly the concepts of "machine learning" and "geometric pattern". Notably, the field of architecture received the highest number of citations, with 365 citations. The study reveals that artificial intelligence technology can be further developed in the field of architecture. It is expected that integrating this technology with the discipline of architecture will yield significant and promising results.

In a separate bibliometric study, the Web of Science database was used to analyze scientific research on digital design in the field of architecture. The analysis results revealed that digital design has shown continuous development in architecture over the years; this highlights the potential of the topic for future research and its integration with other disciplines [62]. Eltaweel and Su [63] conducted a literature review focusing on the application of parametric design in architecture, particularly in daylight and solar radiation applications. The aim of the study was to demonstrate how this approach can significantly improve access to natural light and contribute to energy efficiency.

When bibliometric analysis studies are examined, VOSviewer-based mapping techniques are widely used. These studies generally address scientific research through broader categories such as year, university, institution, author, journal, citation count, and keywords. However, a comprehensive review of these studies is lacking in the existing literature; important details such as the subject matter and the software used are often omitted.

This study stands out not only for its examination of articles obtained through the mapping method but also for its comprehensive analysis conducted within the scope of author-defined classifications. The results of the study are expected to provide valuable contributions to researchers working in this field.

Background of The Study

With the development of computer technologies and the integration of computational methods into architecture,

it has become possible to realize complex designs that go beyond the possibilities of traditional methods. Façades, being the exposed surfaces of the structures, are one of the key construction elements that could leverage such a computational design method. The performance of façades designed using computational methods is of great importance due to their role in protecting buildings from external influences.

The main objective of this study is to obtain findings from research that primarily focuses on façades designed using computational methods and emphasizes façade performance. By analyzing these findings, the study aims to identify performance criteria that have been relatively underemphasized in computational façade design studies, despite the widespread use of these methods today. Thus, it provides a guiding and inspiring framework for researchers who wish to work in this field. This study offers unique value by examining scientific articles within this context and providing guidance for both researchers and designers.

The hypotheses formulated in alignment with the study's objectives are as follows:

- Digital design is frequently encountered not only for visual façade design but also in studies aimed at improving façade performance.
- Many different performance criteria can be taken as basis for façades designed with digital design method.
- The utilization of digital design techniques in studies pertaining to façade performance has exhibited a consistent rise over the years.

MATERIALS AND METHODS

Ensuring optimum building comfort conditions largely depends on the performance of façades. In this context, numerous studies proposing innovative technological solutions utilize digital design approaches. This study examines digital design approaches in relation to façade performance; it aims to identify common performance criteria highlighted in the literature and to analyze current research trends in this field. To this end, selected academic studies within a defined scope were analyzed using VOSviewer and AntConc software.

The Process of Searching

Keyword search has the great advantage of being easy to use because users don't have to learn a sophisticated query language. The user pleasant keyword search provides the user the ability to simply perform queries, hence increasing the overall user experience and allowing the user to quickly access data [64]. Web of Science (WoS) is the first comprehensive international bibliographic database that has become one of the most successful bibliographic data sources typically used for journal selection, research appraisal, and bibliometric analysis [65]. In this study, a search was conducted in the WoS database using the keywords "digital design" and "façade performance", and a total of 96 relevant scientific studies were identified as of May 2023. The main objective was to draw conclusions regarding which performance characteristics these façade studies aim to improve through digital design methods. The obtained academic studies were organized according to their citation counts in the Web of Science database and presented in tables for further analysis.

The Process of Analysis

The obtained studies were initially analyzed using the VOSviewer software, which stands out as one of the most favored tools for bibliometric analysis due to its capability to import data from diverse sources and offer effective visualization [66]. Thus, firstly, the countries where the studies were conducted, authors, keywords, and their interrelationships, both with each other and with citations, were analyzed.

As part of the study, a closer examination was conducted on the top 10 most cited studies to ascertain the predominant research topic. After scrutinizing all 96 articles, classifications were established for performance types, study issues, and software/tools used, and corresponding markings were applied in the tables. The AntConc software was employed to analyze all the information gathered in the tables and facilitate the counting process. AntConc is a corpus analysis toolkit designed for specific use in the classroom. It includes various features that aid in the analysis of linguistic data [67] (Fig. 4).

The collected studies were subjected to bibliometric tabulation and were further analyzed through content

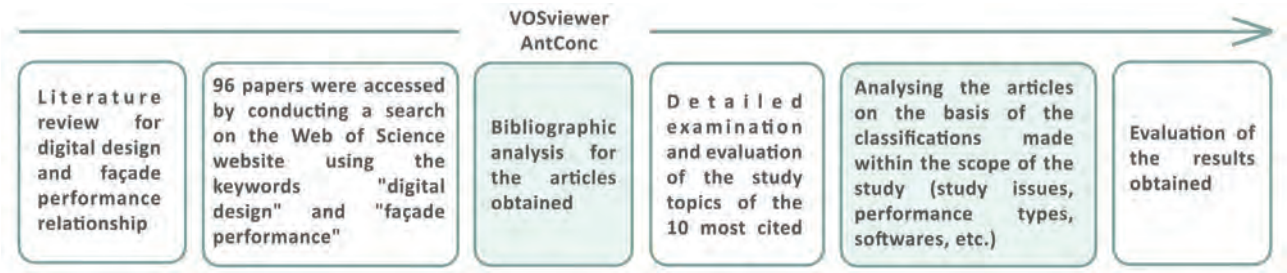


Figure 4. Flow chart of the study.

Table 1. Classifications created in the light of the data obtained by the authors within the scope of the study

	Type of Study Issues	Type of the Performance	Used Software or Tools
1	Adaptive façade	Daylight	Rhinoceros
2	Materials	Energy	Grasshopper
3	Façade	Solar radiation	EnergyPlus
4	Biomimicry	Digital tools	Revit
5	City scale	Geometric quality	Ladybug
6	Windows	Thermal	Radiance
7	Digital twin	Mechanic	Honeybee
8	Prefabricated façades	Sustainability	Diva
9	Sustainable (Green)	Lighting	DAYSIM
10	Lifecycle	Acoustic	Python
11	Pixels	Air quality	ANSYS
12	Building automation	Air temperature	Ecotect
13	Joint system	Humidity	MATLAB
14	Knowledge based engineering	CO2	Abaqus
15	Multi agent systems	Biological growth resistant	Arduino
16	Presentation	Chemical resistance	Autocad
17	PV	Construction	3DSmax
18	Shading systems	Cooling	Galapagos
19	Active noise control		Cyber
20	Active solar system		Façade
21	Automated blinds and dimmable lighting system		Geo-visualization
22	BIM		Kinetic
23	Box counting		Technical implementation
24	Bridge		Visualization

examination within specific classifications. Table 1 displays all subheadings falling within the scope of the classifications created for these analyses.

RESULTS AND DISCUSSION

In this study, 60.4% (n:58) of the 96 studies obtained are articles and 39.6% (n:38) are proceeding papers, which examines numerical design through academic studies on façade performance.

Based on the classification of methods employed in the articles under study, it is observed that 15.6% (n:15) of the studies utilized reviews, 47.9% (n:46) employed in situ/experimentation, and 65.6% (n:63) relied on simulation methods. Notably, in 32.3% (n:31) of the studies, a combination of experimental and simulation methods was utilized. Importantly, it's noteworthy that numerical design methods commonly utilized in studies concerning façade performance are primarily simulation-based.

The top five most used words in titles and abstracts of academic studies are 'façade' (n:47/n:276), 'building' (n:38/n:231), 'design' (n:29/n:247), 'system' (n:23/n:134), and 'performance' (n:16/n:127). In particular, terms in the names such as 'energy' (n:9), 'solar/sun' (n:9) and 'daylight' (n:8) are often used as performance indicators. Moreover, a performance keyword analysis in the abstracts of the summaries showed a predominance of the words 'energy' (n:124), 'solar' (n:43) and 'lighting' (n:36). This comment points out the importance of façades planned according to the sun and the favor of energy-performance studies done with similar considerations.

Analysis of the study areas classified in the WoS database reveals that the top five research area categories are as follows: "engineering" with 39.6% (n=38), "construction technology" with 35.4% (n=34), "architecture" with 20.8% (n=20), "science and technology other topics" with 20.8% (n=20), and "computer science" with 18.8% (n=18).

Only 58.3% (n=56) of the studies received financial support. The National Natural Science Foundation of China (NSFC) ranked first, funding approximately 8% (n=7) of the research. In terms of the number of studies, the People's Republic of China ranked first with 15.6% (n = 15). China was followed by Australia (n = 12), the United States (n=12), the United Kingdom (n=11), Italy (n=9), and Germany (n=8).

China appears to lead in both the number of studies and the level of financial support received. However, it is noteworthy that Australia, the USA, and Italy are among the top five countries in terms of the number of studies despite receiving relatively less financial support. In contrast, it is unexpected that China, which has the highest number of articles, ranks fifth with 59 citations (Fig. 5).

Figure 6 illustrates a robust interconnectedness among five articles [43,68–71]. These studies primarily delve into thermal performance, energy efficiency, and sustainability. The subjects of the studies encompass biomimicry, adaptive

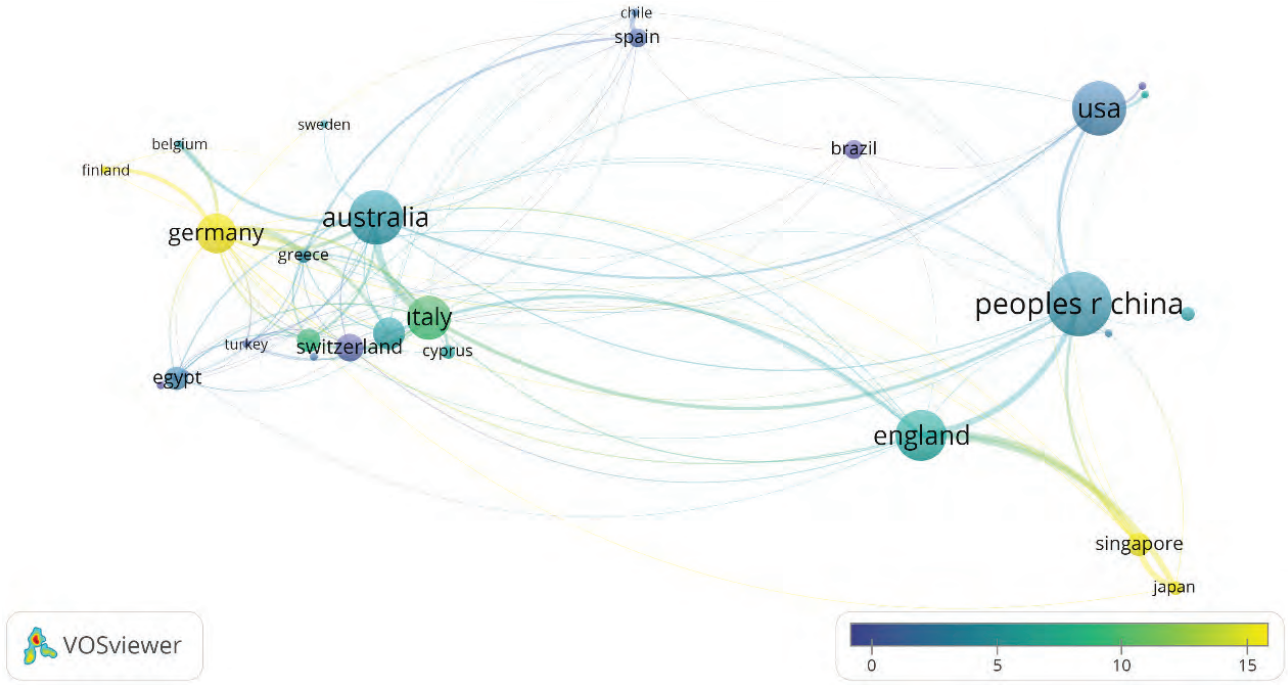


Figure 5. The relationship between the number of articles of countries and citation rate (circle sizes: document; color chart: citations).

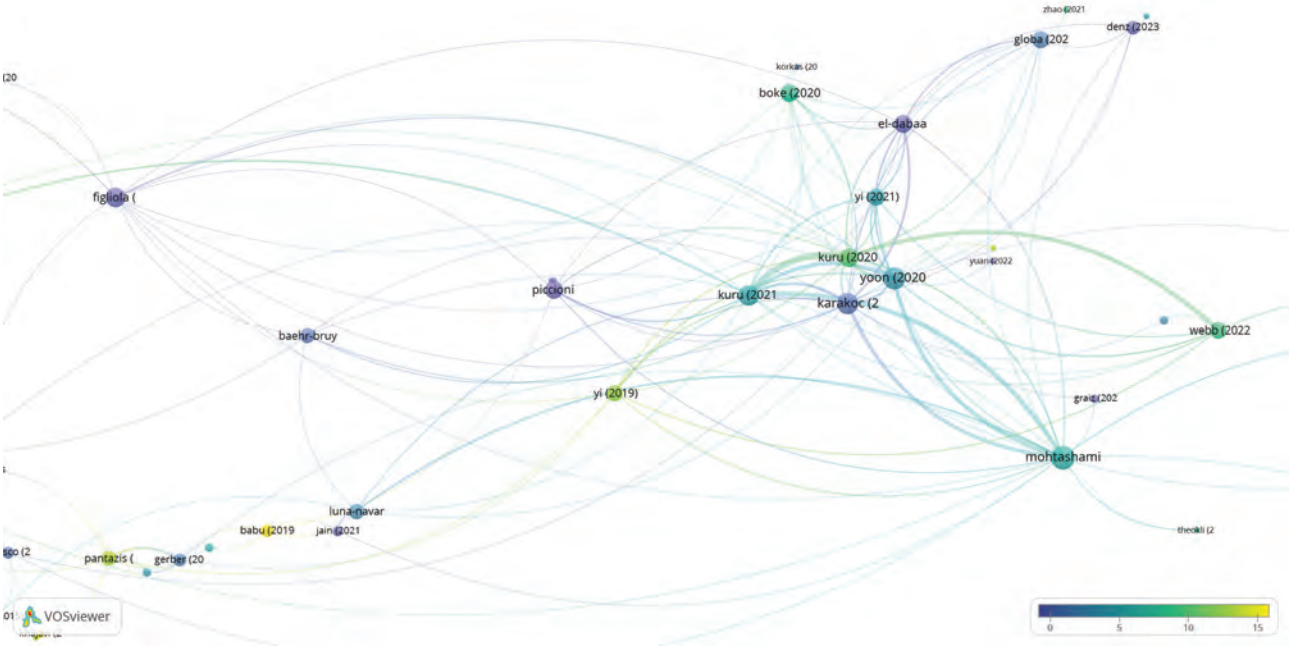


Figure 6. Map of document links by citations (circle sizes: linked documents; color chart: citations).

façades, and materials. Notably, it is observed that these closely connected studies have been predominantly conducted in recent years.

The acquired documents were thoroughly examined within the framework of the classification of 'study issues' and 'performance types' established by the researchers. Subsequently, content analysis was used to identify gaps in the literature and present them.

Analysis: Most Cited Articles

Out of the 96 studies, this article focused on analyzing the 11 most cited ones based on their citations. Table 2

Table 2. Most cited articles and citations [72] [created by author]

	Source	Year	Citation	Journal	Conference	Review	Experi. In-situ	Simula.	Software	Study Issue	Performance type
1	[1]	2019	129	●		-	•	-	Python Minitab Excel	Digital twin BIM	Daylight Air Temperature Humidity
2	[2]	2021	32	●		•	-	-	-	Active Noise Control	Acoustic (Noise)
3	[3]	2008	30	●		-	•	-	-	Shading systems	Daylight
4	[4]	2019	18	●		-	•	•	Radiance Ecotect EnergyPlus Evalglare	Automated blinds and dimmable lighting system	Daylight, Lighting
5	[5]	2017	17		●	-	-	•	Rhinoceros Grasshopper Ladybug Honeybee Energy Plus Radiance DAYSIM	Innovative concrete shading device system	Energy Lighting
6	[6]	2021	14	●		-	-	•	Revit Python Rhinoceros Grasshopper Wallacei-X KUKA SimPro	The reach of the robot arm.	Construction
7	[7]	2021	14	●		-	•	•	Faro Scene 5.5 Revit Matlab	Prefabricated façades Digital twin	Geometric quality
8	[8]	2018	13	●		-	-	•	Rhinoceros MAS design tool	Multi-agent systems	Daylight, Energy
9	[9]	2017	13		●	-	•	-	-	DIY approach Building automation	Lighting Air quality Air temperature Humidity
10	[10]	2019	12	●		-	•	•	Rhinoceros Grasshopper Python GDI	Adaptive façade	Daylight Energy Geometric quality
11	[11]	2018	12	●		-	•	-	-	3D-printed façade	Air temperature

illustrates that among these 11 studies, five were conducted within the last 5 years.

Upon examining the study topics, it becomes evident that the concept of 'digital twin' takes the forefront, with a focus on performance features such as 'daylight', 'air temperature', 'lighting', and 'energy'.

In the study by Khajavi et al. [55], which received the highest number of citations, with 129 citations, a digital twin application based on a sensor network was used to create a real-time digital model of a building. The authors investigated the technical constraints of this method using experimental approach. Without utilizing simulation program. The study emphasized that existing digital twin applications are insufficient for collecting real-time experimental data from buildings. The authors created a digital twin that included more than 25,000 sensor readings for a façade element. Data on light, temperature, and humidity parameters were collected and analyzed within the scope of the study. The aim of the study was to reduce maintenance costs, increase user comfort, and reduce overall management and operating expenses (Fig. 7).

Lam et al. [73], in their study, which received 32 citations, explained the concept of Active Noise Control (ANC) and provided answers to 10 questions regarding its applications in the built environment. The study adopted a review-based approach and used literature sources to highlight areas where the active reduction of environmental noise is possible. Importantly, the study was not experimental or simulation-based. The results indicated that there is a need for further research to enable ANC application in the built environment. Also the study found that there is an increasing interest in intelligent ANC applications due to the complexity involved with the evolution of technology and AI techniques.

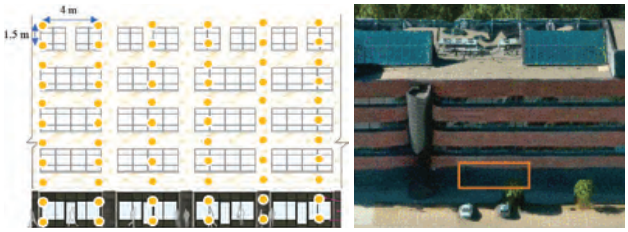


Figure 7. Sensors' mesh on the building and chosen façade for digital twin data collection [55] [created by author].

Figure 8. A scale model and a real-life situation of a sample classroom [74] [recreacted by Cavdar and Vural].

Aghemo et al. [74], in their study, which received 30 citations, developed nine different shading system combinations for the south façade of a sample high school classroom. The study used shading elements with different geometries and locations. The researchers then comparatively evaluated these combinations using measurements obtained from a 1:10 scale model. This method was proposed as an alternative to numerical simulation-based prediction tools; no simulation program was used in the study (Fig. 8).

Babu et al. [75] have 18 citations in their work on the application of automatic shutters and automatic dimming lighting systems in tropical environments. The researchers said that energy savings in lighting can be obtained by "carefully considering the façade orientation, system design and control algorithms using Radiance, Ecotect and EnergyPlus software". Importantly, the study used experimental and modeling approaches.

Similarly, in their study (17 citations), Zani et al. [76] advocated for a balance between transparent, opaque and shaded surfaces in the façades of office buildings in modern architectural design. This balance is necessary to prevent problems such as warmth and too much daylight, and to control the perception of the outer environment. The research team conducted a simulation-based investigation utilizing Rhinoceros, Grasshopper and EnergyPlus software. Here, a computational design approach for a static shading system made of high performance concrete was described. Then the system was optimized by genetic algorithms with respect to numerous parameters.

Tran et al. [77] received 14 citations for their proposed optimization process. This process facilitates the automatic adjustment of the robotic arm's accessibility during the assembly of façade unit elements in the construction industry. The aim of the study was to streamline the robot positioning process and increase construction efficiency. To this end, the researchers examined the limitations of modular robot assembly simulation processes. Achieving the optimal accessibility ratio is expected to reduce the robotic arm's operating time, thereby increasing productivity and reducing risks.

Tran et al. [54], in their study which received 14 citations, presented a new framework for evaluating the geometric quality of prefabricated façades constructed during the construction process. The approach is based on the use of a 3D as-built semantic model and a 3D design model. Based on digital twin methodology, the study enables a quantitative comparison between a 3D as-built digital replica reconstructed from lidar point clouds and a 3D design model. This innovative approach is effective in visually evaluating the quality of prefabricated structures and provides valuable support during the construction process (Fig. 9).

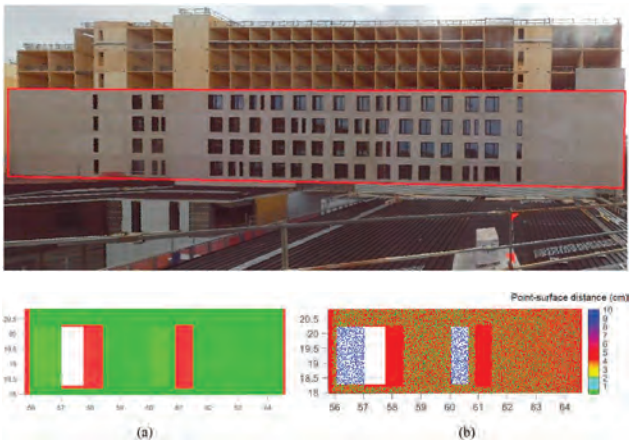


Figure 9. Overlay of the as-built and as-design panels and dense points colourized [54] [recreacted by Cavdar and Vural].

In architecture, digital design paradigms are generally geometry-based, which can have a limited impact on buildings with complex features. Currently, computational design methods capable of predicting performance within the existing digital design process remain limited. Often, the necessary analyses are carried out after the design is completed. However, the new design approach developed by Pantazis and Gerber [27], which has received 13 citations, differs from this common practice. This method offers a design toolset for a multi-factor system that includes a comprehensive evaluation of different environmental performance metrics. This toolset integrates three-dimensional modeling and analysis software.

In Salamone et al. [78] (13 citations) two devices were designed with a Do It Yourself (DIY) approach to control Indoor Environmental Quality (IEQ) and reduce energy usage. The study was based on the development of sensors using wireless communication systems with a low-cost commercial available transmission module and Infrared LED. Most of these sensors were embedded in the internal environment, however the air flow fan sensor was placed within the windows, specifically the façade. The chassis for these devices was made on a 3D printer, using PLA, which is made from renewable resources, as the printing medium. This DIY method allows people to design, build and sell things adapted to their own needs at a lower cost. In this case, the deployment of an automated control system may efficiently analyze the illumination, air quality, air temperature, and humidity.

Yi et al. [79] (12 citations) proposed a responsive building façade through the integration of on-site sensors, building performance simulation (BPS), machine learning, and 3D geometry modeling inside a unified design interface. For this purpose, they used an optimization method that is directly connected to wireless sensor data transmission using the Rhino Grasshopper interface. Therefore, the authors proposed a Tabu based Adaptive Pattern Search Simulated Annealing (T-APSSA) technique. This unique approach allowed for quick performance calculations and created a digital prototype of a responsive façade design.

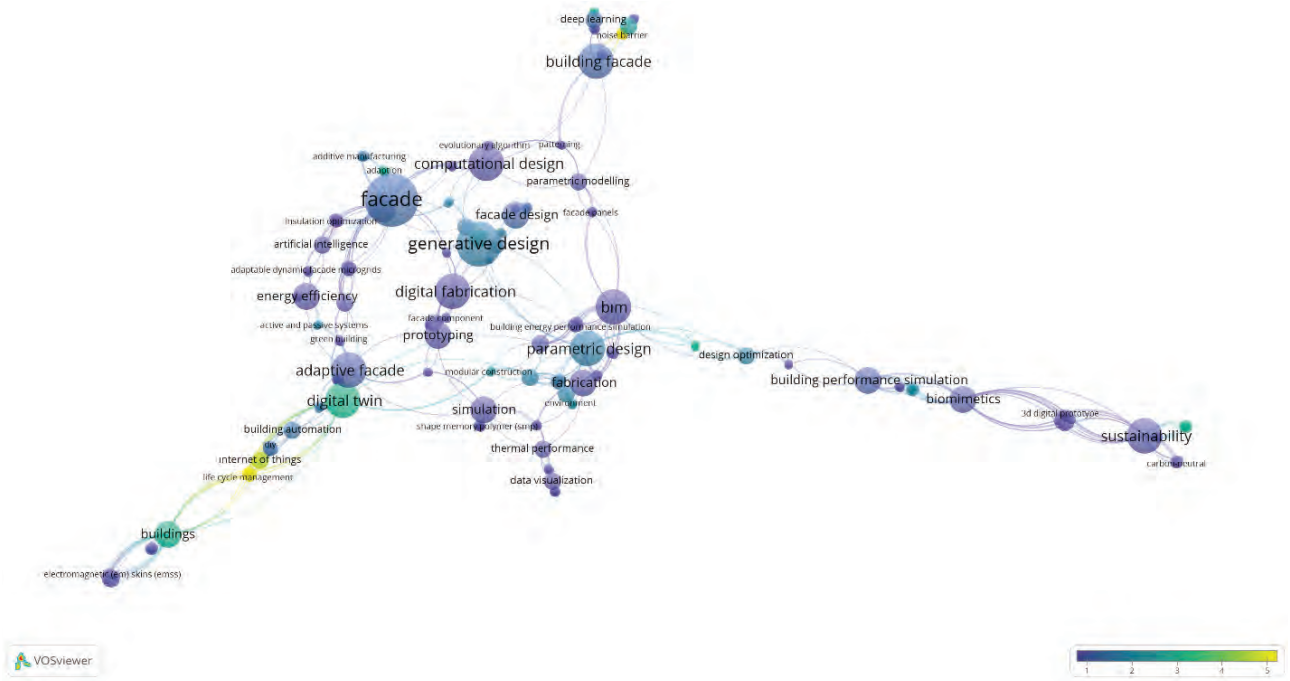


Figure 10. The relationship between the use of keywords and the number of citations (circle sizes: document; color chart: citations).

The 3D printing method is well-suited to improve the thermal performance as it allows to manipulate the size and shape of the material cells to provide insulating materials with low thermal conductivity. Sarakinioti et al. [80] (12 citations) There is an opportunity to explore polymers with low thermal conductivity values where increase of both thermal and waterproofing performance is crucial. Besides, the authors point to the importance of studying the deformations and endurance of materials at extreme thermal conditions.

As seen in Figure 10, the terms 'façade' and 'generative design' appear as the most prominent ones across the publications. However, when examining the citation rate, the phrases "wireless sensor network" and its related terms, "solid modeling," "life cycle management," "real-time systems," "building information modeling," "three-dimensional displays," and "object-oriented modeling," emerge more prominently.

These 96 scientific studies show a chronological increase over the years. It can be seen that there is a significant tendency and the number of studies conducted between 2006 and 2011 is rather small. But the application of digital design technologies to improve facade performance has increased considerably since 2019. It is interesting to see that studies carried out in the period 2021–2022 account for 24% of all research, representing a considerable improvement of the connection between digital design and the architectural sector (Fig. 11).

Analysis: Most Preferred Study Issues

This study, which aims to investigate the role of digital design in articles related to façade performance, examined 96 articles in detail. The studies were comprehensively evaluated according to their thematic categories. As a result, the most common themes were determined to be "adaptive façade" (21.43%), "materials" (14.88%), and "façade" (11.31%).

Articles on adaptive façade studies generally aim to establish clear performance criteria for façades and mostly encompass designs shaped by external environmental conditions. The adaptive façade category also includes studies focusing on kinetic and dynamic façades. In addition, a significant number of studies address façade materials and their performance characteristics.

Studies focusing on conventional façades, which do not adapt to external environmental conditions or include an innovative approach, form a separate category and rank third after the adaptive façade topic. There are also a notable number of studies dedicated to biomimicry, urban scale, and windows, as well as digital twin applications. These are followed by studies on prefabricated façades, sustainability, life cycle management, and pixel-based research (Fig. 12).

Analysis: Most Preferred Performance Types

Analysis of the studies reveals that research focusing on elements or materials in façade design, particularly daylight evaluation, is predominantly structured around performance types such as "daylight" (n=26), "energy" (n=20), and "solar radiation" (n=17) (Fig. 13). These are followed by "digital tools" (n=15), which examine digital environments; "geometric quality" (n=14), which is related to façade design aesthetics; "thermal" (n=13) and "mechanical" (n=11) performance analyses, which address environmental thermal comfort; and "sustainability" (n=7) performance analyses. It is noteworthy that performance categories such as "air quality" (n=4), "humidity" (n=3),

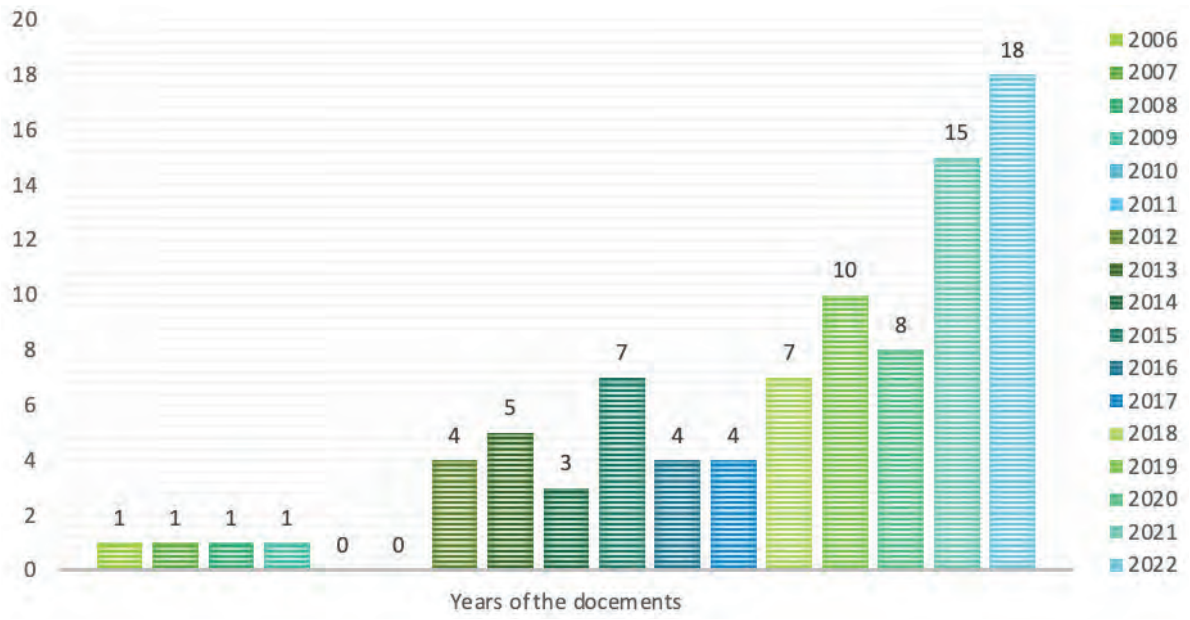


Figure 11. Number of documents by year.

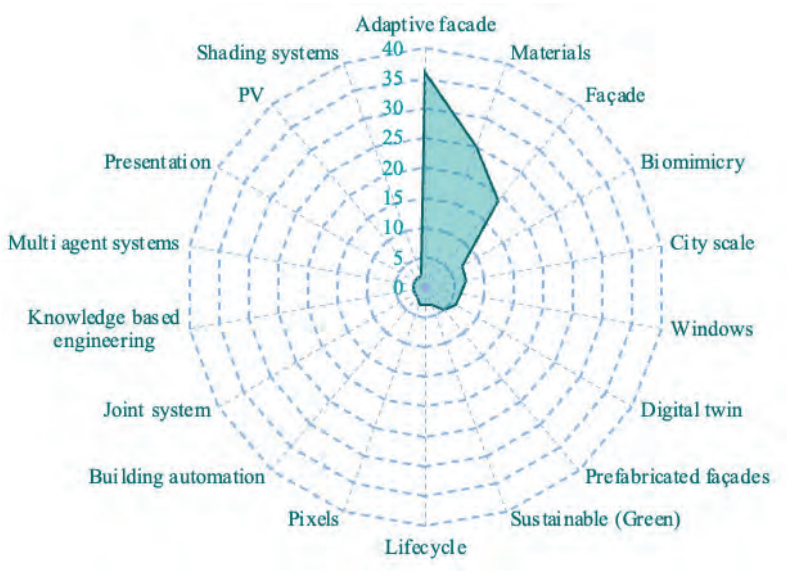


Figure 12. Chart of the study issues of the documents.

and "cooling" (n=1), which are frequently associated with ventilation, are less frequently addressed in the studies. Furthermore, research output on acoustic performance appears to be relatively limited.

Díaz-López et al. [81] conducted a review of scientific studies pertaining to passive cooling strategies. The findings revealed that among the cooling methods employed, natural ventilation, green roof systems, insulated windows, and solar shading emerge as the most commonly utilized. Conversely, studies investigating ventilated façades and evaporative cooling systems remain scarce. Notably, the examination underscored the limited focus on aspects of acoustic comfort and adaptation to climate change—areas that are less emphasized compared to the frequently highlighted subject of thermal comfort and its connection to sustainability.

Analysis: Most Preferred Software-Tool Types

Rogers & Schnabel [82] examined the use of nine prominent instruments used for the digital design process in both education and professional practices in line with expert opinions. Six of the applications, including Rhinoceros and

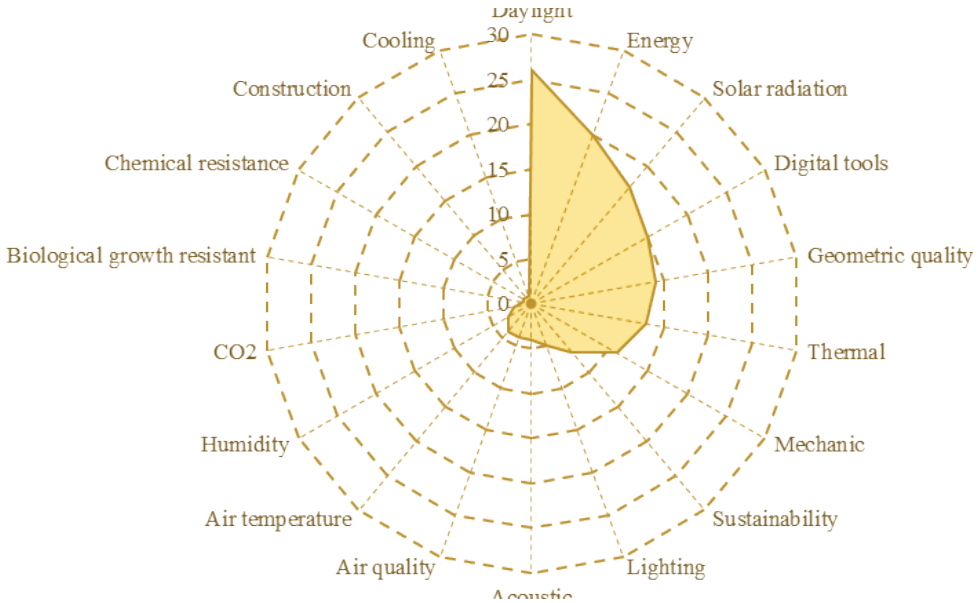


Figure 13. Chart of the performance types used of the documents.

Grasshopper, were found to be more suitable for use in both education and professional practices. In this direction, it is emphasized that new architects' preference of these tools in their design processes will improve creativity.

When it comes to studies on digital design and façade performance, it can be said that the methods to be used will generally focus on simulation programs, softwares and tools. In this direction, 30.2% of the 96 studies examined preferred Rhinoceros, while 28.1% also used the 'Grasshopper' plug-in. It is followed by 'Energy Plus' with 11.5%, 'Revit' with 11.5% and 'LadyBug' with 10.4%. Grasshopper plugins 'Radiance', 'HoneyBee' and 'DIVA' were preferred by 8.3%, 7.3% and 7.3% respectively (Fig. 14).

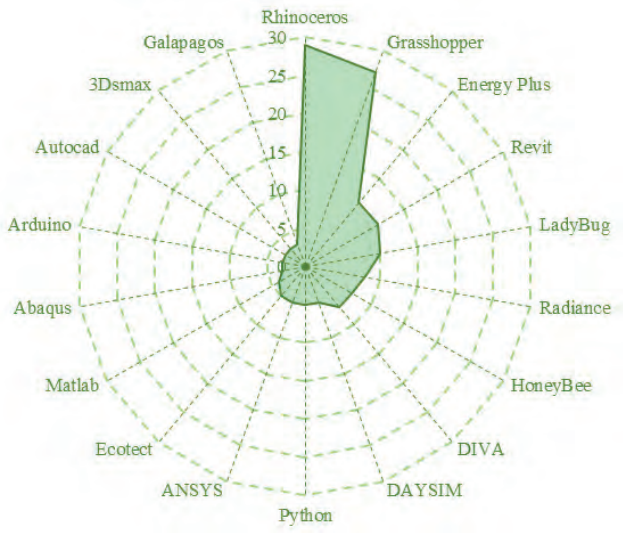


Figure 14. Chart of the softwares/tools used of the documents.

Upon examining the software employed, it becomes apparent that the tools designed to analyze energy, daylight, and solar radiance performance take a prominent role. This correlation aligns directly with the favored performance types in the studies. The realms of adaptive façades, materials, and façade topics exhibit a significant synergy with Rhinoceros, Grasshopper, and EnergyPlus software, particularly in the domains of daylighting, energy efficiency, solar radiation analysis, digital tools employment, geometric quality, and thermal performance optimization. (Fig. 15).

Within the scope of this study, conclusions were drawn based on four evaluation categories: most cited articles, most frequently addressed research topics, performance types, and software/tool types.

- Researchers examining façade performance in the field of digital design generally rely on simulation software. Experimental studies are mostly conducted in research focusing on mechanical [50, 83–86], daylight [27, 33, 34, 46, 47, 75, 79, 87–89], and thermal performance [43, 69, 81, 90–95], and digital tools are also used in these studies. It is noteworthy that articles specifically addressing cooling and ventilation performance are not only limited in number but also largely confined to simulation-based approaches.
- The number of research that are focused on diverse climates is limited; studies focused on specific climates are mostly focused on hot climate types. Most of these research are concerned with façade elements that are intended to lessen solar effects and to function as shading devices. In addition to daylight and temperature, which exert great influence on the indoor comfort, it is important to consider the humidness. This is particularly relevant in humid areas, where natural ventilation techniques can greatly influence both building performance and human health. Up to now, there is still little

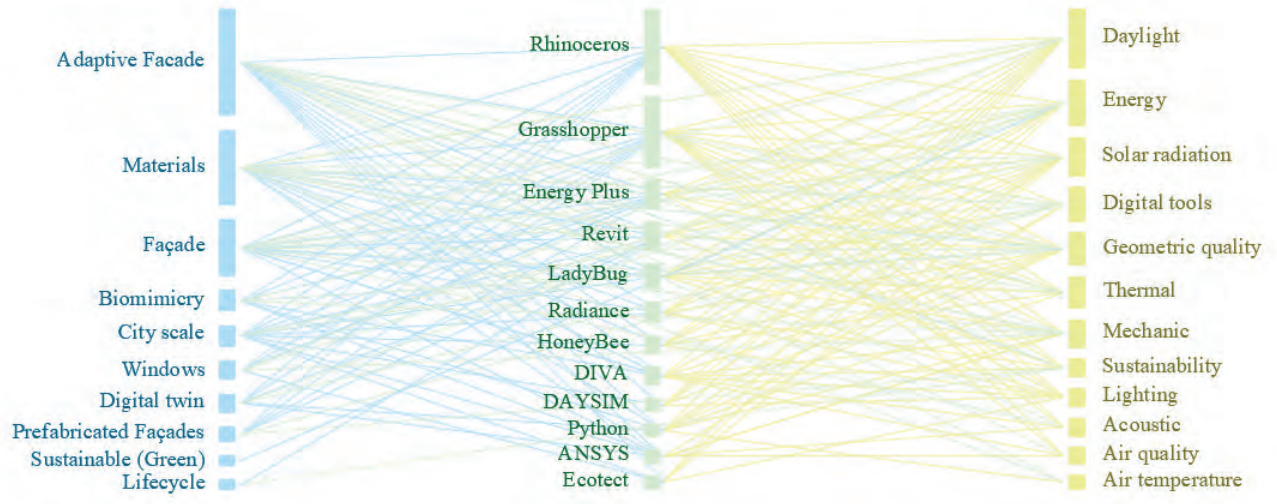


Figure 15. Chart of relationship between software, study subjects and performance types.

research, especially on the performance of ventilation and cooling, where the literature could be strengthened.

- In warmer regions, the main aim is generally to reduce the indoor temperature and improve thermal comfort. These objectives can be attained efficiently by using modern numerical modeling techniques [96]. However, in cold temperature areas, the focus is on avoiding heat loss and maintaining airtightness to improve energy efficiency [97]. Façade design techniques for cold climates have to deal with issues such as frequent temperature changes and humidity and consequently require a different approach than those applied in hot climates [98]. But parametric investigations in cold settings are more restricted. Much research concerns themes such as energy efficient double-skin façades, however there are less studies than in hot regions. Bande et al. [99] talk on the design of innovative facades integrated into buildings in the United Arab Emirates. There is more interest in parametric design in warmer regions. Moreover, while Attia et al. recognize the potential of adaptive façades in cold climates, the total study conducted in this field is comparatively less than the research conducted on warm climates [100].Similarly, studies emphasizing sustainability often focus primarily on energy efficiency and daylight performance [87, 101]. Especially when sustainability is a key evaluation criterion, several factors need to be considered in façade selection. These include façade material properties, CO_2 emissions, breathability, urban context, and other relevant factors.
- With the development of technology, there is a tendency to employ simulation software that supports and complements experimental studies. These simulation methods are progressively becoming more popular than the experimental approaches owing to their more economical viability. While simulations allow for speedy exploration of many scenarios, they may not always capture the entire complexity of real-world conditions as well as physical experiments [102]. Interestingly, the concept of digital twins that connect actual places with simulation software via digital design has gained considerable popularity, especially in recent years [53–55, 103, 104]. The simulation results can be confirmed by comparison with experimental data. For example, Reato et al. [105] employed the twin simulation method to compare their conclusions with experimental data collected from a MEMS electromagnetic actuator. The increasing use of digital twins is seen as promising, particularly for obtaining more precise data for evaluating and comparing facade performance. Furthermore, the use of simulation tools in conjunction with advanced computational methods can increase the reliability of the results. For example, the use of multiple physics approaches allows simulations to capture complex interactions between different domains [104]. In addition, the development of distributed simulation frameworks suggested by Yahiaoui [106] can improve building performance assessments and eliminate the necessity for expensive and time-consuming trials. This approach highlights the fact that simulation tools can supplement, rather than replace, experimental methodologies.
- In studies using simulation software, 8% of researchers used fluid dynamics tools, 3% used acoustic tools, and 13% used mechanical software. In contrast, the usage rates of software tools were 35% for daylight performance and 32% for energy performance. Furthermore, researchers analyzing energy performance tend to base their findings mostly on daylight performance.

CONCLUSION

In today's age of computers and artificial intelligence, architecture has been significantly affected by digital developments. This study examines digital design methods that improve facade performance by interacting with environmental conditions. An analysis of 96 scientific studies selected according to specific criteria shows that although digital design is often associated with aesthetics, only 14.6% of the studies focus on the visual dimension. In contrast, 85.4% of the studies address broader performance factors such as energy, daylight, thermal conditions, acoustics, and sustainability. The research also reveals an increasing interest in digital tools, mechanical properties, CO_2 levels, cyber dimensions, and kinetic systems. Remarkably, studies on digital facade performance have increased over time, with 32.3% of the studies conducted in the last three years.

One of the most important gaps in the literature is ventilation. This shortcoming is especially apparent in regard to indoor air quality and temperature regulation. A lot of studies are on daylighting for energy efficiency and thermal comfort. However, research on managing humidity and temperature by effective air circulation is still restricted. This is especially apparent in hot, humid regions. Further research in this area would contribute significantly to the literature.

The link between architectural design and science is growing stronger. Unlike traditional methods, digital design techniques offer significant opportunities to improve design and create more comfortable living spaces for users. This study aims to guide researchers examining the relationship between digital design and facade performance.

ACKNOWLEDGEMENTS

The dataset analysed in this study is available under the DOI: 10.17632/mjhg9kywwd

This study was produced as part of a doctoral thesis conducted in the Department of Architecture at the Graduate School of Natural and Applied Sciences, Karadeniz Technical University. This thesis study was supported by TÜBİTAK within the scope of the BİDEB 2211- National PhD Scholarship Programs.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Corser R. Fabricating Architecture: Selected Readings in Digital Design and Manufacturing. New York: Princeton Architectural Press; 2010.

[2] Fischer T, Herr CM. Parametric customisation of a 3D concrete printed pavilion. In: Chien S, Choo S, Schnabel MA, Nakapan W, Kim MJ, Roudavski S, editors. Proceedings of the 21st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA); 2016; Hong Kong. p. 549–558. [CrossRef]

[3] Hassan AU, Mohyuddin A, Güleryüz C, Nadeem S, Nkungli NK, Hassan SU, et al. Novel pull–push organic switches with D–π–A structural designs: Computational design of star shape organic materials. Struct Chem 2023;34:399–412. [CrossRef]

[4] Sterk T d'Estrée. Building upon Negroponte: A hybridized model of control suitable for responsive architecture. Autom Constr 2005;14:225–232. [CrossRef]

[5] Yang Y, Fu M, Xing J. Revolutionizing architecture: The synergy of computational design and digital fabrication. Appl Comput Eng 2024;62:1–6. [CrossRef]

[6] Iordanova I. Teaching digital design exploration: Form follows…. Int J Archit Comput 2007;5:685–702. [CrossRef]

[7] Kempenaar A, Brinkhuijsen M, van den Brink A. The impact of regional designing: New perspectives for the Maastricht/Heerlen, Hasselt/Genk, Aachen and Liège (MHAL) region. Environ Plan B Urban Anal City Sci 2019;46:359–376. [CrossRef]

[8] Huang CY, Yang TT, Chen WL, Nof SY. Reference architecture for collaborative design. Int J Comput Commun Control 2010;5:71. [CrossRef]

[9] Bechthold M, King N, Kane AO. Integrated environmental design and robotic fabrication workflow for ceramic shading systems. In: Proceedings of the 28th International Association for Automation and Robotics in Construction (ISARC); 2011; Seoul, South Korea. p. 70–75. [CrossRef]

[10] Chang TW, Huang HY, Datta S. Design and fabrication of a responsive carrier component envelope. Buildings 2019;9:84. [CrossRef]

[11] Yaman M. Different façade types and building integration in energy efficient building design strategies. Int J Built Environ Sustain 2021;8:49–61. [CrossRef]

[12] Gero JS, Sosa R. Complexity measures as a basis for mass customization of novel designs. Environ Plan B Plan Des 2008;35:3–15. [CrossRef]

[13] Aksamija A, Snapp T, Hodge M, Tang M. Re-skinning: Performance-based design and fabrication of building facade components. Res J 2012;4:15–28.

[14] Westphalen D, Koszalinski S. Energy Consumption Characteristics of Commercial Building HVAC Systems. Volume II: Thermal Distribution, Auxiliary Equipment, and Ventilation. Massachusetts: Arthur D. Little, Inc.; 1999.

[15] Salameh M, Taleb H. Courtyard as passive design solution for school buildings in hot area. In: Proceedings of the 2nd World Congress on Civil, Structural, and Environmental Engineering (CSEE'17); 2017; Barcelona, Spain. p. 2–4. [CrossRef]

[16] Salameh M, Touqan B. Traditional passive design solutions as a key factor for sustainable modern urban designs in the hot, arid climate of the United Arab Emirates. Buildings 2022;12:1811. [CrossRef]

[17] Jevremovic L, Turnsek B, Vasic M, Jordanovic M. Passive design applications - industrial architecture perspective. Facta Univ Ser Archit Civ Eng 2014;12:173–182. [CrossRef]

[18] Shao T, Wang L, Ji G. Climate-driven architectural design optimization and exploration, a reverse passive building design approach. In: Proceedings of the 27th CAADRIA Conference; 2022; Sydney. p. 699–708. [CrossRef]

[19] Gajski DD. Principles of Digital Design. International ed. New Jersey: Prentice-Hall Inc.; 1997.

[20] Szalapaj P. Contemporary Architecture and the Digital Design Process. Routledge; 2014. [CrossRef]

[21] Ozden O. Sayısal mimarlık uygulamalarının yapım süreçlerinin irdelenmesi [PhD thesis]. İzmir: Dokuz Eylül Üniversitesi; 2011.

[22] İsbitiren I. Mimarlıkta hesaplamalı tasarım ve üretim yaklaşımları: Pavyonlar üzerine inceleme. Tasarım Enformatiği 2016;2:37–46.
[23] Augugliaro F, Lupashin S, Hamer M, Male C, Hehn M, Mueller MW, et al. The Flight Assembled Architecture installation: Cooperative construction with flying machines. IEEE Control Syst 2014;34:46–64. [CrossRef]
[24] Graiz E, Van de Rie K. Employing surface tooling for thermal modulation. Archit Sci Rev 2022;66:17–255. [CrossRef]
[25] Yetkin S. Sayısal tasarım araçları: Tasarım sürecindeki rolleri bağlamında bir inceleme [MSc thesis]. İstanbul: Fen Bilimleri Enstitüsü; 2011.
[26] Bechthold M, King J, Kane A, Niemasz J, Reinhart C. Integrated environmental design and robotic fabrication workflow for ceramic shading systems. In: Proceedings of the 28th International Association for Automation and Robotics in Construction (ISARC); 2011; Seoul. p. 70–75. [CrossRef]
[27] Pantazis E, Gerber D. A framework for generating and evaluating façade designs using a multi-agent system approach. Int J Archit Comput 2018;16:248–270. [CrossRef]
[28] Bilen K, Urmamen E, Topcu MT, Solmaz I. Evaluations on the energy identity certificate and the usability of calculation method of building energy performance (BEP). Sigma J Eng Nat Sci 2020;11:103–113.
[29] Yüksel Dicle S, Zoroğlu F, Atmaca AB, Yüğrük Akdağ N, Zoreer Gedik G. Evaluation of energy consumption and noise reduction change of a strengthened building: An educational building case. Sigma J Eng Nat Sci 2024;42:1276–1291. [CrossRef]
[30] Nguyen TD, Adhikari S. The role of BIM in integrating digital twin in building construction: A literature review. Sustainability 2023;15:10462. [CrossRef]
[31] Angjeliu G, Coronelli D, Cardani G. Development of the simulation model for digital twin applications in historical masonry buildings: The integration between numerical and experimental reality. Comput Struct 2020;238:106282. [CrossRef]
[32] Akkan Çavdar A, Şanlı Karabak E. Digital twin technology in architecture. In: Aksoy Y, Boyan E, editors. Contemporary Issues in Architecture and Urban Planning 2024: Architecture, Environment, and Interdisciplinarity. İstanbul: DAKAM; 2024. p. 47–65.
[33] Giantini G, Souza LN de, Turczyn D, Celani G. EnvironmentalCeramics. In: Proceedings of the 37th eCAADe and 23rd SIGraDi Conference; 2019. p. 749–758. [CrossRef]
[34] Sharaidin K, Burry J, Salim F. Integration of digital simulation tools with parametric designs to evaluate kinetic façades for daylight performance. In: Digital Physicality-Proceedings of the 30th eCAADe Conference; 2012; Czech Republic: Czech Technical University. p. 701–709. [CrossRef]
[35] Karagianni L, Turrin M, Knaack U, Hordijk T. Additive manufacturing for daylight_Towards a customized shading device. In: Proceedings of the Symposium on Simulation for Architecture and Urban Design; 2016; London. p. 69–76.
[36] Baehr-Bruyère J, Chamilothori K, Vassilopoulos AP, Wienold J, Andersen M. Shaping light to influence occupants' experience of space: A kinetic shading system with composite materials. J Phys Conf Ser 2019;1343:012162. [CrossRef]
[37] Velasco R, Hernández R, Marrugo N, Díaz C. Notes on the design process of a responsive sun-shading system: A case study of designer and user explorations supported by computational tools. Artif Intell Eng Des Anal Manuf 2015;29:483–502. [CrossRef]
[38] Freitas J de S, Cronemberger J, Soares RM, Amorim CND. Modeling and assessing BIPV envelopes using parametric Rhinoceros plugins Grasshopper and Ladybug. Renew Energy 2020;160:1468–1479. [CrossRef]
[39] Sravani T, Prasanna Venkatesan R, Madhumathi A. A comparative LCA study of passive cooling roof materials for a residential building: An Indian case study. Mater Today Proc 2022;64:1014–1022. [CrossRef]
[40] Koukelli C, Prieto A, Asut S. Kinetic solar envelope: Performance assessment of a shape memory alloy-based autoreactive façade system for urban heat island mitigation in Athens, Greece. Appl Sci 2021;12:82. [CrossRef]
[41] Marques R, Eloy S. Customized cork façade. In: ECAADe 2013: Computation and Performance; 2013; The Netherlands. p. 621–626. [CrossRef]
[42] Webb M. Biomimetic building facades demonstrate potential to reduce energy consumption for different building typologies in different climate zones. Clean Technol Environ Policy 2022;24:493–518. [CrossRef]
[43] Kuru A, Oldfield P, Bonser S, Fiorito F. A framework to achieve multifunctionality in biomimetic adaptive building skins. Buildings 2020;10:114. [CrossRef]
[44] Lorenz WE. Combining complexity and harmony by the boxcounting method. In: eCAADe 2013: Computation and Performance; 2013; The Netherlands: Delft University of Technology. p. 667–676. [CrossRef]
[45] Simmons M. The evolution of a specialised practice: Consulting and contracting in the integrated envelope delivery industry. Archit Des 2020;90:24–31. [CrossRef]
[46] Cirillo A, Scofone A. The retrofit of '70s office buildings curtain walls in London. J Phys Conf Ser 2021;2042:1–6. [CrossRef]
[47] Figliola A. Digital workflow for climate resilient building façade generation. Build Res Inf 2023;51:257–278. [CrossRef]

[48] Luna-Navarro A, Hunt GR, Overend M. Dynamic façades – An exploratory campaign to assess occupant multi-domain environmental satisfaction and façade interaction. Build Environ 2022;211:108703. [CrossRef]
[49] Ku K, Chung D. Digital fabrication methods of composite architectural panels for complex shaped buildings. In: Proceedings of the 20th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA); 2015. p. 703–712. [CrossRef]
[50] Schling E, Wan Z. A geometry-based design approach and structural behaviour for an asymptotic curtain wall system. J Build Eng 2022;52:104432. [CrossRef]
[51] de Campos FM, Leite RM, Prudencio CF, Dias MS, Celani G. Prototyping a facade component. In: Architecture in the Age of the 4th Industrial Revolution, eCAADe SIGraDi 2019 Vol.1; 2019; Portugal. p. 179–186. [CrossRef]
[52] Wolf K, Han SO, Doll T, Kurtze L, Bein T, Hanselka H. Performance assessment of an active noise reduction system. In: Volume 2: Automotive Systems; Bioengineering and Biomedical Technology; Computational Mechanics; Controls; Dynamical Systems; 2008. p. 785–791. [CrossRef]
[53] Böke J, Knaack U, Hemmerling M. Prototype of a cyber-physical façade system. J Build Eng 2020;31:101397. [CrossRef]
[54] Tran H, Nguyen TN, Christopher P, Bui DK, Khoshelham K, Ngo TD. A digital twin approach for geometric quality assessment of as-built prefabricated façades. J Build Eng 2021;41:1–13. [CrossRef]
[55] Khajavi SH, Motlagh NH, Jaribion A, Werner LC, Holmstrom J. Digital twin: Vision, benefits, boundaries, and creation for buildings. IEEE Access 2019;7:147406–147419. [CrossRef]
[56] Wallin JA. Bibliometric methods: Pitfalls and possibilities. Basic Clin Pharmacol Toxicol 2005;97:261–275. [CrossRef]
[57] Alsharif AH, Salleh NZM, Baharun R. Research trends of neuromarketing: A bibliometric analysis. J Theor Appl Inf Technol 2020;98:2948–2962.
[58] Donthu N, Kumar S, Pattnaik D, Lim WM. A bibliometric retrospection of marketing from the lens of psychology: Insights from Psychology & Marketing. Psychol Mark 2021;38:834–865. [CrossRef]
[59] Choudhri AF, Siddiqui A, Khan NR, Cohen HL. Understanding bibliometric parameters and analysis. Radiographics 2015;35:736–746. [CrossRef]
[60] Kuzminska OH, Mazorchuk MS, Barna OV, Sydorenko S. Bibliometric analysis in determining the research directions of early career researchers. Inf Technol Learn Tools 2022;91:113–129. [CrossRef]
[61] Özdemir M, Arslan S. Mimarlıkta makine öğrenmesi: Bibliyometrik bir analiz. Online J Art Des 2021;9:194–207.
[62] Erdoğan RE. Mimarlıkta sayısal tasarım araştırmalarının bibliyometrik yöntemlerle analizi ve haritalanması [MSc thesis]. Ankara: Fen Bilimleri Enstitüsü; 2020.
[63] Eltaweel A, Su Y. Parametric design and daylighting: A literature review. Renew Sustain Energy Rev 2017;73:1086–1103. [CrossRef]
[64] Guo L, Shao F, Botev C, Shanmugasundaram J. XRANK: Ranked keyword search over XML documents. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data; 2003; New York, NY, USA: ACM. p. 16–27. [CrossRef]
[65] Pranckutė R. Web of Science (WoS) and Scopus: The titans of bibliographic information in today's academic world. Publications 2021;9:12. [CrossRef]
[66] Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. Prof Inf 2020;29. [CrossRef]
[67] Anthony L. AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. In: Proceedings of the International Professional Communication Conference (IPCC); 2005; IEEE. p. 729–737. [CrossRef]
[68] Mohtashami N, Fuchs N, Fotopoulou M, Drosatos P, Streblow R, Osterhage T, et al. State of the art of technologies in adaptive dynamic building envelopes (ADBEs). Energies 2022;15:829. [CrossRef]
[69] Kuru A, Oldfield P, Bonser S, Fiorito F. Performance prediction of biomimetic adaptive building skins: Integrating multifunctionality through a novel simulation framework. Sol Energy 2021;224:253–270. [CrossRef]
[70] Yoon J, Bae S. Performance evaluation and design of thermo-responsive SMP shading prototypes. Sustainability 2020;12:4391. [CrossRef]
[71] Karakoç E, Çağdaş G. Adaptive architecture based on environmental performance: An advanced intelligent façade (AIF) module. Gazi Univ J Sci 2021;34:630–650. [CrossRef]
[72] Akkan Çavdar A, Vural N. Digital design in academic studies on facade performance: A review. Mendeley Data 2024.
[73] Lam B, Gan WS, Shi DY, Nishimura M, Elliott S. Ten questions concerning active noise control in the built environment. Build Environ 2021;200:1–18. [CrossRef]
[74] Aghemo C, Pellegrino A, Lo Verso VRM. The approach to daylighting by scale models and sun and sky simulators: A case study for different shading systems. Build Environ 2008;43:917–927. [CrossRef]
[75] Babu S, Zhou J, Wan MP, Lamano AS, Sarvaiya JN, Zhang Z, et al. Investigation of an integrated automated blinds and dimmable lighting system for tropical climate in a rotatable testbed facility. Energy Build 2019;183:356–376. [CrossRef]

[76] Zani A, Andaloro M, Deblasio L, Ruttico P, Mainini AG. Computational design and parametric optimization approach with genetic algorithms of an innovative concrete shading device system. Procedia Eng 2017;180:1473–1483. [CrossRef]

[77] Ali AK, Lee OJ, Song H. Robot-based facade spatial assembly optimization. J Build Eng 2021;33:1–11. [CrossRef]

[78] Salamone F, Belussi L, Danza L, Ghellere M, Meroni I. How to control the indoor environmental quality through the use of the Do-It-Yourself approach and new pervasive technologies. Energy Procedia 2017;140:351–360. [CrossRef]

[79] Yi H, Kim MJ, Kim Y, Kim SS, Lee KI. Rapid simulation of optimally responsive façade during schematic design phases: Use of a new hybrid metaheuristic algorithm. Sustainability 2019;11:1–28. [CrossRef]

[80] Sarakinioti MV, Turrin M, Konstantinou T, Tenpierik M, Knaack U. Developing an integrated 3D-printed façade with complex geometries for active temperature control. Mater Today Commun 2018;15:275–279. [CrossRef]

[81] Díaz-López C, Serrano-Jiménez A, Verichev K, Barrios-Padura Á. Passive cooling strategies to optimise sustainability and environmental ergonomics in Mediterranean schools based on a critical review. Build Environ 2022;221:109297. [CrossRef]

[82] Rogers J, Schnabel MA. Digital design ecology: An analysis for an intricate framework of architectural design. In: Proceedings of the 36th eCAADe Conference: Collaborative and Participative Design Vol. 1; 2021. p. 459–468. [CrossRef]

[83] Zhao B, Chen W. Rate-dependent mechanical properties and elastic modulus of ETFE foils used in inflated forming of transparency air-inflated cushion membrane structures. Eng Struct 2021;227:111404. [CrossRef]

[84] Yi H, Kim Y. Prototyping of 4D-printed self-shaping building skin in architecture: Design, fabrication, and investigation of a two-way shape memory composite (TWSMC) façade panel. J Build Eng 2021;43:103076. [CrossRef]

[85] Helbig T, Scheible F, Kamp F, Schieber R. Engineering in a computational design environment - New Terminal 3 at Shenzhen Bao'an International Airport, China. Steel Constr 2014;7:24–31. [CrossRef]

[86] Zhao C, Yang J, Wang X, Ren M, Azim I. Investigation into short term behaviors of embedded laminated glass connections with various configurations. Constr Build Mater 2021;297:123687. [CrossRef]

[87] Spiegelhalter T. Integrated project delivery and total building automation for the nearly net-zero-energy Q1 ThyssenKrupp headquarters. In: Sustainable Human-Building Ecosystems; 2015; Reston, VA: American Society of Civil Engineers. p. 168–175. [CrossRef]

[88] Savvides A, Vassiliades C. Designing urban building blocks around solar planning principles. In: Sustainable Development and Planning IX. Vol. 226; 2017; WITPress. p. 679–690. [CrossRef]

[89] El Daly HMT. Automated fenestration allocation as complying with LEED rating system. Alexandria Eng J 2014;53:883–890. [CrossRef]

[90] Montali J, Sauchelli M, Jin Q, Overend M. Knowledge-rich optimisation of prefabricated façades to support conceptual design. Autom Constr 2019;97:192–204. [CrossRef]

[91] Zhao Y, Chen Y, Li K. A simulation study on the effects of tree height variations on the façade temperature of enclosed courtyard in North China. Build Environ 2022;207:108566. [CrossRef]

[92] Burry J, Salim F, Williams M, Pena De Leon A, Sharaidin K, Burry M, et al. Understanding heat transfer performance for designing better façades. In: Proceedings of the ACADIA. Vol. 13; 2013. p. 71–78. [CrossRef]

[93] Khorasani ML, Burry J, Salehi M. Thermal performance of patterned facades studies on effects of patterns on the thermal performance of facades. In: Proceedings of the Education and Research in Computer Aided Architectural Design in Europe (eCAADe) Vol 3; 2014. p. 267–276. [CrossRef]

[94] Cocho-Bermejo A, Vogiatzaki M. Phenotype variability mimicking as a process for the test and optimization of dynamic facade systems. Biomimetics 2022;7:85. [CrossRef]

[95] Cumo F, Giustini F, Pennacchia E, Romeo C. The "D2P" approach: Digitalisation, production and performance in the standardised sustainable deep renovation of buildings. Energies 2022;15:6689. [CrossRef]

[96] Helmi N, Faggal A, Nessim A. Enhancing energy consumption in Egypt within adaptive facade techniques. J Al-Azhar Univ Eng Sect 2024;19:130–142. [CrossRef]

[97] Liu C, Sun C, Li G, Yang W, Wang F. Numerical simulation analyses on envelope structures of economic passive buildings in severe cold region. Buildings 2023;13:1098. [CrossRef]

[98] Paiho S, Ojanen T, Pinto Seppä I, Paavola M. Critical performance aspects of retrofitting apartment buildings using a multifunctional façade system. Buildings 2019;9:184. [CrossRef]

[99] Bande L, Hamad H, Alqahtani D, Alnahdi N, Ghunaim A, Fikry F, et al. Design of innovative parametric/dynamic façade integrated in the library extension building on UAEU campus. Buildings 2022;12:1101. [CrossRef]

[100] Attia S, Bertrand S, Cuchet M, Yang S, Tabadkani A. Comparison of thermal energy saving potential and overheating risk of four adaptive façade technologies in office buildings. Sustainability 2022;14:6106. [CrossRef]

[101] Tang Y, Gao F, Wang C, Huang MM, Wu M, Li H, et al. Vertical greenery system (VGS) renovation for sustainable arcade-housing: Building energy efficiency analysis based on digital twin. Sustainability 2023;15:2310. [CrossRef]

[102] Yahiaoui A, Sahraoui A. Using systems engineering practices for distributed control and building performance simulation. INCOSE Int Symp 2015;25:1290–1304. [CrossRef]

[103] Benoni A, Salucci M, Oliveri G, Rocca P, Li B, Massa A. Planning of em skins for improved quality-of-service in urban areas. IEEE Trans Antennas Propag 2022;70:8849–8862. [CrossRef]

[104] Zhang J, Fukuda T, Yabuki N. Automatic generation of synthetic datasets from a city digital twin for use in the instance segmentation of building facades. J Comput Des Eng 2022;9:1737–1755. [CrossRef]

[105] Reato FM, Ricci C, Misfatto J, Calzaferri M, Cinquemani S. An alternative multi-physics-based methodology for strongly coupled electro-magneto-mechanical problems. Algorithms 2023;16:306. [CrossRef]

[106] Yahiaoui A. Distributed dynamic simulations of networked control and building performance applications. Simulation 2018;94:145–161. [CrossRef]