

ISSN 1304-7191

E-ISSN 1304-7205



sigma

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

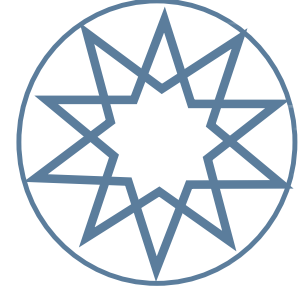
Volume 43

Number 5

Year 2025

YTÜ
PRESS

www.sigma.yildiz.edu.tr



Volume 43 Number 5 Year 2025 - October

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 Molecular Biology and Genetics, Yıldız Technical University, İstanbul, Turkey

Emel AKYOL

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

Eyüp DEBİK

Department of Environmental 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 Materials 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 TULÜ

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

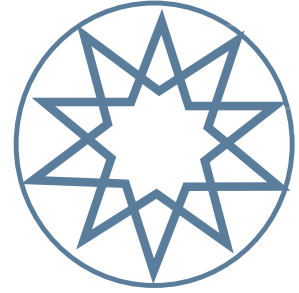
Department of Mechanical Engineering, Yıldız Technical University, İstanbul, Turkey

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



Volume 43 Number 5 Year 2025 - October

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 Environmental 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

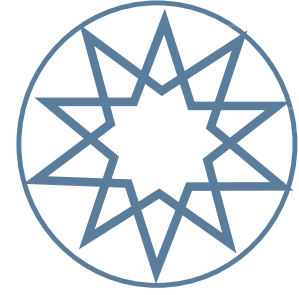
Shengda ZENG

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



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 bimonthly (February-April-June-August-October-December).

Owner: Yıldız Technical University

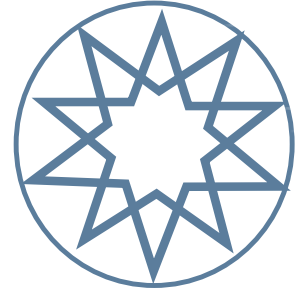
Editor-in-Chief: Ahmet Selim Dalkılıç

Language of Publication: English

Frequency: Bimonthly (6 issues per year)

Publication Type: Online e-version

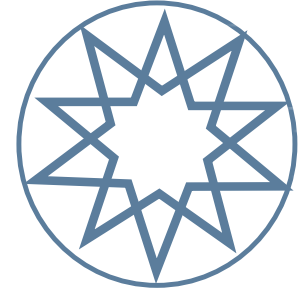
Publisher: Kare Publishing



CONTENTS

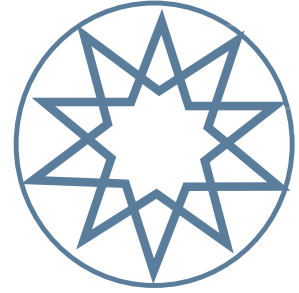
Research Articles

- 1439** Simulations and comparisons of the hay and perez models applied in PVsyst software and their effect on energy output
Sara AL-WAISAWY, Hamdiyah S. JASIM, Ahmed Hussein DUHIS
- 1451** Urban morphology and planning from an entropy perspective: A bibliometric evaluation
Esra BAYAZIT
- 1473** Broken rotor bar fault detection in induction motors through power spectral density to image method
Hatice OKUMUS, Ebru ERGUN
- 1484** Cart KNN: A hybrid distance-based pre-pruned classification and regression method for the early diagnosis of Alzheimer's disease
Pijush DUTTA, Sunita SHARMA, Varun Kumar NOMULA, Gour Gopal JANA, Anubrata MONDAL, Shobhandeb PAUL
- 1495** On the use of different link functions in gamma-pareto regression model: Simulation and application
Nasir SALEEM, Atif AKBAR, Javaria Ahmad KHAN
- 1507** Experimental investigation of thermo-hydraulic performance of multi semicircular rib of dimple hybrid channel
Kishor K. DHANDE, Prakash Santosh PATIL, Lalit N. PATIL, Vikash K. AGRAWAL, Atul A. PATIL
- 1520** Evaluation of the bending fatigue behavior crack propagation in ultra-high-strength steel's resistance spot welding joints, considering the mechanical properties
Mustafa GOKTAS, Bilge DEMIR, Muhammed ELITAS
- 1533** Magnetohydrodynamic flow and effects of Soret, Dufour and thermophoresis through a chemically reacting porous stretching plate
Ankur KUMAR SARMA, Dipak SARMA, Sunmoni MUDOI



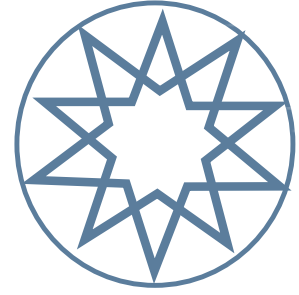
Volume 43 Number 5 Year 2025 - October

- 1547** Trends and future perspectives in wind energy studies: A bibliometric analysis with time-series neural networks
Nihal GENC, Ferdi OZBILGIN, Volkan BASER
- 1572** Real-time tomato flower estimation using deep learning models for e-agriculture applications
Bini D, Pamela D, Shajin PRINCE, K Martin SAGAYAM, Hatıra GÜNERHAN
- 1580** Classification of ancient coins in archaeology using a novel deep learning approach: Bayesian convolutional neural network
Ebru PEKEL ÖZMEN, Soner ÖZMEN, Mehmet Emre SERTKAYA, Tuncay ÖZCAN, Vedat KELEŞ
- 1592** Cat swarm optimized tumor segmentation and an ensemble model for glioblastoma patient survival prediction
Bhagyalaxmi K, Dwarakanath B
- 1607** Examining dimensionality reduction effect of Principal component analysis via hierarchical clustering techniques
Yağmur ŞAN, Tolunay GÖÇKEN
- 1628** A creative combination of time series models for accurate forecasting of monthly rainfall: A case study in Northeast India
D KARTHIKA, K KARTHIKEYAN
- 1643** Exploration of Ulam-Hyers stability for a system of fractional integro-coupled differential equations with integral boundary conditions
Brijendra Kumar CHAURASIYA, Avadhesh KUMAR
- 1653** In-sito co-pyrolysis of tea waste and polypropylene using microwave-assisted pyrolysis: Products analysis, energy consumption, and synergy effect
Husam Talib HAMZAH, Sridevi VELURU, Suhair Abdulhadi MAHDI, Thamer Adnan ABDULLAH, M. TUKARAMBAI, Venkata Rao POIBA



Volume 43 Number 5 Year 2025 - October

- 1663** An integrative approach to agricultural challenges: Predictive modeling, crop alternatives and automation
Swapnil S. JADHAV, Lalit N. PATIL, Vikas S. PANWAR, Sachin P. JADHAV, Digvijay B. KANASE, Sayali S. JADHAV, Neeraj D. PATEL, Khushi V. SALI, Srushti M. JOSHI, Ruchi N. NAGWANSHI
- 1683** Unbiased weighted least squares estimation of Weibull modulus with small samples
Burak BİRGÖREN
- 1693** Designing of triple rings metasurface based perfect absorber for C-band and X-band applications
Jigar M. PATEL, Rahul D. MEHTA
- 1704** The effects of using lithium on biochemical factors of urinary beta-2 microglobulin, creatinine and urea serum in chronic bipolar patients
Seyed Reza MOMENI SHIDEH, Amireh Ebrahimi MOGHADAM, Zainab WETWET, Hayder Abbas Hussein AL MAHDI
- 1712** Mathematical analysis of fractional order Covid 19 epidemic model
Pınar AÇIKGÖZ, Mehmet MERDAN
- 1725** Global solar radiation prediction using machine learning approaches
M. MOHANDDES, Hilal H. NUHA, Satria Akbar MUGITAMA, S. REHMAN, A. AL-SHAILKHI
- 1737** Dynamic water monitoring and management system using labview and machine learning
Rupali DESHMUKH, Rhea GOMEZ, Pramod GARHWAL, Disha PATIL, Nilaj DESHMUKH
- 1751** An abstract approach to sequences and series using soft complex numbers
Atiqe UR RAHMAN, Aram NOORI QADIR, Pishtiwan OTHMAN SABIR
- 1760** Electrokinetic effect on flow of two-phase fluid in a circular tube
Sangeetha J, Ponalagusamy R
- 1785** Investigation of seismic behavior of horizontally curved bridges in comparison with equivalent straight bridges according to the AASHTO LRFD specifications
Esra NAMLI, Turgut ÖZTÜRK



Review Articles

- 1807** **Recent and significant advances in crypto-biometrics: A detailed review**
Priyanka BHATTACHARYYA, G.S.G.N. ANJANEYULU
- 1837** **Exploring metal organic frameworks for efficient drug delivery: A bibliometric review**
Buddhadeb DUTTA, Supratim SUIN
- 1852** **A review on utilisation of nanofluid in solar energy harnessin**
Sanjeev Kumar GUPTA, Vivek RATHORE
- 1886** **Assessing the role of circular economy principles in reducing waste by sustainable manufacturing practices: A review**
Mustafa M. MANSOUR



Research Article

Simulations and comparisons of the hay and perez models applied in PVsyst software and their effect on energy output

Sara AL-WAISAWY¹, Hamdiyah S. JASIM¹, Ahmed Hussein DUHIS^{1,*}

¹Department of Electrical Engineering Techniques, Al-Mussaib Technical College, Al-Furat Al-Awsat Technical University, 54003, Iraq

ARTICLE INFO

Article history

Received: 12 March 2025

Revised: 08 July 2025

Accepted: 10 July 2025

Keywords:

Horizon Band Irradiation;

Photovoltaic System

Performance; Solar Radiation

Modeling

ABSTRACT

This study examines the connections between essential variables that affect solar energy modeling, such as global horizontal radiation, transposition factor, global incident radiation, energy availability, and horizon band radiation. The PVsyst software evaluates two existing solar radiation models—Hay and Perez—to examine their effects on the energy production of a photovoltaic system in Baghdad City, Iraq. While both models forecast identical global horizontal irradiation, the Perez model yields a superior transposition factor and worldwide incident irradiation owing to its incorporation of horizon band irradiation. The Perez model attains a yearly energy availability of 4792.7 kilowatt-hours, surpassing the Hay model's 4687.2 kilowatt-hours. These findings indicate that incorporating horizon band irradiation markedly enhances the precision of energy yield predictions. The Hay model presents simplicity, whereas the Perez model delivers a more thorough and accurate assessment, especially in scenarios with significant diffuse and near-horizon radiation. This study improved photovoltaic system design by emphasizing the significance of comprehensive solar radiation modeling.

Cite this article as: Al-Waisawy S, Jasim HS, Duhis AH. Simulations and comparisons of the hay and perez models applied in PVsyst software and their effect on energy output. Sigma J Eng Nat Sci 2025;43(5):1439–1450.

INTRODUCTION

Output compute models are employed to assess the anticipated energy output for a photovoltaic system, requiring specific input parameters, including global sun irradiation. Given that minor uncertainty in model parameters can cause significant discrepancies in anticipated investment returns, it is imperative to minimize uncertainties arising from model inputs to the greatest extent possible.

Estimating sun radiation on tilted surfaces with varying orientations is crucial for calculating the electrical power

produced by photovoltaic systems, designing solar energy systems, and assessing their long-term performance average [1, 2]. Nonetheless, the existing measurement data are inadequate, as global horizontal and diffuse horizontal radiation readings from pyrometers are sometimes the sole available data in numerous places. Nevertheless, inclined measurements are implemented. The selected tilt angle may not be advantageous for the specific site.

Therefore, the solar radiation received on a slanted surface must be calculated by converting solar radiation from a

*Corresponding author.

*E-mail address: ah.hu.khf@gmail.com

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



level surface to the radiation incident on the specified tilted surface [3-6].

Transposition models that use global and scattered horizontal sunlight have been commonly used within the solar energy field to calculate the sunlight hitting tilted solar panels. Transposition models break down the sunlight on a tilted surface into three parts: direct, diffuse, and sunlight mirrored from the ground. Direct radiation can be calculated by determining the geometry between horizontal and tilted surfaces. The ground reflected radiance can be calculated with the assistance of an isotropic model utilizing simple algorithms. The idea that everything is the same in all directions can work for figuring out how opacity affects things, but it's not as useful for the scattered light part. This is because the way the diffuse light spreads out is complicated and relies on several things, like the angle of the sun and the presence of clouds. The ongoing development and variety of transposition models demonstrate the intricacy of the task.[7-9]

Initial models transformed horizontal diffuse radiation to a tilted plane by presuming that total sky diffuse radiation is isotopically distributed across the sky dome [10-14]. This idea is overly simplistic and does not align with reality. Recent transposition models consider the diffuse component to be anisotropically distributed. Some anisotropic models solely account for an isotropic backdrop and an extra circumsolar region, while others also incorporate horizon-brightening effects. This assumption holds only when clouds are absent, as overcast conditions result in a darker horizon than the zenith [15, 16]. Solar Energy is a renewable, safe, and abundant energy source. Calculating the amount of solar radiation incident upon inclined surfaces with different orientations is required to calculate the building envelope's heat gain and the power generated by photovoltaics [17, 18], design solar systems, and evaluate their long-term average performance. Given the reality that several meteorological/radiometric sites measure global as well as diffuse irradiation on horizontal surfaces, data for inclined surfaces are not accessible. They are calculated using different models than those measured upon horizontal surfaces [19-21].

The total radiation incident on a tilted surface comprises beam, diffuse, and ground reflection. Hourly, straightforward algorithms can accurately compute the direct and reflect components, but the diffused component is more complex, necessitating review and appraisal of the necessary procedures.[22, 23].

Several of these evaluations have focused on computational methods for modeling and evaluating the performance of solar systems. In this context, Connolly et al. [24] provided a comprehensive overview of current computational techniques for assessing the integration of renewable energy initiatives. Several software programs are widely utilized as the foundation for several scientific projects. Wijeratne et al. [25] thoroughly assessed the software tools employed in several research for the simulation

of photovoltaic systems. The efficacy of these PV techniques unequivocally influences the quality of the outcomes achieved in the studies above. Nevertheless, limited research has evaluated the effectiveness of various existing photovoltaic simulation software tools compared to actual photovoltaic system energy output data.

Lee et al. [26] validated four photovoltaic tools (PVsyst, HOMER, and SAM) using actual production data collected from six photovoltaic systems located at the Desert Knowledge the Country Solar Centre, a solar technology demonstration facility situated in Alice Springs, Australia. The systems comprised modules utilizing various photovoltaic technologies: monocrystalline, polycrystalline, amorphous silicon, and cadmium telluride thin-film modules. The installed capacity of the photovoltaic arrays varied from 5.1 kW to 7 kW. The modeling tools demonstrated accuracy, with annual average percentage discrepancies ranging from -1.44% to -4.93%, contingent upon the installation type. HOMER and RETScreen yielded the most precise results, but PVsyst typically produced conservative outcomes. Despite being the most straightforward of the models, Sunny yielded the least precise outcomes.

Axaopoulos et al. [27] evaluated the efficacy of six software tools (TRNSYS, Archelios, Polysun, PVSyst, PV*SOL, and PVGIS) utilizing the actual electrical energy produced by a grid-connected 19.8 kWp solar system situated in Thrace, Greece. Their findings indicated that all instruments generally overestimated the global irradiance received by the photovoltaic modules while dramatically underestimating the electrical energy produced by the system. The tool that demonstrated the most superior overall performance was TRNSYS, followed by Archelios. The findings indicated that software-reliant on a PVGIS irradiation database may exhibit considerable inaccuracies, particularly when analyses are conducted for specific months or brief timeframes.

Freeman et al. [28] evaluated four prominent solar modeling tools (SAM, PVWatts, PVSyst, and PV*SOL Expert) using actual data from nine photovoltaic systems comprising three utility-scale systems and six commercial-scale systems. In this study, they broadened the findings of prior research [29, 30] that solely assessed the SAM tool. The yearly results for the tools exhibited relative errors ranging from 1.4% to -16.2%. In this instance, PVWatts appeared as an anomaly among the nine systems. Excluding the outliers, overall annual errors fell within 8%. On the hourly scale, excluding two particular outliers, all root mean square errors (RMSE) remained below 7% for all tools and systems. This study presented disaggregated errors for various PV systems, rendering it impossible to determine which tool was superior.

De Souza Silva et al. [31] recently performed a comparison analysis of various photovoltaic power simulation software packages (PVsyst, PV*SOL, and HOMER) utilizing production data from the 336.96 kWp photovoltaic plant at the Campinas University (Brazil). The findings indicated

that PV*SOL yielded more conservative outcomes, whilst HOMER produced the most optimistic results. Based on the annual discrepancies between the software tools' outputs and the actual values of the PV plant, PVsyst demonstrated superior performance, with an error rate of 1.02%. In comparison, HOMER was followed by an error of 2.04%. PV*SOL had a greater, albeit more cautious, error of -10.38%.

Converting solar radiation measurements from a horizontal to a tilted surface (and vice versa) is a typical difficulty in solar energy system design. Many models deal with this problem; the first findings date back to [23], which presuppose an isotropic in the sky hemisphere. However, this anisotropic aspect of diffuse irradiance has been a well-known source of inaccuracy associated with that assumption [32-34]. To address this issue, various anisotropic models have been proposed [35-38]. [39, 40] have identified a Perez model (Perez et al., 1990, 1988, 1987, 1986) [41-44] as the most popular and ubiquitous transposition model (for all locations). The Perez model has been employed in numerous studies for various applications. For instance, [45] optimized PV array outputs by modifying the orientation and tilt of the array, while [46] employed the model to analyze building-integrated photovoltaic systems.

Figure 1 shows that the diffuse part of global solar radiation comprises the circumsolar-diffused part, isotropic diffuse radiation from the sky, diffuse irradiation from the horizon, and radiation reflected off the ground (albedo).

To calculate the irradiation on a tilted plane, the two models presented in this analysis consider distinct components that comprise the diffuse radiation.

In 1979, Hay [35] introduced a model to calculate the scattered portion of the radiation on inclined surfaces based on the radiation data obtained on a horizontal surface. The model posits the concept of the transposition factor has two distinct parts. One of the models takes into account the isotropic background. In contrast, the other model includes

a circumsolar component that varies in inverse proportion to the zenith angle (θ_z) as well as the angle of incidence of the direct irradiation (θ_i). Hay has established an index, which is assigned to every component that is considered in the model.

The index is determined by dividing the direct irradiation incident on the horizontal axis (HBI) by the extraterrestrial irradiance (I_o). This factor is designed to convert the impact of the atmosphere on the scattered sunlight, resulting in a decrease in interference on days with clear skies. As a result, the index typically approaches a value of 1. Therefore, the circumsolar component is more critical, as demonstrated in equation (1).

$$TDI = HDI \left[\left(\frac{HBI}{I_o} \right) \frac{\cos \theta_i}{\cos \theta_z} + \left(1 - \frac{HBI}{I_o} \right) \frac{1 + \cos \beta}{2} \right] \quad (1)$$

Where HDI is horizontal diffuse Irradiation, TDI is tilted surface diffuse irradiation, and β indicates the solar panels' inclination angle regarding the ground.

Despite its widespread use in simulation software for solar systems, Hay's model lacks consideration for a critical component termed the horizon, which significantly contributes to the quantity of diffuse radiation on an inclined plane. Thus, Perez's model incorporated the impact of the horizon within the model [11]. Furthermore, the previous models created by Perez failed to consider some scenarios, such as the amplification of circumsolar components in atmospheres with a high fraction of diffuse radiation combined with extreme brightness [47].

The suggested model can be utilized for a broader range of celestial arrangements, in which the horizon and circumsolar elements that intersect with the isotropic dome are assigned specific weights represented by its coefficients F_1 and F_2 , respectively. The coefficients fluctuate based on the degree of clarity, brightness index, and empirical coefficients [44].

In 1987, Perez introduced a revised and streamlined version in which the coefficients. F'_1 and F'_2 Were redefined. The component F_2 can have negative values, indicating the trade-off between the radiance of the horizon and the brightness directly overhead. The phenomenon of horizon brightness is typically observed on days with clear skies, while the zenith component is observed on foggy days [43]. In 1990, modifications were implemented to ensure that the clarity index is not influenced by the angle of the zenith [41].

The finalized version of the model quantifies the scattered radiation on the inclined surface using equation (2).

$$TDI = HDI \left[((1 - F'_1)) \left(\frac{1 + \cos \beta}{2} + F'_1 (a/b) + F'_2 \sin \beta \right) \right] \quad (2)$$

Where a and b are $a = \max(0, \cos \theta)$, $b = \max(0.087, \cos \Phi)$, F'_1 and F'_2 are horizon/zenith and circumsolar

The PV syst. software includes two transposition modules, Hay and Perez, which calculate the global irradiation

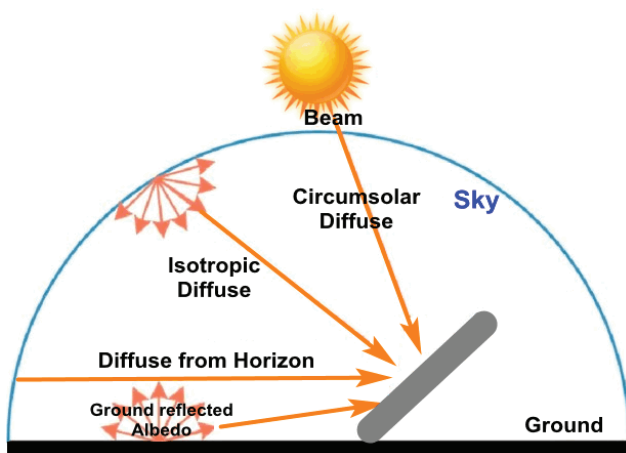


Figure 1. Diffuse component of the global solar radiation.

incidence on the collector plane (GlobInc) based on the global horizontal irradiation (GlobHor).

The transposition factor (TF) quantifies the correlation between the global radiation on the tilted plane (TDI) and the global irradiation on a horizontal plane HBI [39]. In simulation, the term “GlobInc” is the equivalent of “TDI” in theoretical terms, whereas “GlobHor” is the equivalent of “HBI”. The metric that carries more significance in computing the transposition factor (TF) is the (GlobInc), as the (GlobHor) remains constant at all locations.

No authors have utilized the Perez and Hay models through simulation with PV Syst software to analyze the impact of each model on the TF, GlobInc, and energy production in a solar system in Baghdad city. Simulations for two models were conducted with the same system power of 2840 Wp.

The paper structure follows design methodology in Section 2, results, and discussion in Section 3. Finally, the conclusions of the study are presented in Section 4.

MATERIALS AND METHODS

The PV Syst software includes two transposition modules (Hay and Perez) that are used to calculate the (GlobInc) from (GlobHor). The following five steps in the sequence are necessary for simulating PVSyst.

Step 1: Selection of the Site

The PVSYST has the advantage that the software developer automatically links the latitude and longitude data acquired from the NASA-SSE satellite station following the installation site choice.

The site where a PV-based power plant has to be built is quite essential, as it needs to be connected to a data source, and such software requires NASA-SSE satellite data. The place of system implementation was Baghdad. The particular geographical position of Baghdad is depicted in table (1). The meteorological data that was obtained from the NASA website for the location is shown in table (2)

Where DiffHor is Horizontal diffuse irradiation

Step 2: Fixing the Azimuth and Tilt angles

As shown in Figure 2, the tilt angle of 31° is considered the optimal value for achieving maximum irradiance and zero losses concerning optimum. Furthermore, the azimuth angle is set at 0° due to the same rationale.

Step 3: Daily Consumptions

Table 3 presents the energy usage and load data of a small residence. According to the data in this table, the daily energy usage amounts to 8427 Wh, while the overall power demand is 5.995 kW.

Step 4: System Definitions

Once the location's orientation, latitude and longitude, solar radiation statistics, and electricity demand have been collected, the system can be developed, modeled, and simulated. The design and modeling processes cannot begin without first determining the components that will be used. The standalone PV system configuration is shown in figure (4). The main component characteristics necessary for the design of a standalone PV system are shown in table 4

Step 5: Choosing the Transposition Mode

The Hay and Perez models are the two transposition models available to PVSyst. for modeling irradiation. In

Table 1. Geographical location data of Baghdad City

Site	Country	Region	Latitude	Longitude	Altitude	Time Zone	Albedo
Baghdad	Iraq	Asia	33.3406°	44.4009°	45m	3GMT	0.2

Table 2. Baghdad meteorological data obtained from the NASA website

Month	GlobH	DiffHor	Values	GlobH	DiffH
	kWh/m ²	kWh/m ²	Month	kWh/m ²	kWh/m ²
January	91.8	28.5	July	217	63.9
February	112.8	30.8	August	208	53.6
March	154.4	46.2	September	166.5	45.6
April	161.7	60.6	October	123.4	41.2
May	200.9	67	November	89.4	31.2
June	226.8	58.8	December	81.2	27.3
Year				1833.9	554.7

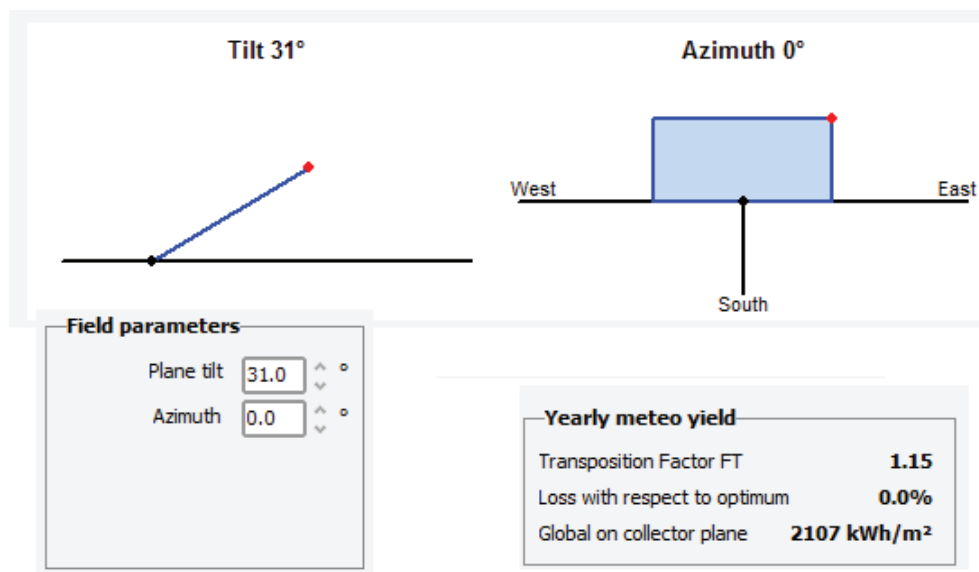


Figure 2. Azimuth and Tilt angles of PV system.

Table 3. The daily consumption of energy for the site

Number	Appliance	Power W	Use Hour/day	Energy Wh/day
8	Lamps (LED or fluo)	9W/lamp	7	504
1	TV	80W/app	5	400
3	Fan	65W/app	8	1560
2	Fridge / Deep-freeze	2.5 KWh/day	12	4999
1	Dish- and Cloth-washer	250 W aver.	2	500
1	Laptop	80W/app	4	320
Stand-by consumers		6 W lot	24	144
Total daily Energy				8427Wh/day
Monthly Energy				252.8 KWh/mth

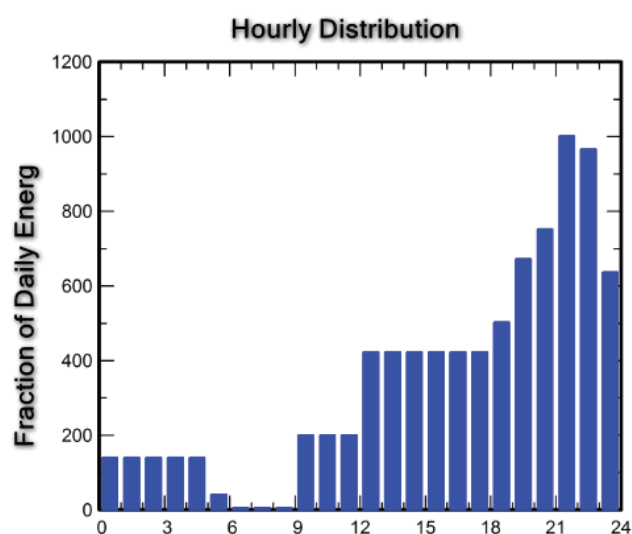


Figure 3. The daily consumption energy for the site.

PVsyst, the Perez 1990 model is the default transposition model used. Selecting a transposition model based on the PV Syst software settings is possible. In this study, two different cases will be compared to determine the TF. Case 1 uses the Hay model, and case 2 uses the Perez model

RESULTS AND DISCUSSION

The study's site is at a longitude of 44.4009°, a latitude of 33.3406°, and an elevation of 45m meters. Data on global irradiation incidents on a horizontal plane is obtained from the NASA meteorology and solar energy website and entered into the PVSyst. Meteorological directory as shown in table (1). The simulation will be for Standalone system with batteries with a system power of 2840 Wp

Two simulations utilizing PV Syst software were conducted employing the Perez and Hay transposition model to compute the irradiance on the plane of the array based

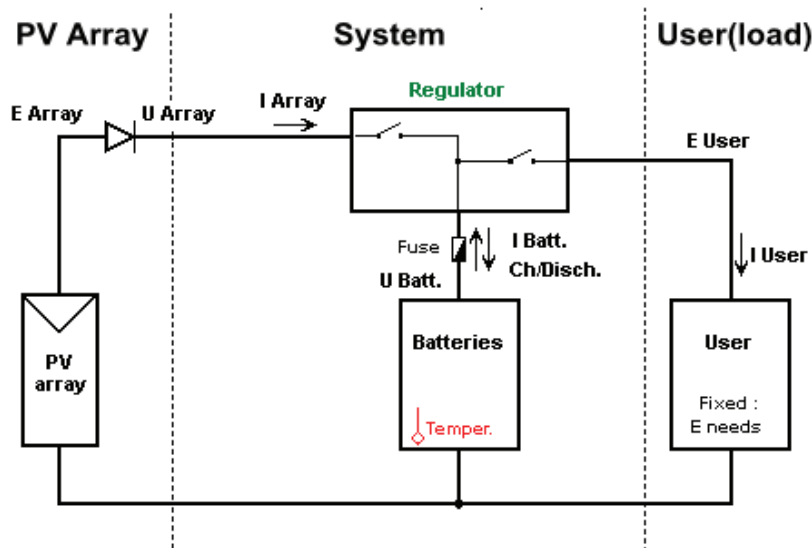


Figure 4. Standalone PV system configuration.

Table 4. illustrates the main component characteristics necessary for the design of a standalone PV system

PV module		Battery	
Manufacturer	LG Electronics	Manufacturer	Electrona
Model	LG 355 N1K-N6	Model	6B50 - Starting
(Original PVsyst database)		Technology	Lead-acid, sealed
Unit Nom. Power	355 Wp	Nb. of units	10 in parallel x 4
Number of PV modules	8 units	Discharging min. SOC	20.0 %
Nominal (STC)	2840	Stored Energy	16.9 Kwh
Modules	4 Strings x 2 In series		
At operating cond. (50°C)		Battery Pack Characteristics	
Pmpp	2605 Wp	Voltage	48 V
U mpp	62V	Nominal Capacity	440 Ah (C10)
I mpp	42A	Temperature	Fixed 20°C
Controller		Total PV Power	
Universal controller		Nominal (STC)	3 Kw
Technology	MPPT converter	Total	8 modules
Temp coeff.	-5.0 mV/°C/Elem.	Module area	14.7 m²

on the monthly and yearly measurements of GlobInc, given that GlobHor remains constant at the exact location. The comparison will assess TF by utilizing each model to see which model influences the output of E_Avail energy.

Case 1: Hay Model

Select the Hay Model in the PV Syst software from the project settings. The simulation results for this case study about the main components of global irradiation are presented in Table 5.

As seen in Table 5, GlobInc is the sum of Beam incident on the collector plane (BeamInc), circumsolar incident on the collector plane (CircInc), sky diffuse incident on the collector plane (DifSIInc), and albedo incident on the collector plane (Alb_Inc). At the same time, the horizon band in the vertical plane (HBndVrt) is 0 kWh/m².

The yearly and monthly main essential results, as shown in table (6)

From table (6), the effective global radiation in the collection plane (GlobEff) is 1957.3 kWh/m². The system will

Table 5. Main components of the global irradiation for case 1

	GlobInc	BeamInc	CircInc	DifSIInc	Alb_Inc	HBndVrt
	kWh/m²	kWh/m²	kWh/m²	kWh/m²	kWh/m²	kWh/m²
January	132.2	83	18.06	29.82	1.31	0
February	151.9	104.7	20.05	25.53	1.612	0
March	180.2	119.9	21.92	36.2	2.205	0
April	164.1	99.3	18.99	43.42	2.309	0
May	186.1	115.4	20.94	46.85	2.869	0
June	200.2	142.4	21.76	32.78	3.239	0
July	196.5	134.5	22.01	36.87	3.099	0
August	204.6	140.4	23.63	37.58	2.971	0
September	186.2	130.4	22.02	31.36	2.378	0
October	150.9	87.1	19.98	42.08	1.762	0
November	124.2	75.5	16	31.48	1.277	0
December	124	80.5	16.78	25.6	1.16	0
Year	2001.1	1313.2	242.16	419.56	26.193	0

Table 6. Annual and monthly main significant results for case 1

	GlobHor	GlobInc	TF	GlobEff	E_Avail
	kWh/m²	kWh/m²	ratio	kWh/m²	kWh
January	91.8	132.2	1.441	130	332.7
February	112.8	151.9	1.346	149.3	377.5
March	154.4	180.2	1.167	176.8	438.1
April	161.7	164.1	1.015	160.2	385
May	200.9	186.1	0.926	181.2	426.4
June	226.8	200.2	0.883	194.9	449.4
July	217	196.5	0.906	191.4	436.5
August	208	204.6	0.983	199.7	455.9
September	166.5	186.2	1.118	182.3	420.3
October	123.4	150.9	1.223	147.9	355.4
November	89.4	124.2	1.389	121.9	301.5
December	81.2	124	1.527	121.8	308.5
Year	1833.9	2001.1	1.091	1957.3	4687.2

produce an annual energy output of 4687.2 kWh; when dividing GlobInc by GlobHor, the resulting ratio is the transposition factor (TF)

Case 2: Perez Model

Select the Perez Model from the project settings in the PV Syst software. The simulation results for this case study about the main components of global irradiation are presented in Table 7.

As shown in table (8), yearly and monthly main significant results

From table (8), the effective global radiation in the collection plane (GlobEff) is 2003.9 kWh/m². The system will produce an annual energy output of 4792.7kWh when dividing GlobInc by GlobHor; the resulting ratio is the transposition factor (TF)

The outcomes show that the significant factors influencing solar power model predictions the (TF), (E_Avail (GlobHor), (HBndVrt), and Global Incident Irradiation (GlobInc)—are inherently interdependent. The efficacy as well as of a solar power system is determined by the cumulative effects of each parameter on the others.

Table 7. Main components of the global irradiation for case 2

	GlobInc	BeamInc	CircInc	DifSIInc	Alb_Inc	HBndVrt
	kWh/m²	kWh/m²	kWh/m²	kWh/m²	kWh/m²	kWh/m²
January	135.7	83	22.46	28.89	1.31	2.856
February	155.2	104.7	22.87	25.97	1.612	4.427
March	184.8	119.9	28.88	33.79	2.205	5.665
April	167.9	99.3	28.75	37.5	2.309	5.226
May	189.9	115.4	32.02	39.6	2.869	6.941
June	204.3	142.4	27.88	30.73	3.239	8.035
July	200.7	134.5	30.33	32.71	3.099	8.056
August	209	140.4	31.11	34.58	2.971	7.397
September	190.3	130.4	27.1	30.43	2.378	6.152
October	155.5	87.1	28.5	38.11	1.762	3.752
November	127.3	75.5	20.64	29.93	1.277	2.375
December	126.7	80.5	19.53	25.47	1.16	2.667
Year	2047.2	1313.2	320.07	387.71	26.193	63.551

Table 8. Annual and monthly main important results for case 2

	GlobHor	GlobInc	TF	GlobEff	E_Avail
	kWh/m²	kWh/m²	ratio	kWh/m²	kWh
January	91.8	135.7	1.479	133.4	341.5
February	112.8	155.2	1.375	152.5	385.2
March	154.4	184.8	1.197	181.3	448.6
April	161.7	167.9	1.038	164.2	393.7
May	200.9	189.9	0.945	185.2	435
June	226.8	204.3	0.901	199	458.2
July	217	200.7	0.925	195.6	445.5
August	208	209	1.005	204.2	465.5
September	166.5	190.3	1.143	186.4	429.1
October	123.4	155.5	1.26	152.5	366
November	89.4	127.3	1.424	125	309.3
December	81.2	126.7	1.559	124.4	315.2
Year	1833.9	2047.2	1.116	2003.9	4792.7

Baseline solar irradiance on a horizontal plane is represented by GlobHor, while solar radiation collected on the tilted collector plane is defined by GlobInc.

HBndVrt, as a component of the Perez model, consistently exhibits more GlobInc than the Hay model. In the Perez model, GlobHor and GlobInc possess 1833.9 kWh/m² and 2047 kWh/m², respectively. GlobHor measures 1833.9 kWh/m² compared to the Hay model, while GlobInc measures 2001 kWh/m².

As a result of HBndVrt direct addition to the GlobInc for the Perez model, both TF and E_Avail are increased. Reduced GlobInc and somewhat diminished system

performance are the outcomes of HBndVrt's exclusion from the Hay model. In the Perez model, GlobInc is 126.7 kWh/m² in December, while HBndVrt is 2.667 kWh/m². Contrasted with GlobInc 124 kWh/m², HBndVrt is 0 kWh/m².

An increase in GlobInc leads to an increase in TF, improving the irradiation that may be used to produce energy. When HBndVrt is included in the Perez model, the TF is marginally greater than in the Hay model. Annual TF in Hay is 1.091, while in Perez, it is 1.116.

As shown in figure (5), the GlobHor, GlobInc, and TF for Perez and Hay models

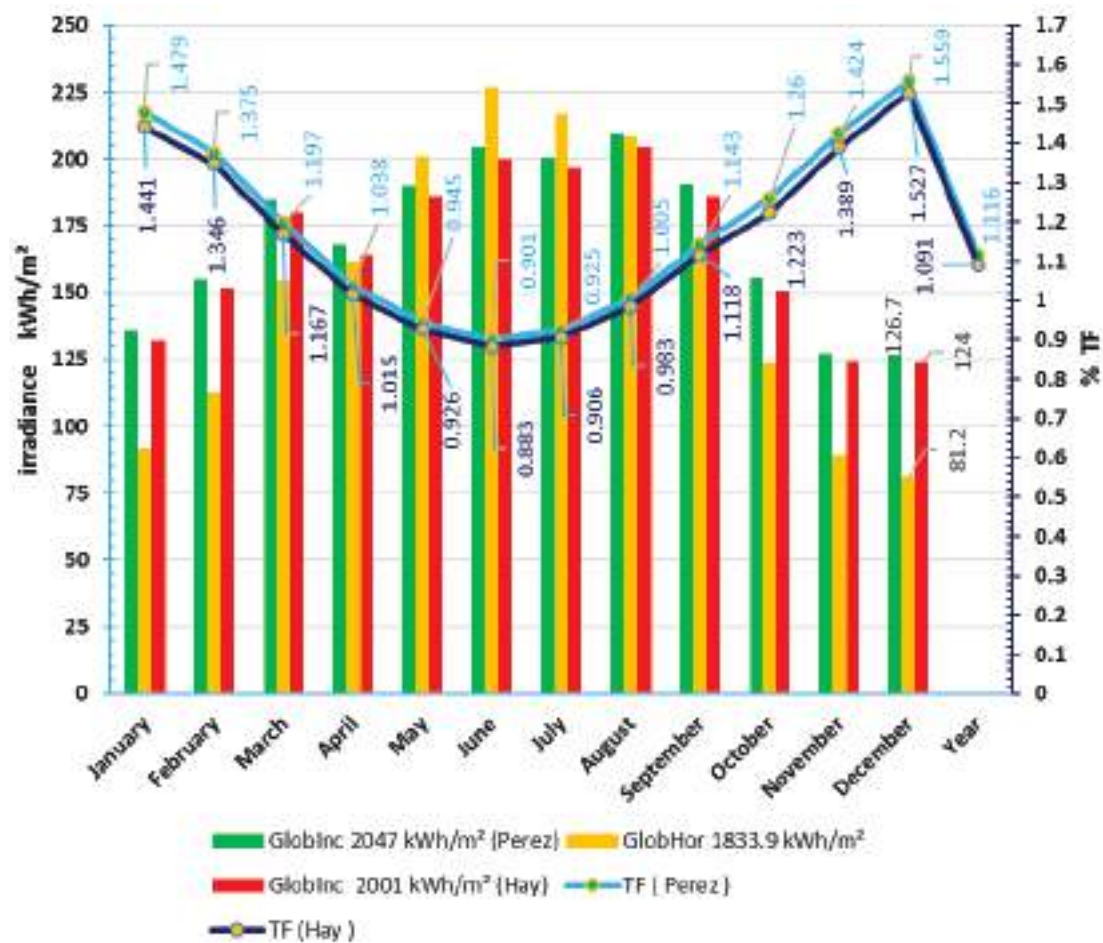


Figure 5. The monthly and yearly of GlobHor, GlobInc, and TF for Perez and Hay models.

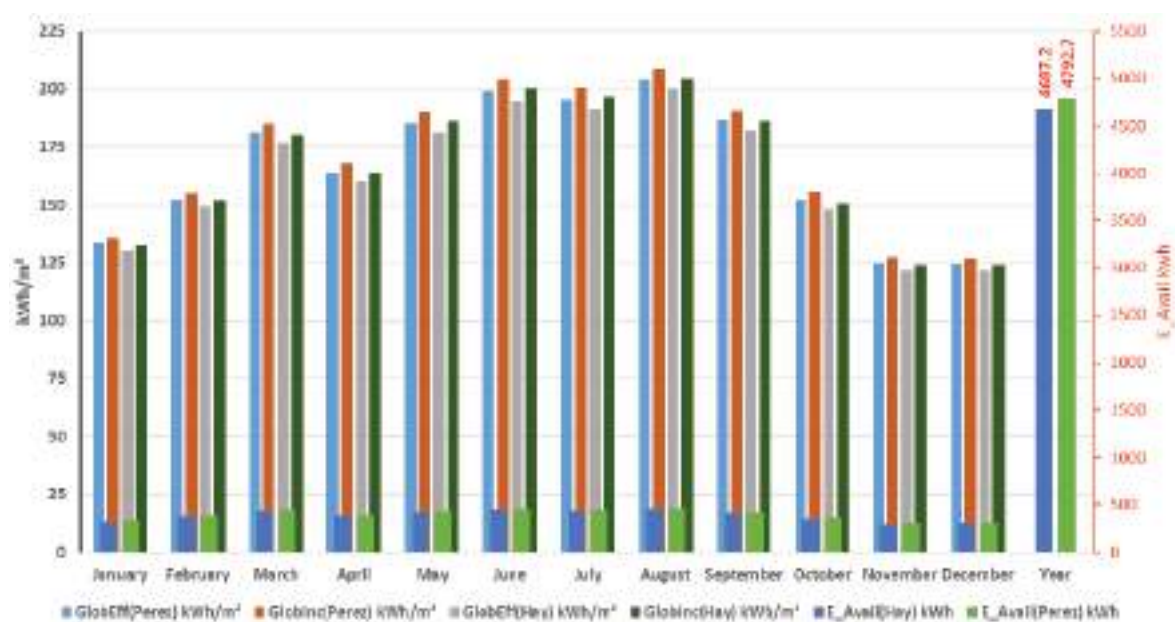


Figure 6. The monthly GlobEff and GlobInc concerning E_Avail yearly for Perez and Hay models.

Figure 6 shows the effect of increasing GlobInc, GlobEff on increasing E_Avail in the Perez model compared with the Hay model

For January, Perez had TF is 1.479 and GlobInc is 135.7 kWh/m², whereas Hay's model E_Avail was directly reliant on GlobInc with TF is 1.441 and GlobInc Equal 132.2 kWh/m². If all other system parameters stay the same, an increase in GlobInc will increase E_Avail.

E_Avail is marginally more in the Perez model than in the Hay model because of the larger GlobInc caused by HBndVrt.

Perez model gives an annual E_Avail of 4,792.7 kWh. While the Hay model is 4687.2 kWh.

During the winter months, such as January and December, the higher TF compensates to the lower GlobHor, thereby maintaining an acceptable GlobInc as well as E_Avail.

These months, GlobInc experiences a discernible influence from HBndVrt. In December, the Perez model had GlobHor is 81.2 kWh/m², TF is 1.559, HBndVrt is 2.667 kWh/m², GlobInc is 126.7 kWh/m², and E_Avail is 315.2 kWh, while the Hay model had GlobHor is 81.2 kWh/m², TF is 1.527, HBndVrt is 0 kWh/m², GlobInc is 124 kWh/m², and E_Avail is 308.5 kWh.

In summer months, such as June and July, the TF experiences a slight decrease due to the increased solar angles. However, the GlobHor increases considerably, ensuring that the GlobInc and E_Avail remain high. For instance, in June, the Perez model yielded GlobHor is 226.8 kWh/m², TF is 0.901, HBndVrt is 8.035 kWh/m², GlobInc is 204.3 kWh/m², and E_Avail is 458.2 kWh, while the other model yielded GlobHor is 226.8 kWh/m², TF is 0.883, HBndVrt is 0 kWh/m², GlobInc is 200.2 kWh/m², and E_Avail is 449.4 kWh. Within the Hay paradigm

CONCLUSION

This study analyzes the impact on essential solar performance metrics, such as TF, GlobInc, and E_Avail, assessed on a monthly and annual basis. The findings indicate that incorporating the horizon band component (HBndVrt) into the Perez model enhances transposition accuracy and produces higher energy yield estimates when compared to the more straightforward Hay model.

The Perez model more effectively predicts sunlight conditions in Baghdad, where haze and dust are common, by providing a more accurate representation of scattered light and the horizon. This underscores the significance of choosing a transposition model that accurately represents the local climate in the design or simulation of photovoltaic systems.

Monthly data reveal how TF compensates for seasonal changes in solar angles, especially in winter months, maintaining energy availability even when GlobHor is reduced. Annual results further emphasize the Perez model's effectiveness, with a higher TF (1.116 vs. 1.091) and annual

energy availability (4792.7 kWh vs. lower values from Hay). These metrics are critical for accurate long-term performance assessments and system sizing.

Ultimately, this study underscores the significance of using more comprehensive sky models, like Perez, to optimize the design and prediction accuracy of solar energy systems, particularly in regions with variable diffuse radiation. The findings reinforce the value of both monthly and annual analysis scales in ensuring robust, context-sensitive solar energy planning.

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] Mehleri ED, Zervas PL, Sarimveis H, Palyvos JA, Markatos NC. A new neural network model for evaluating the performance of various hourly slope irradiation models: implementation for the region of Athens. *Renewable Energy* 2010;35:1357–1362. [\[CrossRef\]](#)
- [2] Arun S, Nagoorvali SK, Kumar KS, Mohan GS. Automation of main bearing bolt and cap loosening machine for automobile crankshaft. *Int J Mech Eng Technol* 2017;8:41–49.
- [3] Allw AS, Duhis AH, Al-Ghanimi A. An efficient of estimation the load profile analysis of photovoltaic system with different shading of local city. *Int J Power Electron Drive Syst* 2022;13:2405–2413. [\[CrossRef\]](#)

- [4] Kudish AI, Evseev EG. Prediction of solar global radiation on a surface tilted to the south. *SPIE* 7046:9–20. [\[CrossRef\]](#)
- [5] 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. [\[CrossRef\]](#)
- [6] Awogbemi O, Von Kallon DV, Kumar KS. Contributions of artificial intelligence and digitization in achieving clean and affordable energy. *Intell Syst Appl* 2024;200389. [\[CrossRef\]](#)
- [7] Riechelmann S, Schrempf M, Seckmeyer G. Simultaneous measurement of spectral sky radiance by a non-scanning multidirectional spectroradiometer (MUDIS). *Meas Sci Technol* 2013;24:125501. [\[CrossRef\]](#)
- [8] Mutlu B, Özyörük B. A research on mathematical model approaches in biomass supply chain. *Sigma J Eng Nat Sci* 2024;42:945–955. [\[CrossRef\]](#)
- [9] Kumar KS, Raju DBN, Arulmani J, Amirthalingam P. Design and structural analysis of liquified cryogenic tank under seismic and operating loading. *Int J Mech Eng Technol* 2016;7:345–366.
- [10] Klein SA. Calculation of monthly average insolation on tilted surfaces. *Solar Energy* 1977;19:325–329. [\[CrossRef\]](#)
- [11] Duffie JA, Beckman WA, Blair N. *Solar engineering of thermal processes, photovoltaics and wind*. Hoboken: John Wiley & Sons; 2020.
- [12] Badescu V. 3D isotropic approximation for solar diffuse irradiance on tilted surfaces. *Renewable Energy* 2002;26:221–233. [\[CrossRef\]](#)
- [13] Javanjal VK, Patil LN, Mahajan KA, Patil AA. Examining the role of nano-sized additives in boosting air conditioner efficiency. *Sigma J Eng Nat Sci* 2025;43:107–120. [\[CrossRef\]](#)
- [14] Kumar KS, Muniamuthu S, Tharanisrisakthi BT. An investigation to estimate the maximum yielding capability of power for mini venturi wind turbine. *Ecol Eng Environ Technol* 2022;23. [\[CrossRef\]](#)
- [15] Grant RH, Heisler GM. Obscured overcast sky radiance distributions for ultraviolet and photosynthetically active radiation. *J Appl Meteorol Climatol* 1997;36:1336–1345. [\[CrossRef\]](#)
- [16] Kumar S, Gaur A, Kamal N, Pathak M, Shrinivas K, Singh P. Artificial neural network based optimum scheduling and management of forecasting municipal solid waste generation–case study: Greater Noida in Uttar Pradesh (India). *IOP Conf Ser Mater Sci Eng* 1478:012033. [\[CrossRef\]](#)
- [17] Cheng CL, Chan CY, Chen CL. An empirical approach to estimating monthly radiation on south-facing tilted planes for building application. *Energy* 2006;31:2940–2957. [\[CrossRef\]](#)
- [18] Kumar S, Sharma SS, Giri J, Makki E, Sathish T, Panchal H. Perovskite materials with improved stability and environmental friendliness for photovoltaics. *Front Mech Eng* 2024;10:1357087. [\[CrossRef\]](#)
- [19] Burlon R, Bivona S, Leone C. Instantaneous hourly and daily radiation on tilted surfaces. *Solar Energy* 1991;47:83–89. [\[CrossRef\]](#)
- [20] Aljanabi M, Al-Kafaji MSS, Duhis AH. Optimum yearly and seasonal tilt angle of solar system in the center of Babylon/Iraq using PVsyst software. *Indones J Electr Eng Comput Sci* 2023;31:626–635. [\[CrossRef\]](#)
- [21] Kumar S, Dave V. Backpropagation algorithm-based approach to mitigate soiling from PV module. In: *Intelligent Computing Techniques for Smart Energy Systems: Proceedings of ICTSES 2018*. Singapore: Springer; 2019. p. 153–161. [\[CrossRef\]](#)
- [22] Duhis AH, Mosheer AD, Abed AM. Optimization of seasonal tilt adjustment photovoltaic system in Karbala, Iraq, by using the albedo benefit. *Indones J Electr Eng Comput Sci* 2023;32:632–642. [\[CrossRef\]](#)
- [23] Bishnoi D. Pressure exertion and heat dissipation analysis on uncoated and ceramic (Al₂O₃, TiO₂ and ZrO₂) coated braking pads. *Mater Today Proc* 2023;74:774–787. [\[CrossRef\]](#)
- [24] Connolly D, Lund H, Mathiesen BV, Leahy M. A review of computer tools for analysing the integration of renewable energy into various energy systems. *Appl Energy* 2010;87:1059–1082. [\[CrossRef\]](#)
- [25] Wijeratne WMPU, Yang RJ, Too E, Wakefield R. Design and development of distributed solar PV systems: do the current tools work? *Sustain Cities Soc* 2019;45:553–578. [\[CrossRef\]](#)
- [26] Lee GR, Frearson L, Rodden P. An assessment of photovoltaic modelling software using real world performance data. CCH Hamburg, Germany: 4339–4343.
- [27] Axaopoulos PJ, Fylladitakis ED, Gkarakis K. Accuracy analysis of software for the estimation and planning of photovoltaic installations. *Int J Energy Environ Eng* 2014;5:1–7. [\[CrossRef\]](#)
- [28] Freeman J, Whitmore J, Blair N, Dobos AP. Validation of multiple tools for flat plate photovoltaic modeling against measured data. *IEEE*:1932–1937.
- [29] Freeman J, Whitmore J, Kaffine L, Blair N, Dobos AP. System advisor model: flat plate photovoltaic performance modeling validation report. Golden (CO): National Renewable Energy Lab (NREL); 2013. [\[CrossRef\]](#)
- [30] Soni P, Dave V, Kumar S, Paliwal H. A comparative study of AI-driven techno-economic analysis for grid-tied solar PV-fuel cell hybrid power systems. *Sci Temper* 2024;15:2248–2257. [\[CrossRef\]](#)
- [31] de Souza Silva JL, Costa TS, de Melo KB, Sakô EY, Moreira HS, Villalva MG. A comparative performance of PV power simulation software with an installed PV plant. *IEEE*:531–535. [\[CrossRef\]](#)
- [32] Ma CCY, Iqbal M. Statistical comparison of models for estimating solar radiation on inclined surfaces. *Solar Energy* 1983;31:313–317. [\[CrossRef\]](#)

- [33] Kumar S, Supriya J, Sharma SS, Paliwal H, Manikanta G, Giri J, et al. Developments, challenges, and projections in solar battery charging in India. *Results Eng* 2024;103248. [\[CrossRef\]](#)
- [34] Thirunavukkarasu M, Sujit K, Ramadass S, Manihoman MP, Elahi Mohammad SM, Rezaul KM, et al. Feature of SiO₂ concentration on solar thermal functional characteristics of flat plate solar collector. *Therm Sci* 2024;28:249–255. [\[CrossRef\]](#)
- [35] Hay JE. Calculation of monthly mean solar radiation for horizontal and inclined surfaces. *Solar Energy* 1979;23:301–307. [\[CrossRef\]](#)
- [36] Klucher TM. Evaluation of models to predict insolation on tilted surfaces. *Solar Energy* 1979;23:111–114. [\[CrossRef\]](#)
- [37] Skartveit A, Olseth JA. Modelling slope irradiance at high latitudes. *Solar Energy* 1986;36:333–344. [\[CrossRef\]](#)
- [38] Temps RC, Coulson KL. Solar radiation incident upon slopes of different orientations. *Solar Energy* 1977;19:179–184. [\[CrossRef\]](#)
- [39] Matagne E, El Bachtiri R. Exact analytical expression of the hemispherical irradiance on a sloped plane from the Perez sky. *Solar Energy* 2014;99:267–271. [\[CrossRef\]](#)
- [40] Myers DR. Solar radiation resource assessment for renewable energy conversion. Golden (CO): National Renewable Energy Lab (NREL); 2012. [\[CrossRef\]](#)
- [41] Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy* 1990;44:271–289. [\[CrossRef\]](#)
- [42] Perez R, Stewart R, Seals R, Guertin T. The development and verification of the Perez diffuse radiation model. Albuquerque (NM): Sandia National Lab (SNL-NM); 1988. [\[CrossRef\]](#)
- [43] Perez R, Seals R, Ineichen P, Stewart R, Menicucci D. A new simplified version of the Perez diffuse irradiance model for tilted surfaces. *Solar Energy* 1987;39:221–231. [\[CrossRef\]](#)
- [44] Perez R, Stewart R, Arbogast C, Seals R, Scott J. An anisotropic hourly diffuse radiation model for sloping surfaces: description, performance validation, site dependency evaluation. *Solar Energy* 1986;36:481–497. [\[CrossRef\]](#)
- [45] Khoo YS, Nobre AM, Malhotra R, Yang D, Rüther R, Reindl T, et al. Optimal orientation and tilt angle for maximizing in-plane solar irradiation for PV applications in Singapore. *IEEE J Photovoltaics* 2013;4:647–653. [\[CrossRef\]](#)
- [46] Zomer C, Nobre A, Cassatella P, Reindl T, Rüther R. The balance between aesthetics and performance in building-integrated photovoltaics in the tropics. *Prog Photovoltaics Res Appl* 2014;22:744–756. [\[CrossRef\]](#)
- [47] Da Silva MK, Narvaez DI, de Melo KB, Costa TS, de Siqueira TG, Villalva MG. Comparative analysis of meteorological databases and transposition models applied to photovoltaic systems. 1.



Research Article

Urban morphology and planning from an entropy perspective: A bibliometric evaluation

Esra BAYAZIT^{1,*}

¹Department of Landscape Architecture, Faculty of Fine Arts and Design, Siirt University, Siirt, 56100, Türkiye

ARTICLE INFO

Article history

Received: 12 March 2025

Accepted: 05 May 2025

Keywords:

Bibliometric Analysis; Entropy;
Scientific Production; Spatial
Entropy; Urban Planning; Urban
Sprawl

ABSTRACT

This study aims to analyze the theoretical and practical reflections of the concept of entropy in the urban studies literature by tracing its conceptual and methodological evolution across two distinct domains. To achieve this, two distinct literature sets retrieved from the Web of Science database Urban Morphology & Entropy (n=56) and Spatial Entropy & Urban Planning (n=643) were comparatively analyzed using bibliometric methods. The analysis was conducted with Bibliometrix and Biblioshiny tools in the RStudio environment, visualizing annual publication output, keyword distributions, citation patterns, thematic maps, collaboration networks, and trend analyses. The findings indicate that China is the most productive country, contributing 35% of the publications within the ‘Spatial Entropy & Urban Planning’ category, which encompasses broader themes such as environmental sustainability, data analytics, and decision support systems. Frequently used keywords across the literature include ‘urban morphology,’ ‘entropy,’ ‘spatial metrics,’ and ‘sustainable development.’ While entropy in the urban morphology literature is primarily associated with structural and formal analysis, it is linked to data science, artificial intelligence, and decision-support mechanisms in urban planning contexts. The unique contribution of this study lies in revealing how the concept of entropy is positioned across different theoretical frameworks and how it is operationalized in urban research at a meta-level. By examining two distinct thematic areas with an interdisciplinary approach, the study traces the evolution of entropy-based research in the urban studies field. Based on the results, it is recommended that future studies enhance database diversity, strengthen interdisciplinary collaboration, and deepen spatial modelling applications.

Cite this article as: Bayazite. Urban morphology and planning from an entropy perspective: A bibliometric evaluation. Sigma J Eng Nat Sci 2025;43(5):1451–1472.

INTRODUCTION

Urban morphology is an interdisciplinary field that examines the physical form, structural patterns, and spatial organization of cities [1-4]. It is regarded as a fundamental

approach to understanding how urban environments are shaped and transformed, as well as the social, economic, and environmental implications of these changes [5]. By analyzing the evolution of cities, their textures, street networks, building configurations, and functional zones,

*Corresponding author.

*E-mail address: esra.bayazit@siirt.edu.tr

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



urban morphology contributes both theoretically and practically to urban design and planning processes [1,6].

However, scholars have noted persistent challenges in achieving consensus on the conceptual foundations and measurement indicators of urban morphology. For instance, Zhang et al. [7] emphasized the need to incorporate ecological considerations such as natural features and ecosystem health. Similarly, [4] proposed an integrative approach that considers both spatial form and function. In this sense, urban morphology remains a critical discipline for understanding, comparing, and sustainably reshaping the physical structure of cities. With the rise of technological advancements and interdisciplinary perspectives, its role in building more livable and sustainable cities has become increasingly important. Nevertheless, the growing spatial complexity of contemporary urban areas demands more advanced analytical tools than traditional morphological methods can provide [6,8-10]. Formal analyses alone are insufficient to capture and quantify the rising structural complexity, underscoring the need for new tools that can quantitatively measure spatial regularity and disorder.

In this context, entropy has emerged as a powerful analytical concept for evaluating urban disorder and spatial heterogeneity [11,12]. Originally developed in the field of thermodynamics, entropy is a universal metric that assesses the degree of disorder or uncertainty in a system, and it has since been adapted across multiple disciplines, including information theory, statistics, probability, machine learning, and urban planning [13-15]. Notably, Shannon's formulation of entropy [16,17] provided a mathematical framework to compute information density and uncertainty levels within systems. For example, Fagerholm et al. [17] defined entropy as a measure of hidden information in a system, shaped by the observer's limitations. Meanwhile, [15] explored entropy as a tool in planning decision-support systems, particularly for representing spatial heterogeneity, land-use diversity, and the irregularity of spatial distributions.

Entropy has been increasingly adopted in spatial analyses to examine the structural complexity and distributional characteristics of urban areas [11,12,18-20]. In urban morphology, entropy has proven useful for analyzing spatial information density, land-use diversity, and urban compactness [21]. In particular, spatial statistical tools such as Shannon Entropy, Moran's I, and Geary's C have been used to evaluate spatial structure, density, and diversity in urban contexts [22-24]. As Jat et al. [25] argued, entropy provides not only a descriptive function but also a comparative and predictive analytical advantage, particularly when used alongside geospatial techniques and remote sensing data.

Recent studies have expanded the scope of entropy by integrating it into digital planning processes, smart city frameworks, and decision-support systems [26,27]. For example, Xiong et al. [28] applied entropy metrics to assess urban sprawl and land-use efficiency, while Altieri et al.

[29] combined entropy with machine learning algorithms to identify spatial anomalies. Despite its growing relevance, however, existing literature remains fragmented, with entropy applied across a wide range of conceptual frameworks and disciplines.

This study aims to address this gap by conducting a comparative bibliometric analysis of two distinct bodies of literature: (1) Urban Morphology & Entropy [n=56]; (2) Spatial Entropy & Urban Planning [n=643].

The primary objective is to evaluate how the concept of entropy has evolved both theoretically and methodologically across these domains. By systematically comparing these two streams, the study reveals the thematic patterns, collaboration networks, and citation structures that shape scholarly discourse. In doing so, it offers both a conceptual mapping of entropy-based urban research and practical insights into its future applications [30-36].

MATERIALS AND METHODS

This study uses bibliometric analysis to analyse academic publications on urban morphology, entropy, spatial Entropy, and urban planning. The Web of Science (WoS) database was selected as the research's data source. WoS is a reliable source of information frequently preferred in the literature due to its comprehensive data and high accuracy rate in academic research [37,38].

Data Collection and Selection Criteria

To analyze the thematic and methodological positioning of entropy in urban research, two separate literature sets were defined using the keyword combinations "Urban Morphology & Entropy" and "Spatial Entropy & Urban Planning". This distinction was made to comparatively examine how entropy is interpreted through different theoretical lenses. The 'Urban Morphology & Entropy' set primarily focuses on morphological and formal analyses of urban structure, while the 'Spatial Entropy & Urban Planning' set reflects a broader planning perspective that includes decision support systems, GIS integration, and technology-oriented methodologies.

To comprehensively investigate developments in the literature, the dataset was constructed using the following keyword combinations:

- (ALL=(urban morphology)) AND ALL=(entropy)
- (ALL=(spatial entropy)) AND ALL=(urban planning)

These combinations were determined to comprehensively capture studies relevant to urban morphology, spatial entropy, and urban planning. The bibliographic search was conducted on 20 February 2025, resulting in 56 documents for Urban Morphology & Entropy and 643 documents for Spatial Entropy & Urban Planning. This dual dataset approach was adopted to reveal how entropy is used differently across disciplinary boundaries and to uncover conceptual overlaps, divergences, and gaps in the literature.

Data Analysis Method

The study used RStudio software to analyse the data obtained, and bibliometric analyses were carried out with the Bibliometrix package and the Biblioshiny visualisation tool. Bibliometrix is a powerful tool that enables the analysis and visualisation of bibliometric data [31]. Biblioshiny, conversely, allows the analysis results to be presented in a more understandable and interpretable form, thanks to its interactive visualisation capabilities.

The bibliometric analysis methods used in the research are as follows [31,34,39-41]:

- Scientific Productivity Trends and Citation Dynamics (annual scientific production, annual citation averages and document types) were analysed.
- Keyword Distribution (Author's Keywords (DE), Word Cloud) and Thematic Trends were analysed.
- Citation Analysis and Co-Citation Networks (Collaboration Network) were analysed.
- Thematic Maps and Trend Analyses (Three-Field Plot, Trend Topic and Factorial Map) were analysed.

Data Processing and Visualisation

The data obtained were pre-processed using R software; missing information was cleaned, and inconsistencies were eliminated. 'Article', 'book', and 'review' categories were included as article types. Only documents written in English were included in the study, and publications in other languages were excluded from the analysis. The study

consists of publications between 1992 and 2024. Different spellings of author names were combined, and keyword normalisation was performed. To increase the reliability of the results, the analyses were repeated more than once, and the consistency of the data set was checked.

The visualised analysis results, which are visualised with the Biblioshiny tool in RStudio, are presented using the following techniques:

- Annual Scientific Production and Average Citations per Year graphs were used to analyse the number of publications and average citations per year.
- Changes in keywords over time were visualised with Word Cloud and Trend Topics visuals.
- Citation and academic collaboration networks were analysed using the Co-citation Network and Collaboration Network analyses.
- The relationships between keywords, countries, and institutions were comparatively evaluated with Three-Field Plot analyses.

This methodological framework was meticulously developed to ensure the study's scientific validity and reliability [42,43,41]. The findings obtained through visualisation techniques provide a comprehensive assessment of scientific productivity and thematic developments in urban morphology, entropy, and Spatial Entropy and urban Planning (Fig. 1).

With this method flow, scientific productivity, academic collaborations, and thematic developments in



Figure 1. Flowchart of the method.

urban morphology and entropy and Spatial Entropy and urban Planning were analysed in detail. The study's comprehensive data set and visualisation techniques will provide a methodological basis for future bibliometric analyses.

RESULTS AND DISCUSSION

A total of 709 documents were initially retrieved from the Web of Science (WoS) Core Collection. Using the Biblioshiny tool in R, a filtering process was applied to include only English-language publications up to the year 2024. Following this refinement, 619 documents under Spatial Entropy & Urban Planning and 56 documents under Urban Morphology & Entropy were selected for comparative analysis. The results are evaluated under the following subheadings:

Scientific Productivity Trends and Citation Dynamics

Table 1 presents the annual scientific production between 1992 and 2024 for the two literature sets (Urban Morphology & Entropy' and 'Spatial Entropy & Urban Planning), as well as their bibliometric characteristics [34, 42–44].

According to the findings, the field of 'Spatial Entropy & Urban Planning' is significantly superior to the field of 'Urban Morphology & Entropy' (47 documents, 5.76% growth), both quantitatively and in terms of development rate, with a total of 619 documents and an annual growth rate of 16.4%. In terms of the average number of citations, 'Spatial Entropy & Urban Planning' produced a higher impact (21.72) and also exhibited a richer literature structure in terms of total sources (246), total references (28,310) and keyword diversity (2264 Author's Keywords). However,

the 'Urban Morphology & Entropy' field has more global partnerships with 31.91% international collaboration level and presents a more specific, thematically concentrated literature. These findings show that urban planning addresses spatial entropy in a multifaceted and interdisciplinary manner. In contrast, the relationship between urban morphology and entropy has developed within a narrower but deeper academic framework (Table 1).

The results also provide insight into the academic depth and collaboration networks of each literature set. Interestingly, the Urban Morphology & Entropy literature has shown a resurgence in recent years, while Spatial Entropy & Urban Planning—despite drawing significant attention in the past—has experienced a relative decline in publication and citation activity [34,45].

The analyses presented in Figure 2 and Figure 3 show the distribution of the average annual number of citations in 'Urban Morphology & Entropy' and 'Spatial Entropy & Urban Planning' between 1992 and 2024. These analyses reveal that Urban Morphology & Entropy has been rising again recently. In contrast, the field of Spatial Entropy & Urban Planning has attracted much attention in the previous years but has recently decreased [44].

In the 'Urban Morphology & Entropy' field, although the annual average number of citations decreased in certain years after 1992, it peaked in 2015. It showed a relatively steady upward trend in the following years. This shows that the academic impact of some pioneering publications has increased over time and produced remarkable outputs in specific periods (Fig. 2).

In Figure 3, high fluctuations in the annual average citation numbers in the field of 'Spatial Entropy & Urban Planning' are observed in earlier years (especially between

Table 1. Comparative bibliometric summary of the urban morphology & entropy and spatial entropy & urban planning literature sets (1992–2024)

	Urban Morphology & Entropy	Spatial Entropy & Urban Planning
Timespan	1992:2024	1992:2024
Total Sources	44	246
Total Documents	47	619
Annual Growth Rate	%5.76	%16.4
Document Average Age	7.34	4.8
Average citations per doc	18.17	21.72
Total References	2343	28310
International co-authorships	%31.91	%24.72
Document Type		
Article	43	568
Proceedings Paper	4	38
Other (Book, Book Review, Review)	-	13
Keywords Plus (ID)	179	1160
Author's Keywords (DE)	153	2264

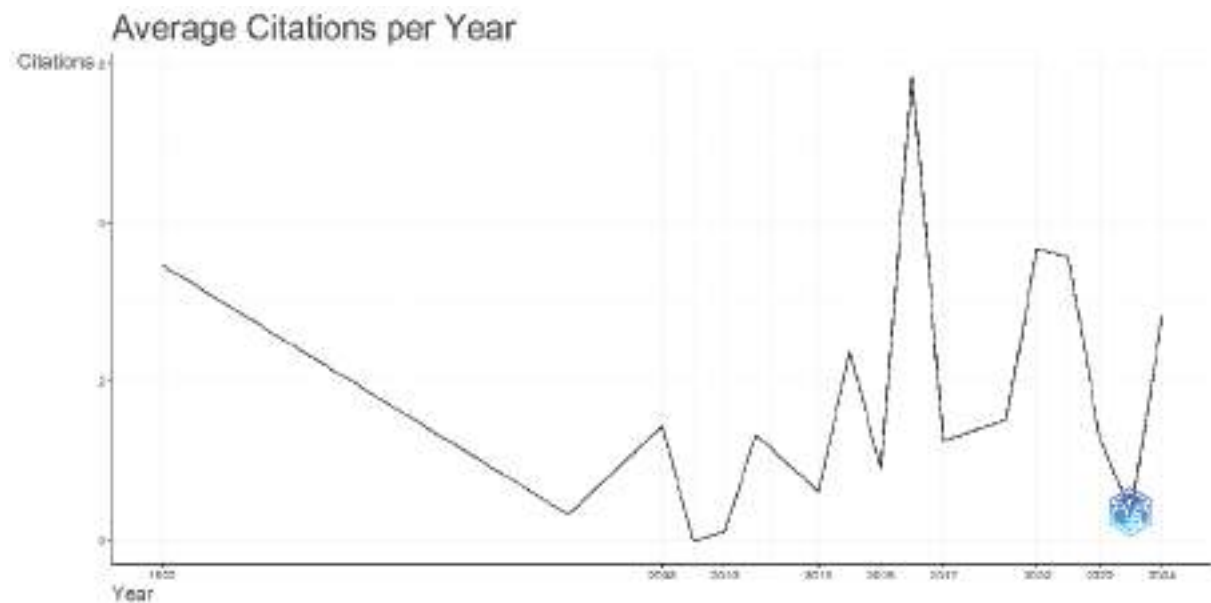


Figure 2. Annual average citation trends in the ‘Urban Morphology & Entropy’ literature (1992–2024), indicating periods of academic influence and thematic growth.

2000 and 2004), indicating that publications in this field have a broader and earlier academic resonance. However, the decline after 2015 suggests that the distribution of citation activity may be more heterogeneous despite the intensive publication production in the field. This suggests that the field is becoming increasingly saturated or that newer topics are attracting interest. The annual citation trends and metrics in the figures show that spatial entropy and urban

planning are emerging as a more dynamic and influential research area [34] (Fig. 3).

In comparative terms, ‘Urban Morphology & Entropy’ shows a trend of later development but steadier citation growth. Meanwhile, ‘Spatial Entropy & Urban Planning’ was heavily cited earlier, but this effect has diminished over time. This reveals that the two fields have different academic visibility and sustainable impact dynamics.

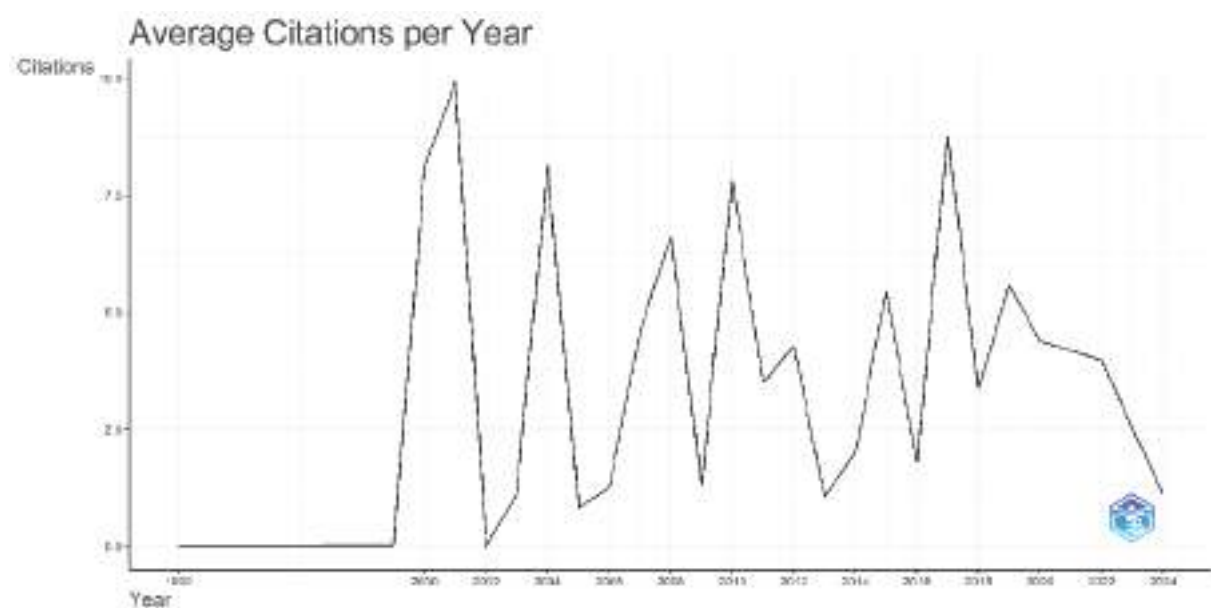


Figure 3. Annual average citation trends in the ‘Spatial Entropy & Urban Planning’ literature (1992–2024), indicating periods of academic influence and thematic growth.

According to the annual scientific production graphs presented in Figure 4 and Figure 5, different levels of publication growth are observed over time in the fields of ‘Urban Morphology & Entropy’ and ‘Spatial Entropy & Urban Planning’ [34,44].

This figure illustrates the average number of citations per year within the Spatial Entropy & Urban Planning literature. The increasing trend after 2016 indicates a growing scientific interest and methodological diversification in this field.

Annual production in ‘Urban Morphology & Entropy’ increased especially after 2015 and gained a remarkable momentum after 2020. However, the total production in this field is still limited, and the number of studies has rarely exceeded 6 per year. This situation shows that this field is mainly limited to original and in-depth research (Fig. 4).

The course of scientific production in the fields of ‘Spatial Entropy & Urban Planning’ shows a much more pronounced upward trend; this rise, which accelerated especially after 2017, continues with a dramatic increase

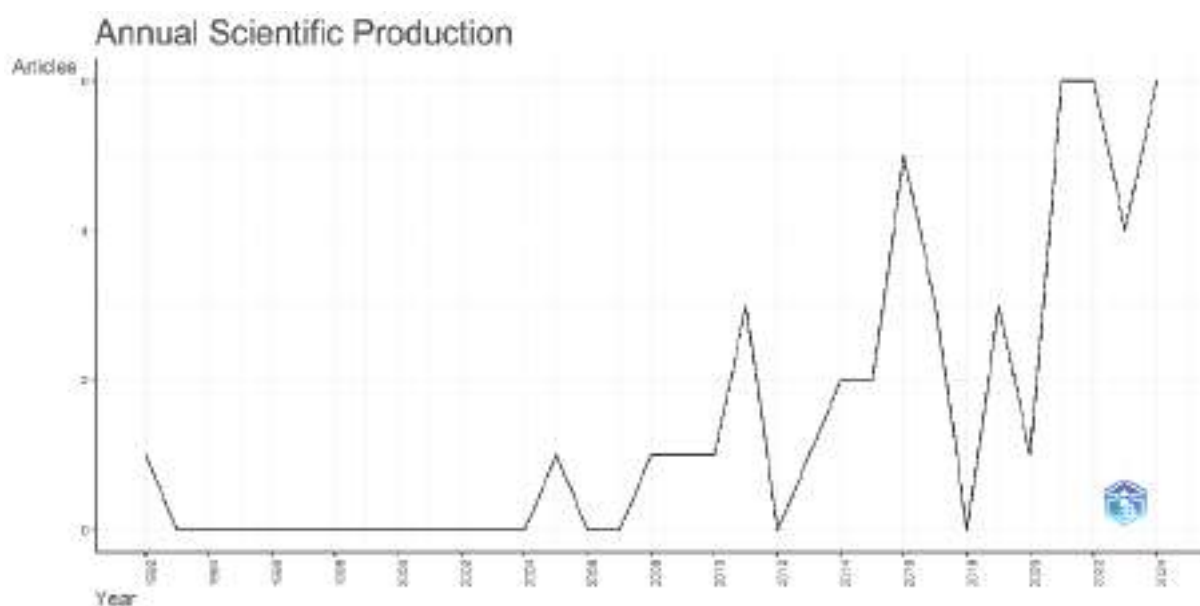


Figure 4. Annual scientific production in urban morphology & entropy (1992-2024).

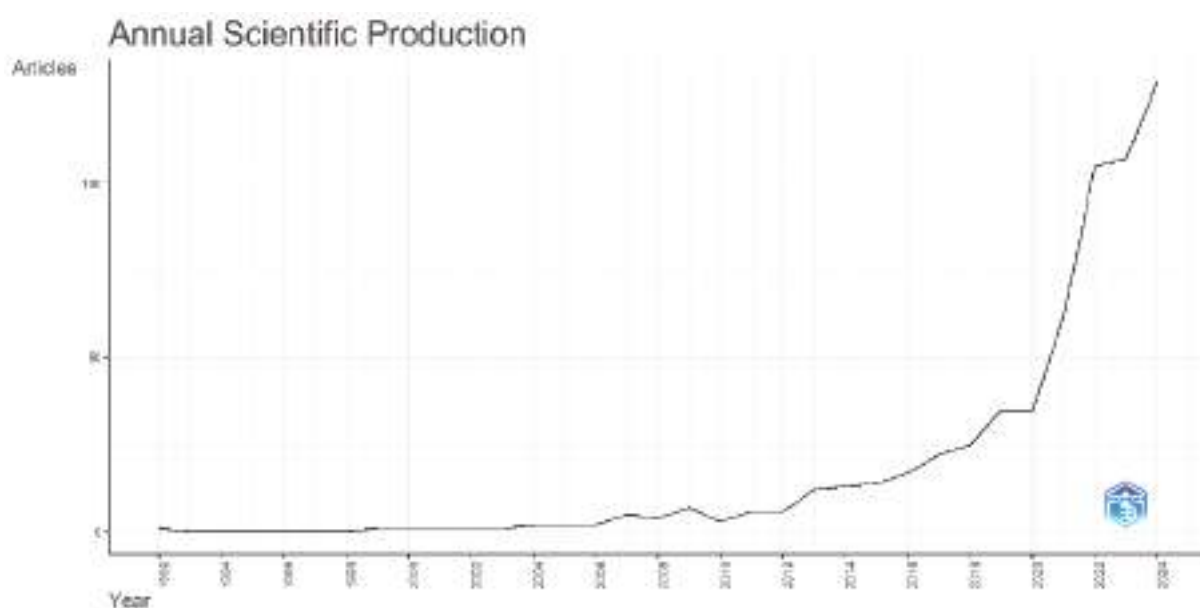


Figure 5. Annual scientific production in spatial entropy & urban planning (1992-2024).

from 2020 onwards. As of 2023, the annual number of articles in this field has exceeded 130. This shows that spatial entropy is increasingly used in urban planning literature and that this field is becoming increasingly attractive for interdisciplinary research. This difference in production between the two fields reveals that 'Urban Morphology & Entropy' corresponds to a more specific area of expertise. At the same time, 'Spatial Entropy & Urban Planning' is associated with wider-scale applications (Fig. 5).

These analyses show that Urban Morphology & Entropy remains a more specific and niche field, whereas Spatial Entropy & Urban Planning has a broader interdisciplinary impact [41,46]. When the differences between the two fields are evaluated in the context of scientific productivity, citation dynamics, international collaboration rates, and keyword distributions, they provide essential clues for interdisciplinary integration and thematic diversity in future research.

These figures compare the annual scientific output of the two literature sets. While Urban Morphology & Entropy demonstrates a steady but limited production, Spatial Entropy & Urban Planning shows a notable increase in publications after 2017, reflecting a shift toward more applied and interdisciplinary studies.

Keyword Distribution and Thematic Trends

Figures 6 and 7 visualise the most frequently used keywords in Urban Morphology & Entropy and Spatial Entropy & Urban Planning as a word cloud. This word cloud analysis visually reveals the thematic foci and common keywords of the Urban Morphology & Entropy and Spatial Entropy & Urban Planning topics, providing a comparative assessment of research trends in both fields. Keyword analysis

shows that different research trends have emerged in both subjects [46].

According to the WordCloud analysis, the most prominent themes in Urban Morphology & Entropy are 'urban morphology', 'entropy', 'morphology', 'urban', 'spatial metrics', and 'sustainability'. These expressions reveal how urban morphology is associated with the concepts of spatial disorder and entropy and how these relationships gain meaning in terms of sustainable urban structures. The density of keywords indicates that the studies focus more on theoretical and morphological analyses (Fig. 6).

It is noteworthy that more applied, technical and geographical components such as 'entropy', 'urban sprawl', 'GIS', 'China', 'land use', 'urbanisation' and 'remote sensing' come to the forefront in the studies on Spatial Entropy & Urban Planning. This indicates that spatial entropy is primarily addressed through geographic information systems, land use planning, and remote sensing technologies. In addition, advanced analysis methods such as 'machine learning', 'deep learning', and 'topsis' reveal the methodological diversity and technical depth of research in this field (Fig. 7).

As a result, although 'entropy' stands out as the basic concept in both fields, theoretical orientation is dominant in urban morphology and entropy studies, while applied and technical orientation is dominant in Spatial Entropy and urban Planning studies. This difference reflects the thematic orientation of the literature and the diversity in the research paradigms of the two fields.

The keyword association networks presented in Figure 8 and Figure 9 reveal the thematic structures and conceptual clusters in the literature on urban morphology, entropy, spatial Entropy, and urban Planning [33,34].

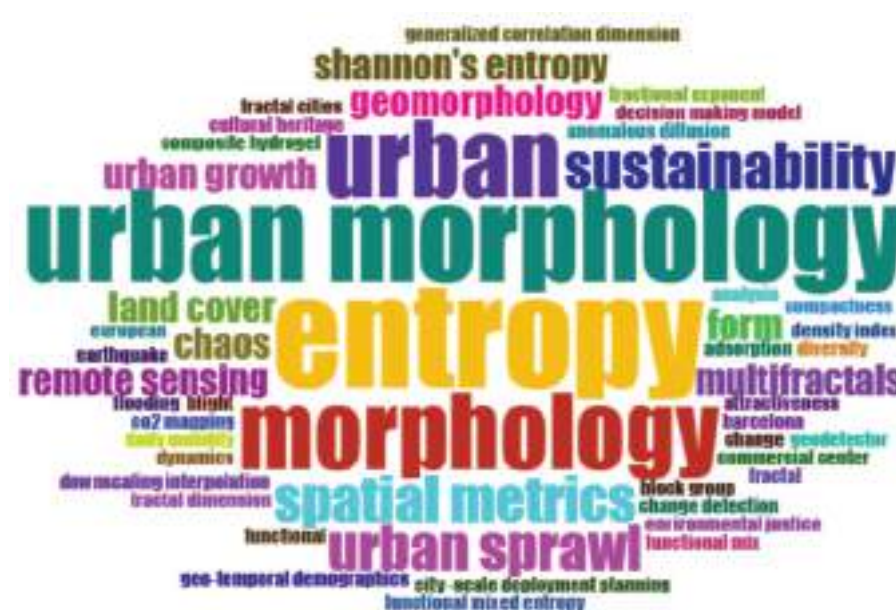


Figure 6. Keyword WordCloud analysis on urban morphology & entropy.

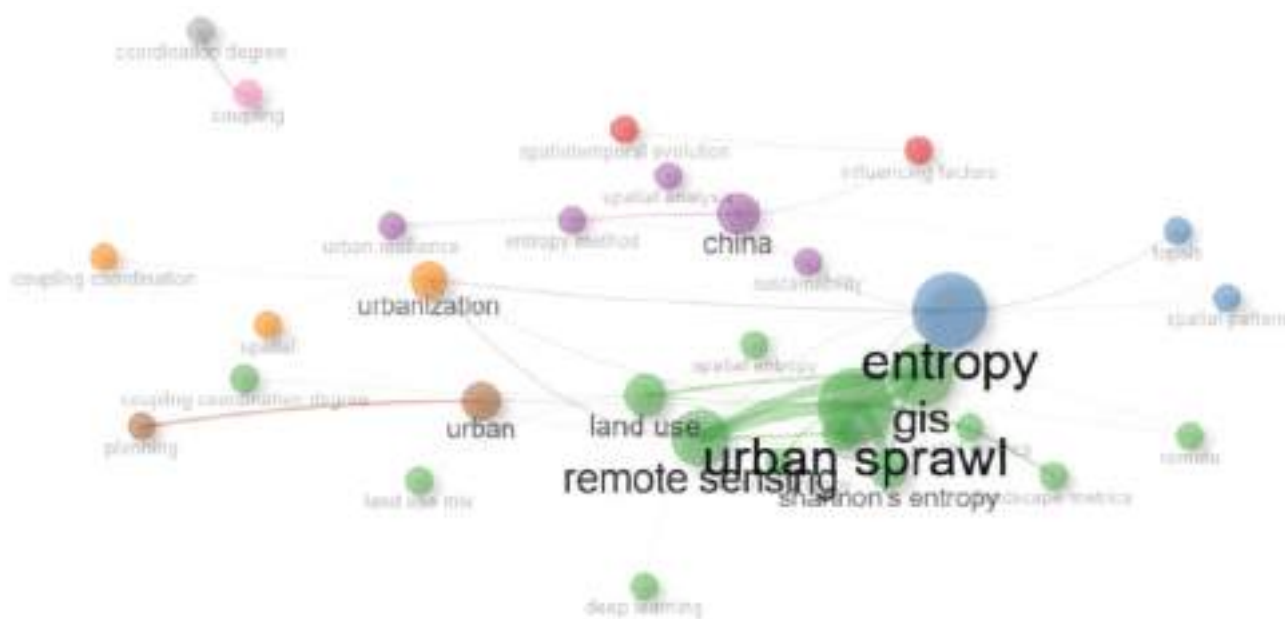


Figure 9. Keyword association network for spatial entropy & urban planning.

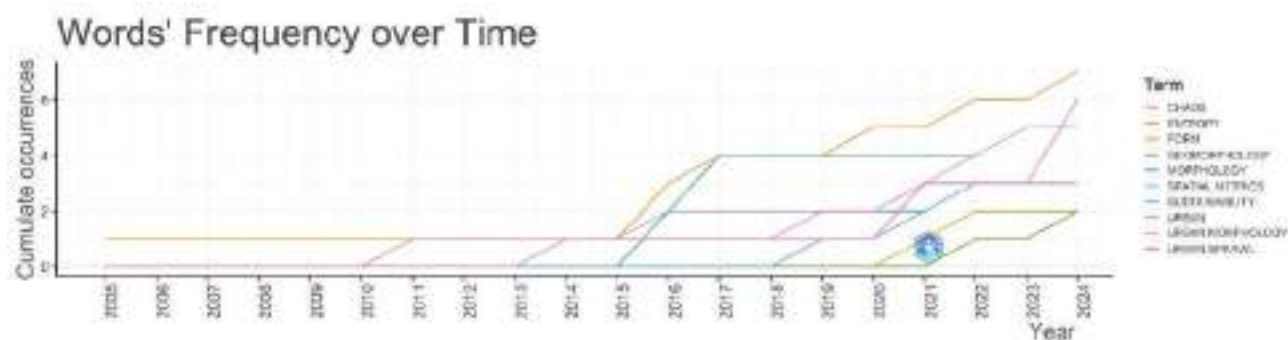


Figure 10. Frequency distribution of keywords in urban morphology and entropy over time.

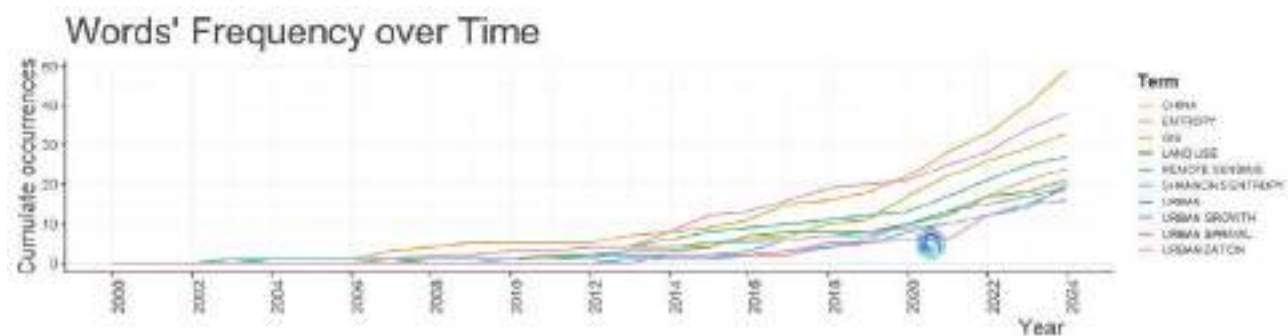


Figure 11. Frequency distribution of keywords in spatial entropy & urban planning over time.

particular, the parallel increase in the terms 'entropy' and 'urban morphology' shows that the studies in which these two concepts are handled together are increasingly intensifying. On the other hand, concepts such as 'chaos' and

'form' follow a relatively more limited usage curve, indicating that such themes are addressed in narrower studies.

The data in Figure 11 reveals that the terms in the Spatial Entropy & Urban Planning literature have increased more

widely and continuously. It is evident that terms such as ‘entropy’, ‘urbanisation’, ‘urban sprawl’, ‘land use’, ‘GIS’, and ‘remote sensing’ exhibit a regular and significant upward trend, particularly since 2016. This trend shows that spatial entropy has found a wider application area, primarily through its integration with geographic information systems and urbanisation studies. In addition, the prominence of the term ‘China’ reveals that a significant portion of the research in this field is China-centred.

Overall, urban morphology and entropy have developed more limitedly and theoretically, whereas Spatial Entropy and urban Planning have grown more comprehensively, appliedly, and interdisciplinarily. This reveals that the two

fields differ in academic interest, geographical density, and methodological diversity [41].

These figures illustrate the temporal evolution of key concepts. The Urban Morphology literature shows consistent usage of a few core terms, while the Spatial Entropy field reflects thematic expansion and adaptation to contemporary planning challenges.

Citation Analysis and Co-Citation Networks

The co-citation network analyses presented in Figure 12 and Figure 13 visualise the scholarly interactions between prominent journals, studies, and research clusters in Urban Morphology & Entropy and Spatial Entropy & Urban

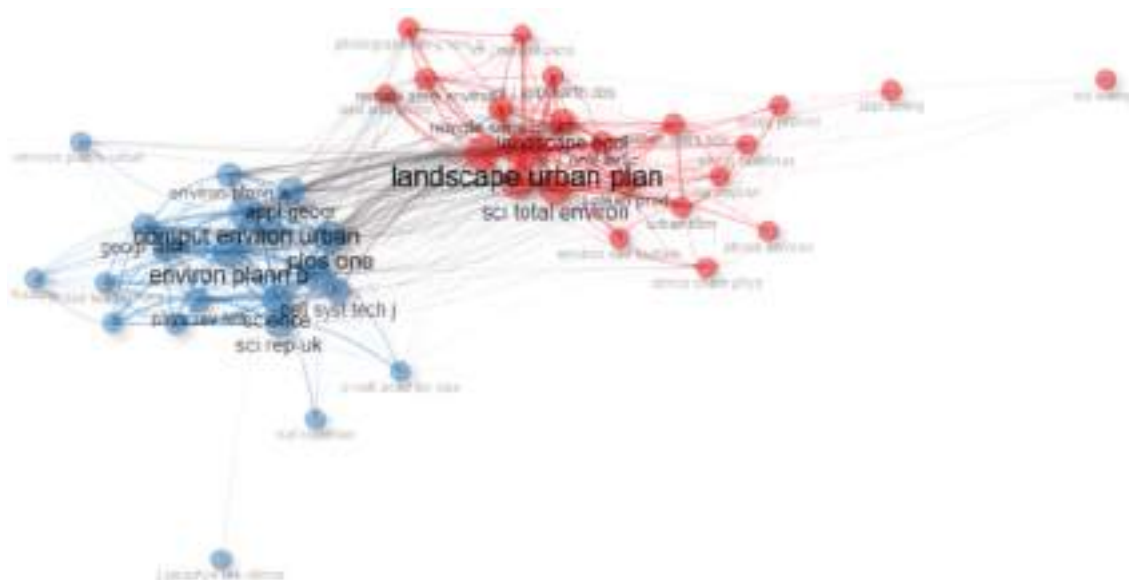


Figure 12. Co-Citation network on urban morphology & entropy.

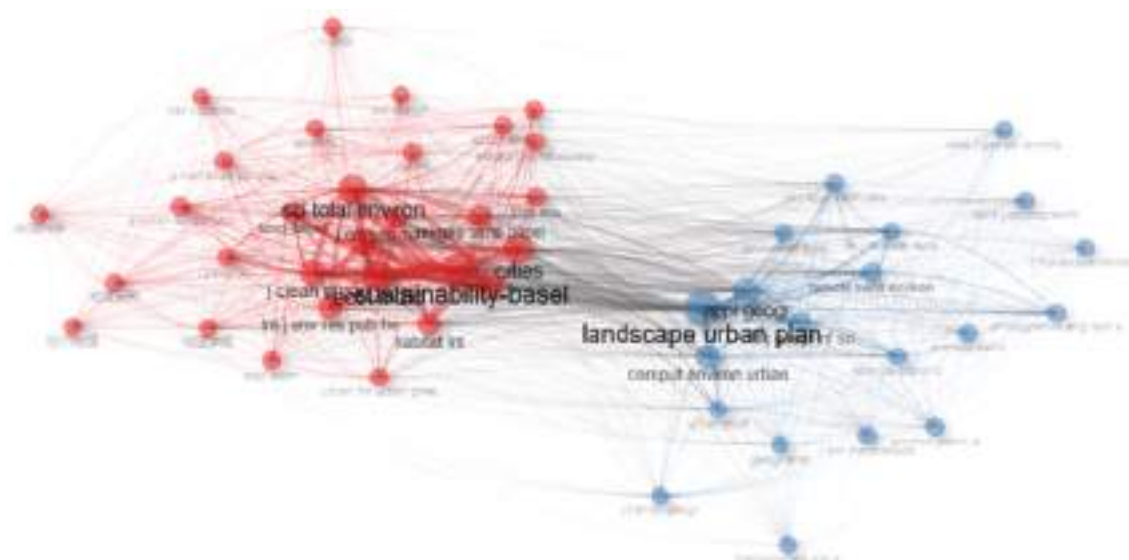


Figure 13. Co-citation network on spatial entropy & urban planning.

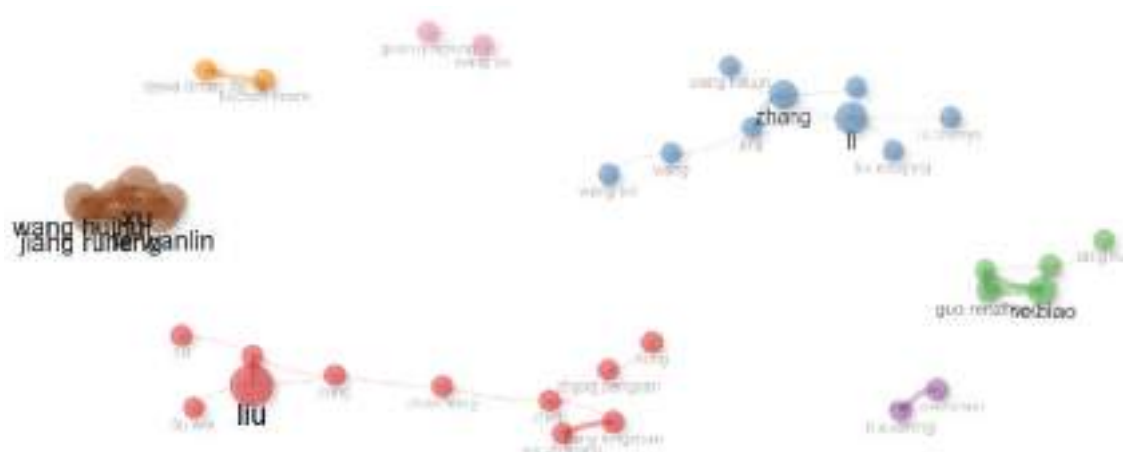


Figure 15. Citation networks in spatial entropy & urban planning.

shows that the Spatial Entropy & Urban Planning literature has diversified rapidly, especially in the last decade, shifting towards a more technical, multidisciplinary, and data-driven research axis.

These findings reveal that urban morphology and entropy are based on more classical theoretical studies. In contrast, Spatial Entropy and urban Planning are more contemporary, technology-based, and open to global research collaborations. The pioneers of both fields provide an essential reference point for future research, guiding the study's contributions to the literature.

These graphs identify influential authors and their citation relationships. Classical theorists dominate the Urban Morphology network, focused on spatial form. At the same

time, the Spatial Entropy domain is driven by contemporary authors from China and other countries with expertise in data-driven urban planning.

Figures 16 and 17 present the global impact of the most cited studies in Urban Morphology & Entropy and Spatial Entropy & Urban Planning. These visualizations are essential in showing which publications have been leading in the literature in both fields [31,48].

The most cited study in Urban Morphology & Entropy, as shown in Figure 16, is the article by Mohajeri published in Renewable Energy, which received 165 citations in total. This is followed by Batty's classic work published in Urban Studies, which provides the basis for urban growth and morphological analyses (118 citations). Other highly cited

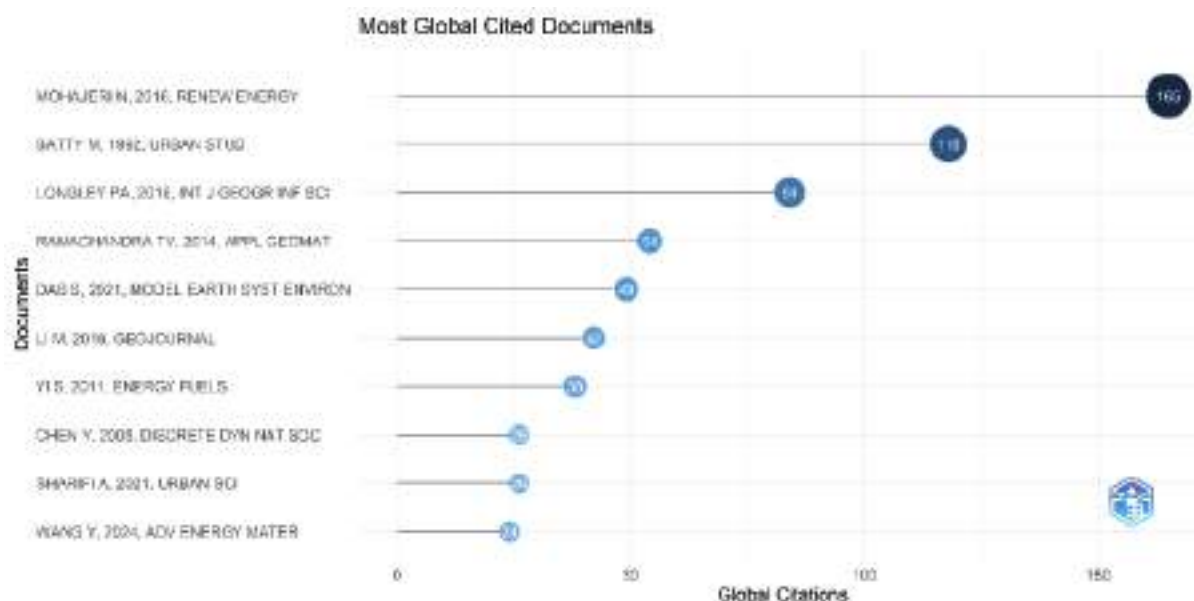


Figure 16. Most cited studies in urban morphology & entropy.

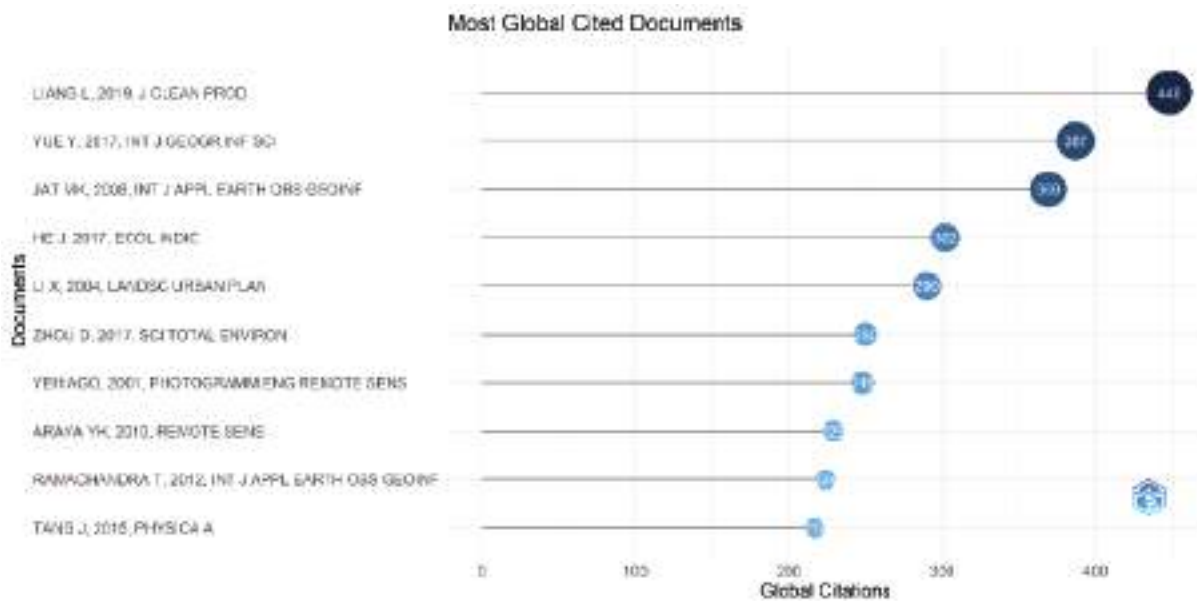


Figure 17. Most cited studies in spatial entropy & urban planning.

studies include spatial statistics, entropic modelling, and energy planning. This shows that entropy is associated with spatial structure in the context of urban morphology and interdisciplinary fields such as sustainability and energy.

Figure 17 reveals the existence of many more highly cited publications in Spatial Entropy & Urban Planning. The most remarkable study is the article published by Liang in the *Journal of Cleaner Production*, which is by far the most cited source in the field with 448 citations. It is followed by Yue (387 citations) and Jat (2008) (369 citations). These articles particularly focus on environmental planning, Geographic Information Systems (GIS), and spatial modelling. In this context, it is observed that the literature in the field of Spatial Entropy & Urban Planning has a more up-to-date, environmentally sensitive, and decision support systems-oriented development line.

These findings reveal that Urban Morphology & Entropy is more theoretically and historically grounded. In contrast, Spatial Entropy & Urban Planning is more applied, focusing on environmental planning and technology-oriented approaches. Furthermore, the latter field is more widely cited in the literature, indicating that it is becoming a more dynamic and rapidly developing field in the global research community. At the same time, this analysis shows that the Spatial Entropy & Urban Planning field appeals to a broader readership and has an interdisciplinary impact. In contrast, the Urban Morphology & Entropy field focuses on more niche and specific topics. These differences emphasise the importance of the comparative framework provided by the bibliometric analysis [31,41,46].

These bar charts display the most cited studies in each field. Urban Morphology is characterized by foundational theoretical work, while Spatial Entropy emphasizes recent,

high-impact studies related to sustainability and thoughtful planning.

Thematic Maps and Trend Analyses

Figure 18 and Figure 19 show the trending topics over time in Urban Morphology & Entropy and Spatial Entropy & Urban Planning, respectively. These analyses provide important clues to understand research trends by revealing the thematic evolution in the respective fields [31,48].

In Figure 18, urban morphology, 'urban', 'entropy', and 'morphology' are among the prominent trending terms in Urban Morphology & Entropy. The fact that these terms are trending shows that this field focuses more on structural morphology and entropic distribution models at the conceptual level. However, in terms of period, these terms seem to be concentrated for a limited number of years and remain largely up to date. This reveals a conceptual continuity in the literature, but the diversity of applications remains limited.

Figure 19 shows that the thematic development in Spatial Entropy & Urban Planning is much more diverse, dynamic, and application-oriented. Here, concepts such as 'entropy', 'urban sprawl', 'GIS', 'urbanisation', 'remote sensing', 'landscape metrics', and 'machine learning' come to the forefront; this shows that spatial entropy issues are addressed in a broad interdisciplinary context. Especially in the post-2019 period, the rise of concepts such as 'coupling coordination', 'urban resilience', 'topsis', 'information entropy', 'driving factors', and 'deep learning' reveals that methodological diversification and planning-oriented applications in this field have increased.

These findings show that urban morphology and entropy are based on basic concepts and theoretical frameworks.

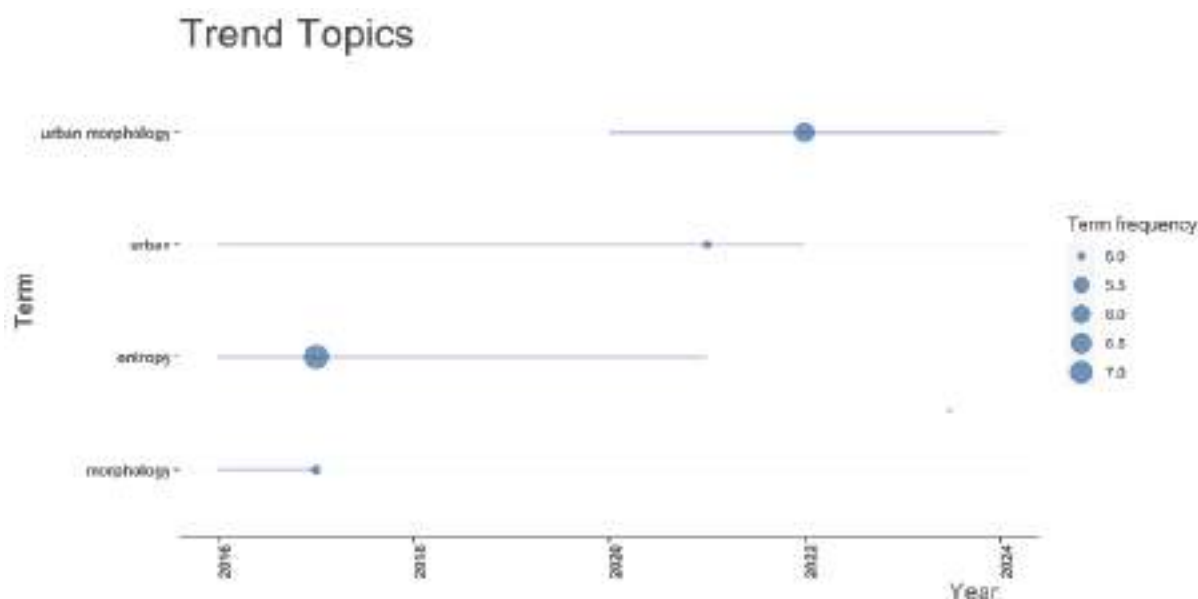


Figure 18. Distribution of trending topics in urban morphology & entropy research areas over time.

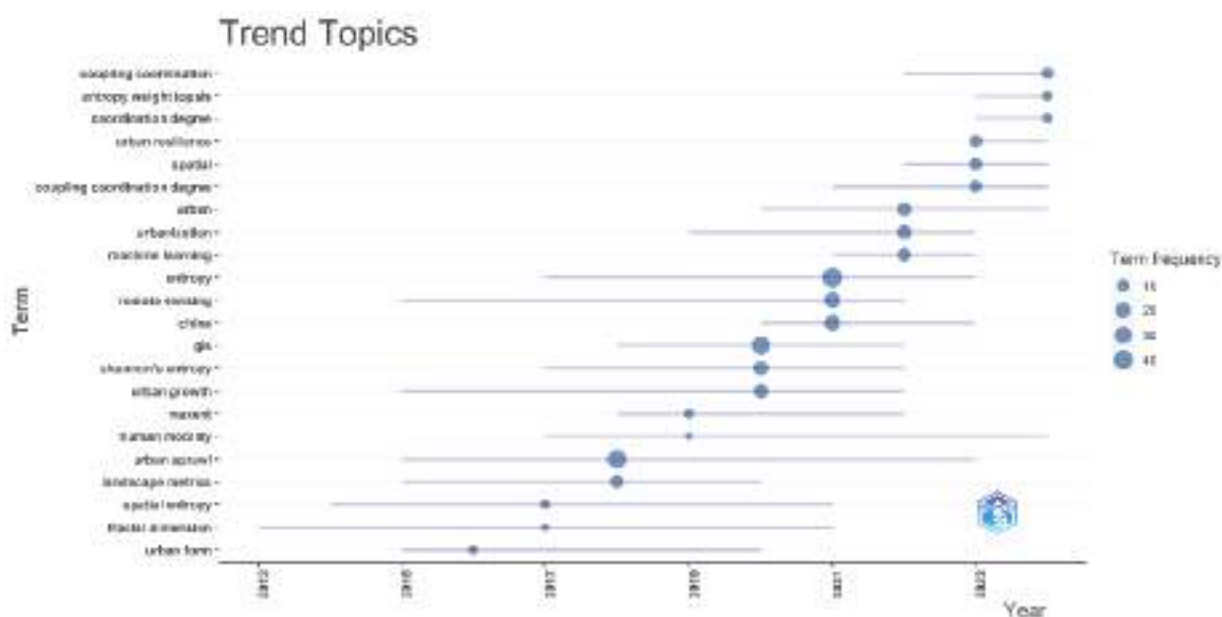


Figure 19. Distribution of trending topics in spatial entropy & urban planning research areas over time.

In contrast, Spatial Entropy and urban Planning are more up-to-date, technology-integrated, data-based, and practice-oriented. Thus, these analyses provide a comparative assessment of the development axes of the two fields over time and reinforce the methodological power of the bibliometric approach [48].

These visuals demonstrate how popular research topics have evolved. Urban Morphology maintains focus on stable theoretical themes, whereas Spatial Entropy exhibits growing attention to emerging technologies and planning tools.

Figures 20 and 21 show a Three-Field Plot analysis that analyses the scientific production on urban morphology and entropy at three primary levels: publishing countries (AU_CO), institutions (AU_UN), and most frequently used keywords (DE). This analysis visualises the relationships between the prolific actors in the literature and the research themes, detailing the geographical and institutional distribution of the discipline [31,45].

According to Figure 20, China, the USA, Poland, Brazil, and Japan stand out among the most productive countries

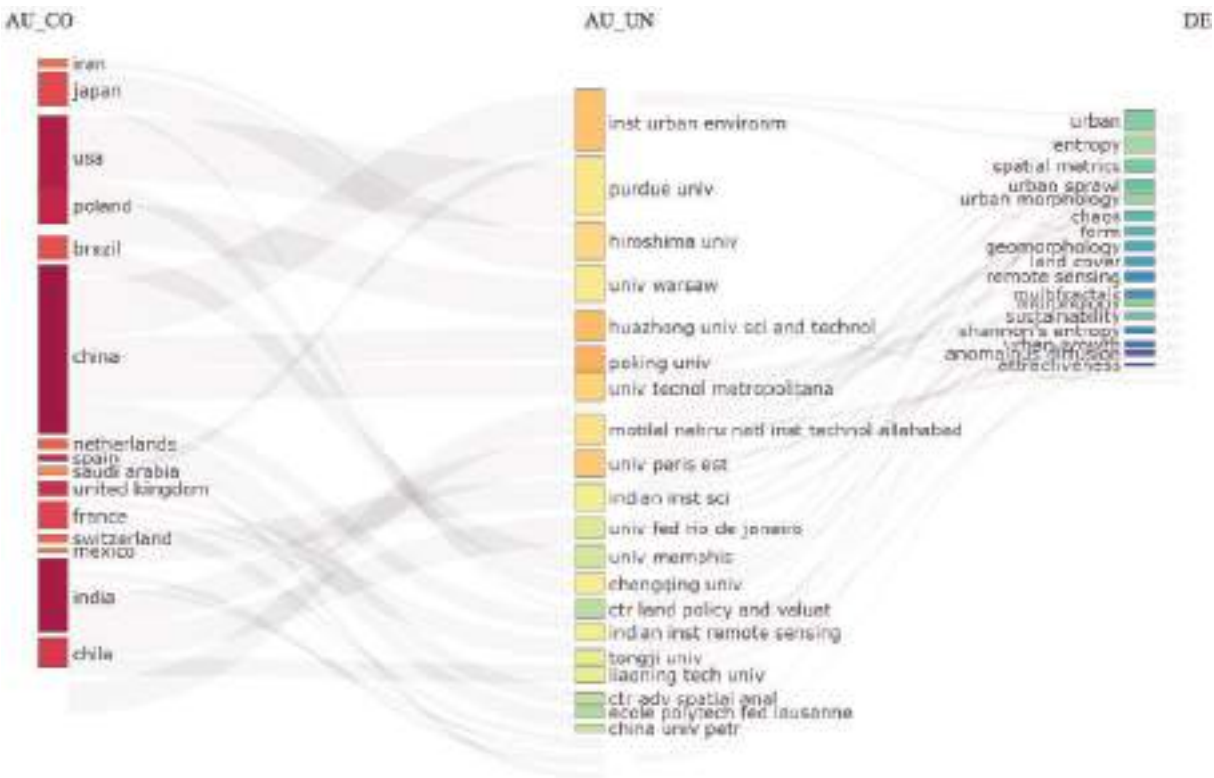


Figure 20. Three-field plot analysis on urban morphology & entropy.

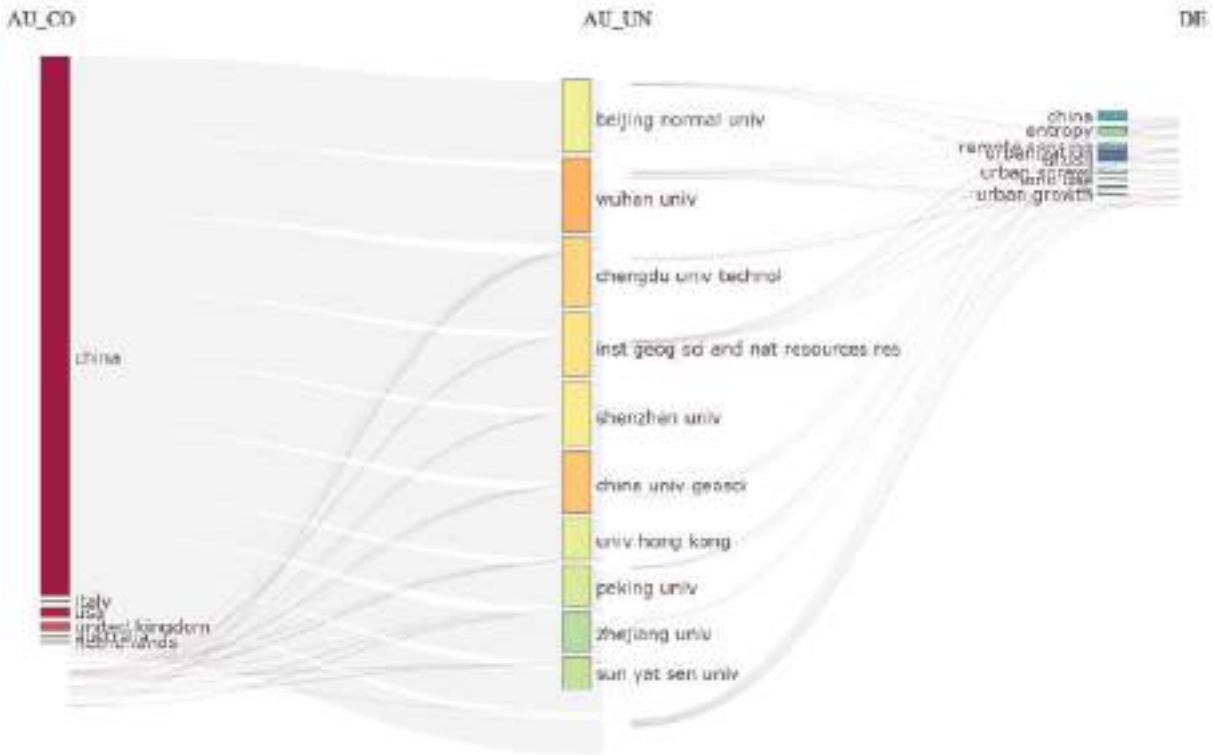


Figure 21. Three-field plot analysis on spatial entropy & urban planning.

in this field. In particular, China is at the forefront of studies related to various keywords through many organisations. This indicates that China has become a central actor in the research on theoretical and applied urban morphology and entropy.

Institutions such as the Institute for Urban Environment, Purdue University, Hiroshima University, and the University of Warsaw stand out at the institutional level. Their research activities are most frequently centred around the keywords 'urban', 'entropy', 'urban morphology', and 'spatial metrics'. This situation shows that specific institutional centres influence the entropic analysis of urban structures and are shaped around specific thematic foci.

Concepts such as 'urban morphology', 'entropy', 'spatial metrics', 'urban sprawl', 'form', and 'land cover' are frequently used at the keyword level. This shows that the formal analysis of spatial structures, urban sprawl, and the evaluation of complexity with entropic metrics are among the prioritised research themes in this literature.

In general, this tripartite structure has allowed the analysis of geographical diversity, institutional activity, and thematic concentrations in the literature together, thus providing a holistic view of the scientific ecosystem of Urban Morphology & Entropy [49].

Figure 21 presents a Three-Field Plot analysis visualising the studies on Spatial Entropy and urban Planning in terms of countries (AU_CO), institutions (AU_UN), and keywords (DE).

The most important finding in Figure 21 is the apparent hegemonic position of China in this research area. China is far ahead in the number of publications at the national and institutional levels. In particular, institutions such as Beijing Normal University, Wuhan University, Chengdu University of Technology, Shenzhen University and the Institute of Geographic Sciences and Natural Resources Research realize a large part of the production in this field. This shows China has become a global centre for spatial entropy and urban planning studies.

At the keyword level, concepts such as 'entropy', 'China', 'remote sensing', 'urbanization', 'land use' and 'urban growth' come to the fore. This shows that spatial entropy is considered together with urban growth, land use and remote sensing techniques, and entropy is integrated into urban planning decisions through these parameters. It reveals that the Spatial Entropy & Urban Planning literature is primarily centred in China, with research themes focusing on rapid urbanization and land use supported by technological tools (especially remote sensing). This is an important indicator for understanding the regional concentration in the field and how entropy is addressed in practical planning contexts.

These figures map the relationships among countries, institutions, and keywords. The Spatial Entropy literature prominently features China as a dominant contributor, reflecting its central role in the global discourse on data-driven urban planning.

Figures 22 and 23 present the conceptual structure of the Urban Morphology & Entropy and Spatial Entropy & Urban Planning literature, visualized using the Factorial Map method. This analysis reveals the thematic clustering of terms and their positions on two basic dimensions (Dimension 1 and Dimension 2), revealing the contextual affinities of subheadings in the literature [31,49]. This map, created by the factor analysis method, shows the spatial relationships and thematic groups of the concept clusters prominent in the literature. The concepts on the two-dimensional plane allow for analyzing thematic diversity and research trends [31].

Figure 22 presents a Factorial Map visualization analyzing the distribution of conceptual constructs in the Urban Morphology & Entropy research area. Phrases such as 'adsorption', 'mechanism', 'composite hydrogel' and 'thermodynamic parameters' in the upper section reveal the interdisciplinary aspect of this field and represent studies where entropy is associated with physical-chemical processes. This shows that entropy in urban morphology is addressed through spatial structure and energy and material cycle-based approaches. Concepts such as 'decision making model', 'deployment planning' and 'temporal mismatch' in the lower right area show the applied use of entropy in decision making, planning processes and time-space mismatches. This points to entropy-based spatial planning approaches. Concepts such as 'urban geography', 'geo-temporal demographics' and 'twitter' in the lower left area represent the new generation of research areas where the relationship between social media, spatial demography and entropy is evaluated. Concepts such as 'urban function', 'spatial interaction probability', 'attractiveness', and 'typical urban unit cluster', which are concentrated in the centre, show that entropy-based spatial analyses are carried out through functional distributions and attractiveness within the city.

Overall, this analysis shows that the urban morphology and entropy literature is multidimensional, encompassing physical environment, energy-based approaches, and social-spatial analyses. Furthermore, this thematic map observes current trends in which different data sources (social media, location-based service (LBS) data) and models are integrated with entropy analysis in an interdisciplinary synthesis (Fig. 22).

Figure 23 presents a Factorial Map visualization analyzing the distribution of conceptual constructs in the Spatial Entropy & Urban Planning research area. Keywords such as 'analysis', 'maxent', 'ecosystem services', 'remote', 'landscape metrics' and 'deep learning', which are in the upper left part of the map, indicate that this field is heavily associated with the assessment of environmental services, remote sensing technologies and artificial intelligence-assisted analysis techniques. These concepts are especially prominent in studies on modelling natural resources, determining sustainability criteria, and analyzing spatial data.

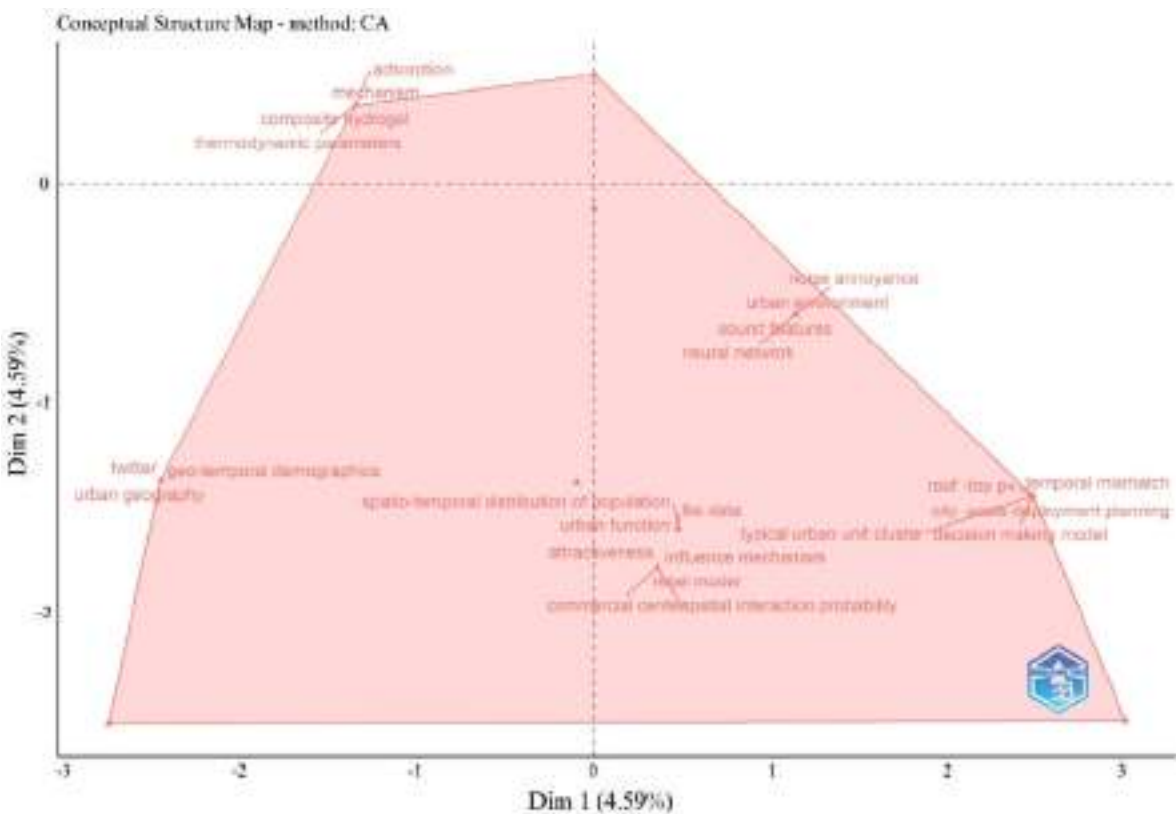


Figure 22. Factorial map analysis on urban morphology & entropy.

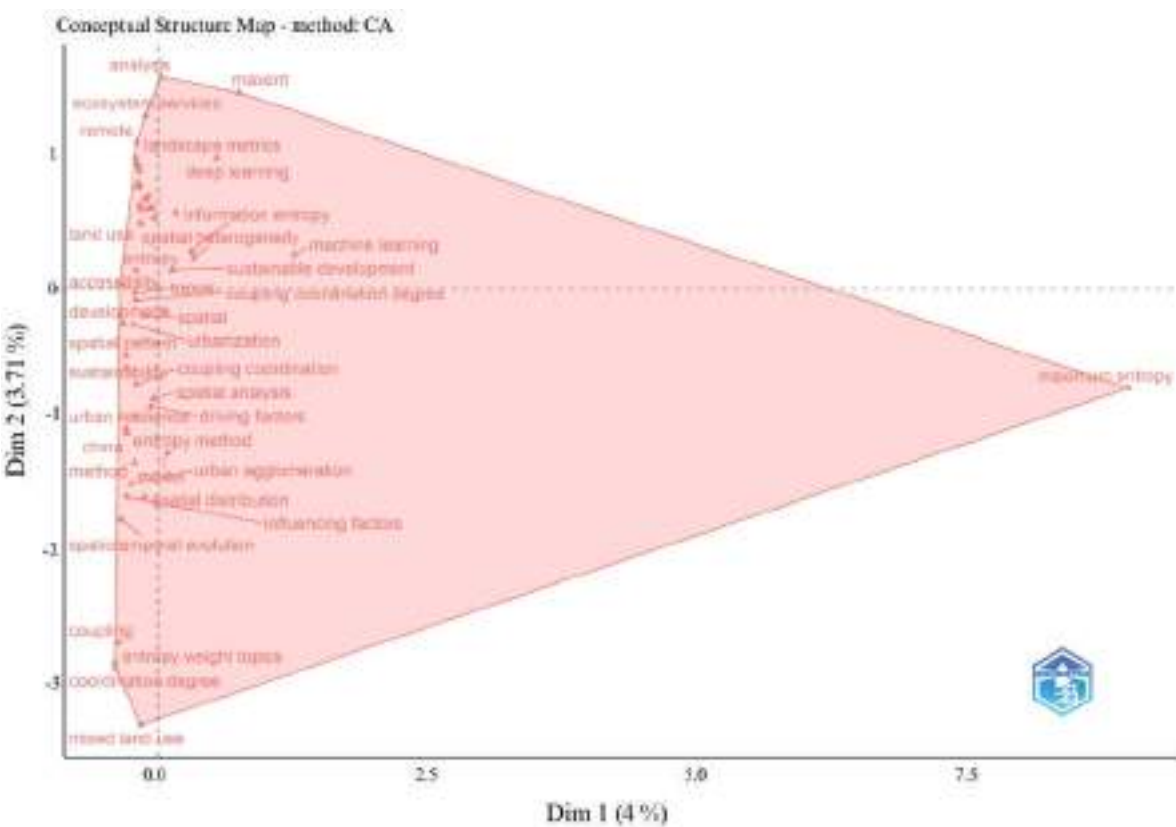


Figure 23. Factorial map analysis on spatial entropy & urban planning.

The cluster in the central region includes terms such as 'entropy', 'land use', 'sustainability', 'urbanization', 'information entropy', 'spatial pattern', 'coupling coordination degree' and 'accessibility'. These terms reveal how spatial entropy is used to measure the complexity of urbanization processes, support sustainable planning practices, and explain land use dynamics. In particular, terms such as 'coupling coordination degree' and 'information entropy' reflect trends in interdisciplinary studies towards integrated assessment of complex systems. Concepts such as 'coordination degree', 'entropy weight tosis', 'spatiotemporal evolution' and 'urban agglomeration' in the lower left region show that multi-criteria decision-making methods and temporal-spatial modelling are intensively used. It is understood that these analyses are used to optimize planning decision processes.

In the upper right part of the map, the concept of 'maximum entropy', which is positioned alone, draws attention. Although this term shows less commonality with other concepts in the conceptual map, it is an important focal point at the theoretical level. The maximum entropy approach is integrated into planning models as a mathematical method for optimal information extraction, especially under uncertainty. The Spatial Entropy & Urban Planning field generally shows an interdisciplinary structure, where remote sensing, artificial intelligence, sustainability, urbanization and decision support systems are intertwined with entropy-based analyses. It can be said that this structure reflects in detail how entropy is positioned not only as a theoretical but also as an applied tool in the urban planning literature (Fig. 23).

These two analyses demonstrate that urban morphology and entropy are more effectively integrated with spatial analysis techniques. Spatial Entropy and urban Planning deal with wider interdisciplinary interaction and technical diversity. These findings clearly illustrate the intersection of the theoretical frameworks of the two fields and the differences in their fields of application [49].

These factorial maps depict the conceptual structure of each literature. Urban Morphology literature clusters around physical form and spatial analysis, while Spatial Entropy integrates themes like sustainability, simulation, and machine learning.

This study takes a comparative approach to the bibliometric analyses conducted under the titles 'Urban Morphology & Entropy' and 'Spatial Entropy & Urban Planning' to understand the evolution of the concept of entropy in the urban studies literature. By examining how entropy is interpreted and used in different contexts, this study reveals the place of entropy-based spatial analyses in urban morphology and planning literature. It analyses thematic relationships and methodological divergences often overlooked in previous studies. In this context, elements of scientific productivity, keyword networks, thematic developments, and intellectual structure are visualized and interpreted for two separate sets of literature [31,44].

The study's methodological framework is based on analyzing the data obtained from the Web of Science (WoS) database using the Bibliometrix package and the Biblioshiny visualization tool in the RStudio environment. This approach is in line with the bibliometric analysis methods used by researchers such as Aria et al. [31,44], and [41]. In the literature, WoS and Scopus databases are among the most common sources for conducting bibliometric studies, and Xu et al. [50], Aria et al. [31], and Zhang et al. [38] are examples of studies that use these databases effectively. Yilmaz et al. [34] visualized earthquake research in Turkey using the WoS database in bibliometric analyses; this study is based on WoS data in entropy and urban morphology-based analyses. In this study, the primary justification for the WoS database's preference was the dataset's comprehensiveness and high academic visibility [51,47,48]. In addition, studies such as Aristovnik et al. [52], Liu et al. [53] and Matsler et al. [43] used the Scopus database; [33] analyzed both databases comparatively and found a high level of correlation [51,54]. These findings indicate no significant content difference between WoS and Scopus databases.

The Bibliometrix package used in the RStudio environment offers advanced analysis techniques such as Two-Field Plot, Three-Field Plot, Thematic Map, Factorial Map, and Dendrogram, which are frequently used in bibliometric analyses [49]. These types of analyses have been previously applied in various studies by Cobo et al. [30], Song et al. [55], Aria et al. [31] and Gonçalves et al. [48]. However, the unique aspect of this research is that it deals with the concept of entropy comparatively through two different sets of literature. The analyses conducted under the titles of 'Urban Morphology & Entropy' and 'Spatial Entropy & Urban Planning' evaluate both thematic developments and the scientific productivity, intellectual structure, and interdisciplinary relations of these two fields from a holistic perspective. In this context, the study is based on the analysis of academic studies on integrating entropy in urban morphology and urban planning.

Many studies use entropy to analyze urban morphology Nazarnia et al. [56], Brigatti et al. [2] or to model issues such as urban sprawl and sustainability in spatial planning [29,57]. Altieri et al. [29] assessed the irregularity in the sprawl patterns of European cities using spatial entropy. Similarly, Brigatti et al. [2] analyzed cities with different morphological structures using entropy-based clustering methods. However, these studies deal with entropy only in a single context and do not cover its multidisciplinary interactions or thematic diversity. Aladag et al. [36] adopted the scoping review method to identify the uses of artificial intelligence applications in project management processes, proposing a theoretical basis for future research. Similarly, this study uses bibliometric analysis to evaluate the place of entropy-based spatial approaches in urban planning and morphological interpretation. Al-Shamayleh et al. [58] emphasize the importance of using integrated software to

support decision-making processes in urban infrastructure management. Similarly, this research systematically addresses the relationship between urban morphology and entropy at the literature level, demonstrating the potential of multi-data and interdisciplinary analysis approaches to contribute to planning. In their study, Yilmaz et al. [34] presented the academic publications after the 6 February earthquakes through bibliometric analyses. However, this study individually evaluated two literatures (Urban Morphology & Entropy and Spatial Entropy & Urban Planning).

At this point, one of the most important contributions of this study is that it reveals how entropy is interpreted in different ways in urban morphology and planning. While the 'Urban Morphology & Entropy' literature is more theoretically oriented and focuses on the physical structure of urban form, the 'Spatial Entropy & Urban Planning' literature provides a more applied framework by incorporating themes such as machine learning, sustainability, climate policies, and decision support systems [15,59,60]. This differentiation is also confirmed by prior studies, including [57] and [17]. It is clearly illustrated in the Factorial Map and Thematic Map analyses, which highlight the thematic affinities and divergences between the two literature sets.

For example, Batty and Galster applied entropy in more conventional morphological contexts, examining urban growth and spatial distribution patterns. In contrast, more recent studies by He & Chen (2024) have incorporated entropy into AI-driven planning and decision-support frameworks, reflecting a methodological evolution in the field.

This growing trend underscores that entropy is no longer confined to structural analysis; instead, it has become a multidisciplinary, data-driven tool widely used in contemporary urban planning practices [29,59].

In the publications titled Urban Morphology & Entropy, it has been observed that concepts such as 'urban morphology', 'spatial metrics', and 'configuration' come to the forefront, and the studies proceed within a narrower thematic framework [56,61]. On the other hand, Spatial Entropy & Urban Planning literature offers a wide range of content with interdisciplinary concepts such as 'machine learning', 'urban resilience', 'maximum entropy', and 'landscape metrics', and associates entropy with decision support mechanisms [62,27,59,15]. This is a critical point where this study differs from the previous literature.

Furthermore, bibliometric analyses show that China has a high level of scientific productivity, especially in the Spatial Entropy & Urban Planning literature, and that most of the publications in this field are produced by China-based institutions. This finding, in line with studies such as [38] and [53], shows that China is leading entropy-based spatial analyses at the global level. The study also provides a methodological contribution to previous studies with different analysis techniques. For example, using bibliometric methods proposed by [30] and [31], analysis tools such as Three-Field Plot, Factorial Map, and Co-Citation Network

visualize intellectual clusters and research relationships. In particular, these analyses have made evident how entropy is integrated with decision support systems in urban planning and spatial structure criteria in morphological analysis.

Another important finding of the study is that entropy is becoming not only a theoretical concept that defines spatial complexity, but also a tool that can be integrated with data-based analyses, artificial intelligence, and geographic information systems in urban planning [23,53]. In this context, spatial entropy may assume more strategic roles in the future in areas such as urban sustainability, spatial distribution of inequalities, and climate adaptation.

This study fills an important gap in the literature and redefines the theoretical position of entropy-based studies in urban research with an interdisciplinary approach. Unlike previous studies on different datasets, this study directly compares two different sets of literature and simultaneously analyzes the theoretical evolution, methodological use, and thematic diversity of entropy. In this respect, the study provides an important framework for how entropy-based spatial analyses can be developed in urban planning and morphological interpretation in the future.

CONCLUSION

This study evaluated the theoretical and practical development of the concept of entropy in urban research by conducting a comparative bibliometric analysis of academic publications under the themes of Urban Morphology & Entropy and Spatial Entropy & Urban Planning. The findings demonstrate that entropy constitutes a central conceptual framework in both fields; however, its applications and research priorities vary significantly. While the Urban Morphology & Entropy literature primarily focuses on the regularity of spatial forms and morphological structures, the Spatial Entropy & Urban Planning literature integrates entropy with data science, sustainability, environmental planning, and artificial intelligence. This divergence illustrates the interdisciplinary evolution of entropy-based analyses and underlines the increasing importance of cross-disciplinary approaches in urban research. Furthermore, the prominence of China in spatial entropy-themed publications signals a shift in the global centers of scientific productivity in this field.

The bibliometric methods employed, such as thematic mapping, three-field plots, collaboration networks, and citation analysis, effectively revealed the structural components, research gaps, and intellectual trends within the literature. In doing so, the study not only provided a systematic assessment of current scholarships but also opened a theoretical discussion about the role of entropy in urban planning and morphological analysis.

From a theoretical standpoint, the findings also contribute to urban systems theory and complexity science by demonstrating how entropy serves as a framework to understand spatial irregularities. In this regard, future

bibliometric evaluations are encouraged to explore emerging directions such as the integration of altimetric or social media-based data to capture broader scholarly and societal impact.

Future research should strengthen the interdisciplinary perspective by integrating entropy not only with urban form analysis but also with themes such as spatial justice, disaster resilience, and environmental sustainability. To further explore the intersection of Urban Morphology & Entropy and Spatial Entropy & Urban Planning, incorporating advanced computational tools like machine learning and big data analytics is highly encouraged.

To broaden bibliometric scope and enhance analytical depth, future studies should consider including additional databases such as Scopus and Google Scholar alongside Web of Science. This would help provide a more comprehensive and diversified understanding of academic contributions across disciplines.

In terms of methodological development, the use of advanced techniques such as spatial regression, simulation modeling, and artificial intelligence can facilitate a deeper understanding of urban complexity. Researchers are also encouraged to apply more sophisticated bibliometric methods, including network analysis, systematic mapping, and meta-analysis, to capture evolving thematic trends with greater precision.

From an applied planning perspective, entropy-based indicators should be more actively utilized within decision-support systems of municipalities and planning offices. These metrics can support data-driven decisions in areas such as green infrastructure planning, spatial equity, and resource allocation, thereby enhancing the functional effectiveness of urban interventions.

Finally, it is recommended that theoretical research continue to deepen the role of entropy not merely as a quantitative tool, but also as a conceptual framework for interpreting the complexity and dynamism of urban systems. Viewing entropy as a way of thinking, rather than solely as a metric, can lead to more nuanced and integrated models of urban analysis and governance.

These conclusions underscore that entropy-based analyses can play both a diagnostic and a guiding role in the disciplines of urban morphology and urban planning. Thus, this study contributes not only to the current literature but also offers a robust conceptual and methodological foundation for future research in entropy-informed spatial studies.

ACKNOWLEDGEMENTS

This article is prepared about “Entropy”, which I encountered during the training program titled “*Arazi Çeşitliliğinin Entropi Temelli Algoritmalar ile Hesaplanması ve Haritalanması*” organized within the scope of TÜBİTAK 2237-A Scientific Education Activities Support. I sincerely thank the valuable academicians and the Training Coordinator, Assoc, who organized the training. Prof. Dr.

Özdemir Şentürk, Prof. Dr. Ahmet Mert, and TÜBİTAK for their contributions and support of this project.

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] Oliveira V. Urban morphology. Oxford: Oxford University Press; 2020.
- [2] Brigatti E, Netto VM, de Sousa Filho FNM, Cacholas C. Entropy and hierarchical clustering: characterising the morphology of the urban fabric in different spatial cultures. *Chaos* 2021;31. [\[CrossRef\]](#)
- [3] Kubat AS. Türkiye Kentsel Morfoloji Ağı (TNUM): araştırma dinamikleri ve eğilimleri. Türkiye Kentsel Morfoloji Ağı (IV. Kentsel Morfoloji Sempozyumu Bildiriler Kitabı, Konya) 2023;1–12.
- [4] Visconti F, Capozzi R. Urban project, a matter of morphology. *Land* 2023;12:1958. [\[CrossRef\]](#)
- [5] Wang J, Huang W, Biljecki F. Learning visual features from figure-ground maps for urban morphology discovery. *Comput Environ Urban Syst* 2024;109:102076. [\[CrossRef\]](#)
- [6] Tse HT. The definition and application of entropy. In: SHS Web of Conferences. Vol. 144. Les Ulis: EDP Sciences; 2022. p. 01016. [\[CrossRef\]](#)
- [7] Zhang P, Ghosh D, Park S. Spatial measures and methods in sustainable urban morphology: a systematic review. *Landsc Urban Plan* 2023;237:104776. [\[CrossRef\]](#)

- [8] Batty M. The size, scale, and shape of cities. *Science* 2008;319:769–771. [\[CrossRef\]](#)
- [9] Yeh AGO, Li X. Measurement and monitoring of urban sprawl in a rapidly growing region using entropy. *Photogramm Eng Remote Sensing* 2001;67:83–90.
- [10] Ahmad S, Chase SC. Style representation in design grammars. *Environ Plann B Plann Des* 2012;39:486–500. [\[CrossRef\]](#)
- [11] Ozdemir I, Norton DA, Ozkan UY, Mert A, Senturk O. Estimation of tree size diversity using object-oriented texture analysis and Aster imagery. *Sensors* 2008;8. [\[CrossRef\]](#)
- [12] Ozkan K, Mert A, Özdemir S. A new proposed GLCM texture feature: modified Rényi Deng entropy. *J Supercomput* 2023;79:21507–21527. [\[CrossRef\]](#)
- [13] Huynh HN. Spatial point pattern and urban morphology: perspectives from entropy, complexity, and networks. *Phys Rev E* 2019;100:022320. [\[CrossRef\]](#)
- [14] Cabral P, Augusto G, Tewolde M, Araya Y. Entropy in urban systems. *Entropy* 2013;15:5223–5236. [\[CrossRef\]](#)
- [15] He W, Chen M. Advancing urban life: a systematic review of emerging technologies and artificial intelligence in urban design and planning. *Buildings* 2024;14:835. [\[CrossRef\]](#)
- [16] Olbrys J. Entropy-based applications in economics, finance, and management. *Entropy* 2022;24:1468. [\[CrossRef\]](#)
- [17] Fagerholm ED, Dezhina Z, Moran RJ, Turkheimer FE, Leech R. A primer on entropy in neuroscience. *Neurosci Biobehav Rev* 2023;146:105070. [\[CrossRef\]](#)
- [18] Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;27:379–423. [\[CrossRef\]](#)
- [19] Shannon CE, Weaver W. The mathematical theory of communication. Urbana: University of Illinois Press; 1949.
- [20] Herold M, Couclelis H, Clarke KC. The role of spatial metrics in the analysis and modeling of urban land use change. *Comput Environ Urban Syst* 2005;29:369–399. [\[CrossRef\]](#)
- [21] Boeing G. Urban spatial order: street network orientation, configuration, and entropy. *Appl Netw Sci* 2019;4:1–19. [\[CrossRef\]](#)
- [22] Ozkan K. Application of information theory for an entropic gradient of ecological sites. *Entropy* 2016;18. [\[CrossRef\]](#)
- [23] Purvis B, Mao Y, Robinson D. Entropy and its application to urban systems. *Entropy* 2019;21. [\[CrossRef\]](#)
- [24] Sepúlveda-Fontaine SA, Amigó JM. Applications of entropy in data analysis and machine learning: a review. *Entropy* 2024;26:1126. [\[CrossRef\]](#)
- [25] Jat MK, Garg PK, Khare D. Monitoring and modeling of urban sprawl using remote sensing and GIS techniques. *Int J Appl Earth Obs Geoinf* 2008;10:26–43. [\[CrossRef\]](#)
- [26] Dewa DD, Buchori I, Sejati AW, Liu Y. Shannon entropy-based urban spatial fragmentation to ensure sustainable development of the urban coastal city: a case study of Semarang, Indonesia. *Remote Sens Appl Soc Environ* 2022;28:100839. [\[CrossRef\]](#)
- [27] Dhanaraj K, Angadi DP. Land use land cover mapping and monitoring urban growth using remote sensing and GIS techniques in Mangaluru, India. *GeoJournal* 2022;87:1133–1159. [\[CrossRef\]](#)
- [28] Xiong J, Wang Z, Ruan M, Yao H, Wei M, Sun R, Yang X, Qi W, Liang F. Current status of neuroimaging research on the effects of acupuncture: a bibliometric and visual analyses. *Complement Ther Med* 2022;71:102877. [\[CrossRef\]](#)
- [29] Altieri L, Cocchi D, Roli G. Measuring heterogeneity in urban expansion via spatial entropy. *Environmetrics* 2019;30:e2548. [\[CrossRef\]](#)
- [30] Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. Science mapping software tools: review, analysis, and cooperative study among tools. *J Am Soc Inf Sci Technol* 2011;62:1382–1402. [\[CrossRef\]](#)
- [31] Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* 2017;11:959–975. [\[CrossRef\]](#)
- [32] Jones T, Huggett S, Kamalski J. Finding a way through the scientific literature: indexes and measures. *World Neurosurg* 2011;76:36–38. [\[CrossRef\]](#)
- [33] Katuk D, Koseoglu E. Affect, architecture and water: bibliometric analysis of the literature. *J Des Plan Aesthet Res* 2023;1. [\[CrossRef\]](#)
- [34] Yilmaz S, Acar U, Akbaba MF, Akbulut T, Akgün C, Akin O, Akpınar B, et al. The scientific outputs of earthquakes in Türkiye: a bibliometric analysis. *Sigma J Eng Nat Sci* 2025;43:513–522. [\[CrossRef\]](#)
- [35] Fan J, Chen Y. University campus architects and their influences from the perspective of bibliometrics. *Int J Arts Humanit Stud* 2022;2:75–87. [\[CrossRef\]](#)
- [36] Aladag H, Güven İ, Ballı O. Contribution of artificial intelligence to construction project management processes: state of the art with scoping review method. *Sigma J Eng Nat Sci* 2024;42:1654–1669. [\[CrossRef\]](#)
- [37] Yang G, Yu Z, Zhang J, Søderkvist Kristensen L. From preference to landscape sustainability: a bibliometric review of landscape preference research from 1968 to 2019. *Ecosyst Health Sustain* 2021;7:1948355. [\[CrossRef\]](#)
- [38] Zhang Y, Ye E, Liu F, Lai N, You X, Dong J, Dong J. The relationship between landscape construction and bird diversity: a bibliometric analysis. *Int J Environ Res Public Health* 2023;20. [\[CrossRef\]](#)
- [39] Chen C. Science mapping: a systematic review of the literature. *J Data Inf Sci* 2017;2:1–40. [\[CrossRef\]](#)
- [40] Zhao X, Zuo J, Wu G, Huang C. A bibliometric review of green building research 2000–2016. *Archit Sci Rev* 2019;62:74–88. [\[CrossRef\]](#)

- [41] Bayazıt E. Academic traces of urban furniture: a bibliometric perspective. *Idealkent* 2024;16:1403–1434. [\[CrossRef\]](#)
- [42] Roldan-Valadez E, Salazar-Ruiz SY, Ibarra-Contreras R, Rios C. Current concepts on bibliometrics: a brief review about impact factor, Eigenfactor score, CiteScore, SCImago Journal Rank, source-normalised impact per paper, H-index, and alternative metrics. *Ir J Med Sci* 2019;188:939–951. [\[CrossRef\]](#)
- [43] Matsler AM, Meerow S, Mell IC, Pavao-Zuckerman MA. A ‘green’ chameleon: exploring the many disciplinary definitions, goals, and forms of “green infrastructure”. *Landsc Urban Plan* 2021;214:104145. [\[CrossRef\]](#)
- [44] Azad AK, Parvin S. Bibliometric analysis of photovoltaic thermal (PV/T) system: from citation mapping to research agenda. *Energy Rep* 2022;8:2699–2711. [\[CrossRef\]](#)
- [45] Eyre-Walker A, Stoletzki N. The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biol* 2013;11:e1001675. [\[CrossRef\]](#)
- [46] Kalkan N, Coşkun S. Bibliometric evaluation of academic studies on water management in xeriscape landscape applications: analysis with R Studio-Biblioshiny software. *Uluborlu Meslek Bilim Derg* 2025;8:41–52. [\[CrossRef\]](#)
- [47] Michelle B, Gemilang MP. A bibliometric analysis of generative design, algorithmic design, and parametric design in architecture. *J Artif Intell Archit* 2022;1. [\[CrossRef\]](#)
- [48] Gonçalves CL, Pereira L, Akkari ACS. Bibliometric mapping of the research trends on software architecture for e-Health systems. *Procedia Comput Sci* 2023;219:1462–1469. [\[CrossRef\]](#)
- [49] Zupic I, Čater T. Bibliometric methods in management and organisation. *Organ Res Methods* 2015;18:429–472. [\[CrossRef\]](#)
- [50] Xu Z, Zhou W, Baltrėnaitė E. Comprehensive bibliometric study of Journal of Environmental Engineering and Landscape Management from 2007 to 2019. *J Environ Eng Landsc Manag* 2019;27. [\[CrossRef\]](#)
- [51] Harzing AW, Alakangas S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics* 2016;106:787–804. [\[CrossRef\]](#)
- [52] Aristovnik A, Ravšelj D, Umek L. A bibliometric analysis of COVID-19 across science and social science research landscape. *Sustainability* 2020;12. [\[CrossRef\]](#)
- [53] Liu J, Kamarudin KM, Liu Y, Zou J. Developing pandemic prevention and control by ANP-QFD approach: a case study on urban furniture design in China communities. *Int J Environ Res Public Health* 2021;18. [\[CrossRef\]](#)
- [54] Archambault É, Campbell D, Gingras Y, Larivière V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. *J Am Soc Inf Sci Technol* 2009;60:1320–1326. [\[CrossRef\]](#)
- [55] Song J, Zhang H, Dong W. A review of emerging trends in global PPP research: analysis and visualization. *Scientometrics* 2016;107:1111–1147. [\[CrossRef\]](#)
- [56] Nazarnia N, Harding C, Jaeger JA. How suitable is entropy as a measure of urban sprawl? *Landsc Urban Plan* 2019;184:32–43. [\[CrossRef\]](#)
- [57] Das Chatterjee N, Chatterjee S, Khan A. Spatial modeling of urban sprawl around Greater Bhubaneswar city, India. *Model Earth Syst Environ* 2016;2:1–21. [\[CrossRef\]](#)
- [58] Al-Shamayleh A, Jew B, Totanji DA, Zakaria S, Nasser M, Tarawneh K. Development of pavement maintenance management system for an urban areas based upon multiple linear regression analysis. *Sigma J Eng Nat Sci* 2025;43:121–132. [\[CrossRef\]](#)
- [59] Roy B, Nuruzzaman K. Quantification of urban expansion in Siliguri urban agglomeration: a model-based approach. *GeoJournal* 2022;87:869–884. [\[CrossRef\]](#)
- [60] Lin J, He P, Yang L, He X, Lu S, Liu D. Predicting future urban waterlogging-prone areas by coupling the maximum entropy and FLUS model. *Sustain Cities Soc* 2022;80:103812. [\[CrossRef\]](#)
- [61] Cengiz S, Görmüş S, Oğuz D. Analysis of the urban growth pattern through spatial metrics: Ankara City. *Land Use Policy* 2022;112:105812. [\[CrossRef\]](#)
- [62] Ramachandra TV, Aithal BH, Sowmyashree MV. Urban structure in Kolkata: metrics and modelling through geo-informatics. *Appl Geomat* 2014;6:229–244. [\[CrossRef\]](#)



Research Article

Broken rotor bar fault detection in induction motors through power spectral density to image method

Hatice OKUMUS^{1,*}, Ebru ERGUN²

¹Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, 61080, Türkiye

²Department of Electrical and Electronics Engineering, Recep Tayyip Erdogan University, Rize, 53100, Türkiye

ARTICLE INFO

Article history

Received: 09 May 2024

Revised: 18 July 2024

Accepted: 04 October 2024

Keywords:

Broken Rotor Bars; Deep Feature Extraction; Fault Diagnosis; Induction Motors; Machine Learning

ABSTRACT

Induction motors play an important role in a variety of industrial applications but are particularly sensitive to electrical faults, such as rotor-related problems such as broken rotor bars. Eliminating such faults is critical to reducing maintenance costs and preventing serious financial losses. This study presents a method based on detailed feature extraction for identifying broken rotor pull-out faults in induction motors. The process is initiated by generating spectrograms from sensor-based signals. However, instead of using these spectrograms directly, the resulting power spectral density data is converted into an optimized image format suitable for processing by pre-classified deep neural networks. To utilize these networks' capabilities, the developed features are fed into nearest neighbor (k-NN) and random forest classifiers for fault detection. The programmatic method was tested on a publicly available dataset of a three-phase step-down motor operating under various load conditions. In particular, the DenseNet201 model's improved features from the mean pooling structure yielded a remarkable accuracy of 99.75% using the random forest classifier. This result demonstrates a powerful and sensitive fault detection tool in induction motors by effectively integrating the conventional circuit techniques with detailed extraction by the proposed method.

Cite this article as: Okumus H, Ergun E. Broken rotor bar fault detection in induction motors through power spectral density to image method. Sigma J Eng Nat Sci 2025;43(4):1473–1483.

INTRODUCTION

Induction motors (IMs) are critical components in industrial environments thanks to their fundamental role in converting electrical energy into mechanical energy and are known for their high efficiency and robustness [1]. However, during operation these motors are subjected to constantly changing stresses such as radial and rotary electromagnetic forces and thermal distortion. These stresses

can lead to various structural failures within the engine. Stator problems are often caused by the weakening of the insulation around the copper windings that can cause electrical short circuits. Rotor failures can occur for several basic reasons. Electrical problems such as overcurrent and electrical discharges can cause rotor bars to overheat and damage. Mechanical stresses such as vibration and misalignment can lead to cracks, while material defects in rotor components can lead to structural weakening. Additionally,

*Corresponding author.

*E-mail address: haticeokumus@ktu.edu.tr

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



thermal factors such as overheating and insufficient cooling cause thermal cycles that lead to wear on rotor materials [2]. Among rotor-related problems, broken rotor bar (BRB) failures are particularly common. These faults can seriously affect engine performance, causing increased vibration, decreased efficiency, overheating and electrical instabilities. Such disturbances can cause unpredictable torque output, which creates significant problems, especially in applications requiring precise control. Furthermore, BRB faults can increase maintenance and repair costs, increase the risk of motor failure, increase spare parts costs, and lead to operational downtime. Therefore, early detection and regular maintenance of these problems are crucial to maintaining the efficiency and lifespan of IMs.

Among BRB fault detection methods, Motor Current Signature Analysis (MCSA) can non-invasively identify faults by analyzing harmonic changes in the current's frequency [3,4]; vibration analysis reveals mechanical defects by evaluating the motor's abnormal vibration patterns[5]; thermal imaging detects overheating areas, visualizing problematic regions[6-8]; and frequency analysis detects faults through changes in operating frequency[2].

With the advancement of technology, tools such as artificial intelligence (AI) [9-11], machine learning (ML) [12,13], and advanced signal processing techniques [14,15] have been incorporated into broken rotor bar (BRB) fault detection methods. By leveraging big data analysis, AI and ML offer the opportunity to identify and predict faults by detecting abnormal behaviors based on data obtained from various sensors such as motor current, vibration, and temperature. Advanced signal processing techniques such as wavelet transform [16] and Fourier transform [17] enable more precise fault detection by revealing details and patterns in motor signals. In light of these advancements, several studies have introduced different methodologies for BRB detection and classification. Lizarraga-Morales et al. have developed an approach based on homogeneity as an index. The approach analyzes one phase of the induction motor startup-transient current, leveraging the homogeneity calculation. The authors also developed a hardware-processing unit based on a field programmable gate array device for online BRB detection and classification. The results demonstrate an accuracy greater than 99.7% for half, one, or two BRBs [18]. Chisedzi et al. have used decision tree classification, deep learning and artificial neural network methods for detecting BRB faults. Arthurs describe the process of identifying the faults by transforming the measured line-current signatures from time domain to frequency domain using discrete Fourier Transform. The decision tree method shows higher accuracy rates of 95% and 98% for 3BRB and 6BRB conditions [19]. Rayyam et al. have presented an approach using a combination of an ant lion optimizer (ALO) and an unscented Kalman filter (UKF) for broken bar fault monitoring in squirrel-cage induction motors (SCIM). The developed ALO-UKF algorithm adjusts the noise matrices for accurate fault detection.

Simulation results showed better performance in broken bar detection compared to traditional methods such as the simple-UKF and EKF algorithms [20]. Ullah et al. carried out a study on fault detection and classification in induction motors using ANSYS Maxwell based simulations. Considering various loading conditions, stator current and leakage flux data were produced in normal and faulty cases (broken bar faults, full pole gap, static eccentricity). A deep neural network (DNN) algorithm was proposed and compared with support vector machines (SVM) and random forest classifiers (RFC) for fault detection and classification. Especially under 100% loading conditions, the DNN algorithm provided higher accuracy with leakage flux data compared to SVM and RFC. However, its performance was not as effective with stator current data [21]. Lee et al. proposed the combination of symmetric uncertainty (SU) and genetic algorithm (GA) to improve feature selection in induction motor fault classification. The SU-GA method was applied to four different engine cases and analyzed with the Hilbert-Huang transform (HHT). SVM classification was performed using SU, GA and SU-GA methods, and the SU-GA method achieved higher accuracy with fewer features. Additionally, simulations with varying levels of white noise confirm the effectiveness of the proposed method in classifying motor faults [22]. A feature fusion residual CNN with a double branch (DBF-CNN) for BRBs in IMs is proposed in [23]. This network utilizes the Hilbert transform to highlight fault features and employs a residual structure to prevent degradation. Furthermore, a double-branch structure extracts global and local features separately, while an attention-based feature fusion method addresses feature loss. Each of these studies contributes to the field of BRB fault detection and highlights the importance of combining traditional diagnostic methods with advanced technological advancements to increase the reliability and efficiency of IMs.

In the field of signal processing, utilizing advances in image analysis by converting signals into visual formats and enhancing their interpretive capabilities is attracting increasing interest. This technique combines the fundamental principles of signal processing with the advanced analytical power of image analysis. Numerous studies in the literature present this approach, encompassing different methods and applications. Some studies that explain this concept are summarized below.

Barrera-Llana et al. proposed a Convolutional Neural Network (CNN) to detect broken rotor bars in squirrel-cage rotors. Using the Finite Element Method Magnetics (FEMM) software, they converted the current-angular position signals of a 28-bar squirrel-cage rotor into images and generated a dataset containing up to six broken bar scenarios at each location. Six different CNN architectures were evaluated in the study: Inception V4, NasNETMobile, ResNET152, SeNET154, VGG16, and VGG19. The VGG19 model performed best with an accuracy rate of

approximately 99% and achieved high precision, sensitivity and F1-score [24].

Shao et al. proposed a deep learning-based machine fault diagnosis method using time-frequency distributions obtained from wavelet transforms. In the study, the performance of the pre-trained models was compared with a CNN model trained from scratch on the same data. The proposed method was tested on three different datasets, and satisfactory results were obtained [25]. Huang and colleagues also detected machine faults using time-frequency distributions obtained from wavelet transforms. The Deep Continuous Convolutional Network (DCCN) uses multi-layer perceptrons with Gaussian masks to parameterize continuous convolution kernels and multiplicative Gabor filters to increase noise resistance. Furthermore, depth-separated convolutions with residual connections are applied to increase computational efficiency. The effectiveness of DCCN has been validated using two laboratory datasets and one public dataset. The method achieved an average macro F1 score of over 98.69% in diagnosing various machine faults and an average score of 86.43% in a special test with reduced training data [26].

Our study also aims to detect BRB faults in electronic devices by adopting the principle of converting signals into visual representations. First, the signals collected from sensors on the electronic device are converted into spectrograms. However, instead of using these spectrograms directly, Power Spectral Density (PSD) data is converted into a visual format. Then pre-trained DNNs are used to extract features from these images. The resulting features are analyzed with k-Nearest Neighbor (k-NN) and Random Forest (RF) classifiers, to obtain accurate fault detection. The effectiveness of this technique has been validated using a dataset obtained from the University of São Paulo, Brazil, which includes normal and BRB fault conditions at four different severity levels and eight different load cases (from no load to full load). Among the 11 pre-trained DNN models, features extracted from the mean pooling layer of the DenseNet201 model, in particular, achieved an impressive accuracy rate of 99.75% when used with the RF classifier.

The remaining sections of this document is organized as follows: Section 2 details the proposed methodology and discusses the pre-trained DNNs along with the used classification techniques. Section 3 presents the performance evaluation. Section 4 concludes the study.

MATERIALS AND METHODS

The proposed method for detection of BRB faults is given in Figure 1. As indicated in the figure, vibration and current signals received from the sensors first undergo a preprocessing phase. During this preprocessing stage, the spectrogram of the signals is generated, followed by the conversion of the resulting PSD values into images. This preprocessing is executed independently for both the current and vibration signals. Following this, deep features

are extracted from the resulting images using pre trained DNNs. The obtained features are combined and fed into RF and k-NN classifiers to assess their effectiveness.

PREPROCESSING

A spectrogram is a time–frequency representation that visualizes the evolution of a signal's frequency components over time [27]. It is obtained by applying the Short-Time Fourier Transform (STFT) to the signal $x(n)$ which reveals how its frequency content changes with time. The STFT can be calculated as shown in Equation (1):

$$STFT\{x(n)\}(m, \omega) = \sum_{n=-\infty}^{\infty} x(n) \cdot w(n-m) e^{-j\omega n} \quad (1)$$

In this equation:

$x(n)$ is the signal being analyzed.

$w(n-m)$ represents the window function, centered at time m .

$e^{-j\omega n}$ is the complex exponential function, where ω denotes the angular frequency.

As previously mentioned, this study does not directly utilize the spectrograms of the signals. Instead, it focuses on the Power Spectral Density (PSD) values derived from the spectrogram process. The PSD is obtained with Equation (2):

$$PSD = \frac{|STFT|^2}{N \times f_s} \quad (2)$$

Where N denotes the sample length while f_s indicates the sampling frequency.

The PSD values obtained are converted into a logarithmic scale, $10 \cdot \log_{10}(PSD)$, significantly improving the visualization of the signal's time-frequency content.

Deep learning models typically require input data in a three-channel (RGB) image format. To meet this requirement, we converted our initial single-channel dataset to a three-channel format by concatenating it. This way, each channel contains the same information; while this may seem unnecessary, it is essential to comply with the model's structural input format requirements. This format simplifies the process of the usage of advanced pre-trained models without the need of specific retraining for single-channel input, allowing for the efficient use of existing deep learning architectures.

Feature Extraction

DNNs are deep learning models commonly used in tasks such as image recognition. Layers like convolutional, pooling, fully connected, and activation functions make up a typical DNN. The network needs more computational power as layers are added because there are more parameters to learn. While training models on larger datasets give better performance, this process can be time-consuming. Therefore, transfer learning has become widespread. Transfer learning allows pre-trained models to

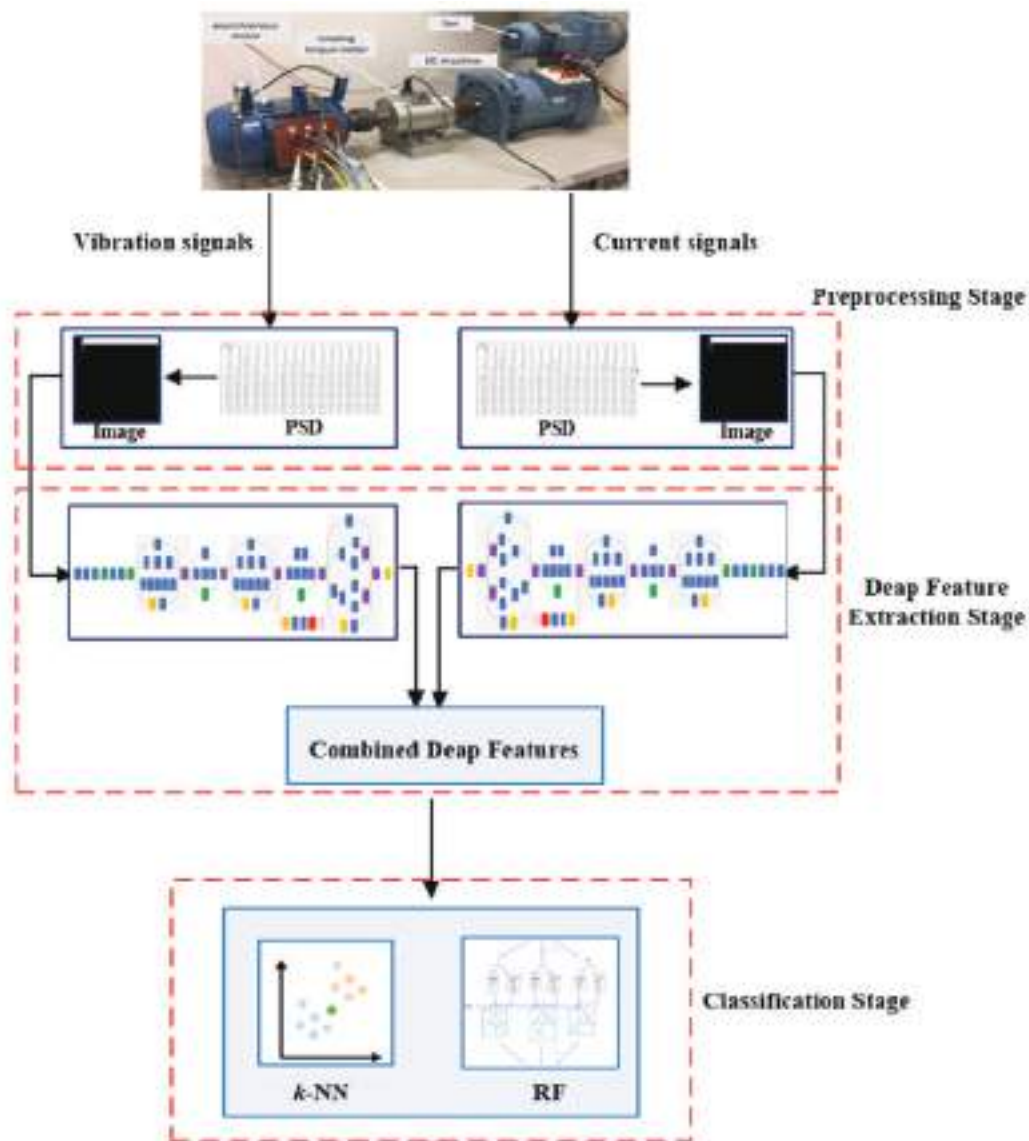


Figure 1. Block diagram of the proposed method.

be reused for different tasks. Furthermore, using knowledge from similar tasks can require less data than training from scratch. Many different DNN architectures have been developed over the years such as VGG, DenseNet, SqueezeNet, ResNet, MobileNet, AlexNet and Inception. In this study, 11 pre-trained models which are VGG-16, VGG-19, DenseNet-201, SqueezeNet, ResNet-50, MobileNetV2, AlexNet, Inception V3, NasNetMobile, DarkNet53, and Inception-ResNet-v2 are used for feature extraction.

The VGG-16 and VGG-19 models, created by the Visual Geometry Group (VGG) at the University of Oxford and first presented by Simonyan & Zisserman in 2014, are renowned for their structural simplicity and reliance on just three 3x3 convolutional layers [28]. The VGG-16 model has 16 layers (13 convolutions, 3 fully connected) while the VGG-19 model has 19 layers (16 convolutions,

3 fully connected). Although they have simple design both models performed well in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and are now among the main reference models for image classification. The VGG models are widely used feature extractors in many applications, particularly in deep learning research, due to their simplicity and practicality.

With 201 levels, DenseNet-201 is an enhanced version of the DenseNet architecture [29]. When compared to other deep architectures, this model has demonstrated exceptional performance on common benchmarks such as the ImageNet dataset, effectively addressing the vanishing-gradient problem and yielding higher accuracy with more economical parameter usage.

Researchers at DeepScale, Stanford University, and the University of California, Berkeley developed SqueezeNet in

2016, which presents the novel idea of “fire modules” [30]. These modules, which are designed to compress (squeeze) input channels before expanding them, effectively replace conventional convolutional layers, significantly reducing the model’s parameter count. As a result, a remarkably small network is produced, which is perfect for installation on devices with limited processing power or in embedded systems.

ResNet-50, a variant of the Residual Network (ResNet) designed by Kaiming He and his team at Microsoft Research, features a total of 50 layers (encompassing 177 sub-layers) [31]. ResNet broke the previous record in the 2015 ILSVRC competition with an error rate of just 3.37%. ResNet-50’s depth and effectiveness have made it a popular model for a variety of computer vision applications, highlighting its importance in the field.

Introduced in 2018 by Google’s research team [32], MobileNetV2 is a distinguished CNN architecture within the MobileNet series, designed to deliver models of high efficiency suitable for devices constrained by their computational power. Inverted residuals and linear bottleneck modules are integrated in this version, which advances from its predecessor, MobileNetV1. Distinct from conventional residual blocks that upscale dimensionality before convolution, MobileNetV2’s innovative bottleneck approach initially compresses dimensionality, conducts convolution, and subsequently restores the dimensionality. This approach minimizes model dimensions and maximizes computing performance, making it very beneficial for mobile systems. Moreover, MobileNetV2 makes use of depthwise separable convolutions, which significantly reduce computational burden and parameter count by splitting a normal convolution into a depthwise convolution and a 1×1 pointwise convolution.

AlexNet is a deep learning architecture introduced by Krizhevsky, Sutskever, and Hinton [33] that significantly advanced developments in computer vision. AlexNet consists of eight main layers; five convolutional layers and three fully connected layers. When pooling, ReLU (Rectangular Linear Unit), normalization, and softmax layers are included, the total number of layers of the architecture reaches 25. Having 62 million trainable parameters, AlexNet is designed to process 227×227 images, highlighting its complexity and capacity in processing visual data. With 62 million trainable parameters, AlexNet is designed to process images of size 227×227 , underscoring its complexity and capacity for handling intricate visual data.

Google’s 48-layer deep, pre-trained convolutional neural network model, Inception V3 network, is an additional development in CNN utility [34]. It performed admirably in the ILSVRC-2015 competition, achieving a top-5 accuracy rate of 92.8%. The use of tiny convolution kernels rather than huge ones is one of Inception V3’s primary architectural innovations. This is achieved by breaking down conventional bigger convolutions into a mix of one- and two-dimensional convolutions, i.e., into sequences

of $1 \times N$ and then $N \times 1$ convolutions. By using a strategic decomposition the model’s parameter count is dramatically reduced, improving its efficiency without compromising performance.

The NasNetMobile, a creation of the Google Brain team unveiled in 2017, embodies a design rooted in neural architecture search (NAS) principles [35]. The regular cell and the reduction cell are two different kinds of building components used in this construction. Multiple layers in normal cells improve feature map resolution, allowing for more in-depth picture analysis. Reduction cells, on the other hand, reduce the amount of data to be processed by reducing the spatial resolution of feature maps, thus ensuring more efficient processing. The final design is built by carefully placing these components in the right order to maximize the neural network’s depth and efficiency.

DarkNet53, an artificial intelligence model developed by Joseph Redmon, serves as the cornerstone of the YOLO (You Only Look Once) algorithm [36]. This model is widely used in deep learning and image processing fields, and shows high performance especially in real-time object detection applications. With the elements from the Deep Residual Network and DarkNet19 architectures, DarkNet53 is based on the sequential use of 1×1 and 3×3 convolutional layers with residual connections. This structure improves the model’s ability to efficiently process and analyze visual data, enabling it to identify objects in images with high accuracy and speed.

The Inception-ResNet-v2 model which stands out for its 164-layer architecture expertly blends the advantages of the ResNet (Residual Networks) and Inception models [37]. The goal of this combination is to improve computer vision capabilities by combining ResNet’s superiority in training very deep networks through residual connections with the Inception architecture’s skill on processing multi-scale picture information. This model which is designed to process 229×229 pixel image inputs, consists of a complex hierarchy of convolutional, pooling and dense layers. The model’s effectiveness in feature extraction and recognition is improved by this deliberate layering, which makes it a powerful tool for a variety of computer vision applications.

CLASSIFICATION METHODS

Random Forest (RF)

RF is a powerful technique that employs multiple decision trees to perform both classification and regression tasks [38]. It builds these trees by randomly sampling n datasets from the original training data using the bootstrap method. At each node of every tree, a subset of m variables is randomly selected, and the best split is determined based on the Gini index [39]. After choosing the optimal split, the tree is further divided into two branches. This process continues until each leaf node represents a single class [40]. The performance of each tree is evaluated by its Out-of-Bag

(OOB) errors, which are a measure of prediction accuracy on data not included in the bootstrap sample. Trees with lower OOB errors are considered more accurate and are given higher weight whereas those with higher errors receive lesser weight. When making predictions, each tree casts a vote weighted by its accuracy. The final classification decision is based on the majority of these weighted votes across all n trees, with the class garnering the most votes being declared the final outcome.

***k*- Nearest Neighbours (*k*-NN)**

The *k*-Nearest Neighbors (*k*-NN) algorithm is a straightforward machine learning method widely used for classification and regression tasks such as image classification, power load forecasting, fault detection and medical diagnosis. In this approach, predictions for a given data point are determined based on the labels or values of its closest *k* neighbors. The algorithm typically identifies these neighbours using a distance metrics such as Euclidean, Manhattan and Minkowski Distance. In classification cases, the most frequent class label among these neighbours is returned as the output; in regression, it calculates the average of these neighbours output values. The effectiveness of this method heavily depends on the correct determination of the '*k*' value.

In this study, the Euclidean distance metric is employed to compute the distances between neighbors. The calculation of Euclidean distance between two points, r and s , in an n -dimensional space is defined as shown in Equation (3) [41].

$$d_E(r, s) = \sqrt{\sum_{i=1}^n (r_i - s_i)^2} \quad (3)$$

RESULTS AND DISCUSSION

In this study, a dataset containing data from an induction motor with broken rotor bar faults, made publicly available by the University of São Paulo in Brazil, was used

[42]. The data set includes a four-pole, three-phase motor with an output power of 1 hp, a supply voltage of 220/380V, and a nominal speed of 1785 rpm. Faults are divided into four levels (from 1 to 4) according to the number of broken rods, and a separate category is defined for fault-free engines (health class).

To comprehensively evaluate these categories, eight different load cases were determined, with torque values ranging from 0.5 Nm to 4.0 Nm, in 0.5 Nm increments. For each condition, 10 experiments were conducted to comprehensively collect both electrical and mechanical signals. The experimental setup for the dataset is shown in Figure 2. Some sample examples of the images transformed from PSD values for healthy and broken rotor bar under different load conditions are given in Figure 3.

Due to the varying sizes of each signal recorded from sensors, in this study, 1001000 data points are used for current signals (I_a), and 140000 data points for vibration signals (V_{ib_acpi}). As stated before there are 8 different loading conditions and 5 fault classes, and each experiment repeated 10 times, resulting in a total of 400 image data. This dataset is randomly split into 80-20% for training and testing. To ensure more stable and accurate results, this process is repeated 10 times.

The classification accuracy results for 11 different DNN models, showing the performance analysis at various layers using *k*-NN and RF methods, along with the standard deviations are given in Table 1. Among these architectures, DenseNet201 (avg_pool) and InceptionV3 (avg1) have given the highest performance with almost perfect accuracies of 99.75% and 99.25%, respectively, with the RF classifier. Meanwhile, InceptionResNetV2, NasNetMobile (global avgpooling2d), and AlexNet (fc8) achieve lower accuracies of 78.25%, 80.62%, and 81.87%, respectively, when using the *k*-NN algorithm. These results show that most models perform better with RF compared to *k*-NN, suggesting that RF's model utilizes the deep features generated by these networks more effectively.

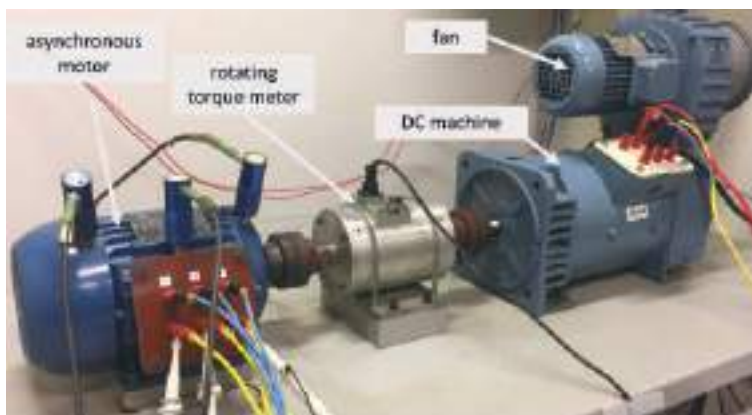
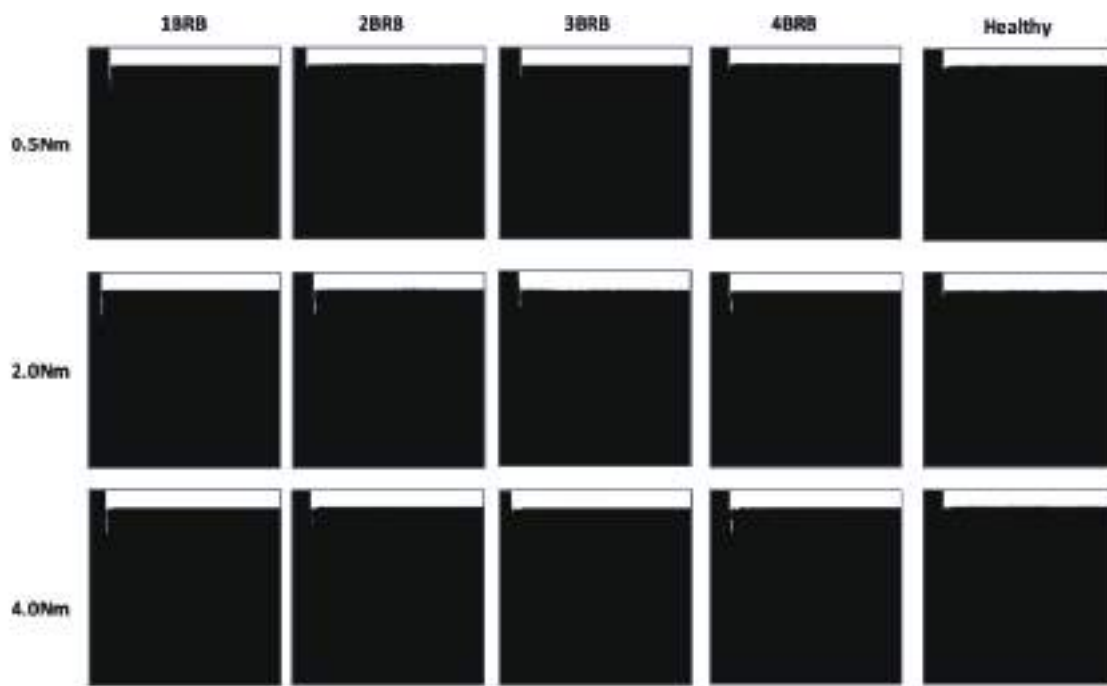


Figure 2. Experimental setup.

Table 1. Classification accuracy results using V_{ib} and I_a signals

Deep feature	k -NN	RF
AlexNet (fc7)	88.37±2.5	97.87±1.56
AlexNet (fc8)	81.87±4.80	96.75±2.14
DenseNet201 (avg_pool)	96±1.64	99.75±0.53
MobileNetV2 (global avg pooling2d)	93.62±2.32	99±0.79
DenseNet201 (fc1000)	95.5±1.21	97.62±1.09
InceptionV3 (avg_pool)	89.12±1.96	99.25±1.05
Vgg19 (fc7)	91.5±2.02	98.75±1.18
Vgg16 (fc7)	94.37±2.52	98.37±1.03
SqueezeNet (pool10)	82.6±3.03	93.12±3.45
DarkNet53 (avg1)	98.50±1.64	98.75±1.56
NasNetMobile (global avg pooling2d)	80.62±3.50	96.12±2.08
Inception-ResNetV2 (avg_pool)	78.25±4.94	92.62±2.24
ResNet50 (fc1000)	89.75±3.67	98.62±1.37

**Figure 3.** PSD-Image transformed samples for healthy and broken rotor bar under different load conditions.

If a closer look is taken at the standard deviation values as illustrated in Figure 4, it can be observed that the k -NN algorithm follows a trend ranging between 1.21 and 4.94, while the RF classifier exhibits standard deviation values between 0.79 and 3.45. These values clearly demonstrate that the RF classifier provides more consistent results in solving the BRB fault detection problem.

In the literature, there are studies that attempt BRB fault detection using only current data. Consequently, this

study has also examined the impact of I_a current signals on fault detection results. When reviewing the results in Table 2, it is observed that classifications carried out using the k -NN method generally provide better outcomes compared when using the combined features. While for the RF method, the use of combined features are found to be more effective. The highest overall result is once again obtained using the DenseNet201 model (avg_pool) combined with RF classification. The standard deviation

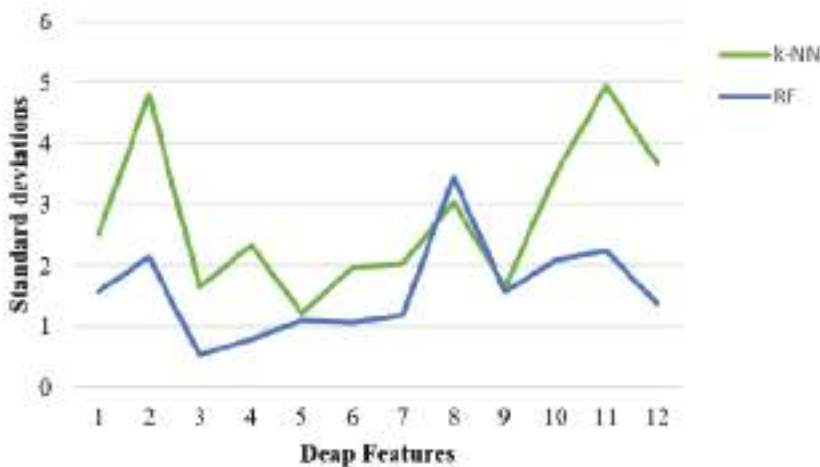


Figure 4. Standard deviation results.

results consistently demonstrate that the k -NN method is significantly more effective when extracting features solely from I_a signals.

The method introduced in this study is evaluated against previous research, as outlined in Table 3, all using the same dataset. References [43,44] and [45] implement a transformation of signals to images, with significant variations in approach. Whereas, reference [23] applies the Hilbert transformation to signals and feeds the transformed data into a dual-branch CNN model. In contrast to other studies in the comparison, reference [23] addresses a three-class problem involving 1BRB, 2BRB, and 3BRB and achieves an accuracy of 99.86%, whereas the others tackle a more complex five-class scenario, including HLT, 1BRB, 2BRB, 3BRB, and 4BRB. Moreover, both References [43] and [44] create spectrogram images from signals. However, the used

methodologies are different; reference [44] uses only current signals and the short time fourier transform (STFT) for spectrogram generation, while reference [43] incorporates both vibration and current signals, employing the continuous wavelet transform (CWT) to produce spectrograms, this approach yields an accuracy score of 100%. Reference [44] takes a different approach by using time-domain gray-scale current imaging techniques coupled with CNN, yielding a 99.53% accuracy.

The proposed methods in this study which employ PSD transformations to image format followed by deep feature extraction, exhibit an effective approach to the classification task. The first proposed model achieves an accuracy of 98.5% using current signals, while the second, utilizing both vibration and current signals, reaches 99.75% accuracy.

Table 2. Classification accuracy results using I_a signals

Deep Feature	k -NN	RF
AlexNet (fc7)	97.50±1.67	97.25±1.54
AlexNet (fc8)	96.37±1.24	96.12±1.71
DenseNet201 (avg_pool)	95.8±2.83	98.5±1.15
MobileNetV2 (global avg pooling2d)	93±2.84	94.6±3.44
DenseNet201 (fc1000)	93.5±2.55	96±3.05
InceptionV3 (avg_pool)	97.5±1.02	98.5±0.99
Vgg19 (fc7)	94.25±2.22	95.62±1.69
Vgg16 (fc7)	95.25±2.49	95.37±1.56
SqueezeNet (pool10)	93.12±3.50	94.5±.71
DarkNet53 (avg1)	98. ±1.58	98.37±1.67
NasNetMobile (global avg pooling2d)	96.12±2.34	96.2±1.86
Inception-ResNetV2 (avg_pool)	97±1.79	96.12±1.50
ResNet50 (fc1000)	95±2.95	96.12±2.32

Table 3. Comparison with the research in the literature

Reference Study	Method	Used signal	Accuracy (%)
[23]	Hilbert transform+double-branch fusion residual CNN	Current signals	99.86
[43]	Spectrogram images with CWT+deap features	V_{ib_acpi} and I_a	100
[44]	Spectrogram images +CNN	Vibration signals	97.67
[45]	Time domain grayscale current imaging technique+ CNN	Current signals	99.53
Proposed	PSD to image transform+ Deap feature extraction	I_a	98.5
Proposed	PSD to image transform+ Deap feature extraction	V_{ib_acpi} and I_a	99.75

CONCLUSION

This study proposes a robust method for detecting BRB faults in IMs using advanced deep feature extraction techniques. By converting PSD data into image formats which can be analyzed by pre-trained DNNs, the proposed approach effectively utilizes different properties of motor signals. The use of k-NN and RF classifiers improves the fault detection process by enabling precise detection of faults that vary under different load conditions. Validation of this approach was performed with publicly available dataset provided the University of São Paulo, Brazil, containing normal and BRB fault cases at four different severity levels and eight different load scenarios. The DenseNet201 model achieved a high classification accuracy of 99.75% with its average pooling layer along with the RF classifier using both vibration and current signals. DenseNet201's densely connected architecture, parameter efficiency, and depth make it a powerful model. The average pooling layer improves performance by providing a comprehensive summary of features, providing robustness to noise, and encouraging balanced learning. These features allow DenseNet201 to outperform other models, making it particularly suitable for complex and sensitive tasks such as BRB fault detection. This research sets a benchmark for future studies aimed on fault detection in IMs. Creating a comprehensive dataset with different motor types, fault conditions and operating scenarios will be beneficial for further development. Also, the accuracy and robustness of fault detection can be greatly increased by using ensemble methods and real-time data processing.

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] Gündoğdu T. Torque capability comparison of induction and interior permanent magnet machines for traction applications. *Gazi Univ J Sci* 2023;1:1. [\[CrossRef\]](#)
- [2] Kumar P, Hati AS. Review on machine learning algorithm based fault detection in induction motors. *Arch Comput Methods Eng* 2021;28:1929–1940. [\[CrossRef\]](#)
- [3] Ince T, Kiranyaz S, Ren L, Askar M, Gabbouj M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans Ind Electron* 2016;63:7067–7075. [\[CrossRef\]](#)
- [4] de Las Morenas J, Moya-Fernandez F, Lopez-Gomez JA. The edge application of machine learning techniques for fault diagnosis in electrical machines. *Sensors* 2023;23:2649. [\[CrossRef\]](#)

- [5] Delgado-Arredondo PA, Morinigo-Sotelo D, Osornio-Rios RA, Avina-Cervantes JG, Rostro-Gonzalez H, Romero-Troncoso RJ. Methodology for fault detection in induction motors via sound and vibration signals. *Mech Syst Signal Process* 2017;83:568–589. [\[CrossRef\]](#)
- [6] Sasikumar B, Venkatasalam K, Rajendran P. Induction motor fault detection and classification using RCNN and SURF based machine learning algorithms and infrared thermography. *Sci Iran* 2024. doi: 10.24200/SCI.2024.61897.7589 [Epub ahead of print] [\[CrossRef\]](#)
- [7] Resendiz-Ochoa E, Enriquez-Ugalde JM, Saucade-Dorantes JJ, Morales-Hernandez LA. Broken rotor bar failures diagnosis with supervised learning and infrared thermography. In: *IEEE 13th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*. Dallas (TX): IEEE; 2017. p.499–504. [\[CrossRef\]](#)
- [8] Trejo-Chavez O, Cruz-Albarran IA, Resendiz-Ochoa E, Salinas-Aguilar A, Morales-Hernandez LA, Basurto-Hurtado JA, et al. A CNN-based methodology for identifying mechanical faults in induction motors using thermography. *Machines* 2023;11:752. [\[CrossRef\]](#)
- [9] Elhaija WA, Al-Haija QA. A novel dataset and lightweight detection system for broken bars induction motors using optimizable neural networks. *Intell Syst Appl* 2023;17:200167. [\[CrossRef\]](#)
- [10] Khishe M. Variable-length CNNs evolved by digitized chimp optimization algorithm for deep learning applications. *Multimed Tools Appl* 2024;83:2589–2607. [\[CrossRef\]](#)
- [11] Yildiz S, Amasyali MF. Iterative ensemble pseudo-labeling for convolutional neural networks. *Sigma J Eng Nat Sci* 2024;42:862–874. [\[CrossRef\]](#)
- [12] Samiullah M, Ali H, Zahoor S, Ali A. Fault diagnosis on induction motor using machine learning and signal processing. *arXiv preprint arXiv:2401.15417*. 2024.
- [13] Yutong G, Khishe M, Mohammadi M, Rashidi S, Nateri MS. Evolving deep convolutional neural networks by extreme learning machine and fuzzy slime mould optimizer for real-time sonar image recognition. *Int J Fuzzy Syst* 2022;24:1371–1389. [\[CrossRef\]](#)
- [14] Hernandez-Ramirez V, Almanza-Ojeda D-L, Cardenas-Cornejo J-J, Contreras-Hernandez J-L, Ibarra-Manzano M-A. Detection of broken bars in induction motors using histogram analysis of current signals. *Appl Sci* 2023;13:8344. [\[CrossRef\]](#)
- [15] Maouche Y, Oumaamar MEK, Boucherma M, Khezzar A. Instantaneous power spectrum analysis for broken bar fault detection in inverter-fed six-phase squirrel cage induction motor. *Int J Electr Power Energy Syst* 2014;62:110–117. [\[CrossRef\]](#)
- [16] Agah GR, Rahideh A, Khodadadzadeh H, Khoshnazar SM, Kia SH. Broken rotor bar and rotor eccentricity fault detection in induction motors using a combination of discrete wavelet transform and Teager–Kaiser energy operator. *IEEE Trans Energy Convers* 2022;37:2199–2206. [\[CrossRef\]](#)
- [17] Ameid T, Talhaoui H, Harzelli I. Broken rotor bar fault diagnosis using fast Fourier transform applied to field-oriented control induction machine: simulation and experimental study. *Int J Adv Manuf Technol* 2017;92:917–928. [\[CrossRef\]](#)
- [18] Lizarraga-Morales E, Rodriguez-Donate C, Cabal-Yepez E, Ramirez ML. Novel FPGA-based methodology for early broken rotor bar detection and classification through homogeneity estimation. *IEEE Trans Instrum Meas* 2017;66:1760–1769. [\[CrossRef\]](#)
- [19] Chisedzi LP, Muteba M. Detection of broken rotor bars in cage induction motors using machine learning methods. *Sensors* 2023;23:9079. [\[CrossRef\]](#)
- [20] Rayyam M, Zazi M. A novel metaheuristic model-based approach for accurate online broken bar fault diagnosis in induction motor using unscented Kalman filter and ant lion optimizer. *Trans Inst Meas Control* 2020;42:1537–1546. [\[CrossRef\]](#)
- [21] Ullah N, Abbas MF, Kazmi SAA, Numan M. Machine learning-based fault classification using stray flux and stator current in induction motor. In: *3rd IEEE International Conference on Artificial Intelligence (ICAI)*. IEEE; 2023. [\[CrossRef\]](#)
- [22] Lee CY, Hsieh YJ, Le TA. Induction motor fault classification based on combined genetic algorithm with symmetrical uncertainty method for feature selection task. *Mathematics* 2020;10:230. [\[CrossRef\]](#)
- [23] Shu Y. DBF-CNN: A double-branch fusion residual CNN for diagnosis of induction motor broken rotor bar. *IEEE Trans Instrum Meas* 2023;72:1–10. [\[CrossRef\]](#)
- [24] Barrera-Llana K, Burriel-Valencia J, Sapena-Bañó Á, Martínez-Román J. A Comparative Analysis of Deep Learning Convolutional Neural Network Architectures for Fault Diagnosis of Broken Rotor Bars in Induction Motors. *Sensors (Basel)* 2023;23:8196. [\[CrossRef\]](#)
- [25] Shao S, McAleer S, Baldi P. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans Ind Inform* 2018;15:2446–2455. [\[CrossRef\]](#)
- [26] Huang X, Xie T, Wu J, Zhou Q, Hu J. Deep continuous convolutional networks for fault diagnosis. *Knowl Based Syst* 2024;292:111623. [\[CrossRef\]](#)
- [27] Khan AS, Ahmad Z, Abdullah J, Ahmad F. A spectrogram image-based network anomaly detection system using deep convolutional neural network. *IEEE Access* 2021;9:87079–87093. [\[CrossRef\]](#)
- [28] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.

- [29] Huang G, Liu Z, van der Maaten L. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2017. p. 4700–4708. [\[CrossRef\]](#)
- [30] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. arXiv preprint arXiv:1602.07360. 2016.
- [31] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016. p. 770–778. [\[CrossRef\]](#)
- [32] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L-C. Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation. arXiv preprint arXiv:1801.04381. 2018. [\[CrossRef\]](#)
- [33] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60:84–93. [\[CrossRef\]](#)
- [34] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016. p. 2818–2826. [\[CrossRef\]](#)
- [35] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2018. p.8697–8710. [\[CrossRef\]](#)
- [36] Redmon J. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2016. [\[CrossRef\]](#)
- [37] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. Proc AAAI Conf Artif Intell 2017;31. [\[CrossRef\]](#)
- [38] Breiman L. Random forests. Mach Learn 2001;45:5–32. [\[CrossRef\]](#)
- [39] Pal M. Random forest classifier for remote sensing classification. Int J Remote Sens 2005;26:217–222. [\[CrossRef\]](#)
- [40] Watts JD, Powell S, Lawrence RL, Hilker T. Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery. Remote Sens Environ 2011;115:66–75. [\[CrossRef\]](#)
- [41] Choubey H, Pandey A. A combination of statistical parameters for the detection of epilepsy and EEG classification using ANN and KNN classifier. Signal Image Video Process 2021;15:475–483. [\[CrossRef\]](#)
- [42] Trembl AE, Flautzino RA, Suetake M, Ravazzoli Maciejewski NA. Experimental database for detecting and diagnosing rotor broken bar in a three-phase induction motor. IEEE DataPort 2020. Available at: <https://iee-dataport.org/open-access/experimental-database-detecting-and-diagnosing-rotor-broken-bar-three-phase-induction> Accessed on Sep 02, 2025.
- [43] Dişli F, Gedikpınar M, Sengur A. Deep transfer learning-based broken rotor fault diagnosis for induction motors. Turk J Sci Technol 2023;18:275–290. [\[CrossRef\]](#)
- [44] Misra S, Kumar S, Sayyad S, Bongale A, Jadhav P, Kotecha K, et al. Fault detection in induction motor using time domain and spectral imaging-based transfer learning approach on vibration data. Sensors 2022;22:8210. [\[CrossRef\]](#)
- [45] Gundewar S, Kane P, Andhare A. Detection of broken rotor bar fault in an induction motor using convolutional neural network. J Adv Mech Des Syst Manuf 2022;16:JAMDSM0020. [\[CrossRef\]](#)



Research Article

Cart KNN: A hybrid distance-based pre-pruned classification and regression method for the early diagnosis of Alzheimer's disease

Pijush DUTTA^{1,*}, Sunita SHARMA², Varun Kumar NOMULA³, Gour Gopal JANA¹,
Anubrata MONDAL⁴, Shobhandeb PAUL⁵

¹Department of Electronics and Communication Engineering, Greater Kolkata College of Engineering and Management, West Bengal, 743387, India

²School of Biotechnology, University Jawaharlal Nehru University, New Delhi, 110067, Delhi

³Principal AI/ML Engineer, Apex IT Services LLC, Atlanta, 10632, USA

⁴Department of Electrical Engineering, Greater Kolkata College of Engineering and Management, West Bengal, 743387, India

⁵Data Scientist, TCS Oval, Newtown, West Bengal, 700160, India

ARTICLE INFO

Article history

Received: 01 April 2024

Revised: 12 June 2024

Accepted: 24 September 2024

Keywords:

Alzheimer's Disease Prediction;
CART-KNN; Confusion Matrix;
Decision Tree; K- Nearest
Neighbors; Pruning

ABSTRACT

Alzheimer's disease is a primary cause of dementia in the aged population furthermore, a significant segment of the global populace experiences metabolic disorders including diabetes and Alzheimer's disease. Alzheimer's disease has a damaging effect on the brain. Due to its detrimental effects on memory and physical functioning, this condition may lead to an increase in the number of inactive individuals as the senior population rises. To identify these diseases early on, researchers have looked into a variety of deep learning and machine learning techniques. Patients with Alzheimer's disease (AD) can recover from it more successfully and with less damage if they receive early diagnosis and therapy. Decision trees are used in conjunction with K-nearest neighbor and classification and regression trees to forecast and recommend the AD. The overall experiment is represented in three stages, in the first stage a comparative study was analyzed with the Proposed model along with its fundamental algorithm k-nearest neighbor algorithm (KNN) and detection threshold (DT), and in the second stage four performance indicators of the confusion matrix were used to demonstrate that our suggested strategy performs better in terms of running time performance metrics, accuracy, precision, and F1 score. In the final stage, the present experiment demonstrates a notable improvement in running time over the conventional k-nearest neighbor technique with an accuracy of 98.23%, the suggested model outperforms other cutting-edge methods for the OASIS dataset. Moreover, the proposed model also has better experimental result also suggests that the proposed CART-KNN model performed better than other general models: K-nearest neighbor and Decision tree by showing smaller root mean square error (RMSE) and mean absolute percentage error (MAPE).

Cite this article as: Dutta P, Sharma S, Nomula VK, Jana GG, Mondal A, Paul S. Cart KNN: A hybrid distance-based pre-pruned classification and regression method for the early diagnosis of Alzheimer's disease. Sigma J Eng Nat Sci 2025;43(5):1484–1494.

*Corresponding author.

*E-mail address: pijushdutta009@gmail.com

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



INTRODUCTION

The most prevalent type of dementia that needs intensive medical attention is Alzheimer's disease (AD). An accurate and timely study of AD prognosis is required to start clinical progress and provide effective patient therapy [1]. AD is a long-term, degenerative brain disease that gradually destroys brain tissue, and impairs thinking and memory [2]. The amount of computing power used by healthcare departments is always growing, and patient data is increased electronically. Although many electronic health records (EHRs) are now easier to access, 80% of the data remains unstructured. Because of this, handling unstructured data with database management software and other conventional techniques is difficult. Some methodologies adopted for EHRs through machine learning and data mining technologies improve the standard and efficiency of healthcare and prescribe proper medicine [3]. Finding undiscovered and meaningful patterns from a large number of pre-existing datasets is a process known as data mining or knowledge discovery. These patterns aid in the comprehension of the historical dataset, the classification of a new set of data, and the creation of data summaries. Data mining categorizes or groups records according to their similarities or differences, hence aiding in identifying deeper patterns within the data [4]. Data mining has been widely applied in various fields over the past few decades, including marketing, retail, finance, stock market forecasting, and medical and healthcare [5]. Machine learning and data mining techniques have been widely employed in medical and healthcare research [6]. These algorithms may be used to categorize distinct subjects based on the similarities in their qualities. The primary goal of this research project is to identify the various phases of Alzheimer's disease (AD) using machine learning and data mining techniques on the Open Access Series of Imaging Studies (OASIS). Finding the most distinctive feature for each of the several phases of AD within the OASIS dataset is also a sub-objective of this research project. Several studies focus on using machine learning to identify blood-based non-amyloid biomarkers for early Alzheimer's disease detection, achieving high sensitivity and specificity levels with novel protein panels [7]. Machine learning, particularly the Random Forest model, achieves 90.6% accuracy in early Alzheimer's detection using neuroimaging biomarkers, emphasizing its robustness and diagnostic advancements in Alzheimer's disease prediction [8]. Random Forest, XGBoost, and Light GBM were used to predict Alzheimer's disease risk in individuals aged 65+ using Finnish register data, showing low predictive accuracies and suggesting additional data sources were needed [9]. Gradient boosting, random forest, and support vector machines utilizing EEG signals and genetic information effectively classifies Alzheimer's disease, achieving high accuracy and offering a comprehensive diagnostic approach beyond imaging limitations [10]. Novel drug discovery for Alzheimer's disease utilized machine learning to identify compounds targeting key proteins like MTOR and BCL2, offering a multi-targeted

therapeutic approach for AD treatment [11]. An explainable machine learning approach for Alzheimer's disease classification achieved high accuracy using SVM models and highlighted key factors like MEMORY and JUDGMENT in AD development [12]. Machine learning-based speech analysis can effectively distinguish between different cognitive impairments in the Alzheimer's disease spectrum and predict cognitive domain performance, offering the potential for early detection and monitoring [13].

In recent research data modeling is quite challenging and complex during big data and short-term prediction. A fusion of CART and KNN models has been proposed to solve this issue. A nonlinear CART model is effective for data porting and mining purposes [14]. KNN, a popular nonparametric regression model for searching similar data in a particular region [15]. A novel hybrid approach of CART and KNN has been experimented with for software fault prediction. The proposed approach achieved an average accuracy of 0.90274 which is much higher than the other nine machine-learning models [16]. This work's main contributions are: to consider the input feature and perform the autocorrelation for finding out the data pattern similarity by CART model. Secondly, combined a regression tree with weighted k-nearest neighbors and distance-based prepruned classification to create a hybrid technique.

The present research is arranged into the following stages: in the literature review section, Research gaps and earlier studies are described. The proposed methodology including dataset description, preprocessing, pruning, and CART-KNN method explained. The final section experimentation and performance assessment of our suggested methodology are covered.

Literature Review

Feature selection and extraction techniques are used in machine learning (ML) algorithms to predict Alzheimer's disease. The Oasis dataset is used as the basis for the classification process. A quick summary of the various methods used in brain image analysis for brain disease diagnosis is given [17]. Based on the findings, this article discusses several important brain disease diagnoses. This work intends to integrate contemporary research on Alzheimer's disease using different AI platforms. The most accurate diagnosis approach may be identified by the authors by utilizing 21 brain illness reviews. Deep convolutional autoencoders investigate AD data analysis where the proposed method extracts MRI characteristics that reflect neurodegeneration and cognitive symptoms from MRI images [18]. With imaging-derived indicators, MMSE or ADAS11 scores can be utilized with over 80% accuracy to predict the diagnosis of AD. To achieve binary classification, a deep neural network with linked layers is employed [19, 20] and to activate each concealed layer distinct activation function was used. A lack of education, high blood pressure, obesity, hearing loss, depression, diabetes, inactivity, smoking, and social isolation are some of the variables that might increase one's risk.

Deep Learning with CNN methods has been proposed to extract MRI features for early Alzheimer's diagnosis, achieving over 97% accuracy, and enhancing detection in its earliest stages [21]. ML models like Gaussian NB, Decision Tree, and Voting Classifier can predict AD disease early with high accuracy, aiding in successful treatment and minimizing harm. Machine learning algorithms can accurately predict Alzheimer's disease with a 96% validation accuracy [22]. A CNN-SVM model was investigated for early Alzheimer's disease prediction using MRI images, achieving 94.44% accuracy on Mild Cognitive Impairment (MCI) subjects [23]. A novel ensemble machine learning approach, incorporating feature selection and adaptive voting, was examined in early Alzheimer's disease detection. The proposed model achieved 93.92% accuracy outperforming traditional methods by 3.39% [24]. A classification model was proposed using SVM, RF, and FNN algorithms for early detection of Alzheimer's disease, enhancing efficiency and reducing processing time in medical data analysis [25]. A speech-based machine learning has been designed for early detection of Alzheimer's disease with a 75.59% accuracy, showing promise for non-invasive diagnosis [26]. Machine learning models like DT, RF, and SVM have been investigated to enable early Alzheimer's disease prediction with up to 80% accuracy using OASIS data [27]. Machine learning models, like Gaussian Naive Bayes, have been proposed to accurately predict Alzheimer's disease early by analyzing neuroimaging and cognitive data with a classification accuracy of 96.92% [28]. Machine learning models like DT, RF, SVM, Voting classifiers, and GB have been investigated for the Prediction of early stages of Alzheimer's disease, achieving 83% maximum accuracy on test data [29]. A novel algorithm has been developed using a modified capsule network on the OASIS dataset, achieving 92.39% accuracy in predicting dementia, and outperforming traditional methods and deep learning classifiers [30].

After reviewing earlier research, we created a hybrid strategy based on decision trees and k-nearest neighbors. The following is a list of the proposed work's primary contributions:

1. a novel method known as distance-based pruning is suggested to prune decision tree nodes. The last part provides a detailed explanation of the distance-based pruning approach's processes.
2. Our suggested method reduces the $O(n)$ running cost of the standard k closest neighbor technique to $O(\log n) + c$ explained in section 3. KNNs are introduced to the decision tree's leaf nodes to lower running costs during the training phase.
3. The CART is created using decision tree's-based leaf nodes, k closest neighbors are placed in place of class labels.
4. To improve the effectiveness of conventional k nearest neighbors, the idea of weights is added in the prediction phase, based on the sigmoid function. The closest neighbors' distance from the site under consideration inversely determines the weights.

WORKING APPROACH

This paper suggests a hybrid method based on weighted k-nearest neighbors, regression trees, and distance-based pre-pruned categorization. The hybrid CART-KNN approach aims to improve prediction accuracy and generalization by combining the decision tree's structured, rule-based model with KNN's instance-based, flexible learning. There are various ways to hybridize CART and KNN, but a common approach involves using CART to preprocess data or make initial predictions and then applying KNN to refine these predictions. The overall process can be performed in three stages: tree-based segmented preprocessing, KNN-based post-processing, and finally prediction of the hybrid model. During the Tree-Based Segmented preprocessing stage CART is used to partition the feature space into distinct regions based on the decision tree's splits. Now KNN algorithm will only consider neighbors within the same region, potentially leading to more localized and accurate predictions. In KNN as a Post-Processing Step after that KNN refines and boosts lower tree node confidence of CART's prediction. During the Hybridization of the Predicted Model, the prediction can be done by either combining predictions by using weighted average prediction obtained by CART and KNN or a voting mechanism where the best prediction is considered by a majority vote from both models. This methodology helps Hybrid CART-KNN to reduce the Overfitting of the distribution of data, improve Locality Sensitivity by handling varying data densities, and flexibility for adapting to different types of data and problems. The run time of the proposed algorithm is reduced to $O(\log n) + c$ from $O(n)$, where each factor has its significance like the notion of $O(n)$ is linear time, while $O(\log n)$ running times divide-and-conquer application because of ideally cutting the work in half every time. m-dimensional points represent independent characteristics in the dataset. A K – Matrix is generated with the following condition; the element $e_{ij} = 0$ if the j^{th} training sample is far from the i^{th} training sample otherwise $e_{ij} = 1$. Following the KNN-matrix computation, the Euclidean distance formula is used to determine the maximum distance Max distance between all points, and a constant parameter called tolerance is added to regulate the creation of decision trees. Each node of DT fulfills the criteria $Max_{distance} * \text{tolerance}$, then the KNN matrix stores the leaf node value rather than class labels. The Weighted K-Tree (WK Tree) creation technique is described in full in this portion of the study publication [31]. These are the steps that explain how to construct a WK-Tree:

Pseudo Code

Let the Input contain, Training Samples X, and Testing Classes Y, and the Output contain Confusion Matrix

Step 1: m dimensional space is considered for all training samples and Euclidean distance is measured by eqn. (1).

Step 2: KNNs construct a matrix by considering all training samples using eqn. (2).

Step 3: Applied CART, a distance-based pre-pruning

Step 4: KNN stores all the leaf nodes without repetition and removes all duplicate data.

Step 5: The decision Tree helps the testing samples to try to reach the leaf node.

Step 6: The weighted label of each test point is assigned.

Step 7: Based on predicted and actual values Confusion matrix is designed.

Step 8: Accuracy, sensitivity, Precision, and F1-Score are calculated with the help of Confusion Matrix

Calculation of the Largest Distance

The approach's initial step involves treating all training samples lying in m -dimensional space and computing the maximum distance between them all. Distance in m dimensions was calculated between all training samples [32]. Equation (1) may be used to determine the Euclidean distance or space,

$$d(P_i, P_j) = \max_{1 \leq i, j \leq n} \sqrt{\sum_{k=1}^m (P_i^k - P_j^k)^2} \quad (1)$$

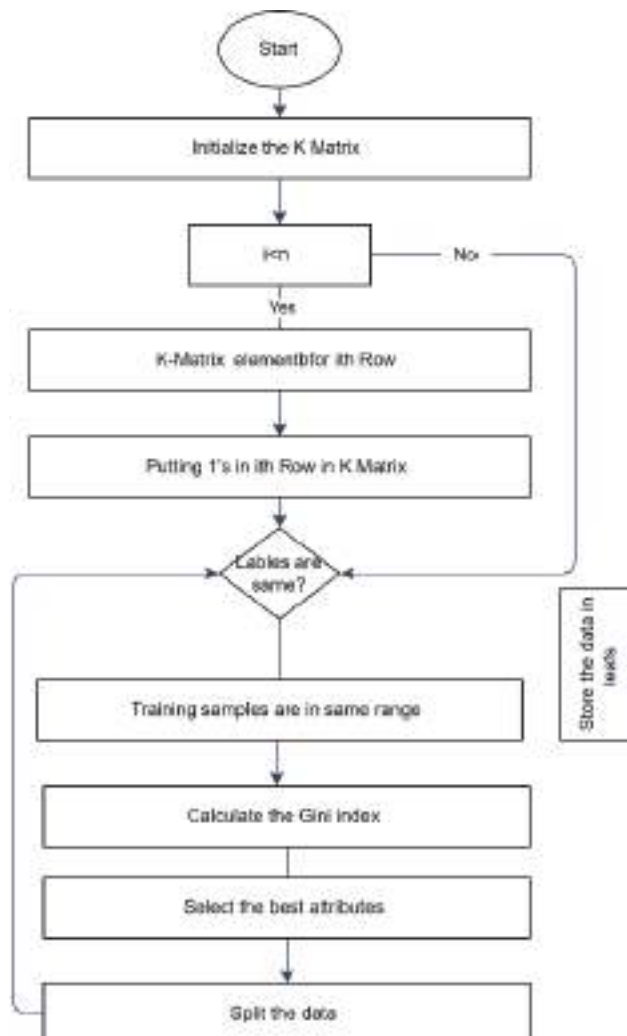


Figure 1. Flow Diagram of the present research.

Iterating over the number of dimensions is the variable k while iterating over the number of training samples is the variable i & j .

KNN Matrix Generation

For every training sample, a $n * n$ matrix of KNN is constructed. The function $\sqrt{n} + c$ is chosen to determine the closest neighbors. Equation (2) displays an example of a $5 * 5$ KNN matrix, where all diagonal components are equal to 1. Position j is regarded as the closest neighbor of position i if the element e_{ij} of the K- matrix is 1, but not the nearest neighbor point if the element e_{ij} of the KNN matrix is 0 (Fig. 1).

$$\text{KNN-Matrix} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \quad (2)$$

Decision Tree Creation

Using distance-based pre-pruning and the Gini index node splitting approach, a decision tree is constructed [33]. Pre- and post-pruning are the two types of DT node pruning techniques. Since pre-pruning based on each leaf node's maximum depth from the tree's root is ineffective, we have instead used pre-pruning based on distance. Distance-based pruning involves multiplying the maximum distance by the constant parameter Tolerance, which is initially set between 0 and 1.

Labeling of Testing Samples

During the labeling step, the testing sample's final label is assigned by applying the weighted k -nearest neighbor technique after traversing the classification and regression trees to reach the leaf node [34]. Using equation (3) weights of each neighbor were calculated.

$$w_i = \left(1.0 - \frac{1}{1 + e^{-d_{ij}}}\right) * L_i \quad (3)$$

Where L_i indicates fault and non-fault-prone classes.

Distance-Based Pruning

The greatest distance between any two points is determined first in distance-based pre-pruning of decision trees, and the decision tree is subsequently pruned using the percentage of the maximum distance. To encompass all locations, the maximum distance is computed [35]. Max distance and the global parameter are computed. Max distance * Tolerance is checked for all nodes, and tolerance is manually set between 1 and 0 according to the point density. Expression of Parameter Max distance is.

Time Complexity

Since training time is a one-time expense, our suggested approach's testing time complexity will be the only thing we talk about [36]. During the prediction phase, the Decision

Tree traverses in $O(\log n)$ time complexity from the root node of each leaf. In our suggested technique, k-nearest neighbors are kept at leaf nodes instead of labels, adding a little constant c to the decision tree's real testing time complexity. Our suggested method offered total run time complexity is $O(\log n + c)$.

Data Preprocessing

There are missing values and redundant data in the raw dataset [37-40]. Features with missing values are extracted and modified as part of data management. In this study, min-max feature scaling methods proposed to scale the feature between the range (0,1) with the help of minimum value & difference of the range and information gain filtering methods used for feature selection carried out in preparation for the KNN algorithm's preprocessing of the dataset.

Data Analysis

Examining the dataset, followed by cleaning, converting, and modeling it into a format that works, is data analysis. Finding useful information that will subsequently be utilized to enhance decision-making is the goal of data analysis [41,42]. The study might be useful, for instance, in assessing the properties of the data and the relationship between co-relation features.

Dataset

This study uses a dataset from the Open Access Series of Imaging Studies (OASIS) [27] to predict Alzheimer's disease. Depending upon cross-sectional MRI and longitudinal MRI the data samples are assigned as dementia and

non-dementia at a certain time or baseline is the first step. 150 people, ages 60 to 96, whose MRI data are included in the study. The dataset contains 72 non-demented patients and 69 were dementia. This distinction did not change during the study. The twelve OASIS longitudinal dataset characteristics are displayed in Table 1. This includes handling missing values, extracting and transforming features, and so on. Nine rows in the SES column have missing values, which may be fixed by removing the corresponding rows [43,44]. The measurement as a whole is now 141.

RESULTS AND DISCUSSION

This article's suggested method was created and tested on a computer with an 8GB RAM corei5 CPU. This method is developed using Anaconda 3, and its performance is evaluated against various machine learning models [45,46]. Before we proceed to the comparative study of proposed methods with the other three machine learning algorithms we find the effectiveness of the CART-KNN with the other two fundamental machine learning algorithms KNN and DT in Table 2. Cohen Kappa ranking is used to determine the degree of agreement between two independent raters and AUC (Area under the curve) for binary classification of the dataset are the main metrics to evaluate.

Accuracy

This study employs a suggested methodology to https://www.kaggle.com/jboisen/mri-andalzheimers?select=oasis_cross-sectional.csv is where you can find this data. Table 3 presents an accuracy performance metric comparison between our suggested method and three SVM, RF, CART-KNN, and decision tree variants (CART). To complete the experiment, set the tolerance value to 0.05 [47].

The accuracy of our suggested method is compared to various machine-learning techniques in Figure 2. The accuracy distribution of our suggested method demonstrates unequivocally that it performs better in this article than alternative machine learning techniques [48].

Table 1. List of twelve attributes

Identification, Age, Years of Education, Gender, Dominant Hand, Socio-Economic Status (SES), Clinical Dementia Rating, Normalize Whole Brain Volume, Mini-mental state examination, Atlas Scaling Factor, Estimated Total Intracranial Volume, Delay.

Table 2. Comparison study of the proposed approach and its fundamental approaches

Methods	Choen Kappa	AUC	Accuracy
KNN	0.887	0.891	90.85%
DT	0.812	0.837	82.55
CART-KNN	0.965	0.98	98.23%

Table 3. A comparative study based on accuracy for constant tolerance =0.05

s	DT	RF	SVM	CART-KNN
Alzheimer's dataset	82.55%	89.64%	84.72%	98.23%
CART:				

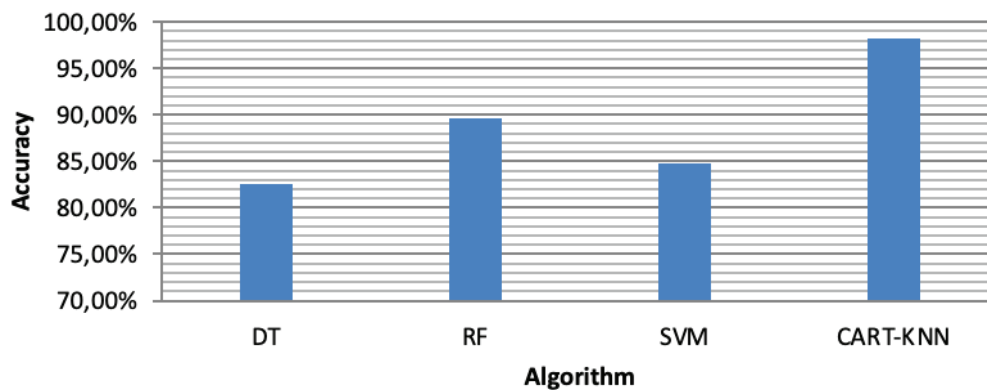


Figure 2. Comparison of accuracy in terms of bar plots.

Other Performance Metrics

The average accuracy, F1-score, and recall, for all methods compared on datasets related to Alzheimer's disease are displayed in Table 4. This article's suggested strategy outperforms all other approaches utilized in the comparison in terms of accuracy and f1-score performance criteria. A graphical depiction of the accuracy, recall, and F1-score comparison research is shown in Figure 3.

Run Time

The run time of the proposed model is $O(\log n) + c$, which is compared with another traditional model in Figure 4.

Table 5 presents a comparison with previous research on the OASIS dataset for AD prediction using several machine-learning algorithms. Kavitha and others. [29] demonstrated 83% accuracy in AD prediction using ML on the OASIS dataset, Basheer et al. Using a combination of deep learning and machine learning, [30] achieved a model accuracy of around 92.39%. Six distinct ML approaches were developed by Uddin et al. [22] and the developers of one of them achieved a maximum accuracy of 96% for the voting algorithm.

Comparison of Prediction Accuracy of the Model

In this part, we assessed the two types of error metrics, MAPE and RMSE. RMSE quantifies the error's variability

Table 4. Comparison of Precision, Recall, and F1-score of the proposed method with other approaches

Techniques	Precision	Recall	F1-score
DT	0.8108	0.7956	0.7928
RF	0.863	0.852	0.851
SVM	0.832	0.827	0.82
CART-KNN	0.972	0.969	0.965

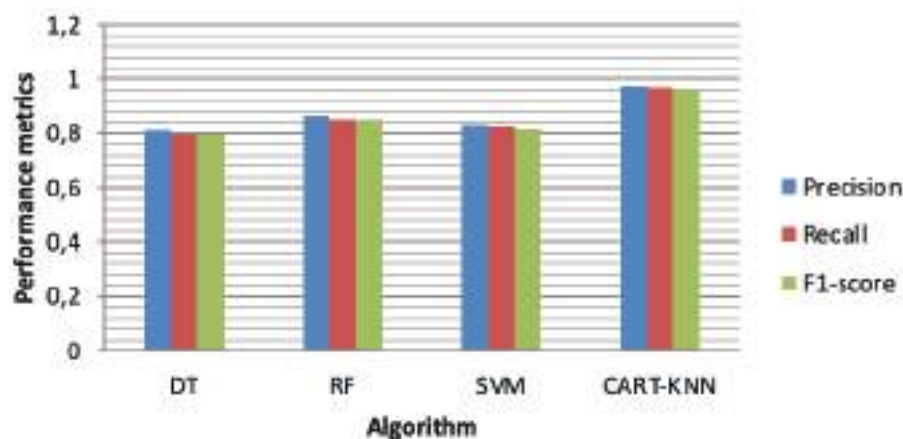


Figure 3. Graphical comparison for other performance metrics.

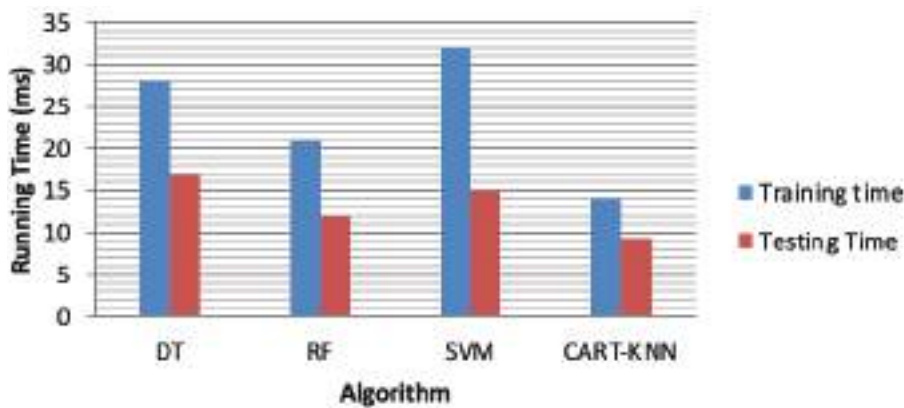


Figure 4. Comparison of Computational time.

Table 5. Comparison with state of art methods based on the OASIS dataset

Reference	Techniques	Outcomes of the algorithm
Kavitha et al. [29]	Gradient Boosting, SVM, DT and Voting	Validation accuracy 83%
Basheer et al.[30]	Machine learning and deep learning models(M-CapNet)	acceptable accuracy of 92.39%
Uddin et al. [22]	Gaussian NB, DT, RF, XGBoost, Voting Classifier, and Gradient Boost	Maximum accuracy obtained 96%
Proposed model	CART- KNN model	Acceptable accuracy about to 98.23%

while MAPE calculates the difference between the predicted and the actual result [49,50]. The following equation can express the following equation:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y}_i)^2}{N}} \quad (4)$$

$$MAPE = \frac{\sum_{i=1}^N \left| \frac{y_i - \bar{y}_i}{\bar{y}_i} \right|}{N} * 100 \quad (5)$$

where N , y_i and $\bar{y}_i$ are several samples, actual and predicted output respectively. A lower value of MAPE and RMSE indicates a better prediction. Figure 5 and Figure 6 represent the optimum parametric condition of the proposed CART -KNN model.

Table 6. Computational value of F1 score for CART-KNN model for different leaf nodes value

Leaf nodes	Testing F1 score	Training F1 Score
0	0.26	0.265
10	0.43	0.46
20	0.47	0.52
30	0.48	0.53
40	0.48	0.54
50	0.48	0.55

The F1 score for the train and testing dataset almost ranges the same lies from 0.25 to 0.54 for different values of no. of the leaf node shown in Table 6, while MAPE is increased for the training dataset by increasing the number of neighbors, but for the testing dataset, MAPE is reversed shown in Table 7.

Table 7. Computational value of MAPE for CART-KNN model for different neighbor value

No of Neighbour	Testing Data	Training Data
0	0.32	0.14
5	0.28	0.14
10	0.25	0.15
15	0.25	0.16
20	0.24	0.19
25	0.23	0.22

Table 8. Comparative study on RMSE and MAPE

Metrics	Algorithm		
	KNN	DT	CART-KNN
RMSE	205278	198634	50678
MAPE	14.98	10.98	8.56

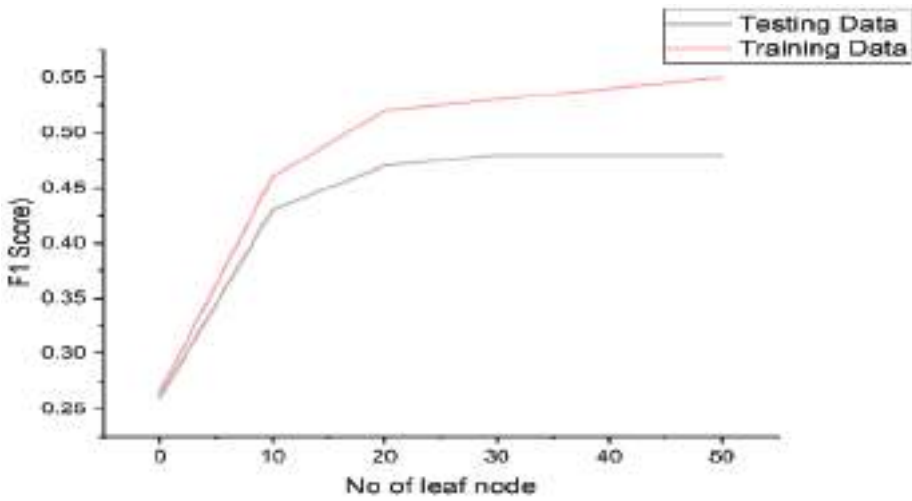


Figure 5. F1 Score for CART-KNN Cross-validation selection.

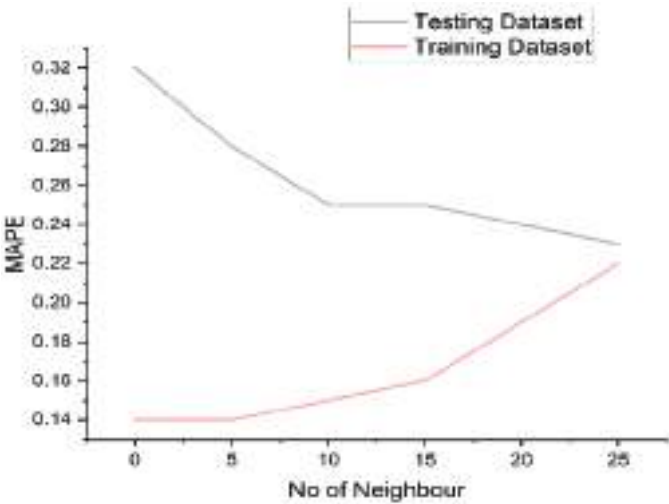


Figure 6. MAPE for CART-KNN Cross-validation selection.

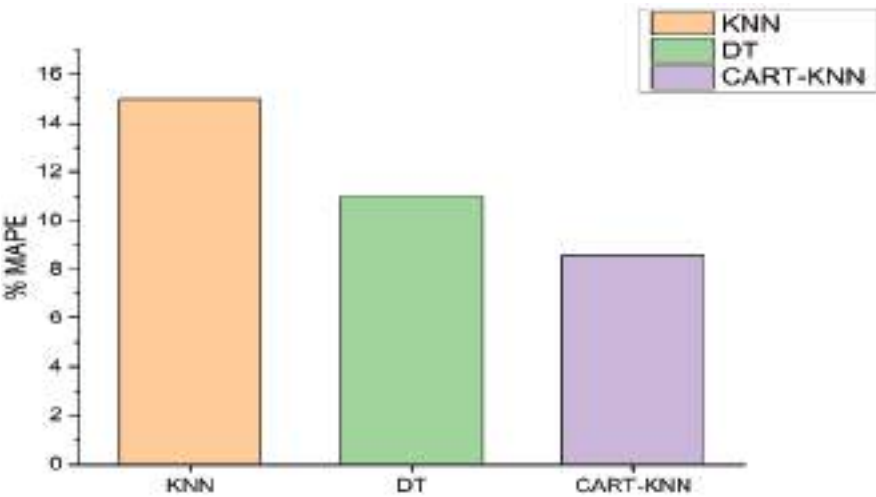


Figure 7. Comparison of % MAPE for all model.

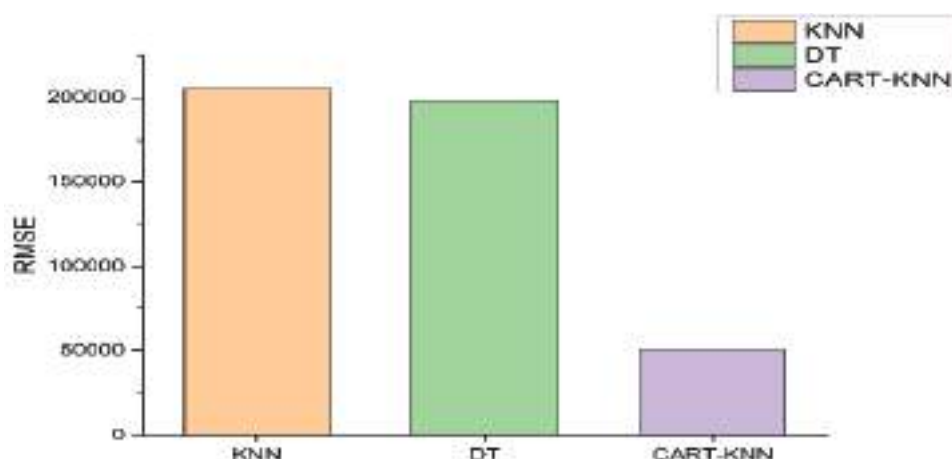


Figure 8. Comparison of RMSE for all model.

Figure 7 and Figure 8 show the prediction performance of the proposed hybrid model and traditional models. The proposed fused CART-KNN model improved the RMSE by 75.31%, and 74.31% while MAPE by 42.85%, and 2.20% respectively concerning traditional DT, and KNN models. Furthermore, Table 8 shows that the hybrid model is closest to the real value and has minimal swings in prediction errors.

CONCLUSION

The hybrid strategy of K nearest neighbor and pre-pruned classification and regression tree (CART) is suggested in this article. To overcome the overfitting problem, a pre-pruned Decision tree is applied in m-dimensional space and stores training samples of leaf nodes. Open Access Series of Imaging Studies (OASIS) datasets dataset for Alzheimer's disease (AD) prediction have been used to test the suggested strategy, and the outcomes are compared with those of other cutting-edge methods based on OASIS datasets and machine learning models. The main conclusions are listed as follows:

1. The robustness of the CART-KNN model allows it to choose parameters based on the features of the time series.
2. The proposed fused CART-KNN model improved the RMSE by 75.31%, and 74.31% while MAPE by 42.85%, and 2.20% respectively concerning traditional DT, and KNN models.
3. The validation accuracy of the proposed CART-KNN is about 98.23% on the test data of AD, which indicates the CART-KNN classifier in the suggested study produces a more advantageous result.

However, this study has also some limitations: The CART-KNN suffered in poor generalization ability due to easy overfit. New hybrid approaches based on conventional machine learning techniques can be created in the future to increase performance and make them suitable for

real-world applications. To make the concept feasible, work on running time complexity reduction can be done

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://www.kaggle.com/jboy-sen/mri-andalzheimers?select=oasis_cross-sectional.csv.

NOMENCLATURE

$d(P_i, P_j)$	Distance between two points
L_i	fault-prone classes
w_i	Weighted value value for i^{th} neighbor
N	no. of samples
y_i	the output of i^{th} number of samples
$\bar{y}_i$	Predicted output of i^{th} samples

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] Murugan S, Venkatesan C, Sumithra MG, Gao X-Z, Ekalliya B, Akila M, et al. DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. *IEEE Access* 2021;9:90319–90329. [\[CrossRef\]](#)
- [2] Akiyama H, Barger S, Barnum S, Bradt B, Bauer J, Cole GM, et al. Inflammation and Alzheimer's disease. *Neurobiol Aging* 2000;21:383–421. [\[CrossRef\]](#)
- [3] Shahbaz M, Ali S, Guergachi A, Niazi A, Umer A. Classification of Alzheimer's disease using machine learning techniques. *Data* 2019;296–303. [\[CrossRef\]](#)
- [4] Yoo I, Alafairet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, et al. Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst* 2012;36:2431–2448. [\[CrossRef\]](#)
- [5] Gupta MK, Chandra P. A comprehensive survey of data mining. *Int J Inf Technol* 2020;12:1243–1257. [\[CrossRef\]](#)
- [6] Berkhin P. A survey of clustering data mining techniques. In: Kogan J, Nicholas C, Teboulle M, editors. *Grouping multidimensional data*. Berlin: Springer-Verlag; 2006. p. 25–71. [\[CrossRef\]](#)
- [7] Dennis E, Manikandan G, Vilma V, Hemalatha S. Alzheimer disease using machine learning. *Int J Sci Res Sci Technol* 2025;58:262. [\[CrossRef\]](#)
- [8] Sowjanya V, Neeha Y, Dishasri L, Neela S, Prasanth Y. Predictive diagnosis of Alzheimer's disease using machine learning. 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), 09-11 May 2024.
- [9] Hao L, Sirpa H, Julian L, Vesa T, Anna-Maija T. Predicting Alzheimer's disease from cognitive footprints in mid and late life: how much can register data and machine learning help? *Int J Med Inform* 2024;190:105540. [\[CrossRef\]](#)
- [10] Wujia Y, Ting-Hsuan S, Kai-Cheng H, et al. Comparative analysis of machine learning algorithms for Alzheimer's disease classification using EEG signals and genetic information. *Comput Biol Med* 2024;176:108621. [\[CrossRef\]](#)
- [11] Fahad M, Alshabrmi F, Aba F, Alkhayl F. Novel drug discovery: advancing Alzheimer's therapy through machine learning and network pharmacology. *Eur J Pharmacol* 2024;976:176661. [\[CrossRef\]](#)
- [12] Alatrany A, Khan W, Hussain A, Hoshang K, Al-Jumeily D. An explainable machine learning approach for Alzheimer's disease classification. *Dent Sci Rep* 2024;14:2637. [\[CrossRef\]](#)
- [13] García-Gutiérrez F, Alegret M, Marquié M, Muñoz N, Ortega G, Cano A, et al. Unveiling the sound of the cognitive status: Machine Learning-based speech analysis in the Alzheimer's disease spectrum. *Alzheimers Res Ther* 2024;16:26. [\[CrossRef\]](#)
- [14] Breiman L. *Classification and regression trees*. Abingdon: Routledge; 2017. [\[CrossRef\]](#)
- [15] Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 1967;13:21–27. [\[CrossRef\]](#)
- [16] Singh M, Chhabra JK. A hybrid approach based on k-nearest neighbors and decision tree for software fault prediction. *Kuwait J Sci* 2023;50.
- [17] Khan P, Kader F, Islam SMR, Rahman AB, Kamal S, Toha MU, et al. Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. *IEEE Access* 2021;9:37622–37655. [\[CrossRef\]](#)
- [18] Martinez-Murcia FJ, Ortiz A, Ramirez J, Castillo-Barnes D. Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders. *IEEE J Biomed Health Inform* 2019;24:17–26. [\[CrossRef\]](#)
- [19] Prajapati R, Khatri U, Kwon GR. An efficient deep neural network binary classifier for Alzheimer's disease classification. In: 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). Piscataway (NJ): IEEE; 2021. p. 231–234. [\[CrossRef\]](#)
- [20] Helaly HA, Badawy M, Haikal AY. Deep learning approach for early detection of Alzheimer's disease. *Cogn Comput* 2022;14:1711–1727. [\[CrossRef\]](#)
- [21] Kadhim KA, Sakran AA, Adnan MM, Salman GA. Early diagnosis of Alzheimer's disease using convolutional neural network-based MRI. *Malays J Fundam Appl Sci* 2023;19:362–368. [\[CrossRef\]](#)
- [22] Uddin KMM, Alam MJ, Jannat JA, Aryal S. A novel approach utilizing machine learning for the early diagnosis of Alzheimer's disease. *Biomed Mater Devices* 2023;1:882–898. [\[CrossRef\]](#)
- [23] Rabeh AB, Benzarti F, Amiri H. CNN-SVM for prediction Alzheimer disease in early step. In: 2023 International Conference on Control, Automation and Diagnosis (ICCAD). Piscataway (NJ): IEEE; 2023. p. 1–6. [\[CrossRef\]](#)
- [24] Irfan M, Shahrestani S, Elkhodr M. Early detection of Alzheimer's disease using cognitive features: a voting-based ensemble machine learning approach. *IEEE Eng Manag Rev* 2022;51:16–25. [\[CrossRef\]](#)
- [25] Varma BSS, Kalyani G, Asish K, Bai MI. Early detection of Alzheimer's disease using SVM, random forest and FNN algorithms. In: 2023 2nd International Conference for Innovation in Technology (INOCON). Piscataway (NJ): IEEE; 2023. p. 1–6. [\[CrossRef\]](#)
- [26] Tripathi T, Kumar R. Speech-based detection of multi-class Alzheimer's disease classification using machine learning. *Int J Data Sci Anal* 2024;18:83–96. [\[CrossRef\]](#)

- [27] Shrivastava RK, Singh SP, Kaur G. Machine learning models for Alzheimer's disease detection using OASIS data. In: Koundal D, Jain DK, Guo Y, Ashour AS, Zaguia A, editors. Data analysis for neurodegenerative disorders. Singapore: Springer Nature; 2023. p. 111–126. [\[CrossRef\]](#)
- [28] Mahjabeen A, Mia MR, Shariful FNU, Mahmud I. Early prediction and analysis of DTI and MRI-based Alzheimer's disease through machine learning techniques. In: Kaiser MS, Waheed S, Bandyopadhyay A, Mahmud M, Ray K, editors. Proceedings of the Fourth International Conference on Trends in Computational and Cognitive Engineering. Singapore: Springer Nature; 2023. p. 3–13. [\[CrossRef\]](#)
- [29] Kavitha C, Mani V, Srividhya SR, Khalaf OI, Tavera Romero CA. Early-stage Alzheimer's disease prediction using machine learning models. Front Public Health 2022;10:853294. [\[CrossRef\]](#)
- [30] Basheer S, Bhatia S, Sakri SB. Computational modeling of dementia prediction using deep neural network: analysis on OASIS dataset. IEEE Access 2021;9:42449–42462. [\[CrossRef\]](#)
- [31] Banga M, Bansal A. Proposed software faults detection using a hybrid approach. Secur Priv 2023;6:e103. [\[CrossRef\]](#)
- [32] Maddipati SS, Srinivas M. A hybrid approach for cost effective prediction of software defects. Int J Adv Comput Sci Appl 2021;12. [\[CrossRef\]](#)
- [33] Priyanka NA, Kumar D. Decision tree classifier: a detailed survey. Int J Inf Decis Sci 2020;12:246. [\[CrossRef\]](#)
- [34] Chen P, Liao BB, Chen G, Zhang S. Understanding and utilizing deep neural networks trained with noisy labels. In: International Conference on Machine Learning. PMLR; 2019. p. 1062–1070.
- [35] Akritidis L, Fevgas A, Manolopoulos Y. An unsupervised distance-based model for weighted rank aggregation with list pruning. Expert Syst Appl 2022;202:117435. [\[CrossRef\]](#)
- [36] Guo L, Fang W, Zhao Q, Wang X. The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. Comput Ind Eng 2021;161:107598. [\[CrossRef\]](#)
- [37] Dutta P, Paul S, Cengiz K, Anand R, Kumar A. A predictive method for emotional sentiment analysis by deep learning from EEG of brainwave dataset. In: Artificial intelligence for neurological disorders. Amsterdam: Elsevier; 2023. p. 25–48. [\[CrossRef\]](#)
- [38] Dutta P, Paul S, Anand R, Majumder M. A predictive method for emotional sentiment analysis by machine learning from electroencephalography of brainwave data. In: Implementation of smart healthcare systems using AI, IoT, and blockchain. Amsterdam: Elsevier; 2023. p. 109–130. [\[CrossRef\]](#)
- [39] Dutta P, Pal S, Kumar A, Cengiz K. Artificial intelligence for cognitive modeling: theory and practice. Boca Raton (FL): Chapman and Hall/CRC; 2023. [\[CrossRef\]](#)
- [40] Dutta P, Paul S, Kumar A. Comparative analysis of various supervised machine learning techniques for the diagnosis of COVID-19. In: Electronic devices, circuits, and systems for biomedical applications. Amsterdam: Elsevier; 2021. p.521–540. [\[CrossRef\]](#)
- [41] Dutta P, Paul S, Obaid AJ, Mukhopadhyay K. Feature selection based artificial intelligence techniques for the prediction of COVID-like diseases. J Phys Conf Ser 2021;1963:012167. [\[CrossRef\]](#)
- [42] Dutta P, Paul S, Sadhu A, Jana GG, Bhattacharjee P. Performance of automated machine learning based neural network estimators for the classification of PCOS. In: Bhattacharyya S, Banerjee JS, De D, Mahmud M, editors. Intelligent human-centered computing. Singapore: Springer Nature; 2023. p. 65–73. [\[CrossRef\]](#)
- [43] Karadurmuş E, Göz N, Taşkın N, Yüceer M. Bromate removal prediction in drinking water by using the least squares support vector machine (LS-SVM). Sigma J Eng Nat Sci 2020;38:2145–2153.
- [44] Alp S, Yiğit OE, Öz E. Prediction of bist price indices: a comparative study between traditional and deep learning methods. Sigma J Eng Nat Sci 2020;38:1693–1704.
- [45] Güneş U, Başhan V, Karakurt AS. Predicting tanker main engine power using regression analysis and artificial neural networks. Sigma J Eng Nat Sci 2023;41:216–225. [\[CrossRef\]](#)
- [46] Das O. Prediction of the natural frequencies of various beams using regression machine learning models. Sigma J Eng Nat Sci 2023;41. [\[CrossRef\]](#)
- [47] Gupta P, Kumar S, Singh YB, Singh P, Sharm SK, Rathore NK. The impact of artificial intelligence on renewable energy systems. NeuroQuantology 2022;20:5012.
- [48] Kumar R, Thakur MA, Rathore NK. Optimizing smart manufacturing with IoT integration and leveraging machine learning analysis. NeuroQuantology 2022;20:1620.
- [49] Dutta P, Paul S, Shaw N, Sen S, Majumder M. Heart disease prediction: a comparative study based on a machine-learning approach. In: Artificial intelligence and cybersecurity. Boca Raton (FL): CRC Press; 2022. p. 1–18. [\[CrossRef\]](#)
- [50] Dutta P, Paul S, Jana GG, Sadhu A. Hybrid genetic algorithm random forest algorithm (HGARF) for improving the missing value imputation in hepatitis medical dataset. In: 2023 International Symposium on Devices, Circuits and Systems (ISDCS). Piscataway (NJ): IEEE; 2023. p. 1–5. [\[CrossRef\]](#)



Research Article

On the use of different link functions in gamma-pareto regression model: Simulation and application

Nasir SALEEM^{1,*}, Atif AKBAR¹, Javaria Ahmad KHAN¹

¹Department of Statistics, Bahauddin Zakariya University, Multan, 60800, Pakistan

ARTICLE INFO

Article history

Received: 21 March 2024

Revised: 13 May 2024

Accepted: 22 July 2024

Keywords:

G-PRM; GLM; Link Functions;
SSR; MSE; AMSE; Reaction Rate
Data

ABSTRACT

The Gamma-Pareto regression model (G-PRM) is appropriate when the response variable follows Gamma-Pareto distribution (G-PD) is used as the generalized linear model (GLM). For the estimation of the G-PRM the Iterative weighted least squares (IWLS) method is used with a specific link function. In this study, we consider G-PRM under different link functions. However, the researchers do not pay much attention to the selection of suitable link functions. In the context of the G-PRM, three link functions are used, which allow for investigating how inverse, identity, and log link functions perform. A Monte Carlo simulation and a demonstration with real data were used to compare the performance of different link functions in G-PRM. The sum squared residual (SSR), mean squared error (MSE) and average mean squared error (AMSE) are used as the evaluation criterion for suitable link function. Both the simulation and real data findings demonstrate that the G-PRM with the identity link function provides efficient results having minimum SSR, MSE and AMSE. Advantages of the paper, choice of suitable link function always significantly impacts the model's performance.

Cite this article as: Saleem N, Akbar A, Khan JA. On the use of different link functions in gamma-pareto regression model: Simulation and application. Sigma J Eng Nat Sci 2025;43(4):1495–1506.

INTRODUCTION

In practice, we often meet the case that the response variables are continuous and follow the GP-D. In that case, the G-PRM and GLM are recommended. Application of GP-D in the form of health care economics, medical science, meteorology, occurrence rate and reaction rate etc. For the reliable results of regression analysis link function always play and important role. The regression model explains a phenomenon the (response variable) based on other phenomena (explanatory variables). Classical regression model is developed with the assumption that the

response variables are normally distributed. This assumption is used for the validity of the test for both the model and its parameters. Regression modeling is usually based on the probability distribution of the response variable. For exponential family distribution the model is usually in the form of GLM i.e. Gamma, Beta, Pareto, Inverse Gaussian, Weibull etc. Gamma distribution is used in homicide data [1] Beta distribution for modeling reading skills data used by [2] Pareto distribution is used in modeling earthquakes, forest fire areas and oil and gas field sizes by [3] presented an application of the Pareto distribution in modeling disk

*Corresponding author.

*E-mail address: nasirsaleem160@gmail.com

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



drive sectors errors among others. To add flexibility to the Pareto distribution, various generalizations of the distribution have been derived, including the generalized Pareto distribution [4] the beta-Pareto distribution [5] and the beta generalized Pareto distribution [6].

Motivation of the Study

The purpose of the present study is work free region. Link functions that work with groups are available, as we have seen in the previous section; there must be some proper statistical measure for the comparison. On the basis of having minimum SSR, MSE and AMSE.

Advantages and Dis-Advantages of the Proposed Method

The merits and demerits of the proposed method. MSE is a commonly used metric that measures the average of the squared differences between the predicted and actual values. It gives more weight to large errors and is sensitive to outliers. MSE is useful when the goal is to minimize the overall error in the model. One of the advantages of MSE becomes a disadvantage when there is a bad prediction. The sensitivity to outliers magnifies the high errors by squaring them. MSE will have the same effect for a single large error as too many smaller errors. The residual sum of squares (RSS) measures the difference between observed data and the model's predictions. It is the portion of variability any regression model does not explain, also known as the model's error. Use of RSS to evaluate how well the model fits the data. RSS is a measure of the overall goodness of fit of the regression model, while MSE is a measure of the average distance between the predicted and actual values. RSS is used to compare different regression models, while MSE is used to evaluate the accuracy of a single model. RSS is a sum squared difference, while MSE is an average of squared error. RSS increases as the number of predictors in the model increases, while MSE decreases as the number of predictions increase.

Organized of the Paper

As it is structured, the paper goes in the following order: literature review in sections. In the next section, the gamma-pareto regression model was then used to estimate the parameters of the GLM of the gamma-pareto regression model. In next section discussed the simulation Study and Application real data (Reaction Rate Data). In last section Conclusion.

LITERATURE REVIEW

[7] explained how the link functions were carried out in the beta ridge regression model. In the case of the appropriate connection functions, the criterion to be evaluated is the smallest MSE. [8] discussed a comparison of some used link functions on the school drop-out rates data of East Java that yielded binomial regression models. In the case of the appropriate linking function, the estimates or criteria include Akaike information criterion (AIC), Bayesian information criterion (BIC), log likelihood (LL) and R-squared. [9] used generalized Weibull linear models with various link functions to survival analysis. Under

various link functions, fit measures of goodness of fit models include deviance, AIC and BIC. [10] described the significance of Beta regression residuals-based control charts using various link functions. To this end, the data on the thermal power plants were utilized through an application. Moreover, three criteria are employed in checking the performance: the average of the run length, the standard deviation of the run length and the median run length. And also test the performance of the proposed control chart in two ways, namely, by monitoring the intercepts and by monitoring the slope coefficients. [11] examine the influential observation detection of the logistic regression with various link functions and employ pearson residuals. And applied an actual life example to analyse the urine calcium oxalate crystals data. [12] used Deviance and Pearson residuals-based monitoring charts with various link functions in tracking logistic regression profiles, to COVID-19 data. [13] compared the link functions relative to fitting logistic ridge regression as an application to urine data. A Monte Carlo simulation study and a real data set are taken into consideration and with scalar mean squared error as the performance evaluation criterion. Similarly, [14] developed a Gamma-Pareto distribution (G-PD). G-PD is a combination of Gamma and Pareto distribution taking a form of Pareto composed in Gamma, with three parameters (α , β , γ). [15] showed that GP-D is the member of exponential family as a condition to develop GLM based on G-PD. They provided some attributes for developing GLM such as mean, variance, and dispersion parameters of response variable defined the link functions for GLM, G-PD. They also provided deviance and AIC for assessing the goodness and parameter estimation of GLM, G-PD. [16,17] invent a simulation scheme for G-PRM used GLM gamma to analyze the relationship between simulated G-PD response variable with explanatory variable. The result showed that goodness of the model only depends on the goodness of fit the response variable to G-PD and the strength of the relationship of response and explanatory variable. [18] discussed about the modeling aims to analyze whether Tropical Rainfall Measuring Mission (TRMM) satellite data is a good estimator for unobserved station's data. Integrating G-PD data with TRMM data to estimate monthly rainfall using GLM. The Truncated Gamma-Pareto Distribution Used to Study Cosmic Rays, Novel Probability Distributions in Astrophysics. [19] a study Gamma- Pareto distributions (IV) and their importance and application and near to relatives and generalizations provide most flexible families of heavy-tailed distributions. It can be used to model income distributions as well as a wide variety of economic distributions. [14] invent a new distribution, namely as a Gamma-Pareto (IV) distribution, after that it called as G-PD (IV) distribution. Many properties of G-PD (IV) are discussed like a limiting behavior, mean, median and mode also moments, skewness and kurtosis. [20] invent the gamma generalized Pareto distribution, a three-parameter model, is used in the current study.

[21] introduced the exponentiated gamma-Pareto distribution. Here the existing distributions, are gamma-Pareto and exponentiated Pareto. The special cases of the exponentiated

gamma-Pareto distribution and its properties, including distribution shapes, limit behavior, hazard function etc. The creation of traffic conflict techniques heavily relies on the safety continuum. Using GGPD, it is possible to explain two crucial model parameters: the threshold and the shifted value. [22] invented a new probability distribution named as weighted gamma Pareto distribution (WG-PD). By using the T-X family and the concept of weighted probability. [23] a study of the new log-gamma-Pareto distribution (LG-PD) it includes as special cases two models such as gamma-Pareto and Pareto distributions. A real-world use of maximum likelihood estimation to estimate model parameters shows its potential. And the properties of LG-PD and its behavior such as mean, median and mode also skewness and kurtosis. [24] describe new q-rung orthopair fuzzy Aczel–Alsina weighted geometric operators under group-based generalized parameters in multi-criteria decision-making problems. [25] discussed the Q-rung orthopair probabilistic hesitant fuzzy hybrid aggregating operators in multi-criteria decision-making problems. [26] explain the importance Bipolar valued probabilistic hesitant fuzzy sets based on Generalized Hybrid Operators in multi-criteria decision-making problems based on TOPSIS. [27] invent Generalized Dice measures of single valued neutrosophic type-2 hesitant fuzzy sets and their application to multi-criteria decision-making problems. [28] discussed hybrid similarity measures of single-valued neutrosophic type-2 fuzzy sets and their application to MCDM based on TOPSIS. [29] discussed the multi-criteria decision making based on vector similarity measures of picture type-2 hesitant fuzzy sets. The applications of G-PD in above mentioned literature used only one variable. Meanwhile, the variable may be affected by other variable(s). To explain the relationship, we need regression model based on G-PD. For non-normal response variable, the regression model is usually in the form of Generalized Linear Model (GLM). [14] mentioned the mathematical relationship between G-PD and gamma distribution (GD). This is reasonable because the G-PD developed from the GD. The existence of this relationship provides a possibility to analyze the G-PD data through GLM Gamma which expand the scope of data distribution which can be modeled. The scope should be wider than Gamma and Pareto distributions individually. In the aforementioned studies, the logit link function is the tool the researchers use to evaluate G-PRM performance. Different link functions, however, are also essential in determining the superiority. In this study, order to get a clear image of the performance of link

functions, we take into consideration three different link functions for G-PRM. We suggest several approaches with various link functions for the parameter estimation for the G-PRM. In addition, a thorough Monte Carlo simulation analysis is carried out under various link functions in order to select a suitable link function that achieves a minimum or average mean squared error (MSE). To prove the effectiveness of the Gamma-Pareto regression model under a specific link function. Data preparation, ensure your data meets the assumptions of the G-PRM, such as positive continuous responses. Model formulation, specify the G-PRM with the chosen link function (e.g., inverse, identity and log). Model estimation, estimate the model parameters using maximum likelihood estimation (MLE). Model evaluation, assess the model's goodness of fit using metrics such as, sum squared error (SSE), Mean squared error (MSE) and Average Mean absolute error (AMAE).

MATERIAL AND METHODS FOR GAMMA-PARETO REGRESSION MODEL AND SIMULATION STUDY

[14] state that the G-PD pdf is provided by,

$$f(y; \alpha, \beta, \gamma) = \frac{\gamma^{-1}}{\beta^{\alpha} \Gamma(\alpha)} \left(\log \left(\frac{y}{\gamma} \right) \right)^{\alpha-1} \left(\frac{y}{\gamma} \right)^{-\left(\frac{1}{\beta}+1\right)} \quad (1)$$

with $\alpha, \beta, \gamma > 0$ and $y > \gamma$.

The mean and variance of G-P distribution are, $E(a(y)) = \alpha\beta$, $V(a(y)) = \alpha\beta^2$ respectively. According to [16,17] With parameters, Eq. (1) can be modified $\alpha = \frac{1}{\phi}$ and $\beta = \mu\phi$. The Gamma Pareto density for y under these conditions is given by

$$f(y; \mu, \phi) = \frac{\gamma^{-1}}{(\mu\phi)^{\frac{1}{\phi}} \Gamma\left(\frac{1}{\phi}\right)} \left(\log \left(\frac{y}{\gamma} \right) \right)^{\frac{1}{\phi}-1} \left(\frac{y}{\gamma} \right)^{-\left(\frac{1}{\mu\phi}+1\right)} \quad (2)$$

With y is continuous and non-negative, $\mu > 0$ and $\phi > 0$. Since the mean and variance of y are $E(y) = \mu$ and $V(y) = \phi V(\mu) = \phi\mu^2$.

For the i th observation, let $x_{i1}, x_{i2}, \dots, x_{ip}$ represent the p non-stochastic regressors. Following that, the G-PRM for the mean of response variable y is provided by [16].

According to [16,17], Link function g in GLM is $g(\mu_i) = X_i^T \beta = \eta_i$ where $\mu_i = E(a(y_i))$, which are presented in Table 1.

Table 1. Different link function of Gamma Pareto Distribution

Link function	Form of link function	Reference
Inverse link function	$\mu = \frac{1}{X'\beta}$	Hanum et al. [2016]
Identity link function	$\mu = X'\beta$	
Log link function	$\mu = \log(X'\beta)$ $\mu = e^{X'\beta}$	

Estimation the GLM Gamma-pareto Regression Model Parameters

Finding the likelihood function's derivative with respect to β_j is the first step in estimating the parameter β_j using maximum likelihood and τ_i is the function of β . By Eq. (2)

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^N \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^N \left[\frac{\partial l_i}{\partial \tau_i} \frac{\partial \tau_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} \right] \quad (3)$$

Now

$$\begin{aligned} \frac{\partial l_i}{\partial \tau_i} &= a(y)b'(\tau) + c'(\tau) = \beta^{-2} \left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) \\ \frac{\partial \tau_i}{\partial \mu_i} &= \frac{1}{\frac{\partial \mu_i}{\partial \tau_i}} = \frac{1}{\frac{\partial \alpha \beta}{\partial \beta}} = \frac{1}{\alpha} \\ \frac{\partial \mu_i}{\partial \beta_j} &= \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \end{aligned}$$

Where $\frac{\partial \mu_i}{\partial \eta_i}$ based on the GLM's link function. So, the score for β_j in GLM Gamma-Pareto is

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^N \alpha^{-1} \beta^{-2} \left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \quad (4)$$

Let l be the log likelihood of the response variable. Lastly, the j th score is presented.

$$U_j = \sum_{i=1}^N \left[\text{var} \left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) \right]^{-1} \left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

The variance U_j is

$$\text{var}(U_j) = \zeta_{jk} = \sum_{i=1}^N \frac{x_{ij} x_{ik}}{\left[\text{var} \left(\log\left(\frac{y_i}{\gamma}\right) \right) \right]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = X^T W X$$

Where,

$$W = \frac{1}{\left[\text{var} \left(\log\left(\frac{y_i}{\gamma}\right) \right) \right]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

Since the estimators of β_j is not in close form.

Iterative weighted least squares (IWLS) were proposed by [30] as a method for estimating β_j .

The resultant estimate by using IWLS is given by,

$$X^T W X b^{(m)} = X^T W Z$$

$$b^{(m)} = (X^T W X)^{-1} (X^T W Z) \quad (5)$$

Where z_i is the adjusted response variable, $X_i^T = (1, X_{i1}, X_{i2}, \dots, X_{ip})$ and $W = \text{diag}(\hat{\mu}_1^2, \dots, \hat{\mu}_n^2)$ is the weighted matrix and $\hat{\mu}_1 = \eta_i = \frac{1}{X' \beta}$, $\hat{\mu}_2 = \eta_i = X' \beta$ and $\hat{\mu}_3 = \eta_i = \log(X' \beta)$. And now, Using W and $\text{var}(U_j)$ for G-P and obtained the iteration for β_j as, i is a number of observations $i=1,2,3, \dots, n$. and j are a number of parameter $j=1,2,3, \dots, p$.

$$\begin{aligned} X^T W X b^{(m)} &= \sum_{k=1}^p \sum_{i=1}^N \frac{x_{ij} x_{ik}}{\left[\text{var} \left(\log\left(\frac{y_i}{\gamma}\right) \right) \right]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)} \\ &\quad + \frac{\left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) x_{ij}}{\left[\text{var} \left(\log\left(\frac{y_i}{\gamma}\right) \right) \right]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \\ z_i &= \sum_{i=1}^N x_{ij} b_k^{(m-1)} + \left(\log\left(\frac{y_i}{\gamma}\right) - \mu_i \right) \frac{\partial \mu_i}{\partial \eta_i} \end{aligned}$$

And finally fitted three models using above methodology.

$$g(\mu_i) = \eta_i = \frac{1}{X' \beta} \text{ and fitted model inverse link, } \hat{\mu}_1 = \eta_i = \frac{1}{X' \beta} \quad (6)$$

$$g(\mu_i) = \eta_i = X' \beta \text{ and fitted model identity link, } \hat{\mu}_2 = \eta_i = X' \beta \quad (7)$$

$$g(\mu_i) = \eta_i = \log(X' \beta) \text{ and fitted model log link, } \hat{\mu}_3 = \eta_i = \log(X' \beta) \quad (8)$$

Many types are available of GLM residuals in literature. Here we used a squared residual.

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for inverse link function in the G-PRM is given by

$$R_{(inverse)} = (y_i - \hat{\mu}_1) \quad (9)$$

$$SR_{(inverse)} = (y_i - \hat{\mu}_1)^2 \quad (9.1)$$

$$SSR_{(inverse)} = \sum (y_i - \hat{\mu}_1)^2 \quad (9.2)$$

$$MSE_{(inverse)} = \frac{\sum (y_i - \hat{\mu}_1)^2}{n-p} \quad (9.3)$$

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for identity link function in the G-PRM is given by

$$R_{(identity)} = (y_i - \hat{\mu}_2) \quad (10)$$

$$SR_{(identity)} = (y_i - \hat{\mu}_2)^2 \quad (10.1)$$

$$SSR_{(identity)} = \sum (y_i - \hat{\mu}_2)^2 \quad (10.2)$$

$$MSE_{(identity)} = \frac{\sum (y_i - \hat{\mu}_2)^2}{n-p} \quad (10.3)$$

The Residuals (R), Squared Residuals (SR) and Sum Squared Residuals (SSR) for log link function in the G-PRM is given by

$$R_{(log)} = (y_i - \hat{\mu}_3) \quad (11)$$

$$SR_{(log)} = (y_i - \hat{\mu}_3)^2 \quad (11.1)$$

$$SSR_{(\log)} = \sum (y_i - \hat{\mu}_3)^2 \quad (11.2)$$

$$MSE_{(\log)} = \frac{\sum (y_i - \hat{\mu}_3)^2}{n-p} \quad (11.3)$$

Simulation Study

This section is aimed at illustrating the work of various link functions by using simulation. Monte Carlo simulation of G-PRM involving the inverse, identity and log link functions has been done in this section. [16,17] invent these link function for G-PRM and define by $\eta_i = \frac{1}{X^T \beta}$ and fitted model inverse link, $\hat{\mu}_1 = \eta_i = \frac{1}{X^T \beta}$, $\eta_i = X^T \beta$ and fitted model identity link, $\hat{\mu}_2 = \eta_i = X^T \beta$, $\eta_i = \log(X^T \beta)$ and fitted model log link, $\hat{\mu}_3 = \eta_i = \log(X^T \beta)$. We used algorithm and simulation schemes developed by [16,17], to generate the response variable which follow a G-PRM is defined as $y_i \sim G - P(\alpha, \beta, \gamma)$, where inverse model $\hat{\mu}_1 = E(y_i) = \frac{1}{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}}$ and identity model $\hat{\mu}_2 = E(y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3}$ and log model

$\hat{\mu}_3 = E(y_i) = \log(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$ with $\alpha = E(y_i) = 10$, and β is the arbitrary values $\beta_0 = 0.05$, $\beta_1 = 0.0025$, $\beta_2 = 0.005$, $\beta_3 = 0.0001$ while $\phi = 0.04, 0.11, 0.17, 0.33, 0.67$ and 2 is the dispersion parameters [31] and [38-43] and γ is minimum value of response variable. Here, the design matrix X generated from normal distribution as $X_{ij} \sim N(-1, 1)$, for $i=1, 2, \dots, n$, and $j=1, 2, 3$. All the generated x-axis are fixed through the whole simulation study. we generate the data set for samples of size $n=25, 50, 100, 200$. These are simulation results that were done on the R software. Performance of link functions of G-PRM is run through the simulation 10000 times. We discussed their performance on the basis of average mean squared error (AMSE) and also suggest which link function is suitable for G-PRM.

In Table 2, AMSE are presented by using different link functions with different dispersion parameters and different sample sizes.

- In this section, for small dispersion level $\phi=0.04$ and sample sizes $n = 25, 50, 100$ the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample size $n = 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.

Table 2. Average mean squared error (AMSE) by using different links functions, G-PRM

ϕ	Sample size n.	Link Functions		
		Inverse link AMSE	Identity link AMSE	Log link AMSE
0.04	25	0.0137	0.0110	0.0137
	50	0.0399	0.0272	0.0398
	100	0.0577	0.0292	0.0342
	200	0.0566	0.0303	0.0280
0.11	25	0.0014	0.0008	0.0014
	50	0.0316	0.0175	0.0239
	100	0.0279	0.0145	0.0164
	200	0.1451	0.0560	0.0550
0.17	25	0.0227	0.0122	0.0201
	50	0.0037	0.0030	0.0037
	100	0.0067	0.0040	0.0058
	200	0.0367	0.0121	0.0142
0.33	25	0.0582	0.0447	0.0597
	50	0.0991	0.0749	0.0956
	100	0.3108	0.1814	0.0041
	200	0.0110	0.0058	0.0017
0.67	25	0.0434	0.0351	0.0137
	50	0.0159	0.0108	0.0130
	100	0.1431	0.0748	0.0677
	200	0.1359	0.0552	0.0110
2	25	0.0629	0.0443	0.0620
	50	0.0574	0.0316	0.0504
	100	0.0822	0.0461	0.0246
	200	0.1833	0.0777	0.0474

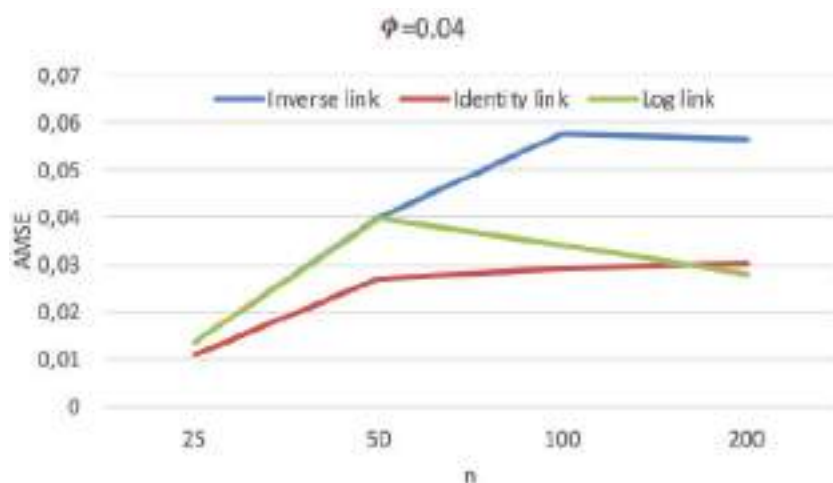


Figure 1. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.04$.

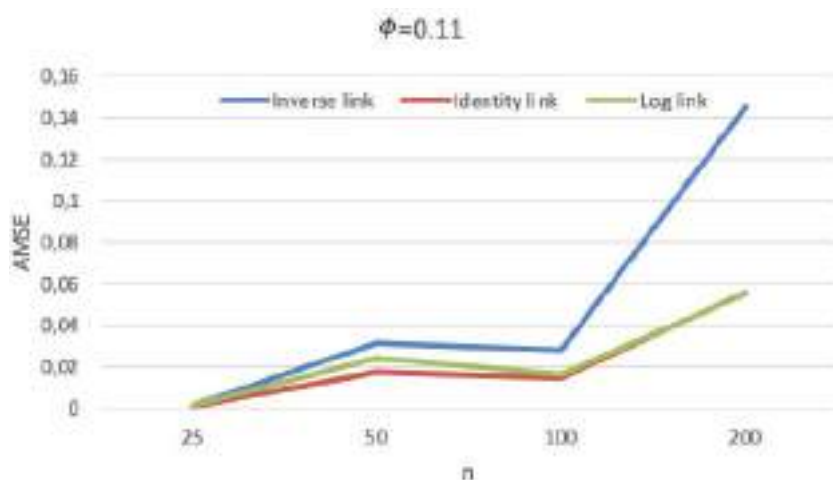


Figure 2. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.11$.

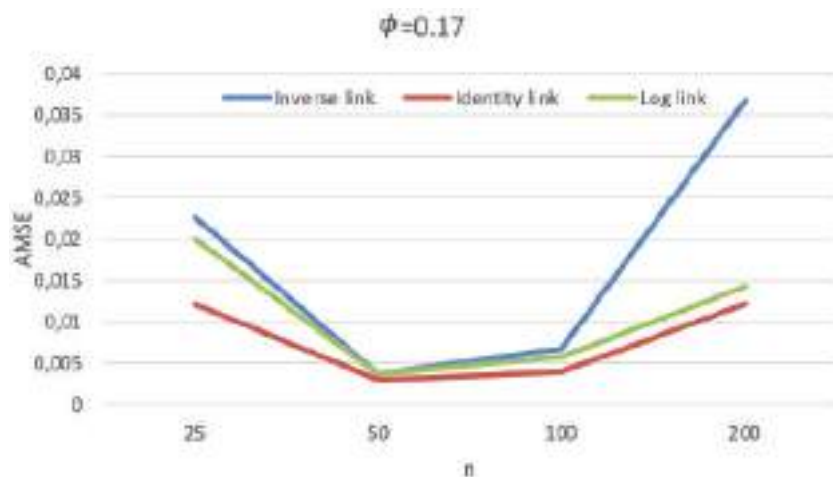


Figure 3. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.17$.

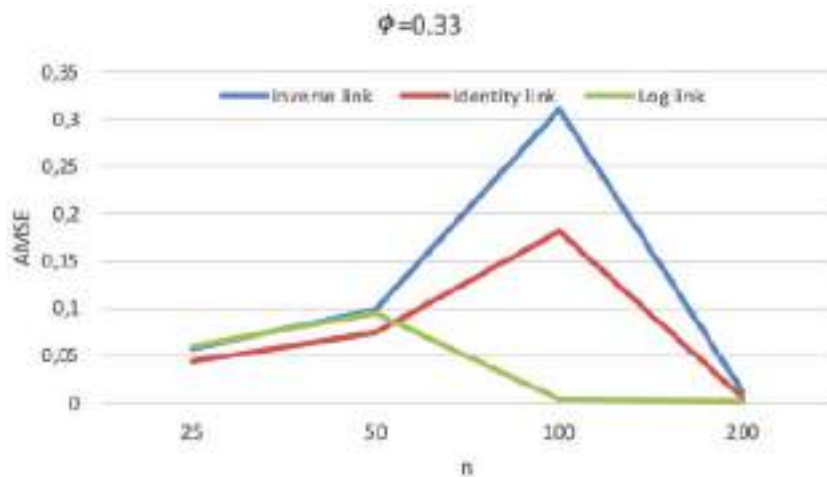


Figure 4. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.33$.

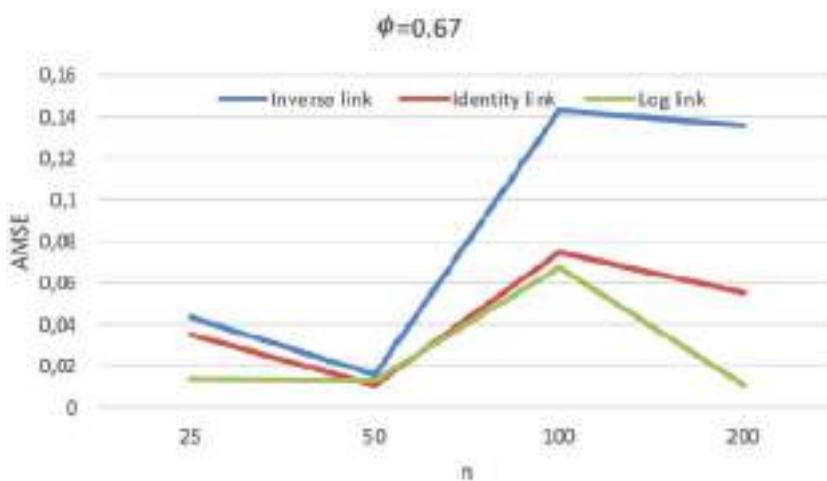


Figure 5. G-PRM under different link functions using AMSE for simulated data, when $\phi = 0.67$.

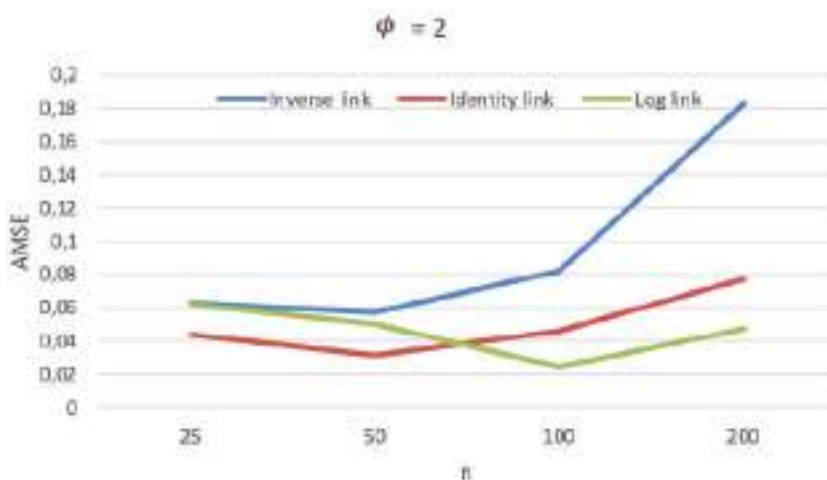


Figure 6. G-PRM under different link functions using AMSE for simulated data, when $\phi = 2$.

- When dispersion level $\phi=0.11$ and sample sizes $n = 25, 50, 100$ the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample size $n = 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- When dispersion level $\phi=0.17$ and for all sample sizes $n = 25, 50, 100$ and 200 the identity link function provides a minimum AMSE as compared to the inverse and log link functions.
- When dispersion level $\phi=0.33$ and sample sizes $n = 25, 50$ the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample sizes $n = 100, 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- When dispersion level $\phi=0.67$ and sample sizes $n = 25$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions. When sample sizes $n = 50$ the identity link function provides a minimum AMSE. For sample size $n = 100, 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- When large dispersion level $\phi=2$ and sample sizes $n = 25, 50$ the identity link function provides a minimum AMSE as compared to the inverse and log link functions. When sample sizes $n = 100, 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.
- It is evident that results are consistent, as AMSE are decreasing with increase of sample size. On the basis of simulation results the identity link function is best for G-PRM.

The graphical results of G-PRM under different link functions using AMSE for simulated data are reported in figures 1-6.

- The figures 1-6, show the performance of different link functions on the basis of AMSE. It is interesting to note that for small and large dispersion levels and sample size's $n = 25, 50$ the identity link function provides a

minimum AMSE as compared to the inverse and log link functions. We can see in all the graphs when sample size's $n = 100, 200$ the log link function provides a minimum AMSE as compared to the inverse and identity link functions.

Application: Reaction Rate Data

Next, we shall see how the various link functions will work when applied to G-PRM using a real-life application. We employed the reaction rate data taken from [32,33]. Then [34,35,36,37] utilized this data set. Data set consist of 24 observations and provide the reaction rate (y) as dependent variable. Three independent variables ($p= 3$) are used to speed up the reaction rate, which are partial pressure of hydrogen (x_1), partial pressure of n-pentane (x_2) and partial pressure of iso-pentane (x_3). As it is mentioned response variable follows a gamma distribution (this is required by following [16,17]). However, this data set is not well fitted to the normal distribution since the trend of the dependent variable is positively skewed also from the distribution fitting test, we observed that the GP distribution is well fitted to this data set, the results are reported in Table 3.

So, the appropriate regression model to determine the reaction rate (y) based on these three explanatory variables such as $p = 3$ explanatory variables, i.e. partial pressure of hydrogen (x_1), partial pressure of n-pentane (x_2) and partial pressure of iso-pentane (x_3) is the G-PR model.

The fitted G-PRM for inverse link function using real data is given by.

$$\hat{y}_i = (0.6700[0.092, S] - 88.689x_1[0.0007, S] + 57.982x_2[0.0002, S] - 138.60x_3[0.0002, N])^{-1}$$

The fitted G-PRM for identity link function using real data is given by.

$$\hat{y}_i = (-3.1750[0.3832, N] + 0.0598x_1[0.0024, S] - 0.0673x_2[0.0025, S] + 0.0046x_3[0.0008, S])$$

Table 3. Distribution goodness of fit tests for Reaction Rate Data

Goodness of fit test		Probability Distribution						
		Gamma	Pareto	Gamma-Pareto	Weibull	Weibull-Pareto	Normal	Normal-Pareto
Anderson-Darling (AD)	Statistic	0.2519	3.1872	0.3299	0.2943	0.9912	1.2462	0.8055
	P-value	0.7538	0.5881	0.8103	0.6288	0.2033	0.0027	0.0390
Cramer-von Mises (CVM)	Statistic	0.0432	0.4570	0.2797	0.0521	0.2392	0.2127	0.1865
	P-value	0.6259	0.0067	0.6922	0.4772	0.3143	0.0033	0.0051
Pearson chi-square (PCS)	Statistic	2.0000	14.880	24.774	6.0000	9.6522	10.667	17.995
	P-value	0.8491	0.0033	0.9605	0.3062	0.2071	0.0584	0.0949

Gamma-Pareto Distribution (GPD).

Table 4. Real Data Coefficients by using different link functions

Regression Coefficients and Standard Error	Link Functions		
	Inverse link	Identity link	Log link
β_0	0.6700	-3.1750	1.0028
(S.E)	(0.0932)	(0.3832)	(0.3434)
β_1	-88.689	0.0598	0.9775
(S.E)	(0.0007)	(0.0024)	(0.0016)
β_2	57.982	-0.0673	-0.4075
(S.E)	(0.0002)	(0.0025)	(0.0011)
β_3	-138.60	0.0046	-0.1104
(S.E)	(0.0002)	(0.0008)	(0.0007)

Table 5. Real Data Fitted Models for different link functions

$Y_{i \text{ response.v}}$	$\hat{\mu}_{\text{inverse}}$	$\hat{\mu}_{\text{identity}}$	$\hat{\mu}_{\text{log}}$
3.541	3.6445	3.1277	2.9818
2.397	2.8500	2.7093	2.3065
6.694	4.5243	5.6077	5.7216
4.722	3.6732	5.6695	5.1428
0.593	1.9297	0.7576	1.1971
0.268	1.6179	0.2461	0.8800
2.797	2.1526	3.4517	2.4241
2.451	1.8564	3.1343	1.9287
3.196	2.9239	3.4936	2.9882
2.021	2.1108	2.7643	1.9187
0.896	2.1208	0.7749	1.2649
5.084	2.6867	5.1504	4.0298
5.686	5.6948	5.4398	5.7639
1.193	1.6180	1.1403	1.1053
2.648	2.4622	3.1403	2.4049
3.303	2.4119	3.0862	2.3485
3.054	2.4319	3.0683	2.3477
3.302	2.4421	3.0945	2.3681
1.271	2.4420	1.7354	1.7310
11.648	10.897	7.5470	11.113
2.002	2.8125	2.4733	2.1626
9.604	11.995	9.8081	18.350
7.754	6.1454	5.6315	6.2097
11.59	14.270	9.7023	18.724

The fitted G-PRM for log link function using real data is given by.

$$\hat{y}_i = \log(1.0028[0.3434, S] + 0.9775x_1[0.0016, N] - 0.4075x_2[0.0011, N] - 0.1104x_3[0.0007, S])$$

where the square brackets contain the standard errors of the estimated parameters. The letter N represents the non-significance and S represents the significance of the regression coefficients.

By using above mentioned three link functions (inverse, identity and log) in G-PRM, we have calculated coefficients

Table 6. Real Data Squared Residuals and sum Squared Residuals and Mean Squared Error

Sr.	SR _(inverse)	SR _(identity)	SR _(log)
1	0.0107	0.1708	0.3128
2	0.2052	0.0975	0.0082
3	4.7075	1.1800	0.9455
4	1.0999	0.8978	0.1771
5	1.7868	0.0271	0.3649
6	1.8222	0.0005	0.3745
7	0.4153	0.4287	0.1391
8	0.3535	0.4669	0.2727
9	0.0740	0.0886	0.0432
10	0.0081	0.5524	0.0105
11	1.5000	0.0147	0.1361
12	5.7472	0.0044	1.1113
13	0.0001	0.0606	0.0061
14	0.1806	0.0028	0.0077
15	0.0345	0.2423	0.0591
16	0.7941	0.0470	0.9110
17	0.3870	0.0002	0.4989
18	0.7394	0.0431	0.8723
19	1.3712	0.2157	0.2116
20	0.5633	16.817	0.2861
21	0.6569	0.2222	0.0258
22	5.7209	0.0417	76.493
23	2.5875	4.5051	2.3849
24	7.1826	3.5632	50.899
SSR	97.715	29.691	136.55
MSE	4.8857	1.4845	6.8227

and standard errors of reaction rate data which are presented in Table 4. The impact of link functions can be observed that how link functions change the effect of regressors. In our case, the use of identity link makes the intercept negative but rest of two provide a positive intercept as β_0 . Similarly, the use of inverse link for β_1 provide a very large negative co-efficient. For β_2 , inverse link provides a large co-efficient but this time with positive effect, other two link functions provide competitively very small negative coefficient. Inverse and log link function generate negative coefficients for β_3 but coefficients generated by inverse link function is very large and other one is very small. Identity link function provide positive small coefficient. Standard errors (S.E) by inverse link functions are smallest then others. In Table 5, we fitted a three G-PRM for inverse, identity and log link functions. And the estimated values of Y_{response} , fitted models are $\hat{\mu}_{1\text{inverse}}$, $\hat{\mu}_{2\text{identity}}$ and $\hat{\mu}_{3\text{log}}$. It is evident that G-PRM $\hat{\mu}_{2\text{identity}}$ has a minimum SSR and MSE as compare to others $\hat{\mu}_{1\text{inverse}}$ and $\hat{\mu}_{3\text{log}}$.

CONCLUSION

In this paper, we evaluated the performance of different link functions to estimate G-PRM and identify the best link function. For this purpose, inverse, identity and log link functions are considered. To compare the performance of different link functions, we conducted a Monte Carlo simulation study an empirical application Reaction rate data. It has widespread applications in the fields of industry, medical science Biopsy, rainfall data, chemical science and engineering etc. The Gamma-Pareto regression model is a powerful tool for modeling continuous positive data, and the choice of link function can significantly impact the model's performance. Here are some advantages of our study using different link functions in the Gamma-Pareto regression model. The log link is commonly used in Gamma-Pareto regression as it provides a natural way to model the mean of the response variable. It also ensures that the predicted values are always positive. Advantage easy to interpret, and the estimated coefficients can be exponentiated to obtain relative changes in the mean. The identity link is similar to the log link but does not

involve any transformation. It is useful when the response variable has a large range of values. Advantage the model is simpler to implement. The inverse link is useful when the response variable has a large range of values and the relationship between the mean and variance is not strong. Advantage provides more flexible modeling of the mean-variance relationship. The choice of link function in Gamma-Pareto regression depends on the specific characteristics of the data and the research question. Each link function has its advantages, and selecting the appropriate one can lead to more accurate and interpretable results. In the simulation study, we used different dispersion level dispersion levels $\phi=0.04, 0.11, 0.17, 0.33, 0.67$ and 2 and different sample sizes $n = 25, 50, 100$ and 200 . we evaluated the performance of the G-PRM under finite sample sizes, dispersion parameters and different link functions. The performance evaluation of these link functions has been done by using SSR, MSE and AMSE. From the simulation results, we observed that the performance of identity link function in G-PRM provides minimum AMSE as compared to inverse and log link functions for all simulation cases. From the reaction rate data identity link function in G-PRM provides a minimum SSR and MSE as compared to inverse and log link function. We also noticed from simulation results that the identity link function gives better performance as compared to the other link functions. Based on real-life and simulated results, we may suggest that identity link function is appropriate whenever practitioners want to apply G-PRM. In general, we recommend the use of identity link function in G-PRM.

Recommendations of research to be conducted in the future, some dimensions still require exploration. In this research, the performance of various link functions is being discussed based on SSR, MSE and AMSE. These can be extrapolated to the influence diagnostics that find the various GLM residuals in various link functions in the G-PRM. Even further generalized to GLM influence diagnostics, where one parameter is estimated in a biased manner (Ridge), modified ridge estimation, Liu estimation, modified Liu estimation and Stein estimation.

FUNDING

No funding resource is available for this study.

DECLARATIONS CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- [1] Altun E, Korkmaz MC, El-Morshedy M, Eliwa MS. The extended gamma distribution with regression model and applications. *AIMS Mathematics*, 2021.
- [2] Pammer K, Kevan A. The contribution of visual sensitivity, phonological processing, and nonverbal IQ to children's reading. *Scientific Studies of Reading*. 2007 ;11(1):33-53.
- [3] Burroughs SM, Tebbens SF. Upper-truncated power law distributions. *Fractals*. 2001; 9(02):209-22.
- [4] Pickands III J. Statistical inference using extreme order statistics. *the Annals of Statistics*. 1975; 119-31.
- [5] Akinsete A, Famoye F, Lee C. The beta-Pareto distribution. *Statistics*. 2008 ;42(6):547-63.
- [6] Mahmoudi E. The beta generalized Pareto distribution with application to lifetime data. *Mathematics and computers in Simulation*. 2011 ;81(11):2414-30.
- [7] Mustafa S, Amin M, Akram MN, Afzal N. On the performance of link functions in the beta ridge regression model: Simulation and application. *Concurr Comput Pract Exp* 2022;34:e7005. [\[CrossRef\]](#)
- [8] Prasetyo RB, Kuswanto H, Iriawan N, Ulama BSS. A comparison of some link functions for binomial regression models with application to school drop-out rates in East Java. *AIP Conf Proc* 2019;2194:012001. [\[CrossRef\]](#)
- [9] Afsana-Al-Sharmin, Islam MA. Generalized Weibull linear models with different link functions. *Adv Appl Stat* 2017;50:367–384. [\[CrossRef\]](#)
- [10] Amin M, Noor A, Mahmood T. Beta regression residuals-based control charts with different link functions: An application to the thermal power plants data. *Int J Data Sci Anal* 2024;22:1–13. [\[CrossRef\]](#)
- [11] Amin M, Fatima A, Akram MN, Kamal M. Influential observation detection in the logistic regression under different link functions: An application to urine calcium oxalate crystals data. *J Stat Comput Simul* 2024;94:346–359. [\[CrossRef\]](#)
- [12] Cheema M, Amin M, Mahmood T, Faisal M, Brahim K, Elhassanein A. Deviance and Pearson residuals-based control charts with different link functions for monitoring logistic regression profiles: An application to COVID-19 data. *Mathematics* 2023;11:1113. [\[CrossRef\]](#)
- [13] Hadia M, Amin M, Akram MN. Comparison of link functions for the estimation of logistic ridge regression: An application to urine data. *Commun Stat Simul Comput* 2024;53:4121–4137. [\[CrossRef\]](#)
- [14] Alzaatreh A, Famoye F, Lee C. Gamma-Pareto distribution and its applications. *J Mod Appl Stat Methods* 2012;11:7. [\[CrossRef\]](#)
- [15] Hanum HH. Modeling extreme rainfall with gamma. *Appl Math Sci* 2015;9:6029–6039. [\[CrossRef\]](#)
- [16] Hanum H, Wigena AH, Djuraidah A, Mangku IW. Developing generalized linear model of Gamma-Pareto Distribution. *Far east Mathematical Sciences Journal*. 2016.
- [17] Hanum H, Wigena AH, Djuraidah A, Mangku IW. Modeling gamma-Pareto distributed data using GLM gamma. *Glob J Pure Appl Math* 2016;12:3569–3575.
- [18] Hanum H, Wigena AH, Djuraidah A, Mangku IW. The application of modeling gamma-Pareto

- distributed data using GLM gamma in estimation of monthly rainfall with TRMM data. *Sriwijaya J Environ* 2017;2:40–45. [\[CrossRef\]](#)
- [19] Alzaatreh A, Ghosh I. A study of the Gamma-Pareto (IV) distribution and its applications. *Commun Stat Theory Methods* 2016;45:636–654. [\[CrossRef\]](#)
- [20] De Andrade TA, Fernandez LMZ, Silva FG, Cordeiro GM. The gamma generalized Pareto distribution with applications in survival analysis. *Int J Stat Probab* 2017;6:141–156. [\[CrossRef\]](#)
- [21] Alzaghal A. The exponentiated gamma-Pareto distribution with application to bladder cancer susceptibility. *J Stat Appl Probab Lett* 2020;7:39–56. [\[CrossRef\]](#)
- [22] Dar AA, Ahmed A, Reshi JA. Weighted Gamma-Pareto distribution and its application. *Pak J Stat* 2020;36:287–304.
- [23] Ashour SK, Said DR, Fahim MA. The log-Gamma-Pareto distribution. *Int J Sci Basic Appl Res* 2014;16:357–374.
- [24] Özlü Ş. New q-rung orthopair fuzzy Aczel–Alsina weighted geometric operators under group-based generalized parameters in multi-criteria decision-making problems. *Computational and Applied Mathematics*. 2024; 43(3):122.
- [25] Özlü Ş. Q-rung orthopair probabilistic hesitant fuzzy hybrid aggregating operators in multi-criteria decision-making problems. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*. 2023; 27(3):366–73.
- [26] Özlü Ş, Al-Quran A, Riaz M. Bipolar valued probabilistic hesitant fuzzy sets based on Generalized Hybrid Operators in multi-criteria decision-making problems based on TOPSIS. *Journal of Intelligent & Fuzzy Systems*. 2024; 1–20.
- [27] Özlü Ş. Generalized Dice measures of single valued neutrosophic type-2 hesitant fuzzy sets and their application to multi-criteria decision-making problems. *International Journal of Machine Learning and Cybernetics*. 2023;14(1):33–62.
- [28] Özlü Ş, Karaaslan F. Hybrid similarity measures of single-valued neutrosophic type-2 fuzzy sets and their application to MCDM based on TOPSIS. *Soft Computing*. 2022; 26(9):4059–80.
- [29] Özlü Ş. multi-criteria-decision making based on vector similarity measures of picture type-2 hesitant fuzzy sets. *Granular Computing*. 2023; 8(6):1505–31.
- [30] Dobson AJ. An introduction to generalized linear models. Boca Raton: Chapman and Hall/CRC; 2002. [\[CrossRef\]](#)
- [31] Amin M, Amanullah M, Cordeiro GM. Influence diagnostics in the gamma regression model with adjusted deviance residuals. *Commun Stat Simul Comput* 2017;46:6959–6973. [\[CrossRef\]](#)
- [32] Huet S, Bouvier A, Poursat MA, Jolivet E, Bouvier AM. Statistical tools for nonlinear regression: A practical guide with S-PLUS and R examples. New York: Springer; 2004. p. 233.
- [33] Huet S, Bouvier A, Poursat MA, Jolivet E, Bouvier AM. Statistical tools for nonlinear regression: A practical guide with S-PLUS and R examples. New York: Springer; 2006. p. 233.
- [34] Amin M, Amanullah M, Aslam M, Qasim M. Influence diagnostics in gamma ridge regression model. *J Stat Comput Simul* 2019;89:536–556. [\[CrossRef\]](#)
- [35] Amin M, Qasim M, Amanullah M. Performance of Asar and Genç and Huang and Yang's two-parameter estimation methods for the gamma regression model. *Iran J Sci Technol Trans A Sci* 2019;43:2951–2963. [\[CrossRef\]](#)
- [36] Amin M, Qasim M, Yasin A, Amanullah M. Almost unbiased ridge estimator in the gamma regression model. *Communications in Statistics-Simulation and Computation*. 2022;51(7):3830–50.
- [37] McCullagh P. Generalized linear models. London: Routledge; 1989. [\[CrossRef\]](#)
- [38] Saleem N, Akbar A, Laeeq S, Ahmad S, Hanum H. Influential observations detection in the gamma-pareto regression model under different link functions with standardized and adjusted deviance residuals: simulation and application. *Sigma*. 2025;43(3): 760–776.
- [39] Saleem N, Akbar A, Rehman Irshad AU, Ozdemir, B, Ahmad Khan J. Influence Diagnostics in Gamma-Pareto Regression Model with Cook's Distance for Standardized and Adjusted Pearson, Deviance, and Likelihood Residuals: Simulation and Application. *Journal of Mathematics*. 2025(1);3440237.
- [40] Saleem N, Akbar A. Empirical evaluation of the gamma-Pareto regression residuals for influence diagnostics. *Journal of Statistical Computation and Simulation*. 2024;94(16):3573–3598.
- [41] Saleem N, Akbar A, Imon AR, Khan JA. Multicollinearity in Binomial Regression: A comparison between CERES and PR Plots for detection. *Journal of Statistical Modeling & Analytics (JOSMA)*. 2024;6(1).
- [42] Saleem N, Akbar A, Imon AR, Al Mamun, ASM. Detection of Outliers in Binomial Regression Using CERES and Partial Residual Plots. *Journal of Statistical Modeling & Analytics (JOSMA)*. 2022;4(2).
- [43] Saleem N, Akbar A, Sharif SB, Al Mamun ASM, Imon AHMR, & Ahmad S. Performance Comparison between CERES and PR Plots in Detecting the Heteroscedasticity Problem using Liver Cancer and Simulated Data. *Global Journal of Engineering and Technology*, 2022;7(3).



Research Article

Experimental investigation of thermo-hydraulic performance of multi semicircular rib of dimple hybrid channel

Kishor K. DHANDE^{1,*}, Prakash Santosh PATIL², Lalit N. PATIL³, Vikash K. AGRAWAL³,
Atul A. PATIL¹

¹Department of Mechanical Engineering, Dr. D.Y. Patil Institute of Technology, Pune, 411033, India

²Department of Mechanical Engineering, JSPM's Rajarshi Shahu College of Engineering, Pune, 411033, India

³Department of Automation and Robotics, Dr. D.Y. Patil Institute of Technology, Pune, 411033, India

ARTICLE INFO

Article history

Received: 30 March 2024

Revised: 12 June 2024

Accepted: 24 September 2024

Keywords:

Compound Channel; Heat Transfer Augmentation; Multi Semicircular Rib; Optimization Techniques

ABSTRACT

To enhance the performance of gas turbine, it is necessary to operate at high inlet gas temperature, which required more effective cooling method for turbine blades. The ribs, protrusion and dimples are efficient turbulators for heat transfer augmentation. In the present study new compound structure of multi semicircular rib-dimple are presented experimentally and validated with simulation/optimization techniques to achieve pressure drop and heat transfer characteristics. Multi semicircular rib with 10 and 20 dimples are investigated and placed at the bottom surface of a rectangular channel. The experiment was conducted by considering the data such as: Reynolds numbers (Re) 12,000 to 30,000, Rib pitch (P) to height (e) ratio = 8 to 10, rib height (e) to channel hydraulic diameter (Dh) ratio = 0.15 and dimple depth (δ) to dimple diameter (d) ratio = 0.2. It is found that the combination of multi semicircular rib with 20 dimples shows the maximum thermal performance over multi semicircular rib with 10 dimples and rib alone channel. Friction loss found more in compound channel with 20 dimples whereas less in the 10 dimples. Compound channel with 20 dimples enhanced 30 % more heat transfer over 10 dimples and 25 % additional than the rib alone channel, also multi semicircular rib shows the least thermal performance.

Cite this article as: Dhande KK, Patil PS, Patil LN, Agrawal VK, Patil AA. Experimental investigation of thermo-hydraulic performance of multi semicircular rib of dimple hybrid channel. Sigma J Eng Nat Sci 2025;43(5):1507–1519.

INTRODUCTION

To get more output and efficiency from a gas turbine; it should operates at very high inlet gas temperature. These temperatures are not sustained by blade material. To protect the blade life; various cooling techniques are used;

the internal cooling method involves protrusion, dimple, ribs, fin, vortex generator and groove. Rib turbulators and dimple alone are frequently used in the gas turbine blade cooling or other cooling applications. These methods are useful to generate recirculation, vortices, separation of

*Corresponding author.

*E-mail address: dhandekishor183@gmail.com

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



flow and attachment/reattachment in the flow channel that enhanced the involvement of fluid molecules and consecutively heat transfer rate.

In the early years; many researchers focused on different types of rib geometries. J. C. Han [1] reported parallel and V-shaped broken ribs and shows that the 60° parallel or V-shaped broken ribs indicates more heat transfer enhancement than the 45° parallel or V-shaped broken ribs. Jongmyung Park et al. [2] numerically presented 45 deg. angled rib for flow structural at Reynolds number from 18300 to 48000. Some studies show that due to angled rib turbulators large-scale secondary flow induced also mixing of flow develops, which increased the local Nusselt numbers. Patil and Borse [3] focused on various cooling techniques for gas turbine blade, concluded that there is still scope to use W Shape, Semicircular and Multi Semicircular shape ribs. Luca Baggetta et al. [4] studied 450 angled ribs using the liquid crystal thermography and show that existence of intersecting rib, percentage rate of heat transfer enhancement was less than the pressure drop. Abraham and Vedula [5] presented V and W rib for converging channel at Reynolds number ranging from 5000 to 30000 and found same average Nusselt number of V and W shape ribs. Promvonge and Thianpong [6] experimentally presented rectangular, triangular and wedge shapes ribs at staggered and in-line arrangement, shows triangular rib with staggered arrangement performed well over the other ribs. Jae-Won Seo et al. [7] presented optimization of a boot-shaped rib at Reynolds number 20,000 using 3-dimensional RANS analysis and MOGA techniques. Concluded that the heat transfer oriented design shows more FNu and Ff than to the friction factor oriented design. Min Kim et al. [8] presented an optimal design to find out appropriate rib geometry of angled rib using advanced response surface method. Observed the maximum thermal performance at rib angle (α) - 54.670 deg. and pitch ratio (P/e) - 6.8.

In some another cooling technique, dimples are used for heat transfer enhancement due to less pressure drop advantage. Yu Rao [9] studied various shapes of dimple such as: spherical, teardrop, elliptical and inclined elliptical at Reynolds number 8500-60,000. Result shows that the teardrop dimples performed well against regular spherical dimples; also the elliptical dimples indicate lowest performance than the spherical dimples. Lu Zheng et al. [10] studied ridged dimple and show that reduction in the wall temperature and enhanced the heat transfer. Saeed Nazari et al. [11] studied staggered, square and triangular arrangement of dimples by taking three depths of dimples: ($\delta/d = 0.25, 0.375$ and 0.5). Result indicates that dimple depth ratio (δ/d) = 0.25 which shows the superior performance than other tested depth ratio. Also staggered and triangular arrangement of dimple performed well at low and high dimple depth. Ayush Gupta et al. [12] studied the performance of plate heat sink using dimples and protrusions. Show that maximum performance of heat sink observed at a staggered arrangement when dimple pitch ratio (s/d) kept 2.5 and dimple depth ratio 0.5.

To fulfil the today's industry need and also to achieve more enhancement from cooling techniques; researchers are focused on combination of two ribs (hybrid structure), Sahu et al. [13] focused on hybrid ribs, which is combination of triangular and semicircular ribs. The rib pitch to height ratio (p/e) kept 6.6 to 53.3 and Reynolds number from 12,640 to 52,410, they show that hybrid rib performance was highest at pitch ratio of 13.3. Pongjet Promvonge et al. [14] presented combination of chamfered-V-groove and punched-V-rib for solar air heater duct. Range of Reynolds number (Re) varies from 5300 to 23,000. It indicates that the combined punched-V-rib and chamfered-V groove shows 14-15% more performance compared to chamfered-V-groove and combined solid-V-rib. Qi Jing et al. [15] studied different dimple/protrusion arrangements at Reynolds number 10,000 to 100,000 concluded that protrusion produced the largest Nu compared with dimple, also performance of combined dimple-protrusion was found in between dimple and protrusion alone case.

Nowadays, few researchers focused on combination of two methods for cooling the gas turbine blades, such as ribs with protrusions, pin fin with dimples, ribs with dimples and dimples with protrusions. Patil et al. [16] presented compound and rib alone channel for gas turbine cooling, indicated that W rib compound channel shows improved performance than W rib channel. Yonghui Xie et al. [17] studied numerically hybrid structure of dimple with secondary protrusions, different configurations of dimple-protrusion was tested and concluded that dimple channel with secondary protrusions shows higher thermal performance than ordinary dimple channel. Also Xie studied combination of two techniques such as: pin-fin-dimple/protrusion, dimple/protrusion structure in a U shaped duct. It shows that the pin fin-dimple/protrusion structure enhanced more heat transfer than the dimple/protrusion structure [18]. Rao and Zhang [19] presented experimentally combination of miniature V-shaped rib with dimple at Reynolds number (Re.) up to 60,000 indicated that the hybrid structure shows more heat transfer enhancement than the dimple structure. Inderjot Kaur and Prashant Singh [20] studied combination of V-shaped protrusion/concavity with miniature V-rib. Concluded that the hybrid structure of V-rib with V-protrusion shows the improved thermal hydraulic performance (THP) of range 2.15.

In literature, major work was found on mainly the dimples or ribs alone, only some investigators studied combination of different shape of ribs or two different methods such as hybrid structure, rib-dimple, rib-protrusion, dimple-protrusion [16,19]. The aim of the existing study is to develop advanced cooling technique for gas turbine blade, which is substantially the combination of two different enhancement techniques. In the present work mostly focused on combination of multi semicircular shape ribs with spherical dimples for thermal performance, also Multi semicircular rib alone and compound channel with 10 and

20 dimples are presented to get optimal configuration of ribs with dimple pattern.

Combustion turbines, also referred to as gas turbines, are a kind of internal combustion engine that produces mechanical energy from liquid fuels like natural gas [21]. Usually, an electric generator or other machinery such as ships, airplanes, and other vehicles are powered by this mechanical energy. The compressor, the combustion chamber, and the turbine are the three primary parts that make up a gas turbine's basic mechanism [22]. The compressor takes in ambient air and compresses it to a high pressure to start the operation. In order to get the air ready for combustion, this stage raises its temperature and pressure. Axial-flow compressors, which have several stages of moving and stationary blades, are used in modern gas turbines to get great compression efficiency. After entering the combustion chamber, the high-pressure air from the compressor is combined with fuel and ignited. The generated gas's temperature and pressure are greatly increased during the combustion process, and the chamber's construction allows it to withstand extremely high temperatures and steady combustion. There are several different kinds of gas turbines, such as microturbines, aeroderivative gas turbines, and heavy-duty gas turbines. Heavy-duty gas turbines are built to run continuously under heavy loads and are commonly seen in power plants that generate electricity. Smaller power plants and mechanical drive applications, including powering compressors or pumps, employ aeroderivative gas turbines, which are evolved from jet engines. Compared to heavy-duty turbines, they are more flexible and lighter. Due to their tiny size and excellent efficiency, microturbines are small-scale turbines that are utilized in combined heat and power (CHP) and distributed generation systems.

EXPERIMENTATION

Experimental Detail and Methodology

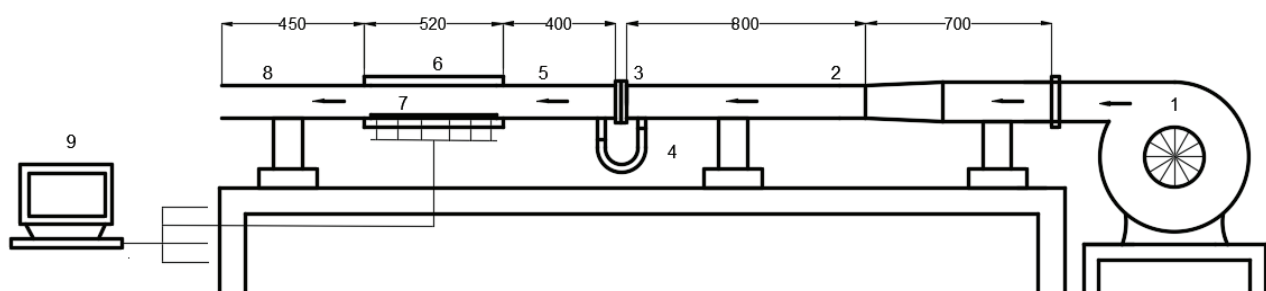
An experimental set up was designed and manufactured to measure the pressure loss and heat transfer in

multi semicircular rib and compound channel, as shown in Figure 1. Main parts of the set up was a centrifugal blower, entrance/developing section, an orifice plate assembly, rectangular duct with test section and exit section. To prevent the heat loss from the rectangular channel, it was made from Bakelite sheet and the entire length of set up was 2900 mm, out of that 1200 mm - entrance section, 520 mm - test region and 450 mm - exit region. The aspect ratio and hydraulic diameter of channel was 4:1 and 36 mm respectively. In the experimental set up a blower was connected to 700 mm flow developing region after that entrance section attached to ensure fully developed flow in the rectangular channel then after test section assembly connected. Lastly the exit section was connected to the rectangular duct to decrease the end effects of flow.

Air mass flow rate was adjusted through variable frequency drive (VFD) and measured with an orifice plate assembly. The mica flat plate heater was used to supply uniform heat flux and controlled through voltage variable of 6 Amperes rating. Pressure loss in the test section was recorded using digital micro manometer connected along the inlet and outlet pressure taps. Inlet and outlet temperature of air was recorded using the thermocouples placed at respective sections. Also the surface temperatures of test plate are recorded using K-type thermocouples inserted at various locations of test surface and output of thermocouples are connected to the data acquisition system. All instruments was calibrated from NBL accredited Laboratory and at steady state condition all experimental data was recorded, Figure 2 shows the photo graphs of experimental set up.

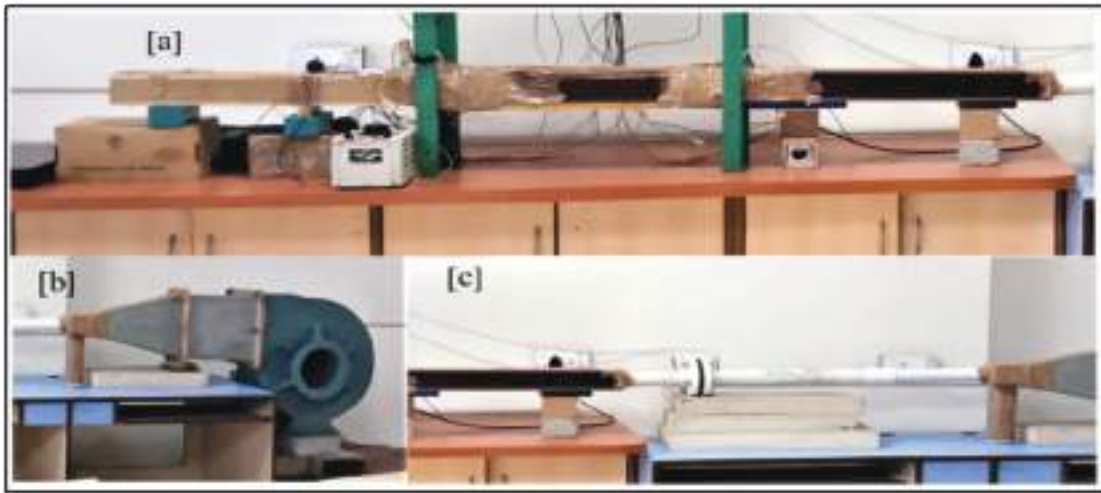
Test Section

For test plate 12 mm aluminum sheet was used and the multi semicircular ribs are pasted on the surface of test plate using adhesive material. Dimples on test plate surface are carved using a special designed ball nose tool of diameter 20 mm and depth of dimple kept 4 mm ($\delta/D = 0.2$). The mica flat plate heater was used to supply a constant heat input and placed between the base plate and a test plate. To reduce the heat loss from the channel, proper glass wool



1- Centrifugal blower 2 - developing/straighten section 3 - orifice plate assembly 4 - U tube manometer 5 - Inlet region 6- rectangular duct 7- Test plate 8 - exit region 9 - data acquisition system

Figure 1. Schematic representation of experimental setup.



a- Rect. duct with exit section, b- Blower with honey comb section, c- Inlet section and developing section with orifice plate

Figure 2. Experimental set up.

insulation was provided between heater and base plate. The test plate with base plate and heater are bolted and placed in a bakelite made rectangular channel. Figure 3 shows the photographs of rectangular duct with different test sections.

Rib Geometry

The multi semi-circular shape rib was made from 45 mm diameter aluminum pipe, first outer and inner

diameter of 3003 aluminium pipe was prepared on CNC machine, then using special purpose machine required size ribs are prepared. Rib width and height was kept 5 mm each, rib height to hydraulic diameter ratio (blockage ratio e/D_h) of the rectangular duct was 0.150 and pitch to height ratio of rib (P/e) was 10 to 12. Special ball nose tool of 20 mm diameter was designed for dimples and carved on the test plate surface by coordinate-measuring machine



Figure 3. Rectangular channel with different test plates.

Table 1. Multi semicircular rib-dimple pattern for cooling application

Cases	Configuration	Number of ribs/dimples
1	Multi semicircular rib channel	10 ribs
2	Dimple channel	20 dimples
3	Multi semicircular rib - dimple compound channel	5 ribs + 10 dimples
4	Multi semicircular rib - dimple compound channel	5 ribs + 20 dimples

(CMM), also dimple depth to dimple print diameter ratio (δ/D) was kept 0.2. In this study multi semicircular rib and compound channel are tested at Reynolds number (Re) 12000 to 30000, for finding the best combination of multi semicircular rib-dimple pattern for cooling application as shown in Table 1.

Data Reduction

In the present work the values of heat transfer coefficient, Nusselt number and friction factor are determined by considering the measured experimental data [23]. The procedure to calculate the different parameters is given below: The Reynolds number (Re) is obtained from equation (1) recommended by Promvong and Thianpong (2008), where μ - dynamic viscosity of fluid (air) and V - velocity of air passing through duct.

$$Re = \frac{\rho V D_h}{\mu} \quad (1)$$

Equation (2) was used for calculation of average heat transfer coefficient (h_a) proposed by Han and Zhang (1992), where T_s - test plate average surface temperature and T_b - bulk temperature of air.

$$h_a = \frac{Q_{net}}{A(T_s - T_b)} \quad (2)$$

The net heat supplied (Q_{net}) was computed from the total heat supplied minus heat loss from the test section, also found maximum heat loss below 13% from the test channel. The average Nusselt number was computed by equation (3) proposed by Luca Baggetta et al. (2017) where h_a - average heat transfer coefficient, D_h - rectangular channel hydraulic diameter and k - thermal conductivity of air. Nusselt number results of plain plate are normalized with modified Dittus-Boelter/Sieder and Tate correlation (4).

$$Nu = \frac{h_a D_h}{k} \quad (3)$$

$$\frac{Nu}{Nu_o} = \frac{h D_h / k}{0.021 Re^{0.8} Pr^{0.4}} \quad (4)$$

Frictional factor was calculated using Eq. (5) considering the pressure loss along the test channel and it is required to evaluate the thermal performance. Plain duct experimental friction factor result was normalized with Blasius correlation [24] and calculated using equation (6).

$$f = \frac{(P_i - P_e)}{[(4(L/D_h)(1/2 \rho V^2))]} \quad (5)$$

$$\frac{f}{f_o} = \frac{f}{0.079 Re^{-1/4}} \quad (6)$$

Thermal performance (η) is computed by equation (7) proposed by Luca Baggetta et al. (2017) and for calculation

considered the friction factor ratio and Nusselt number ratio.

$$\eta = \frac{(Nu/Nu_o)}{(f/f_o)^{1/3}} \quad (7)$$

When determining the thermal performance in the experimental investigation of the multi semicircular rib - dimple hybrid channel, several key assumptions are typically considered. The system is assumed to be in a steady-state, meaning that the temperatures and heat fluxes are constant over time. This simplifies the analysis by eliminating the need to account for transient effects [24]. A uniform heat flux is applied to the channel walls. This assumption ensures that the heat input is consistent along the length of the channel, facilitating a more straightforward analysis of the heat transfer characteristics. The thermophysical properties of the working fluid, such as thermal conductivity, viscosity, and specific heat, are assumed to be constant. This assumption is valid if the temperature range within the channel is not large enough to cause significant variations in these properties. The contribution of thermal radiation to the overall heat transfer is assumed to be negligible compared to convection. This is a reasonable assumption for typical fluid flow conditions in heat exchangers and cooling channels where convective heat transfer dominates. The flow is assumed to be fully developed both hydrodynamically and thermally. This means that the velocity and temperature profiles have stabilized and do not change along the length of the channel, except at the entrance and exit regions.

Uncertainty Analysis

The experimental uncertainty was calculated using the method recommended by Kline and McClintock [25]. The maximum error found in pressure drop, flow rate and temperature measurement are $\pm 3.4\%$, $\pm 3.1\%$ and $\pm 0.7^\circ\text{C}$ respectively. The ambiguity in the Nusselt number was observed up to 9.5% and the maximum experimental ambiguity observed up to 7.7% in the friction factor [26]. Uncertainty analysis is an essential aspect of experimental research, as it quantifies the degree of confidence in the measured data. This analysis allows researchers to understand the reliability and accuracy of their experimental results, accounting for possible errors or uncertainties in measurements.

Normalize Smooth Duct Results with Correlations

To validate the experimental data of plain duct, its results are verified with correlation results. The experimental results of Nusselt number was verified with the results of modified Dittus-Boelter and Sieder-Tate correlation [27] as shown in equation (8, 9). Experimental friction factor for smooth duct was obtained from the recorded data and normalized with the result obtained from modified Blasius correlation [28] as shown in equation (10).

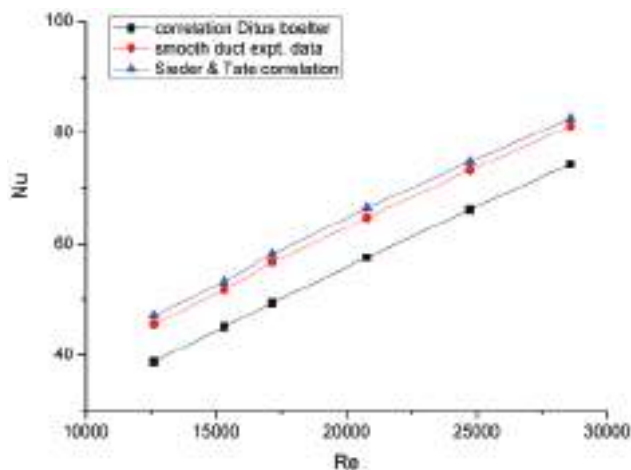


Figure 4. Deviation in the experimental and correlation value of plain duct.

$$Nu = 0.0230 Re^{0.8} Pr^{0.33} (\mu/\mu_w)^{0.14} \quad (8)$$

$$Nu = 0.023 Re^{0.8} Pr^{0.4} \quad (9)$$

$$f = 0.079 Re^{-1/4} \quad (10)$$

Figure 4 presents the plot of correlation and experimental value of Nusselt number for plain duct against Reynolds number. From graph concluded that experimental results was nearer to the Sieder-Tate correlation than Dittus bolter correlation. The average error in the Nusselt number was found between ± 8.5 to 10.3% , also the friction factor error was found 6.2 to 8.5% than Blasius correlations. Therefore it is presumed that experimental data are in good agreement with the correlation data indicated in the literature survey.

RESULTS AND DISCUSSIONS

Nusselt number and friction factor are calculated from collected experimental data and presented at a range of Reynolds number from 12000 to 30000. The result of friction factor and Nusselt number of multi semicircular rib and compound channels are compared with each other to determine the optimal rib-dimple configuration for cooling of gas turbine blade.

Heat Transfer in Ribbed Duct

This section discusses the rib alone and compound channel effect on the enhancement of heat transfer. Figure 5 presents the Nusselt number for rib and dimple alone channel. It indicates that multi semicircular rib performed better than plain and dimple duct, augmented average 96 and 68 % heat transfer than plain and dimple duct respectively. Rib improved heat transfer because ribs disturbed, promote flow mixing and induced diverse secondary flows. In case of multi semicircular rib its two small curvature shape

created two different more heat transfer area in front of ribs and two low heat transfer region behind the ribs, due to that observed heat transfer enhancement in multi semicircular rib [29].

Multi semicircular ribs improved the heat transfer but also increased pressure loss, to get additional enhancements generally increased number of ribs but other side pressure loss also improved by higher percentage. To achieve more enhancements in the heat transfer by virtue of minor rise in pressure loss, currently used combination of two methods such as rib -dimple, protrusion-dimple, rib- protrusion. but enhancement of dimple compound structure depends on number of dimples and depth of dimple.

In the existing study presented combination of rib-dimple structure, as dimple has less friction over other enhancement methods. In the current work two types of cases studied, in first case two dimples are followed by multi semicircular rib, means total 10 dimples and five ribs are replaced by ten ribs as shown in Figure (6 c). In second case four dimples followed by multi semicircular rib means total 20 dimples and five ribs tested. In this channel dimples position are such that two dimples kept horizontally and

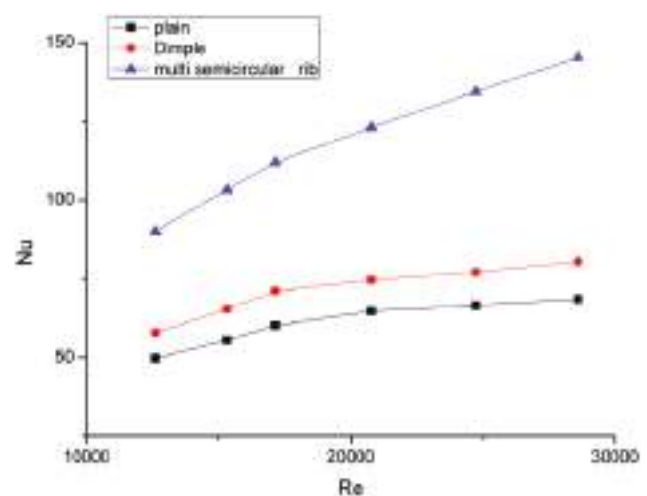


Figure 5. Variation of Nusselt number with Reynold's number (multi semicircular rib).

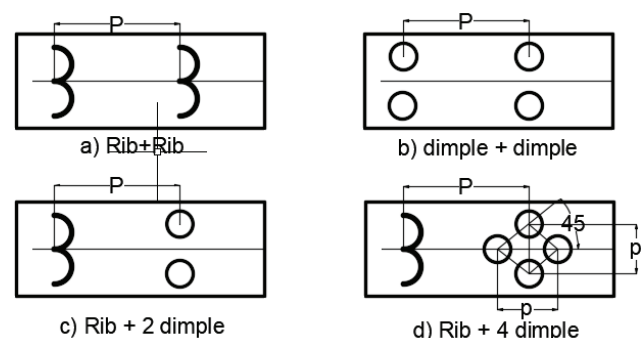


Figure 6. Rib, dimple alone and rib - dimple position.

vertically each, also pitch between dimples are same (horizontal and vertical) and made 45 deg. inclination angles with each other as presented in Figure (6 d).

Figure 7 shows result of Nusselt number for tested geometries, indicates that Nusselt number was maximum for compound duct with 20 dimples than rib alone and 10 dimples compound channel. It means that compound channel shows more enhancement in heat transfer related to the rib alone channel, on the other hand 10 dimples compound channel heat transfer observed less compared to the multi semicircular rib. Enhancement of heat transfer using dimples depends on many factors such as number of dimples, position diameter and depth of dimple of dimple. Compound channel with 20 dimples indicates more heat transfer and compound channel with 10 dimples shows less heat transfer compared to the rib channel. Compound channel with 20 dimples shows 25 and 30 % more heat transfer than multi semicircular rib and 10 dimples compound duct respectively. In case of 20 dimples compound duct position of dimples such that, behind and in front of every half curvature portion of rib trailed by one dimple each, also behind and in front of meet point of two curvature shape followed by one dimple each. Due to such position of dimples more mixing of primary and secondary vortices observed. In this compound structure the secondary flows creating from the ribs will undergo additional improvement due to the dimples and the joint effect of the flow separation layer of dimples and the secondary flows produced by the ribs will result in improvement of heat transfer. It is observed from the present work and earlier work of Patil et al. (2021) that heat transfer was maximum in the compound channel compared to the rib alone channel.

In case of 10 dimples compound channel less enhancement found because of less number of dimples applied in channel also turbulence created by multi semicircular rib was not fully mixed with dimples. Enhancement of 10 dimples compound duct found somewhat less than rib channel but very less than 20 dimples compound channel. From this result dimples involvement in the enhancement are seen, as one rib replaced by two dimples result differed by small percentage than rib channel, but if increased the number of dimples and change position of dimple then found more augmentation in the heat transfer. To increase the number of dimples has limitation, as if only increased the number of dimples, pressure drop also increased so consider optimum number of dimples. Enhancement through dimples depends on depth, diameter, position and quantity of dimples. In the present compound channel work optimum values of these parameters are considered. Figure 8 shows the plausible flow structure for dimple surface having diameter 20 mm and depth ratio $\delta/D = 2$.

Friction Factor in Ribbed Duct

In this section, the friction factors for rib and compound channels are discussed, starting with the rib channel

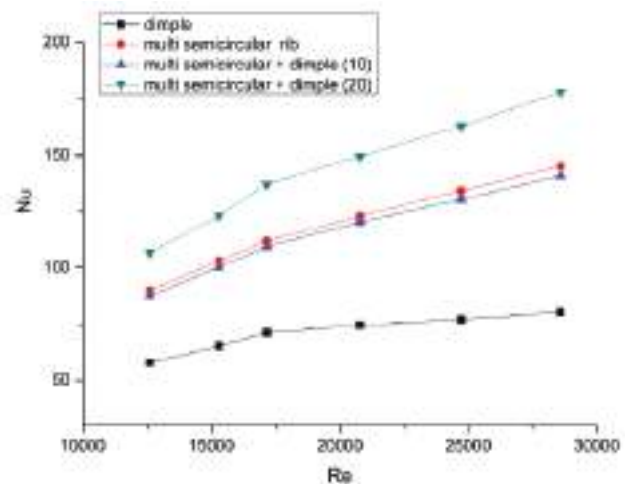


Figure 7. Nusselt number versus Reynolds number for different configurations.



Figure 8. Plausible flow structure of dimple.

followed by the compound channel. The friction factors for the multi-semicircular rib and dimple channels are presented in Figure 9. It was found that the multi-semicircular rib channel exhibited, on average, 200% and 102% more friction compared to the plain and dimple ducts, respectively. This increase in friction is attributed to the ribs creating disturbances and retardation in the flow. The multi-semicircular ribs, with their two curvatures, offered an increased area to the flow, which resulted in higher turbulence and, indirectly, an increased pressure drop. In comparison, the dimple channel showed less friction than the rib channel.

Figure 10 shows the friction factor for rib, dimple, and compound channels, displaying a progressively reducing trend as the Reynolds number increases. This occurs due to the suppression of the viscous sublayer; as the Reynolds number increases, the boundary layer thickness decreases, resulting in improved fluid velocities within the boundary layer. The compound channel with 20 dimples exhibited a better friction factor than the multi-semicircular rib and the compound channel

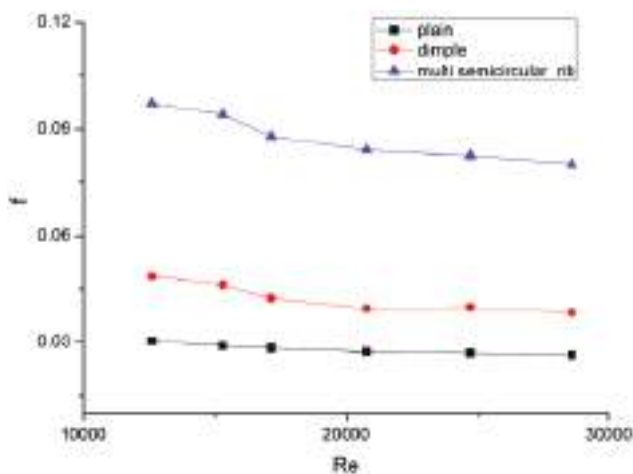


Figure 9. Friction factor variation with Reynolds number (ribs alone and dimple channel).

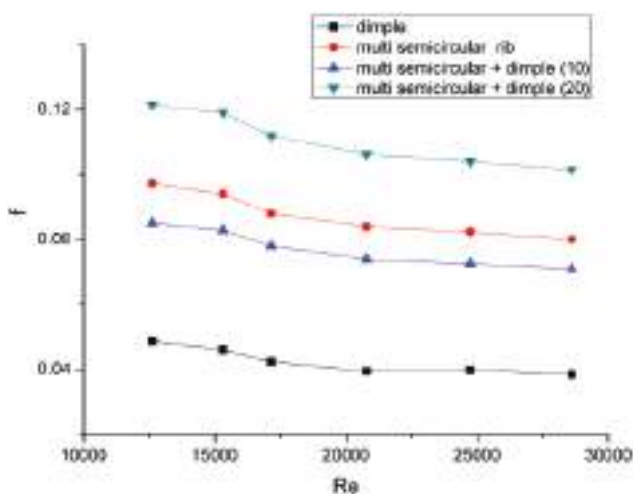


Figure 10. Variation of friction factor with Reynolds number (all tested configurations).

with 10 dimples. In the case of the compound duct with 10 dimples, five ribs are replaced by ten dimples. Since dimples create less friction than ribs, this configuration results in lower friction. Dimples enhance the flow area at the narrowest cross-section, leading to a decrease in flow velocity and a reduction in turbulent mixing in the mainstream flow area, which helps reduce the pressure drop. The compound channel with 20 dimples showed 19% more friction than the rib channel, while the compound channel with 10 dimples showed 11% less friction than the rib channel. Additionally, the compound channel with 20 dimples exhibited 36% more friction than the compound channel with 10 dimples. These findings were also observed by Yonghui Xie et al. (2017).

In case of 20 dimples compound channel four dimples applied after each rib if one or two dimples used friction found may be less but to get further enhancement applied more dimples, so combined effect of four dimples along with rib found more friction. Friction loss in dimples occurred due to depth of dimple, as depth increased friction also increased. The turbulent mixing in the downstream half of the dimple was significantly augmented because of flow attachment and impingement mainly at the rear rim of the dimple, which can consequently, increased the friction. In the present compound channel optimum parameters considered due to that found less growth in the friction over rib channel but other side heat transfer improved by more percentage. From this concluded that to achieve more percentage heat transfer enhancement, increase small percentage of friction, this principle applied in the current compound channel.

Validations of Experimental Results

Experimental results are validated through simulation software STAR-CCM [30]. The realizable $k-\epsilon$ turbulence model with hexahedral mesh was used and validated using standard data obtained from literature. Figure 11 presents velocity contours of multi semicircular rib, 10 dimples and 20 dimples compound channel. Also Figure 12 presents Nusselt number contours of multi semicircular rib with 4 dimples (20 dimples compound channel) and indicates that Nusselt was more near rib portion and dimple corner portion. Rib backside area shows less Nusselt number compared to the front side area due to frontal area effect.

Figure 13 shows the variation in experimental and numerical Nusselt number for compound channel and multi semicircular rib. Simulation results observed near about 8 % higher than experimental results, also 20 dimples compound channel shows highest Nusselt number, as same trend of results observed during experimental work. In case of semicircular rib simulation results are 7 to 8 % higher than experimental results, as difference between both results found less, so it is presumed that experimental results are validated, same trend of simulation results observed by Kaur and Singh (2020).

Figure 14 shows the result of experimental and numerical friction factor of 20 dimples compound channel and multi semicircular rib. Found numerical results 10-11 % less than experimental results, also 20 dimples compound channel shows highest friction than other channel, same results were observed during experimentation. Friction factor of multi semicircular rib found less against 20 dimples compound channel, also seen less variation among experimental and numerical results of friction factor for compound and rib channel. Variance between experimental and numerical results occurred because of heat loss and instrument error in actual experimentation, whereas in numerical analysis an ideal condition was assumed.

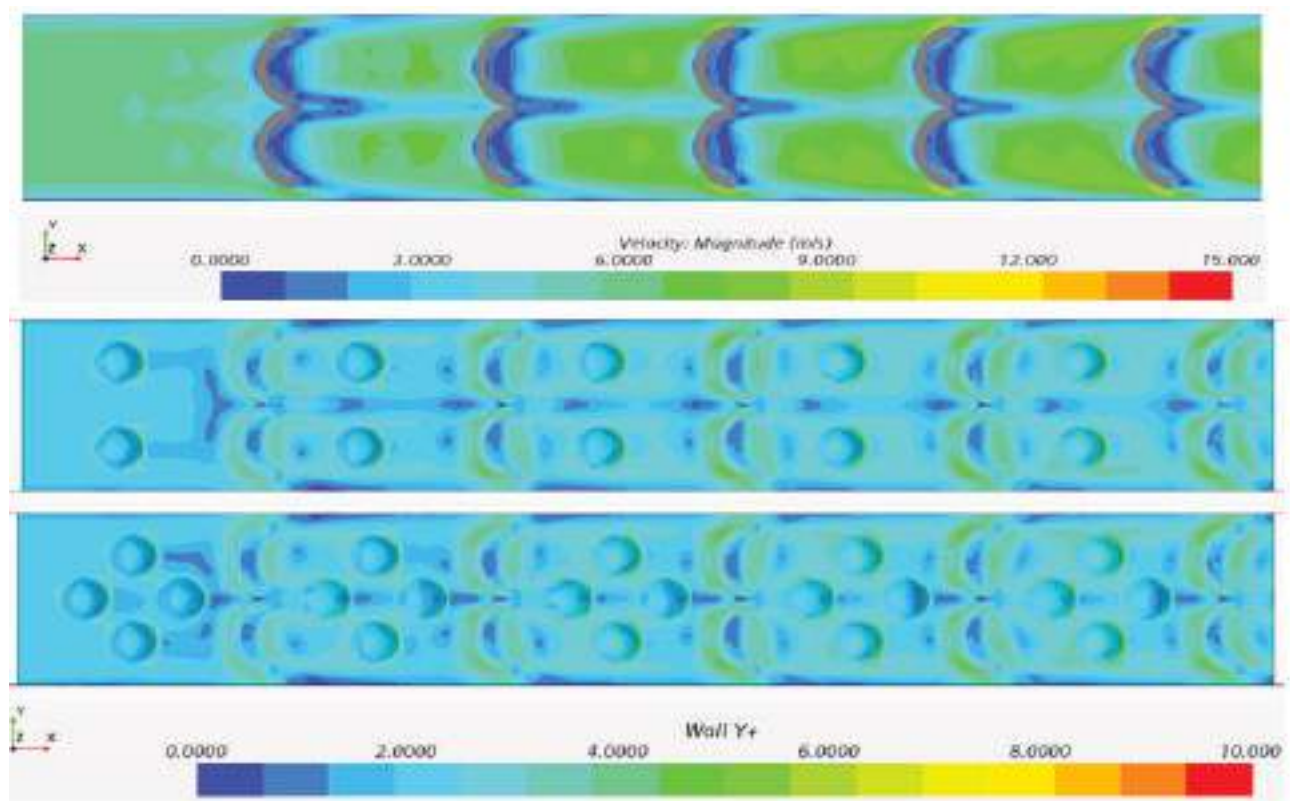


Figure 11. Velocity contours of multi semi rib and compound channel.

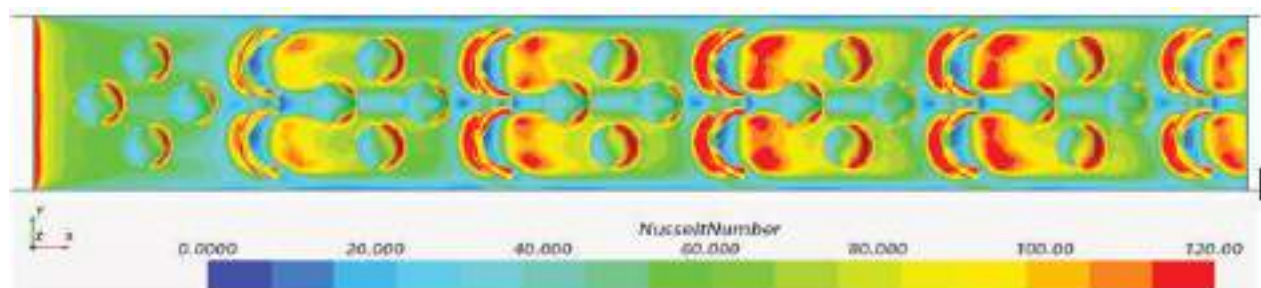


Figure 12. Nusselt number contours of multi semicircular rib with 20 dimples.

Optimization Technique

To select best configuration for thermal performance, applied multiple criteria decision-making (MCDM) techniques. In the engineering problem, to decide the optimum parameters from many variables; various methods are used such as: response surface method (RSM), TOPSIS, COPRAS, genetic algorithm. Among the MCDM techniques, TOPSIS method was a well-known optimization technique [31].

Technique for order Preference by Similarity to ideal Solution (TOPSIS) was one of the numerical methods among the multiple-criteria decision making techniques. It is suitable practical method with a simple mathematical model, also various benefits of TOPSIS methods such as

rationality, simplicity, good computational efficiency, comprehensibility and capability to check the relative performance of each alternative in a simple mathematical form. Total seven steps in TOPSIS method and the flow chart of the optimization process is shown in Figure 15.

Outcomes of TOPSIS method are shown in Table 2, indicates that 20 dimple compound channel performed best than 10 dimples compound channel, multi semicircular rib and dimple alone channel. From result table concluded that optimization results are near about same as experimental results, found less than 10% difference.

Experimental result of multi semicircular rib and compound channel are validated using CFD and TOPSIS method, found same trend of results, also seen less

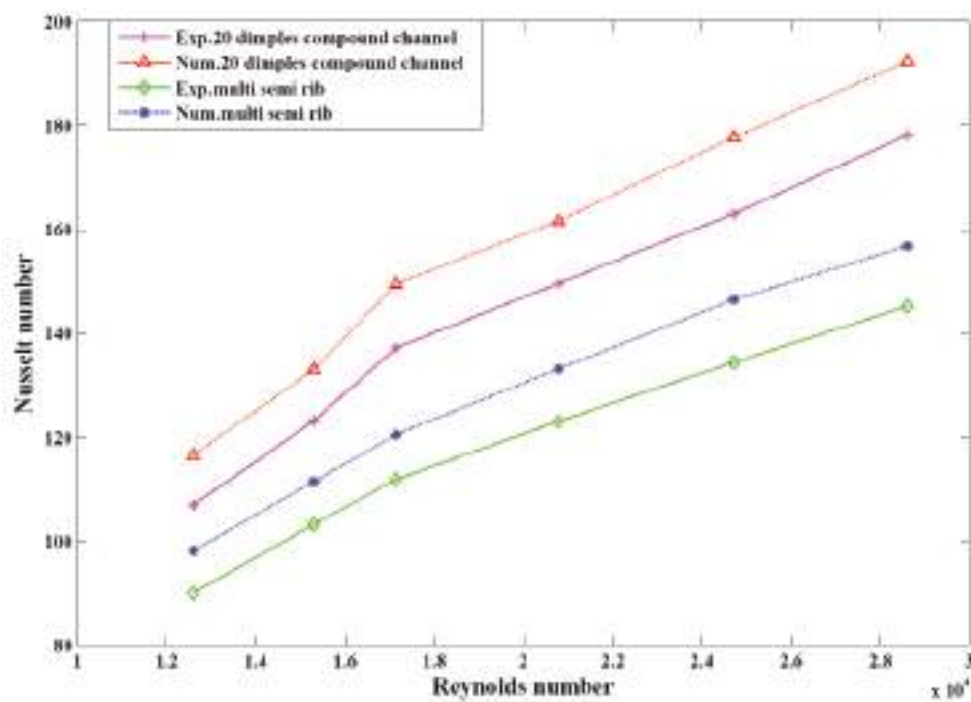


Figure 13. Variation in experimental and numerical result of Nusselt number.

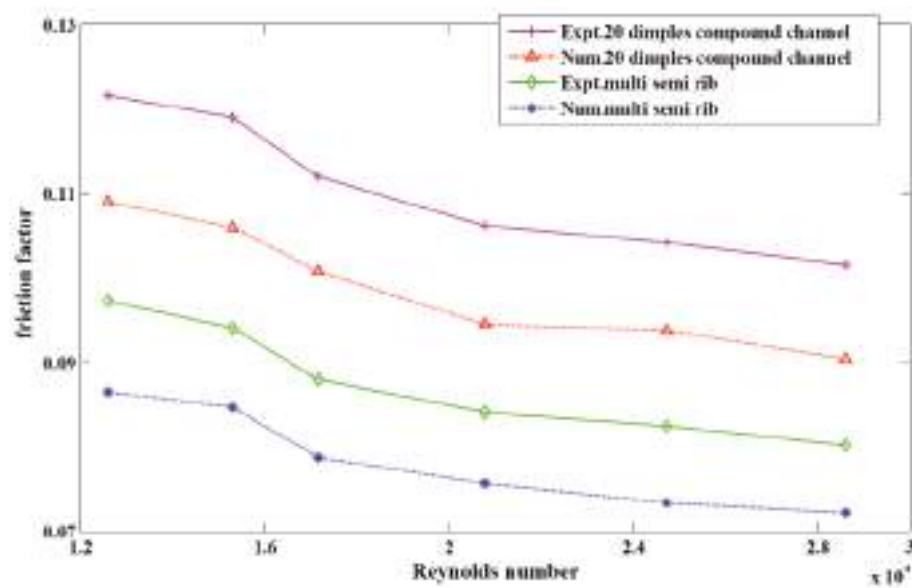


Figure 14. Variation in experimental and numerical result of friction factor.

difference between these results, so said that experimental results are validated [32].

Thermal Performance of Configurations

Thermal performance was determined by taking account the result of heat transfer and frictional losses. Figure 16 shows the thermal performance of tested structures, the compound channel with 20 dimples performed

Table 2. Outcomes of TOPSIS method

RANK		Ci
20 dimples compound channel	1	0.7431
10 dimples compound channel	2	0.5619
Multi semicircular Rib	3	0.4813
Dimple channel	4	0.3270

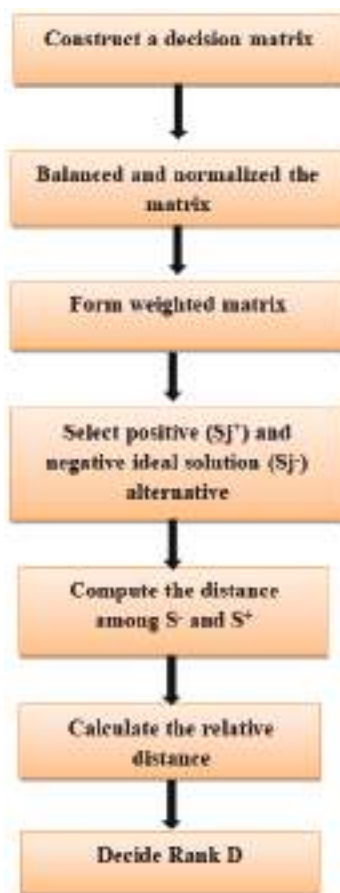


Figure 15. Optimization process flow chart (TOPSIS method).

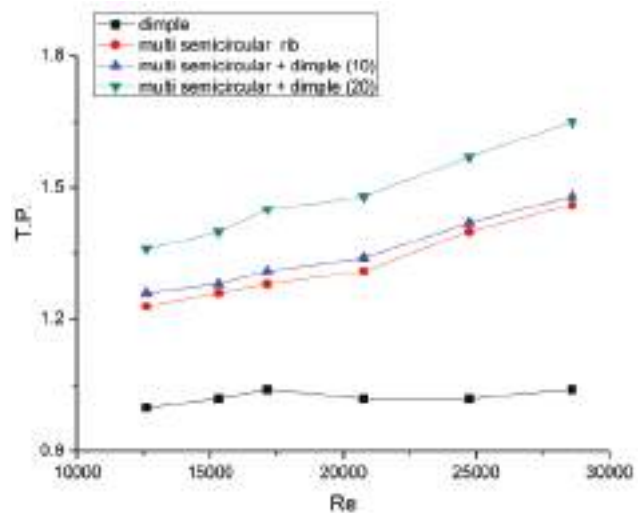


Figure 16. Thermal Hydraulic Performance of different configurations.

better than the multi semicircular rib and 10 dimples compound channel. It shows additional heat transfer enhancement than the frictional losses compared to the all tested configurations. The compound channel with 10 dimples have less heat transfer over multi semicircular rib but due to a small friction it performed better than the rib channel, likewise multi semicircular rib indicates lowest performance because of more friction.

Validation of CFD Results with Experimental Data (Nusselt Number vs Reynolds Number)

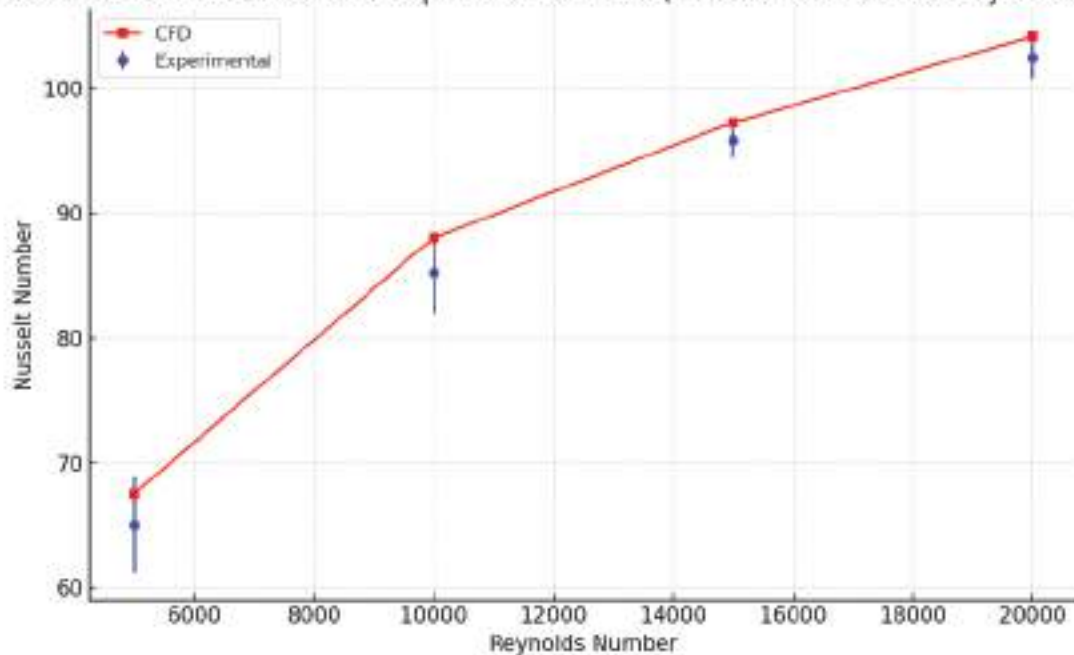


Figure 17. Error bars representing the percentage error between the CFD and experimental results.

The error bar plot comparing the experimental and CFD results for the Nusselt number as a function of Reynolds number. The plot includes error bars representing the percentage error between the CFD and experimental results as shown in Figure 17.

CONCLUSION

The study of heat transfer and flow friction characteristics in a rectangular channel with a multi semicircular rib and compound channel, using experimental methods, simulations, and optimization techniques, revealed that a compound channel with 20 dimples significantly enhanced heat transfer by 38% and 10% compared to plain and dimpled channels, respectively, but also increased friction by 29% and 16%. The 20-dimple compound channel outperformed the multi semicircular rib and 10-dimple compound channels, achieving up to 25% more heat transfer and 30% higher friction. While the multi semicircular rib offered slightly better heat transfer than the 10-dimple compound channel, the latter performed better overall due to lower friction. Additionally, increasing the number of dimples in the compound channel improved heat transfer but also increased 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] Han JC, Zhang YM. High performance heat transfer ducts with parallel broken and V-shaped broken ribs. *Int J Heat Mass Transf* 1992;35:513–523. [\[CrossRef\]](#)
- [2] Park J, Park S, Ligrani PM. Numerical predictions of detailed flow structural characteristics in a channel with angled rib turbulators. *J Mech Sci Technol* 2015;29:4981–4991. [\[CrossRef\]](#)
- [3] Patil PS, Borse SL. Recent studies in internal cooling of gas turbine blade: a review. *Int J Appl Eng Res* 2018;13:7131–7141.
- [4] Baggetta L, Satta F, Tanda G. A possible strategy for the performance enhancement of turbine blade internal cooling with inclined ribs. *Heat Transf Eng* 2019;40:184–192. [\[CrossRef\]](#)
- [5] Abraham S, Vedula RP. Heat transfer and pressure drop measurements in a square cross-section converging channel with V and W rib turbulators. *Exp Therm Fluid Sci* 2016;70:208–219. [\[CrossRef\]](#)
- [6] Promvong P, Thianpong C. Thermal performance assessment of turbulent channel flows over different shaped ribs. *Int Commun Heat Mass Transf* 2008;35:1327–1333. [\[CrossRef\]](#)
- [7] Seo JW, Afzal A, Kim KY. Efficient multi-objective optimization of a boot-shaped rib in a cooling channel. *Int J Therm Sci* 2016;106:122–133. [\[CrossRef\]](#)
- [8] Kim KM, Lee H, Kim BS, Shin S, Lee DH, Cho HH. Optimal design of angled rib turbulators in a cooling channel. *Heat Mass Transf* 2009;45:1617–1625. [\[CrossRef\]](#)
- [9] Rao Y, Feng Y, Li B, Weigand B. Experimental and numerical study of heat transfer and flow friction in channels with dimples of different shapes. *J Heat Transf* 2015;137:031901. [\[CrossRef\]](#)
- [10] Kanargi OB, Lee PS, Yap C. A numerical and experimental investigation of heat transfer and fluid flow characteristics of a cross-connected alternating converging-diverging channel heat sink. *Int J Heat Mass Transf* 2017;106:449–464. [\[CrossRef\]](#)
- [11] Nazari S, Zamani M, Moshizi SA. Comparative study on the influence of depth, number and arrangement of dimples on the flow and heat transfer characteristics at turbulent flow regimes. *Heat Mass Transf* 2018;54:2743–2760. [\[CrossRef\]](#)
- [12] Gupta A, Kumar M, Patil AK. Enhanced heat transfer in plate fin heat sink with dimples and protrusions. *Heat Mass Transf* 2019;55:2247–2260. [\[CrossRef\]](#)
- [13] Sahu MK, Pandey KM, Chatterjee S. Thermo-hydraulic performance of rectangular channel roughened with combined semi-circular and triangular ribs. *Heat Mass Transf* 2019;55:2889–2900. [\[CrossRef\]](#)
- [14] Promvong P, Skullong S. Heat transfer in solar receiver heat exchanger with combined punched-V-ribs and chamfer-V-grooves. *Int J Heat Mass Transf* 2019;143:118486. [\[CrossRef\]](#)
- [15] Jing Q, Xie Y, Zhang D. Effects of channel outlet configuration and dimple/protrusion arrangement on the blade trailing edge cooling performance. *Appl Sci* 2019;9:2900. [\[CrossRef\]](#)



Research Article

Evaluation of the bending fatigue behavior crack propagation in ultra-high-strength steel's resistance spot welding joints, considering the mechanical properties

Mustafa GOKTAS¹, Bilge DEMİR¹, Muhammed ELITAS^{2,*}

¹Department of Mechanical Engineering, Karabuk University, Karabuk, Türkiye

²Department of Mechanical Engineering, Bilecik Şeyh Edebali University, Bilecik, Türkiye

ARTICLE INFO

Article history

Received: 18 June 2024

Revised: 14 August 2024

Accepted: 10 October 2024

Keywords:

Bending Fatigue; Dual-Phase Steel; Mechanical Properties; Numerical Analysis; Resistance Spot Welding

ABSTRACT

Resistance spot welded joints of dual-phase steels are subjected to cyclic loads in the automotive industry. So, fatigue behavior in the joints is an important issue to consider. This study focused on designing a new fatigue model suitable for real conditions not addressed in existing literature and optimizing welding parameters for resistance spot welded DP1000 steel. Experimental and numerical analyses were used to evaluate the fatigue and mechanical performance of the resistance spot welded joints of DP1000 steel. Resistance spot welding applications were performed in different weld currents and constant weld time parameters. The joint samples were subjected to optical and scanning electron microscope image analysis and mechanical testing for tensile shear, hardness, and fatigue. The numerical analysis also simulated nugget formation, tensile shear, fatigue stress formation, and fatigue tests. In addition, fracture patterns of fatigue samples were defined by scanning electron microscope and stereo microscope. Also, the fatigue test data were evaluated and modeled as an interpretation of crack initiation and propagation. As a result, weld metal hardness values increased by approximately 100-200 HV compared to the base material at different welding currents. The best load-bearing capacity (16.607 kN) and fatigue life performance (900,000 cycles) were obtained at 8 and 9 kA, respectively. Newly, the fracture occurred at higher cycles in the current study than in the literature. A neutral axis line was formed on the fracture surface due to bending, and it was very close to the outer surface of the sheet. Chevron traces indicating brittle fracture were observed. In addition, experimental and numeric analysis studies were compatible. Fatigue behavior is an important factor in the automotive industry. Current studies show that fatigue behavior increases joint life.

Cite this article as: Goktas M, Demir B, Elitas M. Evaluation of the bending fatigue behavior crack propagation in ultra-high-strength steel's resistance spot welding joints, considering the mechanical properties. Sigma J Eng Nat Sci 2025;43(4):1520–1532.

*Corresponding author.

*E-mail address: muhammed.elitas@bilecik.edu.tr

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



INTRODUCTION

Advanced high-strength steel (AHSS) is an auto body-in-white material that reduces vehicle weights by 25%. For a decrease of 1% in car weight, the fuel consumption amount can be reduced by 0.66% in regular performance [1,2]. The best-known type of AHSS is dual-phase (DP) steel. DP steels have high strength, good ductility, continuous yielding behavior, good corrosion resistance, and satisfying press performance [3,4]. As general AHSS, DP sheet steels are also produced by a thermal process [5]. Therefore, they are susceptible to repeated exposure to heat. However, this situation is inevitable for DP steels due to heat formations in the welding process, particularly Resistance Spot Welding (RSW) in the automotive manufacturing system [6]. Consequently, it is important to understand and anticipate the thermal process and its effects well to ensure the correct product performance [2,7].

Research showed that every welding process might show internal defects, cracks, porosity, residual stresses [4,8–11], etc., and metallurgical changes that negatively affect the fatigue life, strength, and other properties [12,13]. As these defects and residual stresses are exposed to repetitive loads at service conditions, extra stress concentrations may occur at these points. The stress concentration leads to an increase in applied stress. So, fatigue cracks appear, and with the propagation of the cracks through all parts, rupture occurs in the structural parts of the automobile [4,14,15]. Kishore et al. [16] stated that unlike tensile-shear fracture, which started at the location of minimum hardness, fatigue crack started at the location of maximum stress concentration, which experiences both metallurgical notches because of a sharp gradient of hardness and mechanical notch from the inherent geometry of the spot welds. Researchers [17,18] generally have been interested in the fatigue performance of RSW joints under different modes of loads but rarely mixed-mode loading. Akbulut [17] conducted a numerical investigation to observe the effects of plate thickness and nugget diameter on the fatigue life of tensile shear specimens. He used both Morrow's mean stress and Coffin-Manson approaches in the analysis and verified the performance of both methods using experimental data available in the literature. He also determined the elastic and plastic strains required for the finite element calculations and obtained a variation in fatigue life predictions as a function of selected geometry parameters. Several researchers have reported that the T-shaped structural parts solution specified for automobile RSW couplings is accurate. As reported by Samadi et al. [19], Lee et al. [20], Ouyang et al. [21], Lv et al. [22], and Kim et al. [23], joints of T-shaped structural parts can be subjected to mixed mode loads, especially in automobiles. Samadi et al. [19] evaluated the residual stress and fatigue life of T-joints welded with a multi-pass weld in different directions. They found that the compressive residual stress in the specimen gradually increased from single-pass to

double-pass and triple-pass welds, and the fatigue life of the specimen also gradually improved as the number of passes increased. They also found that multi-pass welding in different directions affected the residual stress and fatigue life of the material. Lee et al. [20] proposed the 3D FEM fatigue analysis method in their study. They carried out three-dimensional non-steady heat conduction analysis and thermoelastic-plastic analysis to simulate welded pipe members and clarified the overall fatigue behaviour of T-type steel pipe members under high cycle loading. They reported that the proposed 3D fatigue FEM analysis method was useful as a numerical tool for calculating the remaining fatigue life with crack. Ouyang et al. [21] investigated the static three-point bending properties and cyclic bending fatigue performance of three-dimensional five-directional braided T-beam composite (3D5DBTC) at room temperature and recorded the load-displacement hysteresis loop curves and stiffness degradation curves to reveal the relationship between stiffness degradation and damage evolution. They reported that there were three distinct stages corresponding to matrix cracking, interface debonding and fibre breakage, respectively, throughout the fatigue loading, and that the matrix cracking and resin-yarn interface debonding occurred at the flange, while the fibre breakage occurred in the web. Lv et al. [22] provided a method for simulating high-cycle crack propagation in T-tube joints and compared the results of the simulations with experimental data. They found that the surface stress calculated from the tubular joint models using the coordinate mapping method was close to the experimental data and validated the simulation method by comparing the crack propagation rate and crack growth process between the simulation and experimental results. Kim et al. [23] investigated fatigue analysis methods to account for changes in local stress distribution and degradation of the transfer function due to fatigue damage in a multi-point spot-welded structure in the time and frequency domains. They reported that the local stress and fatigue life change dramatically with the accumulation of fatigue damage and that vibration fatigue analysis, which considers changes in the response of the spot-welded structure due to fatigue damage, is the most reasonable approach under irregular vibration loading.

As vehicles are subjected to cyclic loading conditions, it is necessary to evaluate the performance of the combinations in fatigue conditions. As with all areas of weakness, the problem of auto body fatigue must be solved because it results in brittle fracture. Therefore, the performance of auto bodies joined by welding under dynamic conditions is particularly crucial and requires detailed research. [16,24,25]. This text discusses some important literature based on work conducted on the fatigue properties of RSW-welded dual-phase steels. Soomro et al. [15] investigated the effects of double pulse welding on the fatigue properties of RSW DP590 steel. They found that samples produced using the double pulse RSW method could withstand more cycles (approximately 10-23%) than samples produced using the

single pulse RSW method. Kishore et al. [16] examined the fatigue behavior of RSW-welded DP600 and interstitial free (IF) steel. They found that the fatigue sample failed in the HAZ of the IF steel with transgranular striations on the fracture surface. Banerjee et al. [18] researched the fatigue property of RSW DP590 steel concerning nugget size, microstructures, and notch geometry. They reported that fatigue life under low load was better in joints with smaller nuggets than in larger nuggets. Ordóñez et al. [24] studied the effect of overloading on the fatigue strength of RSW joints made of DP980 steel. They found that the stress concentration factor caused by the spot-welding geometry decreased the fatigue life. Fujimoto et al. [25] studied the effect of shot peening on the fatigue strength of RSW joints of 980 MPa steel sheets and found that shot peening after RSW improved the fatigue strength of the joints (the fatigue limit of shot-peened RSW joints increased to approximately twice that of non-peened types) and that crack initiation and propagation were delayed in the region where compressive residual stress was not imparted by shot peening. Janardhan et al. [26] studied how work hardening affects the fatigue behavior of RSW DP600 steel. They concluded that prestraining DP600 steel sheets does not significantly impact the fatigue behavior of spot welds, and combinations fail in the interfacial region of the HAZ. Thierry et al. [27] performed a fatigue test using Lap-shear and T-peel samples representing a wide range of steel strengths joined by RSW. They found that the crack occurred for high loading outside the HAZ, starting in the base material. Ghanbari et al. [28] investigated the fatigue behaviour of RSW in ferrite-martensite DP steel joints with one and three spot welds and hybrid joints using the experimental parameters of welding time, electric current intensity and compressive force. The experimental design was carried out using the Taguchi statistical method. They found that the effect of electric current on the fatigue life of the joints was greater than other parameters, but its effectiveness decreased as it increased, and the adhesive on the hybrid joints also had a significant effect on increasing their fatigue life and fatigue strength. Xie et al. [29] researched the fracture characteristics, macrostructure, microstructure, microhardness distribution and fatigue morphology of DP780 RSW joints. They found that the applied load, the microhardness gradient in the HAZ and the tilt angle on the fracture surface caused by macrodeformation influenced the low-cycle fatigue life. They also found that the applied load was the controlling factor in the low-cycle fatigue test.

Studies on RSW junctions of dual-phase steel in the open-access literature have generally focused on rotational fatigue testing assisted by tensile-compression fatigue tests. However, RSW welded specimens are subject to more than just tensile compression or rotational fatigue. They can also undergo plain-bending fatigue. Unlike previous studies, this study aimed to explain the plane-bending fatigue behavior of DP1000 automotive sheet steel with RSW. For the first time in the literature, a fatigue model design suitable for real conditions has been studied in this research. In addition, determining the optimum spot welding parameters of DP1000 steel sheets with minimum welding problems is of practical importance. This study was organized experimentally and theoretically on RSW samples of DP1000 sheet steel joints with tensile shear and hardness tests. As a result of the T-shaped part's investigations, the stress values occurring on the piece were examined and simplified to an ideal experimental model with similar boundary conditions. Although there are studies on the fatigue properties of RSW steels, no study has been found on fatigue fracture modeling. This study covers novelties, such as explaining fatigue issues by constructing new graphic modeling. Moreover, the fatigue fracture model has been evaluated in detail. Data on the fatigue life of the combinations is necessary for modern and reliable design. Results could be used to improve the design of automotive components subjected to fatigue loading.

MATERIALS AND METHODS

The development of DP1000 steel aims to apply it to structural parts that protect the cabin when the vehicle crashes and satisfy the needs of low carbon equivalents for RSW for heavy gauges up to 2 mm in the platform [13]. DP1000 is AHSS. It is widely used in the automotive industry because of its high strength, energy absorption during impact, good formability, and weldability. The DP1000 is used in vehicle body, safety parts, and chassis [30]. 230VAC heat input, 50Hz frequency, and 2500VA power capacity from the SPECTROLAB model LAVFA18A were used to determine the chemical composition of the DP1000. The chemical composition and microstructure image of commercial DP1000 sheet steel used in the experimental studies is given in Table 1. The DP1000 consists of martensitic and ferritic phases with ultimate tensile strength of 980 MPa and yield strength of 720 MPa [30].

Table 1. DP1000 sheet steel chemical composition (wt. %) with its microstructure

Material DP 1000	C	Si	Mn	P	S	Cr	Ni
	0.136	0.203	1.57	0.021	0.003	0.022	0.039
	Al	Ti	V	Nb	Cu	Fe	Co
	0.044	0.001	0.009	0.021	0.01	97.897	0.021

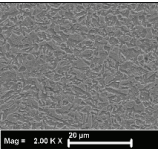




Figure 1. The overlapped RSW joint (EN ISO 14273 standards - tensile shear and Fatigue test).

Table 2. Physical and chemical properties of the RSW electrode used

Alloy	Chemical Composition Weight %	Electrical Conductivity m/Ωmm ²	Thermal Conductivity (W/mK)	Tensile Stress (MPa)	Hardness (HV 30)
CuCrZr	Cr: %0.7-0.12 Zr: %0.06-%0.15 Cu > 95	> 43	320	490	> 160

Samples were prepared 100x30x1.2 mm according to EN ISO 14273 standards[31]. Two parts of the samples were overlapped and then welded within ISO 5821 standards (Fig. 1). F-type 6 mm radius electrodes were used. The physical and chemical properties of the RSW electrode used are given in Table 2. The RSW process was applied at a constant 4 kN-3.3 bar electrode pressure for five different welding currents (6-7-8-9-10 kA) and 12 cycles of welding time (1 cycle=0.02 s). Tensile shear tests were conducted on the SHIMADZU AGS-X Universal test machine with a 100 kN loading capacity using 2 mm.min⁻¹ crosshead speed. Three samples were tested for each welding parameter to measure the average tensile shear force value.

For image analysis and hardness tests, the cross-sections of the RSW specimens were prepared by following standard metallographic procedures (grinding-1200 and

polishing-1μ paste-fabric, then etching using 2% nital solution). Micro-macro structure image analysis of the samples was accomplished using optical, stereo, and scanning electron (SEM) microscopes. Microhardness measurements were performed using a Qness Vickers hardness tester with HV 0.3 (Fig. 2).

Experimental force and stress-cycle curves were analyzed and obtained by finite element analysis (FEA). JmatPro program was used for the material model. Fatigue properties were obtained by utilizing the chemical properties of the material using the JmatPro program. Stress analysis for modeling was performed and verified with ANSYS and Simufact finite element programs. Fatigue analysis was performed with ANSYS finite element software (using the fatigue module). Forces in amplitude values were obtained with ANSYS and Simufact.

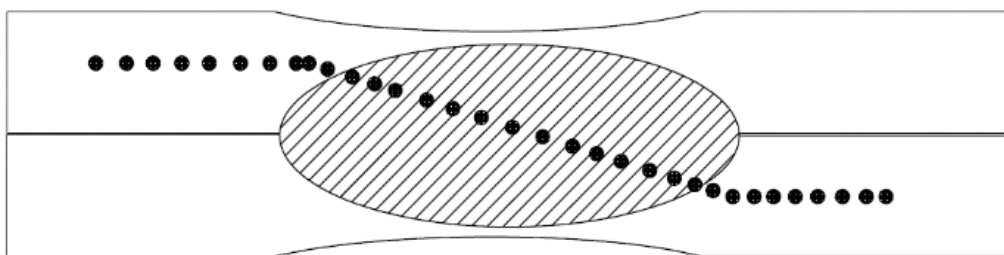


Figure 2. Hardness measurement line on the cross-section of RSW specimens.

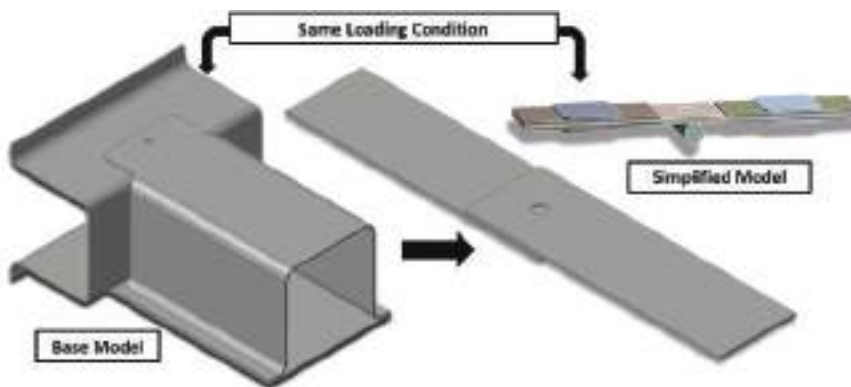


Figure 3. Model of fatigue analysis in ANSYS environment.

First, the model of a T-shaped automobile structure with an RSW joint was subjected to loading in the ANSYS environment for analysis. As a result of the analysis performed on the T-shaped part, the stress values on the part were examined and simplified to an ideal model that could be tested with similar boundary conditions. The created model is shown in Figure 3. A sheet of the same material of the same thickness was placed between the specimen and the jaws to prevent force concentrations in the handles of the fatigue device. The stress and fatigue life values obtained from the analysis of the two models were compared. Then, the model was tested experimentally under the same boundary conditions, and the experimental and

simulation values were compared. The fatigue test machine is shown in Figure 4. DP1000 steel RSW fatigue samples were subjected to dynamic-fatigue loadings at stress ratio (R) ~ -0.5 , frequency (f) = 10 Hz, and seven (7-1 mm) amplitude levels. Three samples were tested for each welding parameter. The arithmetic mean of the results obtained for each parameter was calculated.

RESULTS AND DISCUSSION

Weld Nugget and Hardness

The change of nugget diameters and heat affected zone (HAZ) of RSW samples for the different weld currents (7, 8, 9, and 10 kA) and the nugget images of the samples (7, 8, and 9kA) are given in Figure 5. It was observed that the melting zone-weld metal was not formed sufficiently at the 6 kA welding current. The nugget size becomes larger with increased welding current[13]. Meanwhile, a 1 unit increase in welding current from 7 kA to 9 kA caused an increase of 0.8, 0.13 mm in nugget diameter. However, when the welding current increased from 9 to 10 kA, a 0.26 mm decrease occurred. When the welding current was increased from 7 kA to 8 kA and from 8 kA to 9 kA, the nugget diameter values increased by approximately 15% and 2%, respectively. However, when it was increased from 9 kA to 10 kA, a decrease of roughly 4% was observed. So, the change of welding current between 7 kA and 8 kA caused a higher increase in nugget diameter. That shows that nugget diameter and HAZ thickness increase with the increase of welding current, peak at 9kA welding current, and decrease after this value. These results could be explained by splashing and nugget liquid metal loss due to high weld heat input with high weld current. The nugget diameter increased up to a definite current value. Still, due to the increased welding current and high heat input, the nugget diameter decreased with molten metal splashing and increased indentation amounts [13,32].

The weld microhardness of the RSW samples is given in Figure 6. The hardness values of nuggets are much higher



Figure 4. Schematic view of the flexural fatigue test system specially designed for this study and test machine image.

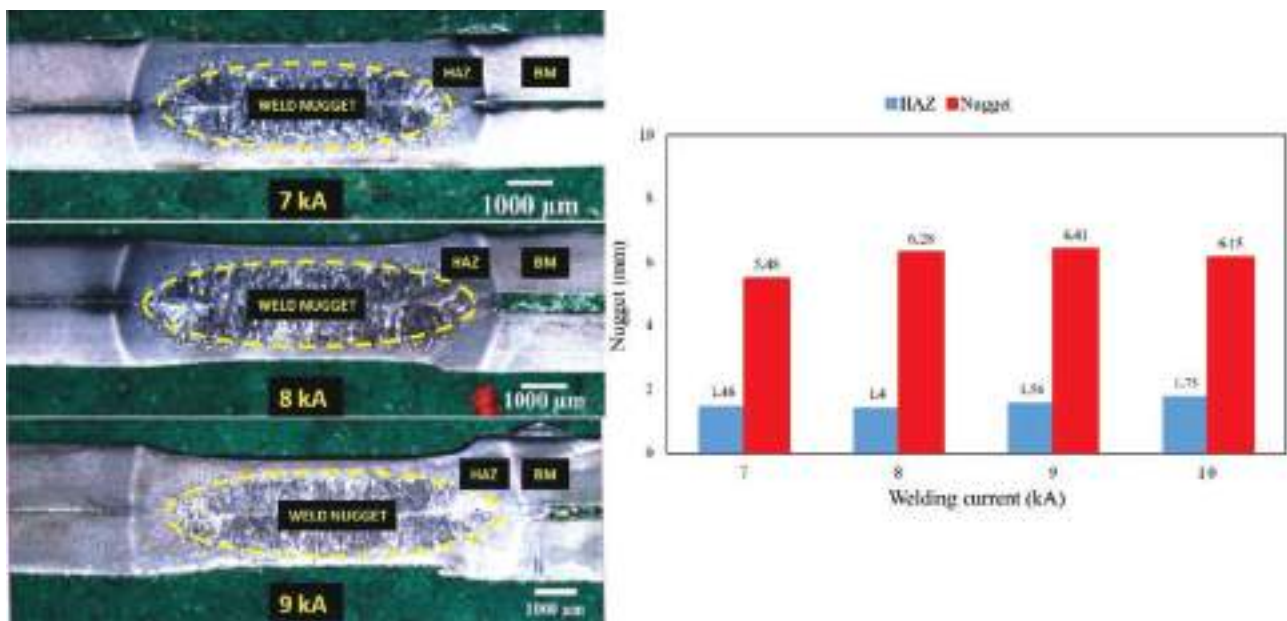


Figure 5. The graphic of HAZ length and weld nugget diameters of RSW samples vs. weld current with their images (7, 8, and 9 kA, respectively).

than the base material. The hardness increases in the weld metal compared to the base material, which is approximately 100-200 HV. However, Pakkanen et al. [30] found 150 HV higher hardness at weld metal than in base material. The higher hardenability of base material could interpret these results depending on alloying elements (Table 1.) The alloying elements such as silicon and manganese can increase the hardenability of weld metal[4].

Moreover, the cooling rate after RSW is considerably higher than the critical cooling rate required for martensite formation [33]. Therefore, the heating and cooling cycles produce

much harder martensite than base material dual-phase steel. Badkoobeh et al. [4] stated that the weld metal of all samples was composed of a lath martensite structure because of the high cooling rate. The base material microstructure of all samples was made up of ferrite and martensite phases. Pakkanen et al. [30] reported similar results in their studies.

Hardness is developed due to the martensite ratio; the temperature gradient between A_1 (the temperature of the austenite-to-pearlite eutectoid transformation) and melting temperature increases from HAZ to the fusion zone. Nikoosohbat et al. [34] explained the significant change in

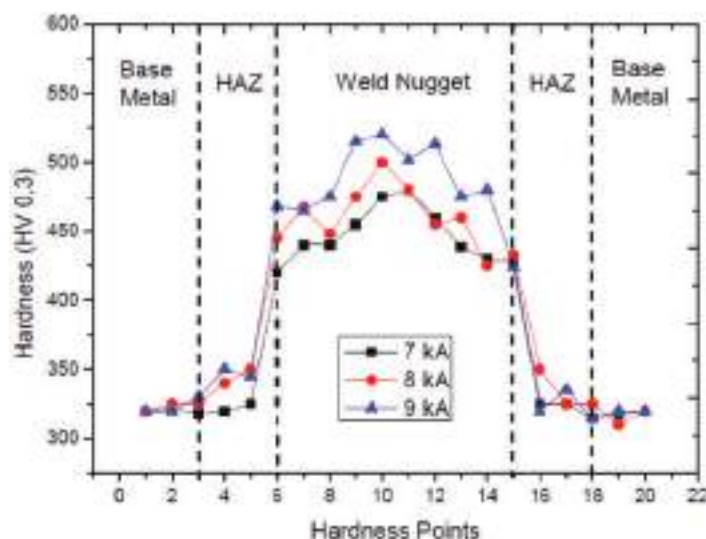


Figure 6. Hardness values of samples welded at different parameters belonging to different welding regions.

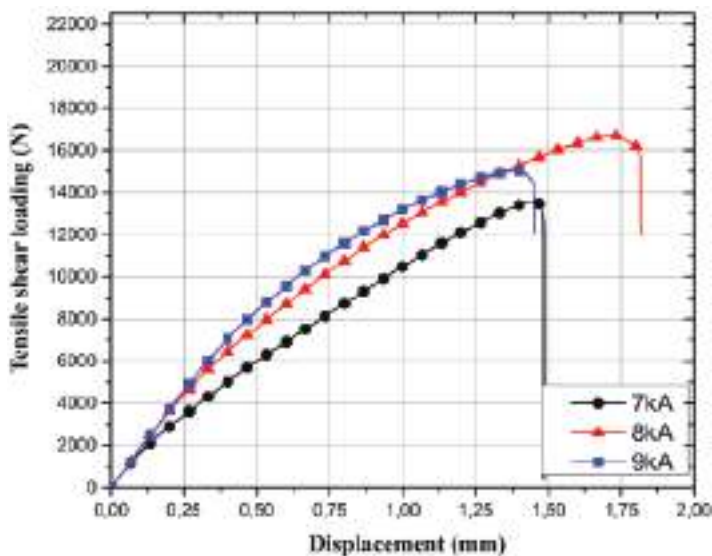


Figure 7. Tensile shear test results of the samples.

hardness values in the joint zones (weld metal, HAZ, and base material). They interpreted severe microstructural changes in the heat-affected areas due to thermal changes in the weld cycle during the RSW process. In the literature, there are more examples for weld metal, HAZ, and weld materials, respectively, such as 500 HV - 300 HV - DP980 [35], 375- 315 HV, and 470-325 HV- DP1000 (low alloyed and high alloyed) [13]. Also, Chabok et al. [13] obtained the weld metal hardness value in the range of 350-450 HV. However, this study found weld metal hardness values in the range of 325-425 HV. The reason for the differences could be interpreted as the differences in the welding parameters and the materials used [13].

Tensile Shear Testing

Figure 7 shows that the highest load-bearing capacity was obtained from 8 kA welding currents. In addition, the load-bearing capacity increases by increasing the welding current, but it peaks at one point and then decreases. An increase in tensile shear force is caused by the nugget becoming larger. By increasing the nugget size, the bonding area in the interface of two sheets is increased, and the strength of the combination improves. So, the tensile shear force increased [4]. Chabok et al.[13] observed a gradual increase in tensile shear force with increased nugget size. In addition, it shows that the load-carrying capacity of the 9kA sample is high, but it shows early failure and breakage. Splashing and this situation may cause some discontinuities.

There are several studies on DP1000 sheet steel in the literature. Elitas and Demir [36] reported that the 3 bar - 7 kA welding current samples showed the highest tensile shear load-bearing capacity. On the other hand, the lowest one was a 4 bar - 5 kA welding current sample. Therefore, as other researchers have pointed out, it is undesirable because splashing can reduce the joint's tensile shear-bearing

capacity and energy absorption. Chabok et al. [13] obtained the maximum tensile shear force values as 23 kN and 15 kN for the high and low carbon contents of DP1000, respectively. At this point, the nugget diameter is also stated as 7 mm. Elitas and Demir [36], who examined weld pressure, reported the peak value as approximately 22 kN, depending on the electrode pressure and welding current value. Badkoobeh et al. [4] found the maximum tensile shear force values between 10.7 kN-17.2 kN for different Si content of DP1000. The maximum tensile shear force values of the samples obtained at 7, 8 and 9 kA welding currents were 13.110 kN, 16.607 kN, and 14.706 kN, respectively. There are some similarities and differences in the results obtained in the literature due to chemical composition, sheet thickness, and RSW welding parameters.

Fatigue Behavior Fracture

The fatigue test results for the RSW weld joints were recorded as force-cycle graphics, where the force is the average force value until the sample breaks. These results were used to calculate fatigue stress and draw S-N diagrams. This study also showed that these force-cycle diagrams could be interpreted to explain the fatigue crack formation. For example, Figure 8 shows the force-cycle graphics of the 2 and 6 mm amplitude of fatigue test results and indicates the model developed to interpret the force-cycle diagram to explain fatigue crack propagations. This model shows in detail that at Stage 1: Fatigue crack initiation and initial growth, Stage 2: Crack growth along with the sheet thickness, and Stage 3: Crack growth across the sheet width (Figure 8c). This figure also shows the fatigue fractured sample images.

The mode of fracture propagation seen in Figure 8c is defined as Type-2 fracture mode in the literature. Samples exhibit partial pull-out type fracture in which some part of the nugget is pulled out from the sheet. During the fatigue test,

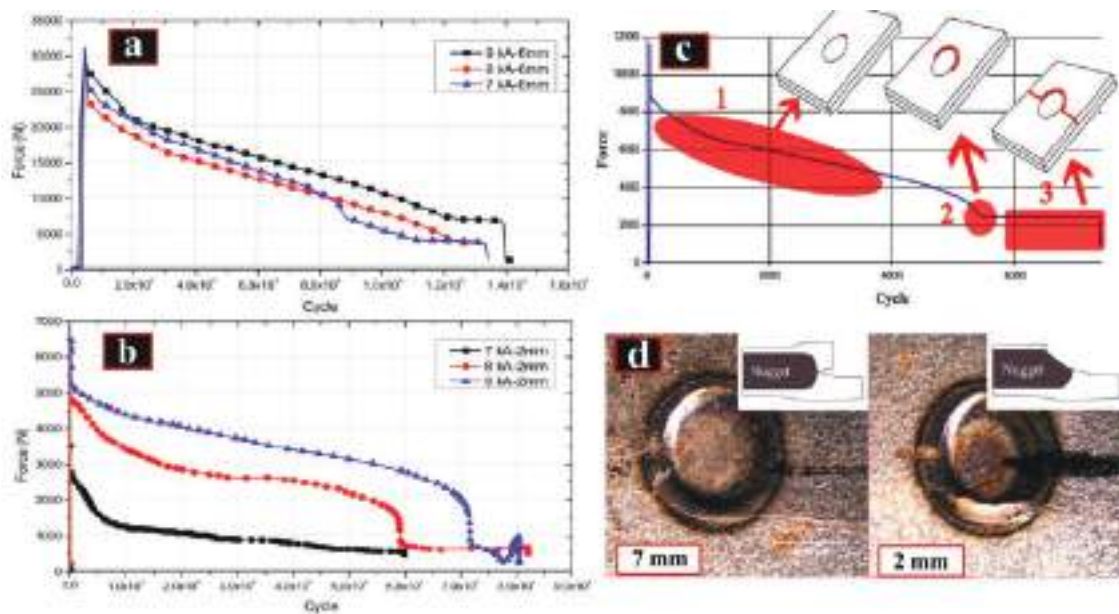


Figure 8. Force-cycle graphs of fatigue tests at (a) 6 mm and (b) 2 mm amplitudes (c) Comparative modeling of the force-cycle graph obtained at 8 kA sample with the breakage stages (2 mm amplitude) and (d) fractured nuggets of the 8kA sample.

the crack started from the root of the notch and then grew a short distance throughout the nugget perimeter, i.e., through HAZ to weld metal. The crack then propagated in the transverse direction throughout the base material sheet width on reaching the surface. The fracture surface shows two different regions: (i) the middle area of the nugget, which shows striations, and (ii) the bottom area, which has a flat surface [15].

The nugget size is significant to the strength of the RSW joints [4,37]. The Simufact FEA analysis of bending

fatigue stress formation in the RSW-joined DP1000 steel samples is given in Figure 9. The model explained above in Figure 8 and the fatigue stress formation analysis results with the Simufact FEA program in Figure 9 are compatible. There is a directly proportional relationship between fatigue life and nugget size [15,24]. As the nugget size increases, the stress concentration factor around the notch root decreases, thus causing the fatigue initiation life to improve in Stage 1. [24].

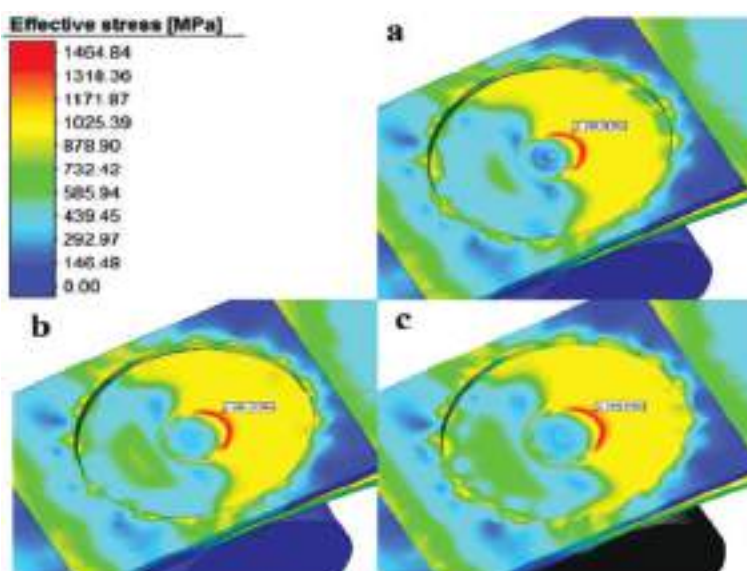


Figure 9. Analysis of fatigue stresses formation on DP1000 steel samples joined with: (a) 7kA; (b) 8kA; (c) 9kA welding current with the Simufact FEA program.

Figure 9 shows nugget interfaces obtained at the Simufact FEA program. As a result, it was observed that the stresses started at the nugget root and branched around the nugget during loading (also feasible in our modeling in Figure 8). As a result of the analysis, it was observed that the highest stress occurred at the 8 kA RSW sample. The maximum forces were obtained at 1143 N for the 8 and 9 kA samples and 1141 N for the current in the 7 kA sample.

The fatigue life of spot welds decreases as the shear component of the loading conditions increases due to an increase in the amplitude value. However, fracture tendency increases in interface mode [15,16]. It was observed that the fatigue life of the RSW DP1000 steel sheet samples decreased as the amplitude values increased. In the fatigue tests, all samples with an amplitude of 1 mm have transcended the fatigue limit of 1×10^6 cycles. Also, samples welded with different welding currents exhibit similar fatigue life in low-cycle fatigue tests. The sheet's strength is more critical than the weld strength in low-cycle fatigue. The 8 and 9 kA samples exhibited close to each other in all cycles. The best life performance was obtained from 9 kA welded samples depending on the nugget diameter size. As a result of the studies, it is seen that there is a directly proportional relationship between the weld diameter and fatigue life. These results could be explained by the increase in the fatigue life by increasing the weld diameter, reducing the stress concentration factor in the weld notch [18].

Fatigue curves created with these data are shown in Figure 10 via force-cycle and stroke (amplitude)-cycle. As can be seen from the graphics, the sample welded with 9 kA showed a slightly higher fatigue strength performance than other samples (7 and 8 kA).

Researchers have investigated the fatigue performance of RSW joints under different modes of loads generally and mixed-mode loading rarely. As reported by Kim et al. [23], joints of T-shaped structural parts can be subjected

to mixed mode loads, especially in automobiles. For parts subjected to mixed mode loads, fatigue damage may occur below the regular cycles. The literature reported that the notch effect might cause this situation due to RSW and the heterogeneous microstructure formed by the weld zones (base material, HAZ, and weld metal) [38]. The samples obtained at 4 kN constant electrode force and 7, 8, and 9 kA welding current values were fractured at approximately 500,000, 700,000, and 900,000 cycles, respectively. However, Ordonez et al. [24] found that failure occurred after approximately 100,000 cycles for RSW DP980 steel. In the current study, fracture was observed to occur at higher cycles. In this context, a contribution to the literature has been made. Additionally, the general evaluation of the results reveals significant developments, particularly for the automotive industry. It appears that most spot weld fractures in vehicles result from cyclic loading in the form of vibrations. [15]. Fatigue behavior is an important factor in the automotive industry. Current studies show that it increases joint life.

After the fatigue tests, the fracture surfaces of the test samples observed in the macro-micro-scale are shown in Figure 11 and Figure 12. The branching striations due to the excess strain are seen in Figure 11a. Figure 11b shows a smoother surface due to the abrasion caused by the low strain and the high number of cycles. The neutral axis line formed due to the bending of the fractured surface is seen. As it moves away from the neutral axis line, tensile or compression stresses increase in the sheet metal section depending on the bending moment. Since the neutral axis is very close to the outer surface of the sheet, the maximum tensile stress is realized at the nugget, as shown at the crack initiation point in Figure 11a. The micro image of the weld joint area is shown in Figure 12. The crack starts in the welding zone and propagates through the sheet, and rupture occurs at the sides with the cycles. The rupture zone can be seen in

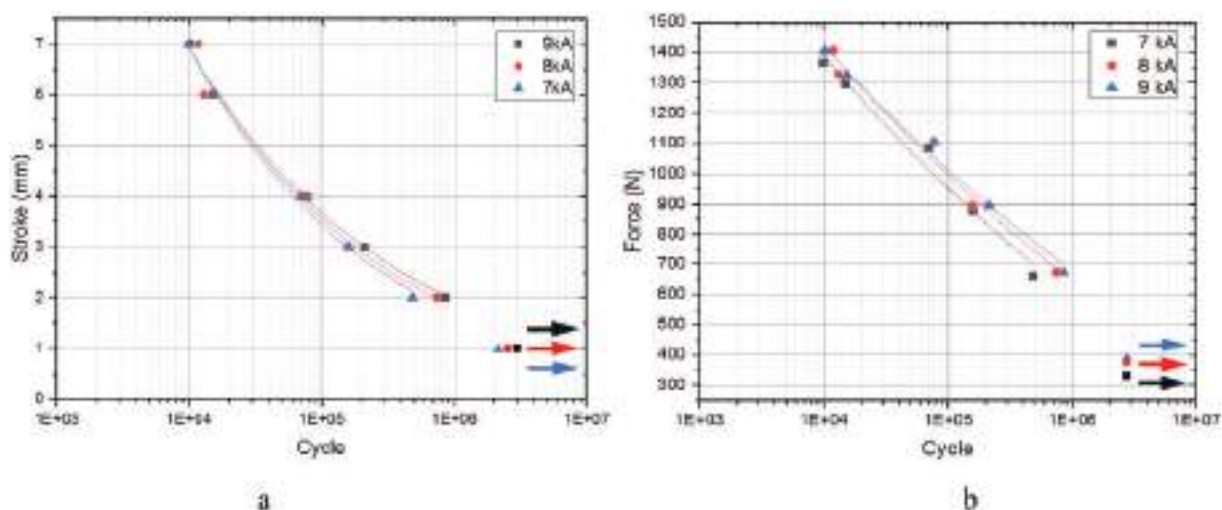


Figure 10. Fatigue test results of samples welded with different welding currents (a) Stroke-Cycle, (b) Force-Cycle.

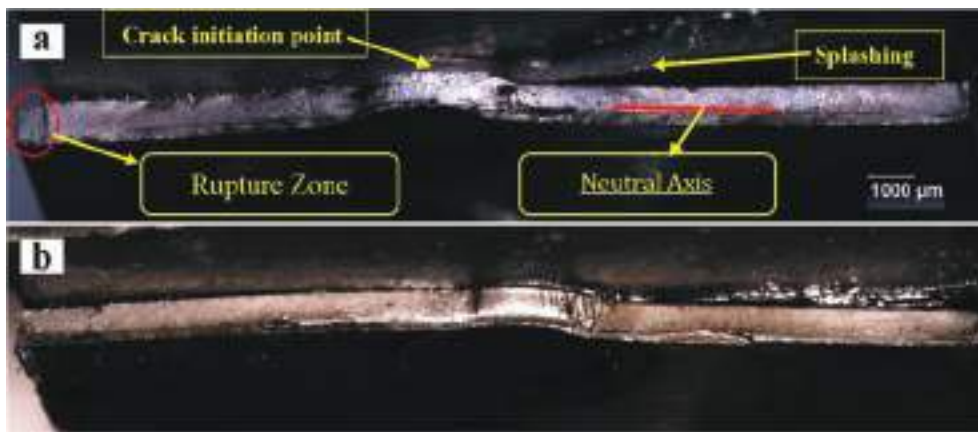


Figure 11. Macro images of damaged surfaces of the 8-kA sample (a) The branching striations, (b) A smoother surface.

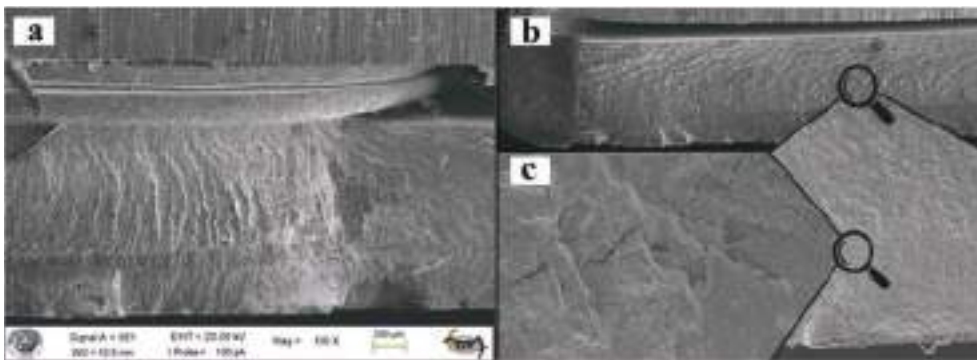


Figure 12. SEM images of damaged surfaces of the 8-kA sample; (a) SEM image of the weld zone on the damaged surface; (b) SEM image of the area where the rupture occurred on the damaged surface; (c) Image of intergranular fractures on the fracture surface of the sample damaged at low amplitude.

Figure 12a. Janardhan et al. [26] reported that crack initiation occurs in the base material under cyclic loading from the HAZ in the case of tensile shear loading. Also, Thierry et al. [27] concluded that the crack occurred for high loading outside the HAZ, starting in the base material. In another study by Thierry et al. [39], fatigue failure occurred in the base material. However, Jamasri et al. [40] found that the fatigue cracks started at the nugget crack due to the excessive sheet separation. Then, the crack propagation occurred through the thickness of the thinner sheets in the HAZ. In experimental studies, it can be said that the exact definition of crack initiation can only be obtained differently depending on the material and weldment parameters [31].

The chevron pattern is an important indicator of brittle fracture. Fractures may progress within the grain boundary or via the cleavage mechanism within the grain. A typical feature of fracture surfaces is the V-shaped chevron pattern. Figure 12 shows chevron traces caused by crack propagation due to cyclic loads. These traces indicate the initial crack location. The crack starts from the weld zone and propagates along sheet metal. Eventually, the rupture

occurs. Especially when considering the distances between the chevron traces, it was observed that the weld zone was at high amplitudes, and the chevron traces in sheet metal were observed. Decreasing the amplitude values reduced the distance between the two cracks (striation). It has been observed that while the chevron trace angles increase in the weld zone, the chevron angles and the base material narrow. It is thought that this is due to the retardation of crack propagation due to the increased hardness in the weld zone. The crack propagation between the grains is expected to cause perpendicular stress on the surface.

Brittle fractures commonly occur in martensitic structures in relatively high-alloyed DP steels. [41]. The weld metal of DP1000 steel primarily consists of the martensite phase, which makes it brittle. Therefore, a brittle fracture mode was observed in the RSW samples.

CONCLUSION

This study focused on designing a new fatigue model suitable for real conditions, which is not addressed in

existing literature, and optimizing welding parameters for resistance spot welded DP1000 steel. The following conclusions can be drawn:

1. As the welding current increased to 9 kA, the nugget diameter increased, but at higher values, the nugget diameter decreased due to splashing and nugget liquid metal loss.
2. The martensite phase, which contributes to the hardness, increased significantly in the weld metal compared to the base material. So, weld metal hardness values increased by approximately 100-200 HV compared to the base material at different welding currents.
3. As the nugget diameter increased, the stress concentration factor decreased. Consequently, it was observed that fatigue life increased.
4. The maximum tensile shear force value (16.607 kN) and the best fatigue life (900,000 cycles) were obtained at 8 and 9 kA weld samples, respectively.
5. It was recently observed that fractures occurred at higher cycle numbers in the current study than in previous literature.
6. The crack starts from the notch's root and grows a short distance along the nugget perimeter. The crack then propagated in the transverse direction along the base material sheet width on reaching the surface.
7. It was observed that the chevron traces caused by the crack propagation developed due to cyclic loads starting from the crack welding zone and propagating along with the sheet metal, and finally, a rupture occurred. High amplitude values make these traces more explicit.
8. Fatigue behavior is an important factor in the automotive industry. Current studies show that fatigue behavior increases joint life. Additionally, the results of this study will establish a foundation for future research on bending fatigue and the optimization of welding parameters.

ACKNOWLEDGMENTS

The authors thank Karabuk University Coordinators of Research Projects for supporting the projects numbered Fyl-2020-2123 and KBÜBAP-17-KP-463, Simufact company, TOFAS AS, and Mr Canderin Önder from Netform Engineering.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data supporting this study's findings are available within the article. Raw data supporting this study's findings are available from the corresponding author upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest concerning this article's research, authorship, and/or publication.

ETHICS

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

REFERENCES

- [1] Goktas M. An investigation of the effect of corrosive environment on the fatigue life of DP1000 steel junctioned with resistance spot welding [Doctorial Thesis]. 2020.
- [2] Rao SS, Arora KS, Sharma L, Chhibber R. Investigations on Mechanical Behaviour and Failure Mechanism of Resistance Spot-Welded DP590 Steel Using Artificial Neural Network. *Trans Indian Inst Met* 2021;74:1–20. [\[CrossRef\]](#)
- [3] Zhao D, Wang Y, Zhang P, Liang D. Modeling and experimental research on resistance spot welded joints for dual-phase steel. *Materials* 2019;12:1108. [\[CrossRef\]](#)
- [4] Badkoobeh F, Nouri A, Hassannejad H, Mostaan H. Microstructure and mechanical properties of resistance spot welded dual-phase steels with various silicon contents. *Mater Sci Eng A* 2020;790:139703. [\[CrossRef\]](#)
- [5] Roodgari MR, Jamaati R, Aval HJ. A new method to produce dual-phase steel. *Mater Sci Eng A* 2021;803:140695. [\[CrossRef\]](#)
- [6] Rajarajan C, Sivaraj P, Balasubramanian V. Role of welding current on mechanical properties and microstructural characteristics of resistance spot welded dual phase steel joints. *Phys Met Metallogr* 2020;121:1447–1454. [\[CrossRef\]](#)
- [7] Elitas M. Effects of welding parameters on tensile properties and failure modes of resistance spot welded DC01 steel. *Proc Inst Mech Eng Part E J Process Mech Eng* 2023;237:1607–1616. [\[CrossRef\]](#)
- [8] Elitas M, Demir B. Residual stress evaluation during RSW of DP600 sheet steel. *Mater Test* 2020;62:888–890. [\[CrossRef\]](#)
- [9] Kasman Ş, Ozan S. Characterization of friction stir welded AA 3003-H24 aluminum alloy plates. *Sigma J Eng Nat Sci* 2022;40:620–629. [\[CrossRef\]](#)
- [10] VVS KR. Investigation on effect of tig welding parameters on dissimilar weld joints of aisi 304 and aisi 310 steels using response surface method. *Sigma J Eng Nat Sci* 2021;39:80–96.
- [11] Öcalır Ş, Eşme U, Boğa C, Külekci MK. Investigation of mechanical and metallographic properties of two different aluminum alloys joined with friction stir welding method using different welding parameters. *Sigma J Eng Nat Sci* 2020;38:1333–1349.

- [12] Pandya KS, Grolleau V, Roth CC, Mohr D. Fracture response of resistance spot welded dual phase steel sheets: Experiments and modeling. *Int J Mech Sci* 2020;187:105869. [\[CrossRef\]](#)
- [13] Chabok A, van der Aa E, Pei Y. A study on the effect of chemical composition on the microstructural characteristics and mechanical performance of DP1000 resistance spot welds. *Mater Sci Eng A* 2020;788:139501. [\[CrossRef\]](#)
- [14] Reza Kashyzadeh K, Farrahi G, Ahmadi A, Minaei M, Ostad Rahimi M, Barforoushan S. Fatigue life analysis in the residual stress field due to resistance spot welding process considering different sheet thicknesses and dissimilar electrode geometries. *Proc Inst Mech Eng Part J Mater Des Appl* 2023;237:33–51. [\[CrossRef\]](#)
- [15] Soomro IA, Pedapati SR, Awang M, Alam MA. Effects of double pulse welding on microstructure, texture, and fatigue behavior of DP590 steel resistance spot weld. *Int J Adv Manuf Technol* 2023;125:1271–1287. [\[CrossRef\]](#)
- [16] Kishore K, Kumar P, Mukhopadhyay G. Microstructure, tensile and fatigue behaviour of resistance spot welded zinc coated dual phase and interstitial free steel. *Met Mater Int* 2022;28:1–21. [\[CrossRef\]](#)
- [17] Akbulut M. Parametric investigation into fatigue life behaviour of spot welded tensile shear test samples. *Proc Inst Mech Eng Part C J Mech Eng Sci* 2021;235:4119–4128. [\[CrossRef\]](#)
- [18] Banerjee P, Sarkar R, Pal TK, Shome M. Effect of nugget size and notch geometry on the high cycle fatigue performance of resistance spot welded DP590 steel sheets. *J Mater Process Technol* 2016;238:226–243. [\[CrossRef\]](#)
- [19] Samadi F, Mourya J, Wheatley G, Khan MN, Nejad RM, Branco R, et al. An investigation on residual stress and fatigue life assessment of T-shape welded joints. *Eng Fail Anal* 2022;141:106685. [\[CrossRef\]](#)
- [20] Lee J-H, Kim J-S, Kang S-U, Hirohata M, Chang K-H. Fatigue life analysis of steel tube member with T-shaped welded joint by FEM. *Int J Steel Struct* 2017;17:833–841. [\[CrossRef\]](#)
- [21] Ouyang Y, Wang H, Gu B, Sun B. Experimental study on the bending fatigue behaviors of 3D five directional braided T-shaped composites. *J Text Inst* 2018;109:603–613. [\[CrossRef\]](#)
- [22] Lv W, Ding B, Zhang K, Qin T. High-Cycle Fatigue Crack Growth in T-Shaped Tubular Joints Based on Extended Finite Element Method. *Buildings* 2023;13:2722. [\[CrossRef\]](#)
- [23] Kim J-H, Lee D-C, Lee J-H, Ham Y-S, Kang K-W. Fatigue analysis of spot-welded automobile components considering fatigue damage-induced stiffness degradation in time and frequency domains. *Int J Precis Eng Manuf* 2017;18:389–397. [\[CrossRef\]](#)
- [24] Ordoñez JH, Ambriz RR, García C, Plascencia G, Jaramillo D. Overloading effect on the fatigue strength in resistance spot welding joints of a DP980 steel. *Int J Fatigue* 2019;121:163–171. [\[CrossRef\]](#)
- [25] Fujimoto H, Ueda H, Ueki R, Fujii H. Improvement of fatigue properties of resistance spot welded joints in high strength steel sheets by shot blast processing. *ISI Int* 2016;56:1276–1284. [\[CrossRef\]](#)
- [26] Janardhan G, Dutta K, Mukhopadhyay G. Influence of work hardening on tensile and fatigue behavior of resistance spot-welded dual-phase steel. *J Mater Eng Perform* 2023;32:624–637. [\[CrossRef\]](#)
- [27] Thierry D, Vucko F, Luckeneder G, Weber B, Dosdat L, Bschorr T, et al. Fatigue behavior of spot-welded joints in air and under corrosive environments: Part I: Materials, specimen and test results in air. *Weld World* 2016;60:1211–1229. [\[CrossRef\]](#)
- [28] Ghanbari HR, Shariati M, Sanati E, Nejad RM. Effects of spot welded parameters on fatigue behavior of ferrite-martensite dual-phase steel and hybrid joints. *Eng Fail Anal* 2022;134:106079. [\[CrossRef\]](#)
- [29] Xie L, Shi B, Xiao Z, Ren J, Li D. Fatigue Characteristics of DP780 Steel Spot Welding Joints with Different Static Fracture Modes. *Mater Trans* 2021;62:191–197. [\[CrossRef\]](#)
- [30] Pakkanen J, Vallant R, Kičin M. Experimental investigation and numerical simulation of resistance spot welding for residual stress evaluation of DP1000 steel. *Weld World* 2016;60:393–402. [\[CrossRef\]](#)
- [31] Göktaş M, Demir B, Elitas M. An investigation on bending fatigue in a corrosive environment of dual-phase 1000 sheet steel RSW joints and damage model via experiment and numeric analysis. *Proc Inst Mech Eng Part E J Process Mech Eng* 2024;09544089241257899. [\[CrossRef\]](#)
- [32] Kimchi M, Phillips DH. *Resistance Spot Welding Fundamentals and Applications for the Automotive Industry*. Ohio: Morgan & Claypool; 2017. [\[CrossRef\]](#)
- [33] Holtschke N, Jüttner S. Joining lightweight components by short-time resistance spot welding. *Weld World* 2017;61:413–421. [\[CrossRef\]](#)
- [34] Nikoosohbat F, Kheirandish S, Goodarzi M, Pouranvari M, Marashi SPH. Microstructure and failure behaviour of resistance spot welded DP980 dual phase steel. *Mater Sci Technol* 2010;26:738–744. [\[CrossRef\]](#)
- [35] Pouranvari M, Marashi SPH. Key factors influencing mechanical performance of dual phase steel resistance spot welds. *Sci Technol Weld Join* 2010;15:149–155. [\[CrossRef\]](#)
- [36] Elitas M, Demir B. The effects of the welding parameters on tensile properties of RSW junctions of DP1000 sheet steel. *Eng Technol Appl Sci Res* 2018;8:3116–3120. [\[CrossRef\]](#)
- [37] Kishore K, Kumar P, Mukhopadhyay G. Resistance spot weldability of galvanized and bare DP600 steel. *J Mater Process Technol* 2019;271:237–248. [\[CrossRef\]](#)

-
- [38] Holovenko O, Ienco MG, Pastore E, Pinasco MR, Matteis P, Scavino G, et al. Microstructural and mechanical characterization of welded joints on innovative high-strength steels. *Metall Ital* 2013;3:3–12.
- [39] Thierry D, Vucko F, Luckeneder G, Weber B, Dosdat L, Bschorr T, et al. Fatigue behavior of spot-welded joints in air and under corrosive environments: Part II: Fatigue under alternating and combined corrosion and fatigue load. *Weld World* 2016;60:1231–1245. [\[CrossRef\]](#)
- [40] Ilman MN, Soekrisno R. Corrosion fatigue behavior of resistance spot welded dissimilar metal welds between carbon steel and austenitic stainless steel with different thickness. *Procedia Eng* 2011;10:649–654. [\[CrossRef\]](#)
- [41] Pouranvari M, Marashi SPH. Critical review of automotive steels spot welding: process, structure and properties. *Sci Technol Weld Join* 2013;18:361–403. [\[CrossRef\]](#)

- [16] Patil PS, Dhande KK, Borse SL. Experimental investigation of heat transfer and pressure drop using combination of ribs and dimples. *Aust J Mech Eng* 2023;21:628–640. [\[CrossRef\]](#)
- [17] Xie Y, Shen Z, Zhang D, Ligrani P. Numerical analysis of flow structure and heat transfer characteristics in dimpled channels with secondary protrusions. *J Heat Transf* 2016;138:031901. [\[CrossRef\]](#)
- [18] Xie Y, Shi D, Shen Z. Experimental and numerical investigation of heat transfer and friction performance for turbine blade tip cap with combined pin-fin-dimple/protrusion structure. *Int J Heat Mass Transf* 2017;104:1120–1134. [\[CrossRef\]](#)
- [19] Rao Y, Zhang P. Experimental study of heat transfer and pressure loss in channels with miniature V rib-dimple hybrid structure. *Heat Transf Eng* 2020;41:1431–1441. [\[CrossRef\]](#)
- [20] Kaur I, Singh P. Heat and flow characteristics of V-shaped protrusion/concavity combined with miniature V-ribs. *Numer Heat Transf A Appl* 2020;78:359–377. [\[CrossRef\]](#)
- [21] Ferreira SB, Pilidis P. Comparison of externally fired and internal combustion gas turbines using biomass fuel. *J Energy Resour Technol* 2001;123:291–296. [\[CrossRef\]](#)
- [22] MacIsaac B, Langton R, Belobaba P, Cooper J, Seabridge A. *Gas turbine propulsion systems*. Chichester: John Wiley & Sons; 2011. [\[CrossRef\]](#)
- [23] Rajawat RS, Singh KK, Mishra VN. Approximation by modified Bernstein polynomials based on real parameters. *Math Found Comput* 2024;7:297–309. [\[CrossRef\]](#)
- [24] Mishra VN, Rajawat RS, Sharma V. On generalized quantum Bernstein polynomials. In: Mishra RK, Patel MK, Kumar SD, editors. *Advances in pure and applied algebra*. Berlin: De Gruyter; 2023. p.149–160. [\[CrossRef\]](#)
- [25] Kline SJ. The purposes of uncertainty analysis. 1985. Available at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d07dee57ac-defe1fe87862cb7f9c74650c7c2f72> Accessed on Aug 26, 2025.
- [26] Raiz M, Rajawat RS, Mishra VN. α -Schurer Durrmeyer operators and their approximation properties. *Ann Univ Craiova Math Comput Sci Ser* 2023;50:189–204. [\[CrossRef\]](#)
- [27] McAdams WH, Drew TB, Bays GS Jr. Heat transfer to falling-water films. *Trans ASME* 1940;62:627–631. [\[CrossRef\]](#)
- [28] Nakayama Y. *Introduction to fluid mechanics*. Oxford: Butterworth-Heinemann; 2018. [\[CrossRef\]](#)
- [29] Pandey S, Rajawat RS, Mishra VN. Approximation properties of modified Jain-Gamma operators preserving linear function. *Palestine J Math* 2023;12:169–182.
- [30] Zastawny M, Lardeau S. Application of Simcenter STAR-CCM+ for analysis of CFD sensitivities in NASA juncture flow simulation. In: *AIAA aviation 2020 forum*. Reston (VA): American Institute of Aeronautics and Astronautics; 2020. [\[CrossRef\]](#)
- [31] Raiz M, Rajawat RS, Mishra LN, Mishra VN. Approximation on bivariate of Durrmeyer operators based on beta function. *J Anal* 2023;32:311–333. [\[CrossRef\]](#)
- [32] Agrawal VK, Patil LN, Chavan KV, Nimbalkar UD. A computational analysis of heat transfer in solid and vented disc brakes: CFD simulation and thermal performance assessment. *Multiscale Multidiscip Model Exp Des* 2024;7:4735–4749. [\[CrossRef\]](#)



Research Article

Magnetohydrodynamic flow and effects of Soret, Dufour and thermophoresis through a chemically reacting porous stretching plate

Ankur KUMAR SARMA^{1,*}, Dipak SARMA¹, Sunmoni MUDOI¹

¹Department of Mathematics, Cotton University, Pan Bazar, 781001, India

ARTICLE INFO

Article history

Received: 19 July 2024

Revised: 21 August 2024

Accepted: 03 November 2024

Keywords:

Magnetohydrodynamic Flow;
Soret and Dufour Effects;
Thermophoresis; Porous
Medium

ABSTRACT

This study investigates the novel sequel of Soret, Dufour, and thermophoresis on a chemically reacting porous stretching plate. With a transverse magnetic field, the fluid model is two-dimensional. Using the similarity transformation, the non-linear governing partial differential equations are converted to solvable ordinary differential equations, which are numerically solved using bvp4c. Graphs depict velocity, temperature, and concentration curves for various parameters. The values of skin friction, Nusselt number, and Sherwood number are shown in tables. The velocity profile decreases with the rise in Magnetic parameter and porosity parameter. The temperature profile escalates with the rise in thermophoretic parameter. While the concentration profile decreases with the rise in thermophoretic parameter. Also, the concentration profile escalates with the rise in Soret number. The Soret effect of a chemically reacting porous stretching plate causes species migration due to temperature gradients, whereas the Dufour effect creates heat due to concentration gradients, affecting temperature and concentration profiles. Thermophoresis moves particles over temperature gradients, impacting species distribution and reaction rates. These factors complicate heat, mass transmission, and chemical processes. In comparison with previous studies, the current results are considered to be quite excellent. Current research applies to industrial appliances, food, cosmetics, and biomedical sectors.

Cite this article as: Kumar Sarma A, Sarma D, Mudoi S. Magnetohydrodynamic flow and effects of Soret, Dufour and thermophoresis through a chemically reacting porous stretching plate. Sigma J Eng Nat Sci 2025;43(4):1533–1546.

INTRODUCTION

Magnetohydrodynamics (MHD) is a field that studies the behavior of electrically conducting fluids in the presence of magnetic fields, such as plasmas, liquids, and some gases. MHD finds diverse applications across industries and scientific fields. It enables the development of innovative

power generation systems, such as MHD generators that convert heat directly into electricity without moving parts, revolutionising energy production from fossil fuels, nuclear reactors, or renewable sources. MHD's influence extends to aerospace propulsion, where it offers potential for propulsion systems in spacecraft and aircraft, leveraging the interaction between magnetic fields and conducting fluids

*Corresponding author.

*E-mail address: mth2291004_ankur@cottonuniversity.ac.in

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



for thrust. Its role in astrophysics is pivotal, aiding in comprehending celestial phenomena like solar flares, magnetospheres, and the dynamics of cosmic plasma. MHD's significance in fusion research involves confining and controlling high-temperature plasmas in fusion reactors through magnetic fields, contributing to advancements in clean energy production. Moreover, its applications span geophysics, industrial processes like metal casting and electromagnetic pumps, as well as fundamental research in plasma physics and fusion endeavours.

The Soret and Dufour effect are two related phenomena in fluid dynamics and heat transfer that describe the behaviour of mass and heat transfer in a fluid when there are gradients in temperature and concentration. These effects are particularly relevant in situations involving non-isothermal (temperature-varying) and non-isobaric (pressure-varying) flows. Swiss scientist Charles Soret named the thermal diffusion effect, also known as the Soret effect, in 1879, while L. Dufour introduced the diffusion thermos effect, also known as the Dufour effect, in 1873. A concentration gradient in a fluid allows molecules or particles to move owing to a temperature gradient, known as thermal diffusion or Soret effect. Comprehending the Soret effect is essential for managing the dispersion of particles and solutes in fluids, refining industrial procedures, and deciphering natural occurrences where temperature gradients impact particle behavior. In systems where accurate adjustment of concentration profiles is necessary, it is crucial. The Dufour effect, also known as thermo-diffusion, is the phenomenon where a temperature gradient in a fluid causes a change in the concentration of different components or species in the fluid. It is important to comprehend the Dufour effect in systems where mass transfer and heat transfer happen simultaneously. It makes processes more efficient and easier to control when heat and mass diffusion interact. This is done by helping engineers understand and predict how fluid mixes will behave in different situations.

The movement of particles in a fluid medium due to temperature gradients is known as thermophoresis. When there's a variation in temperature within a fluid, it can cause particles within that fluid to migrate from regions of higher temperature to regions of lower temperature. This movement occurs due to the interaction between the particles and the temperature gradient in the surrounding medium. Various factors, including the size and properties of the particles, the temperature gradient, and the fluid characteristics, can influence the motion of particles in thermophoresis. Fields like physics, chemistry, and engineering often study it in the context of colloidal systems, where understanding the behavior of particles within a fluid has implications. In many scientific and industrial operations, managing and forecasting particle behavior requires an understanding of thermophoresis. Environmental research relies on thermophoresis as it enhances pollution dispersion models and serves engineering applications that require precise manipulation of particle distributions. Thermophoresis

has applications in a variety of fields, including the manipulation of particles in microfluidic devices, the study of aerosols and particle transport, and the development of certain materials and coatings. This phenomenon allows for improved heat management by changing fluid characteristics and thermal boundary layers, which is critical for streamlining industrial operations. It manipulates particle trajectories by interacting with magnetic fields to regulate particle deposition and migration, which is essential in applications like coatings and pollution management. Thermophoresis also changes the structure and stability of the flow, which makes it possible to precisely control how the fluid acts in MHD-driven systems and improves their usefulness in areas such as electromagnetic pumping and plasma processing.

Cortell [1], Ibrahim [2], and Shateyi [3] disregarded the outcomes of the Soret and Dufour effects in heat and concentration transport, citing variations in the consequences reported by the Fick's and Fourier laws. Eckert and Drake [4] noted that the Soret effect affects isotope separation and the combination of low (He, H₂) and medium molecular weight gases (H₂, air). Afterwards, many researchers, including Sattar and Alam [5], extensively studied their significance in heat transfer problems. Chamkha and Rashad [6] investigated the Soret and Dufour effect on mixed convection flow around a vertical cone with an angular velocity. Mortimer and Eyring [7] used simple transition states to construct the Soret and Dufour phenomenon. Tak, Khan, and Mathur [8] studied double diffusion inside a Darcy porous media on a vertical permeable surface. The Soret and Dufour impact on MHD mixed convection over an infinite vertical porous plate was analyzed by Ahmed and Sengupta [9]. In porous media, Moorthy and Senthilvadivu [10] quantitatively studied the Soret and Dufour impact on MHD-free convection with varied velocities. Reddy and Chamkha [11] solved the Soret and Dufour effects of MHD convective flow across a stretched sheet in porous media numerically. Jha and Gambo [12] presented their investigation of the sequel to Soret and Dufour on unsteady free convection, where the bounding wall has ramped temperature and the species has ramped concentration, respectively. Ashraf, Hayat, Alsaedi, and Shehzad [13] investigated Soret and Dufour's effects on an Oldroyd-B fluid in the presence of convective boundary conditions. They used the homotopy analysis method to describe this phenomenon. Gbadeyan, Idowu, Ogunsola, Agboola, and Olanrewaju [14] studied the Soret and Dufour effects on mixed convection heat and mass transfer flow over a stretching surface filled with viscoelastic fluid. They compared the Runge-Kutta shooting technique with the Ham method for the zero magnetic field parameter. Srinivasacharya and RamReddy [15] studied thermo-diffusion and diffusion-thermo on an exponentially stretched surface using a quiescent fluid. Siddique, Nadeem, Awrejcewicz, and Pawtowski [16] examined Soret and Dufour on an unstable MHD second-grade nanofluid over an exponential surface. Another

work by Mahdy, A. [17], used a different boundary layer of non-Darcian mixed convective flow to show what happens when there is suction and injection along with double diffusion. Srinivasacharya and RamReddy [18] researched the cross diffusion effect on an exponentially extended sheet in a non-Darcy quiescent fluid using the Keller Box method. In MHD mixed convection heat and mass transfer flow with radiation, hall current, Soret, and Dufour, Shateyi et al. [19] described the magnetic field's effect. Shaheen et al. [20] examined the sequel of Soret and Dufour to a 2D Casson nanofluid in a deformable cylinder. The variables they used were the Arrhenius activation energy and chemical reaction. Considering Hall current and heat flux, Quader and Alam [21] used the FDM technique to show the double diffusion effect on unsteady natural convection. Subhakar, Gangadhar, and Reddy [22] used a moving, non-isothermal vertical plate to understand the influence of Soret and Dufour on heat generated or absorbed. Balla et al. [23] used Hillesdon and Pedley's model to demonstrate nanofluid bioconvective flow and further inquired about the sequel of Soret and Dufour on it. Reddy, Rao, and Vijaya [24] considered a viscoelastic fluid to show the Soret and Dufour effect passing through an infinitely extended sheet. Rashed [25] worked on the thermophoresis effect, variable velocity in non-Darcian flow, the occurrence of Soret, Dufour, and thermal radiation. A number of researchers [26–40, 41–65] have recently focused on the investigation of diverse impacts, such as thermophoresis and magnetic fields, on various fluid flow characteristics and geometry.

We have expanded upon the work of Ishak [66] to address unsteady MHD forced convection over a stretching plate. We apply the Soret, Dufour, thermophoresis, and chemical reaction sequel to our work, focusing on a 2D steady flow. To the author's knowledge, no one has yet examined the combined effects of thermophoresis, Dufour, chemical reaction, Soret, and joule heating on a continuous MHD flow through a porous stretched plate. The examination of all these combined impacts in fluid flow was the main focus of this investigation. Also, when compared to previous studies, the new results are considered to be quite good.

MATHEMATICAL FORMULATION

We consider a 2D coordinate system. The surface is extended along the x-axis with velocity $U_w = ax$, keeping the origin stationary. The transverse magnetic field M_0 is applied externally 90° to x-axis in positive direction of y-axis. The y-axis is considered normal to stretching plate, while x-axis is parallel to stretching plate. The flow problem is graphically demonstrated in Figure 1, where u and v are the velocity of the fluid along x and y direction respectively. In the course of our analysis, we have taken into consideration the following assumptions on the flow problem:

- Fluid is contained in a porous media and is thought to be electrically conducting with electrical conductivity (σ_e).

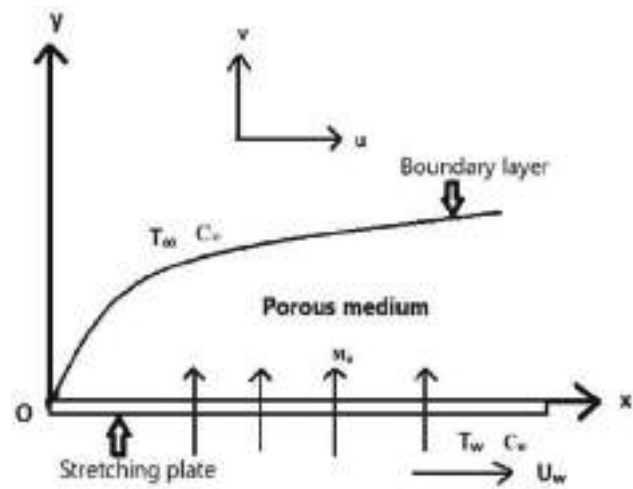


Figure 1. Geometry of the flow problem.

- Magnetic field induced is neglected and the magnetic Reynolds is considered to be small.
- Fluid is considered to be steady, 2D, laminar and incompressible.
- Fluid is taken to be chemically reacting with a first order chemical reaction.
- The flow issue assumes no external electric field.
- Hall effect and polarisation of charges are neglected.

These assumptions with Boussinesq's approximation provide the continuity, momentum, energy, and concentration equations [6]:

Continuity equation:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad (1)$$

Momentum equation:

$$u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = \nu \frac{\partial^2 u}{\partial y^2} - \frac{\sigma_e M_0^2 u}{\rho} - \frac{\nu u}{k_0} \quad (2)$$

Energy equation:

$$u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} = \alpha \frac{\partial^2 T}{\partial y^2} + \frac{\sigma_e M_0^2 u^2}{\rho c_p} + \frac{D_M K_T}{c_s c_p} \frac{\partial^2 C}{\partial y^2} \quad (3)$$

Concentration equation:

$$u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} = D_M \frac{\partial^2 C}{\partial y^2} - K^*(C - C_\infty) + \frac{D_M K_T}{T_M} \frac{\partial^2 T}{\partial y^2} - \frac{\partial}{\partial y} (V_t (C - C_\infty)) \quad (4)$$

where, $V_t = \frac{\nu K_T}{T_r} \frac{\partial T}{\partial y}$ is the thermophoretic velocity.

The relevant BCs [66] are:

$$y = 0: \quad u = U_w, \quad v = 0, \quad T = T_w, \quad C = C_w, \text{ and} \quad (5)$$

$$y \rightarrow \infty: \quad u \rightarrow 0, \quad T \rightarrow T_\infty, \quad C \rightarrow C_\infty \quad (6)$$

It is determined that the boundary conditions are selected in such a manner that, initially close to the plate at $y = 0$, $u = U_w$, which is the extended velocity of the surface, $v = 0$, is sought. However, the temperature and concentration of the fluid are considered at the surface, which are denoted by the symbols T_w and C_w accordingly. Also, far from plate at $y \rightarrow \infty$, $u \rightarrow 0$, and the temperature and concentration of the fluid are considered to be the ambient temperature and concentration, which are denoted by the symbols T_∞ and C_∞ accordingly.

We consider:

$$T_w = T_\infty + bx, \quad C_w = C_\infty + dx \quad (7)$$

The continuity equation (1) is fulfilled, by using a stream function ψ such that $u = \frac{\partial \psi}{\partial y}$ and $v = -\frac{\partial \psi}{\partial x}$.

Using similarity variables and nondimensional values, the nonlinear PDEqs (2)-(4) are turned into nonlinear ODEqs:

$$\begin{aligned} \eta &= \left(\frac{\alpha}{v}\right)^{\frac{1}{2}} y, \quad \psi = (avx^2)^{\frac{1}{2}} f(\eta), \quad \theta(\eta) = \frac{T - T_\infty}{T_w - T_\infty}, \\ M &= \frac{\sigma_e M_0^2}{\rho a}, \quad S_p = \frac{v}{k_0 a}, \quad Ec = \frac{a^2 x}{b C_p}, \quad \alpha = \frac{\kappa}{\rho C_p}, \quad K = \frac{K^*}{a}, \\ \zeta &= \frac{K_T(T_w - T_\infty)}{T_r}, \quad Pr = \frac{\mu C_p}{\kappa}, \quad Sc = \frac{v}{D_M}, \\ Du &= \frac{D_M K_T(C_w - C_\infty)}{v C_s C_p(T_w - T_\infty)}, \quad Sr = \frac{D_M K_T(T_w - T_\infty)}{v T_M(C_w - C_\infty)} \end{aligned}$$

The transformed nonlinear ODEs are:

$$f''' + ff'' - f'^2 - Mf' - S_p f' = 0 \quad (8)$$

$$\theta'' + Pr Mec f'^2 + Pr Du \phi'' + Pr(f\theta' - \theta f') = 0 \quad (9)$$

$$\phi'' - Sc K \phi + Sc Sr \theta'' + Sc(f\phi' - \phi f') - Sc \zeta(\phi' \theta' + \phi \theta'') = 0 \quad (10)$$

Also, transformed beginning and boundary conditions:

$$\begin{aligned} f(0) &= 0, \quad f'(0) = 1, \quad \theta(0) = 1, \quad \phi(0) = 1 \quad \text{and} \\ f'(\infty) &\rightarrow 0, \quad \theta(\infty) \rightarrow 0, \quad \phi(\infty) \rightarrow 0 \end{aligned} \quad (11)$$

The coefficient of skin friction, Nusselt and Sherwood number are as follows:

$$Re_x^{\frac{1}{2}} C_f = f''(0), \quad Re_x^{-\frac{1}{2}} Nu_x = -\theta'(0) \quad \text{and} \quad Re_x^{-\frac{1}{2}} Sh_x = -\phi'(0) \quad (12)$$

Method of Solution

Solution in MATLAB bvp4c analyses linear or nonlinear boundary value issues. It can forecast $y(x)$ for each x in $[a, b]$ while accounting for BCs at each step. BCs at infinity are adjusted for individuals who can logically address

the present problem using this technique. By applying this method, the BCs at ∞ are adjusted to a position where solving the present issue becomes feasible. The three-stage Lobatto IIIa formula is a finite difference code in Bvp4c. Shampine et al. recommended Lobatto IIIa [67]. This collocation polynomial gives uniform fourth-order accurate solutions across the interval by partitioning it into subintervals with a mesh of points. The solver calculates numerical error for each subinterval. If the answer fails tolerance, the mesh is adjusted and the procedure repeated. In addition to the original mesh points, a grid point solution assumption is always needed. An initial estimate at the beginning mesh points and step size adjustments attain the necessary accuracy. To use the finite difference based solution bvp4c, the equations (8), (9), and (10) are changed as follows.

$$\begin{aligned} f &= y_1, \quad f' = y_1' = y_2, \quad f'' = y_2' = y_3, \quad \theta = y_4, \quad \theta' = y_4' = y_5, \\ \phi &= y_6, \quad \phi' = y_6' = y_7, \quad y_3' = -y_1 y_3 + y_2^2 + M y_2 + S_p y_2 \end{aligned} \quad (13)$$

$$y_5' = Pr[-MEcy_2^2 - Du y_7' - (y_1 y_5 - y_4 y_2)] \quad (14)$$

$$y_7' = Sc[Ky_6 - Sry_5' - (y_1 y_7 - y_6 y_2) + \zeta(y_7 y_5 + y_6 y_5')] \quad (15)$$

Also, the BCs (11) are changed accordingly

$$y_1(0) = 0, \quad y_2(0) = 1, \quad y_4(0) = 1, \quad y_6(0) = 1 \quad (16)$$

$$y_2(\infty) = 0, \quad y_4(\infty) = 0, \quad y_6(\infty) = 0 \quad (17)$$

The MATLAB solver bvp4c uses the above modified results to do the numerical computation. To fulfil convergence criteria, precision is tested to the sixth decimal point. This approach was explained in depth by Shampine *et al.* [67]. This scheme's algorithm is shown in Fig. 2 [68].

Validation

When comparing the current equations with the equations described by Ishak [66], the values of $-\theta'(0)$ for various values of Pr show good agreement with the present results. It is likely that the bvp4c technique will be used to compare the results for $-\theta'(0)$, disregarding the other parametric values for different values of Pr as shown in Table 1.

Table 1. Comparison of $-\theta'(0)$ for several values of Pr

Pr	Ishak [66]	Present values	Error
1	0.9548	0.9548	0.0000
2	1.4715	1.4715	0.0000
3	1.8691	1.8691	0.0000

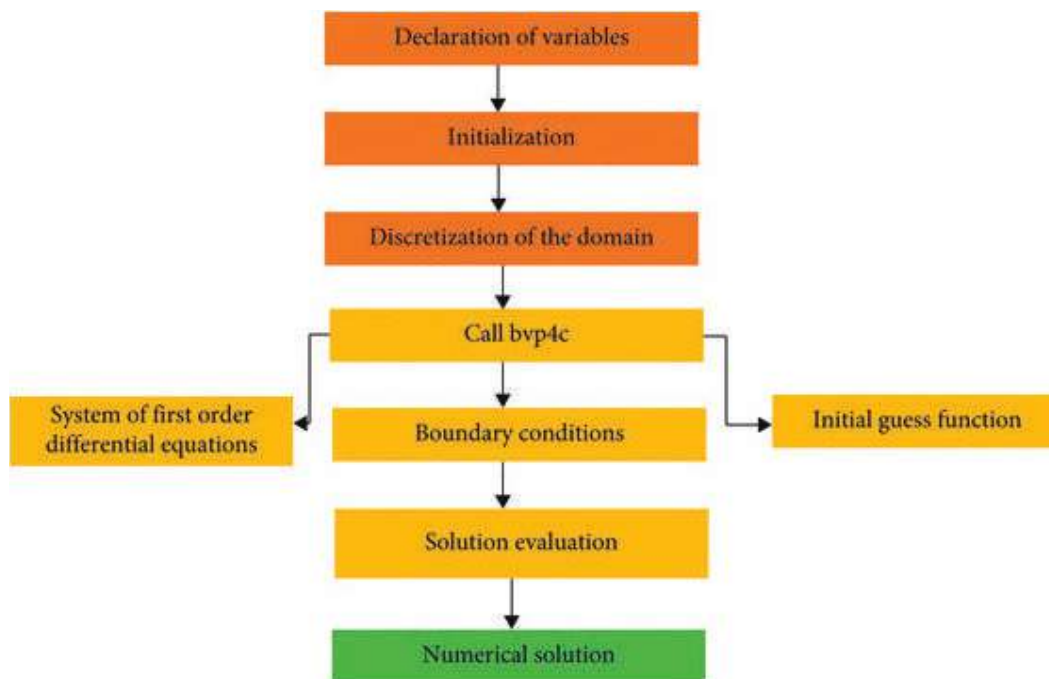


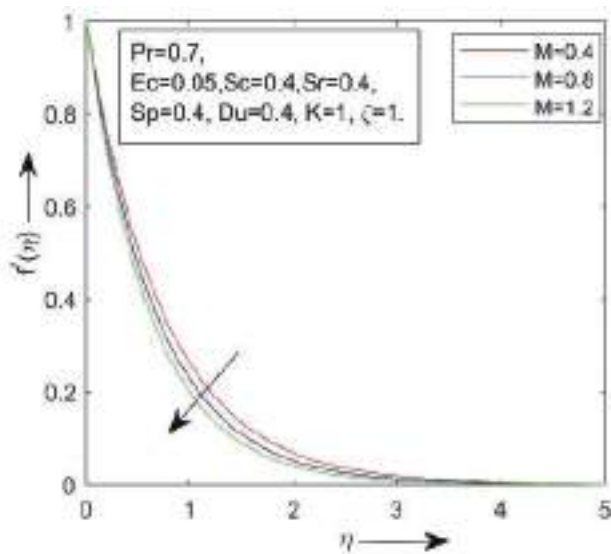
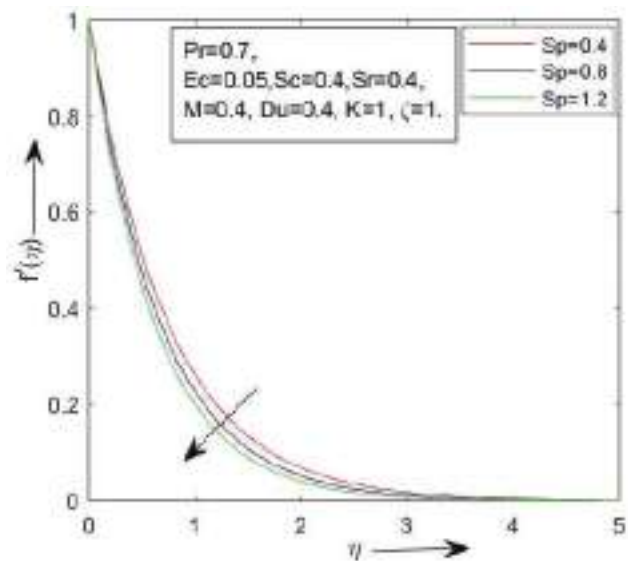
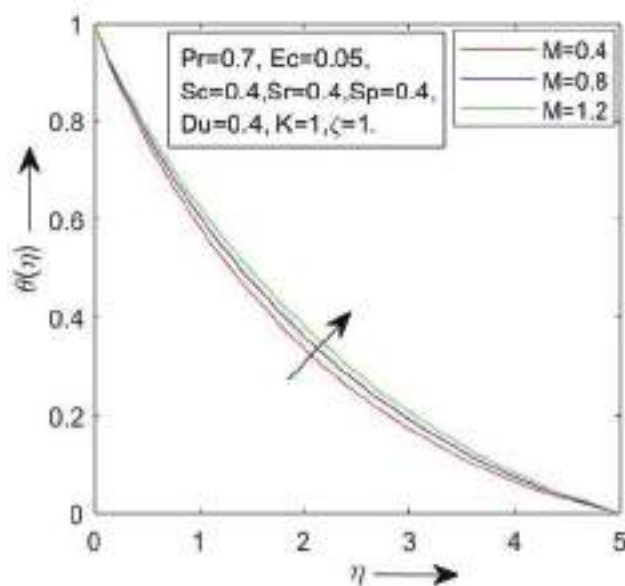
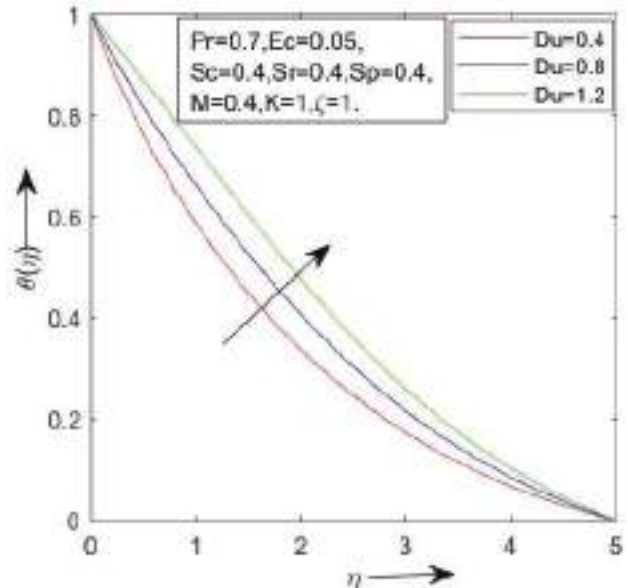
Figure 2. bvp4c algorithm.

RESULTS AND DISCUSSION

Numerical computations for the problem parameters are presented graphically from Figure 3 to 13. Also, the variation of the skin friction, Nusselt and Sherwood number are displayed via Tables 2 to 6. The standard values of the parameters are taken as $Pr = 0.7$ (air), $Ec = 0.05$, $M = 0.4$, $S_p = 0.4$, $Sc = 0.4$, $Sr = 0.4$, $Du = 0.4$, $K = 1$ recognised by Ishak [66], Elbashbeshy *et al.* [69]. Additionally, M is chosen to be minimal because generated magnetic field, Hall current, and ion slip are all factors that contribute to the magnetic field's perceived weakness.

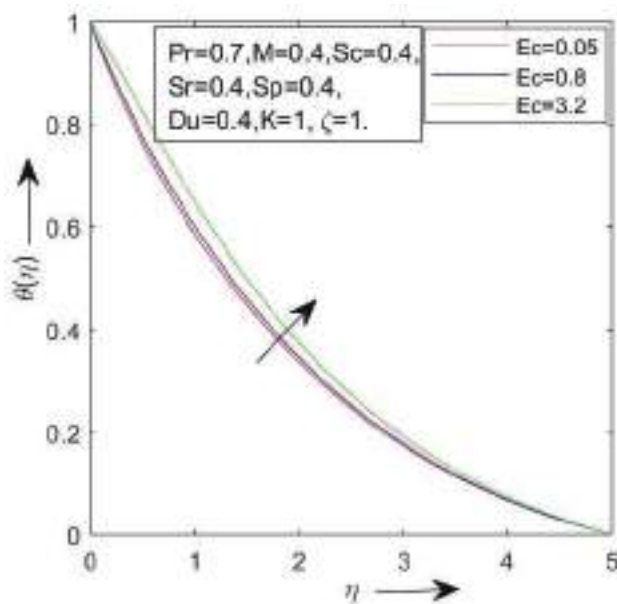
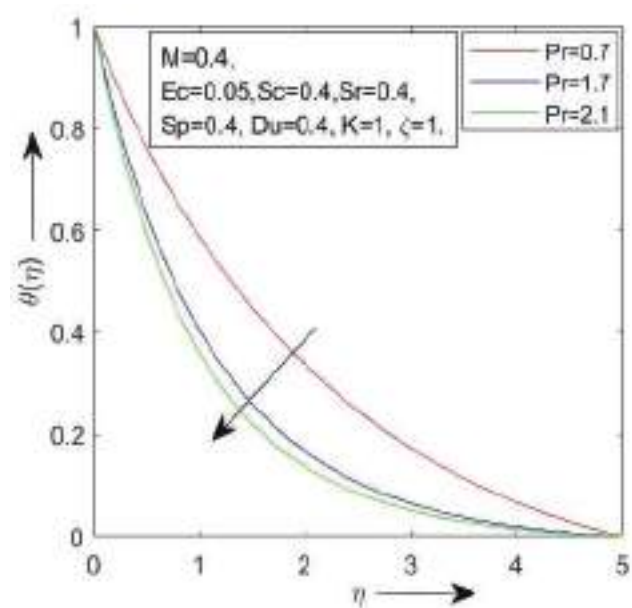
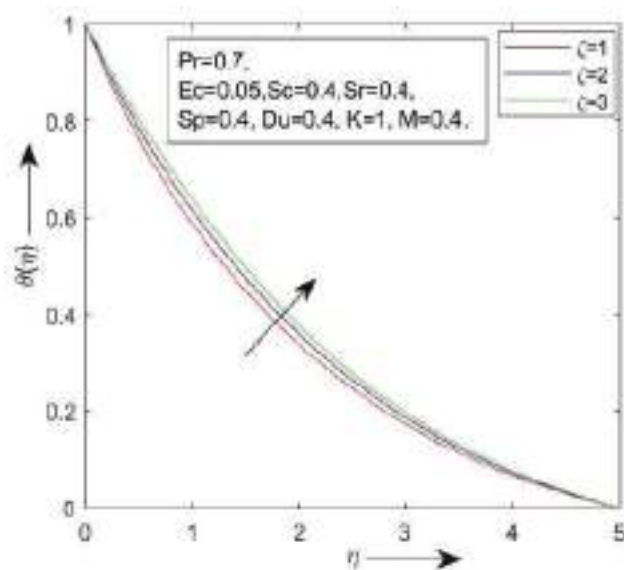
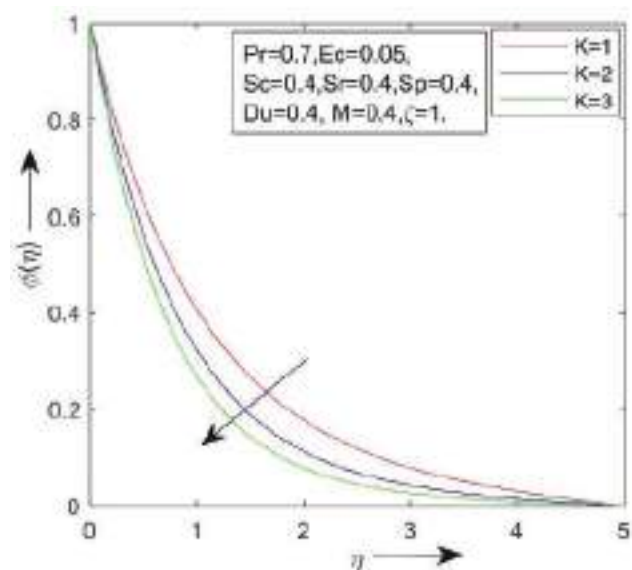
Figure 3 demonstrates that the velocity profile declines as M (Hartman number) rises. This is due to the effect of Lorentz force, which detards the motion of fluid flow. This phenomenon affects magnetohydrodynamics (MHD) applications, such as nuclear reactor cooling systems, where the magnetic field controls fluid flow. In electro-magnetic pumps and fusion reactor plasma confinement, the magnetic field regulates ionised fluid flow. Figure 4 illustrates the sequel of S_p on the flow profile. It shows that as S_p escalates, the fluid flow declines. This is due to the fact, as the occurrence of porous medium causes resistance to the fluid flow and as a result the flow profile reduces. Controlling flow via porous structures is important in oil recovery, groundwater filtering, and bioengineering. For improved oil recovery, adjusting porosity may affect reservoir fluid extraction. In groundwater filtration, water flow is controlled via soil or filters. Figure 5 shows the impact of M on the temperature profile. As M escalates, the fluid's temperature rises. Owing to the fact that as the

magnetic field rises the fluid particles starts vibrating frequently as such increasing the fluid's temperature. Figure 6 illustrates the impact of Du on the temperature profile. As Du escalates, the fluid's temperature rises. Higher Du values enhance boundary layer thickness and fluid temperature, resulting in fast energy transfer. Figure 7 displays the impact of Eckert number on temperature. The fluid's temperature escalates as Ec increases, indicating a decrease in enthalpy and rise in kinetic energy. Viscous dissipation converts kinetic energy into heat, raising fluid temperature as Eckert number increases. For performance, efficiency, and safety, temperature management is crucial in high-speed flow systems including aerospace vehicles, turbomachinery, microfluidics, biomedical devices, and industrial operations like polymer extrusion and lubrication. Figure 8 depicts the temperature curve and the sequel of Pr . When Pr increases, temperature decreases, demonstrating a link between velocity and thermal boundary layer thickness. Escalating values of Pr are thought to imply that thermal diffusivity is dominating, and hence the thermal boundary layer is thinner than the velocity boundary layer. From Figure 9 it is seen that as ζ (thermophoretic parameter) escalates, the fluid's temperature rises. Owing to the fact that, as increase in ζ means that particles are more responsive to the temperature gradient. This leads to enhanced movement of particles from hot to cold regions. Evidently, fluid's temperature escalates. As the thermophoretic parameter increases, particles move away from hot locations, reducing heat loss. This raises local temperatures, which helps with combustion, material processing, and industrial furnace activities that require focused heat,

Figure 3. Velocity profile vs M .Figure 4. Velocity profile vs S_p .Figure 5. Heat profile vs M .Figure 6. Heat profile vs Du .

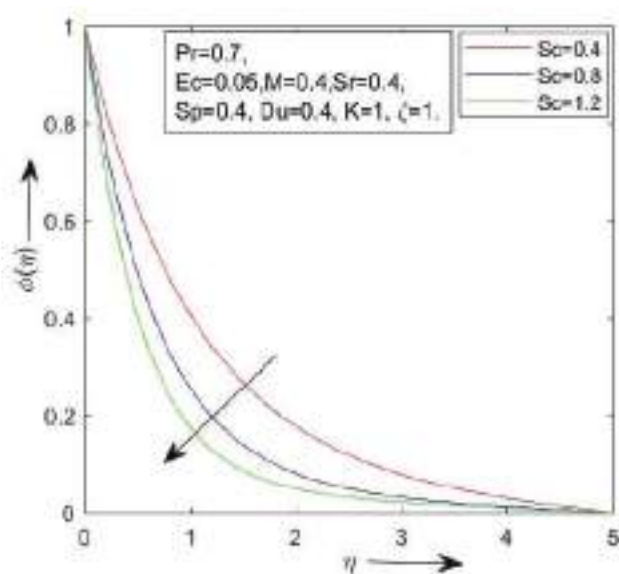
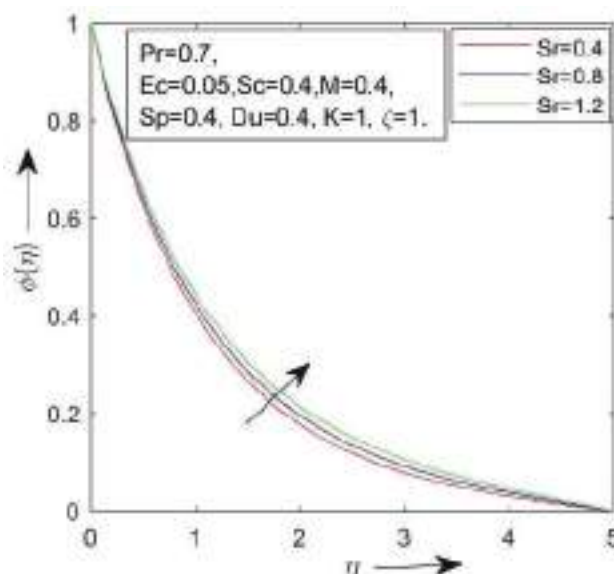
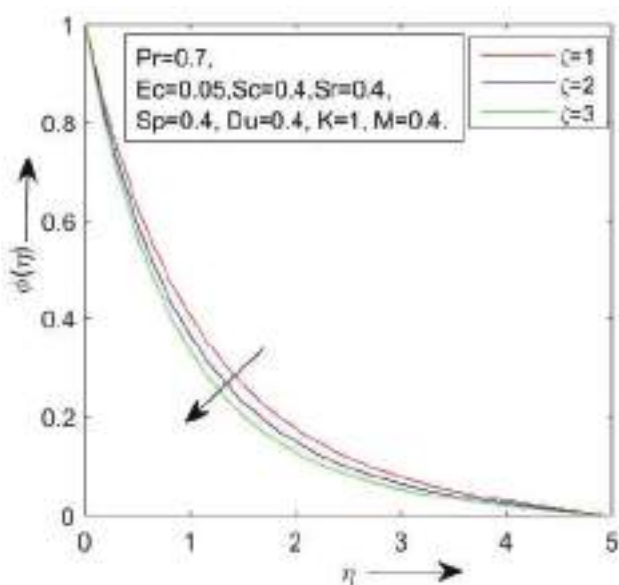
thermal insulation, or heat transmission. The impact of K causes a considerable decrease in fluid concentration, as seen in Figure 10. The concentration of the fluid falls as K rises because it represents the reaction rate, or the speed at which reactants turn into products. Reactants are consumed more quickly at a higher reaction rate, which lowers the concentration of reactants in the fluid. If the products do not refill the reactants or if the reactants are not continually supplied, then this increased consumption lowers the overall concentration of the reacting species. Fluid concentration gradually decreases as a result of the quicker depletion of reactant concentration brought on by

the faster chemical processes. This is consistent with the physical fact that species consumption leads to a decrease in species concentration. From, Figure 11 illustrates that as Sc escalates, the fluid's concentration decreases. It is due to the fact that as Sc rises, it rises the fluid's viscosity and hence decreasing fluid's concentration. As the Schmidt number increases, momentum diffusivity takes precedence over mass diffusivity. The fluid's solutes and particles diffuse more slowly than its viscous diffusion. An increased Schmidt number reduces diffusing species concentration, altering chemical mixing, pollution management, and heat transfer systems. Figure 12 shows that for superior

Figure 7. Heat profile vs Ec .Figure 8. Heat profile vs Pr .Figure 9. Heat profile vs ζ .Figure 10. Concentration profile vs K .

values of Sr , the concentration profile escalates. Greater Soret numbers correlate to greater temperature gradients, which in turn produce more convective flow. The fluid's concentration escalates with a hike in the Soret number because this number quantifies the effect of thermal diffusion, where a temperature gradient induces a concentration gradient. A greater Soret number implies better thermodiffusion, which moves particles or solutes from low to high temperatures. Because of migration, solutes accumulate in some places, increasing their concentration. Essentially, the Soret effect redistributes the solutes more

effectively with a higher Soret number, enhancing the local concentration where the temperature gradient drives the solutes to accumulate. Hence, concentration distribution escalates. From Figure 13 it is seen that as ζ escalates, fluid's concentration declines. The thermophoretic parameter, which quantifies how well particles move in response to a temperature gradient, decreases as the value rises. Particles move from hot to cold areas faster as it rises, creating a net outflow from areas with higher concentrations. Although the distribution becomes more even as a result of this shift, the concentration in the formerly higher-density areas

Figure 11. Concentration profile vs Sc .Figure 12. Concentration profile vs Sr .Figure 13. Concentration profile vs ζ .

decreases. The enhanced thermophoretic effect dilutes local concentrations, particularly in close proximity to heat sources, leading to a general decrease in the solute or particle concentration in the fluid as they disperse more evenly. A higher thermophoretic parameter increases particle movement in response to temperature gradients, lowering concentration in hotter locations. This idea applies to particle movement and deposition in pollution control, manufacturing, cooling systems, and biological applications.

The variation in the skin friction coefficient, Nusselt and Sherwood number are shown in Tables 2 to 6. From Table 2, it can be noticed that the magnitude of skin friction escalates with the upsurge in M and S_p . From Table 3, it is noticed that magnitude of Nusselt number (i.e., the rate of heat transfer) escalates with the upsurge in Pr , whereas it decreases with upsurge in M , S_p and Ec . From Table 4, We can observe that the rate of heat transfer decreases with the upsurge in Sc , Du , K and ζ , whereas it increases with the rise in Sr . From Table 5, it can be seen that the magnitude of Sherwood number decreases with the rise in M , S_p and Ec ,

Table 2. Skin friction with the variation of M and S_p .

Pr	M	S_p	Ec	Sc	Du	Sr	K	$-f''(0)$
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	-1.3417
	0.8							-1.4833
	1.2							-1.6125
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	-1.3417
		0.8						-1.4833
		1.2						-1.6125

Table 3. Nusselt number with the variation of Pr, M, S_p and Ec

Pr	M	S_p	Ec	Sc	Du	Sr	K	$-\theta'(0)$
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.5633
1.0								0.6926
1.7								0.9398
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.5633
	0.8							0.5315
	1.2							0.5046
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.5633
		0.8						0.5354
		1.2						0.5118
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.5633
			1.50					0.4416
			2.50					0.3576

Table 4. Nusselt number variation w.r.t. Sc, Du, Sr, K and ζ .

Pr	M	S_p	Ec	Sc	Du	Sr	K	ζ	$-\theta'(0)$
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.5633
				0.8					0.4794
				1.2					0.4195
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.5633
					0.8				0.4084
					1.2				0.2587
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.5633
						0.8			0.5731
						1.2			0.5835
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.5633
							2		0.5147
							3		0.4679
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.5633
								2	0.4931
								3	0.4331

Table 5. Sherwood number with the variation of Pr, M, S_p and Ec

Pr	M	S_p	Ec	Sc	Du	Sr	K	$-\phi'(0)$
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.9689
1.0								0.9936
1.7								1.0418
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.9689
	0.8							0.9528
	1.2							0.9393
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.9689
		0.8						0.9537
		1.2						0.9408
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	0.9689
			1.50					0.9435
			2.50					0.9261

Table 6. Sherwood number with the variation of Sc, Du, Sr, K and ζ

Pr	M	S _p	Ec	Sc	Du	Sr	K	ζ	$-\phi'(0)$
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.9689
				0.8					1.4529
				1.2					1.8345
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.9689
					0.8				0.9390
					1.2				0.9105
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.9689
						0.8			0.9273
						1.2			0.8832
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.9689
						2			1.1825
						3			1.3544
0.7	0.4	0.4	0.05	0.4	0.4	0.4	1	1	0.9689
								2	1.0915
								3	1.1807

whereas it escalates with the rise in Pr. From Table 6, it can be observed that magnitude of Sherwood number escalates with the rise in Sc, K and ζ , whereas it decreases with the rise in Du and Sr.

CONCLUSION

The novel aspect of the current research is to study the sequel of Soret, Dufour and thermophoresis in a chemically reacting stretching porous plate in occurrence of a transverse magnetic field. The following results can be concluded from the above study:

- The flow profile decreases with the rise of M and S_p.
- The heat profile escalates with the rise of M, Du, Ec and ζ . While it decreases with the rise of Pr.
- Concentration profile decreases with the rise of K, Sc and ζ , whereas it escalates with the rise in Sr.
- Magnitude of skin friction coefficient escalates with the rise in M and S_p.
- Magnitude of heat transfer rate upsurges with the rise in Pr and Sr, whereas it decreases with the rise in M, S_p, Ec, Sc, Du, K and ζ .
- The magnitude of mass transfer rate upsurges with the rise in Pr, Sc, K and ζ , whereas it decreases with the rise in M, S_p, Ec, Du and Sr.

The current investigation should primarily focus on future work on the irregular behavior of the problem and its relevance to entropy production and activation energy. Significant sequel of thermal radiation, heat source, thermophoresis, and chemical reaction parameters on flow profiles can be found in a broad range of industrial and technological applications, including metallic coating, crystal growth, electromagnetic pumps, power generators, MHD accelerators, and reactor cooling. The study's conclusions might be used to a variety of processes, including

paper manufacture, suspensions, heat exchanger technology, material processing, drying, and surface evaporation of water bodies. It is also possible to expand this work to a three-dimensional flow issue and investigate the varied flow effects for non-Newtonian fluids, including Maxwell, Carreau, and Williamson fluids.

NOMENCLATURE

a,b,d	Constant,
M ₀	Constant Magnetic field, (Nm/A)
C _p	Specific heat at constant pressure, ($\frac{J}{kgK}$)
C _s	Concentration susceptibility,
K*	Rate of first order chemical reaction,
f	Dimensionless stream function,
Ec	Eckert number,
k ₀	Permeability of porous medium, (m ²)
M	Magnetic parameter (Hartmann number),
Pr	Prandlt number,
S _p	Porosity parameter,
ζ	Thermophoretic parameter,
D _M	Coefficient of molecular diffusivity,
D _T	Coefficient of thermal diffusivity,
K _T	Thermal diffusion ratio,
T _r	Reference temperature, (K)
T _M	Mean fluid temperature,
Sc	Schmidt number,
Sr	Soret number,
Du	Dufour number,
K	Chemical reaction parameter,
T	Fluid's temperature, (K)
C	Species concentration, (kg/m ³)
u	fluid's velocity through x-axis, (m/s)
v	fluid's velocity through y-axis, (m/s)
(x,y)	Cartesian coordinates

Greek Symbols

ρ	fluid's density, (kg/m ³)
μ	fluid's viscosity, (Pa s)
σ_e	Electrical conductivity, ($\frac{1}{\Omega m}$)
η	Dimensionless similarity variable,
κ	Thermal conductivity, ($\frac{W}{mK}$)
ν	Kinematic viscosity, (m ² /s)
α	Thermal diffusivity, (m ² /s)
ψ	Stream function,
θ	Dimensionless temperature,

Superscript

' Derivative w.r.t η

Subscript

w	Properties at the plate
∞	Free stream condition

Abbreviation

MHD	Magnetohydrodynamics
ODE	Ordinary differential equation
PDE	Partial differential equation
BC	Boundary Conditions
2D	Two-Dimensional

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] Cortell R. Suction, viscous dissipation and thermal radiation effects on the flow and heat transfer of a power-law fluid past an infinite porous plate. *Chem Eng Res Des* 2011;89:85–93. [\[CrossRef\]](#)
- [2] Ibrahim FS, Elaiw AM, Bakr AA. Effect of the chemical reaction and radiation absorption on the unsteady MHD free convection flow past a semi-infinite vertical permeable moving plate with heat source and suction. *Commun Nonlinear Sci Numer Simul* 2008;13:1056–1066. [\[CrossRef\]](#)
- [3] Shateyi S. Thermal radiation and buoyancy effects on heat and mass transfer over a semi-infinite stretching surface with suction and blowing. *J Appl Math* 2008;2008:1–12. [\[CrossRef\]](#)
- [4] Eckert ERG, Drake RM. Analysis of heat and mass transfer, Jr. McGraw-Hill; 1972. *AICHE J* 1972;18:670. [\[CrossRef\]](#)
- [5] Sattar AM, Alam MM. Thermal diffusion as well as transpiration effects on MHD free convection and mass transfer flow past an accelerated vertical porous plate. *Indian J Pure Appl Math* 1994;25:679–688.
- [6] Chamkha AJ, Rashad AM. Unsteady heat and mass transfer by MHD mixed convection flow from a rotating vertical cone with chemical reaction and Soret and Dufour effects. *Can J Chem Eng* 2014;92:758–767. [\[CrossRef\]](#)
- [7] Mortimer RG, Eyring H. Elementary transition state theory of the Soret and Dufour effects. *Proc Natl Acad Sci U S A* 1980;77:1728–1731. [\[CrossRef\]](#)
- [8] Tak S, Sunder Mathur R, Gehlot R, Kumar, Khan A. MHD free convection-radiation interaction along a vertical surface embedded in Darcian porous medium in presence of Soret and Dufour's effects. *Therm Sci* 2010;14:137–145. [\[CrossRef\]](#)
- [9] Ahmed N, Sengupta S. Thermo-diffusion and diffusion-thermo effects on a three dimensional MHD mixed convection flow past an infinite vertical porous plate with thermal radiation. *Magnetohydrodynamics* 2011;47:41–60. [\[CrossRef\]](#)
- [10] Moorthy MBK, Senthilvadivu K. Soret and Dufour effects on natural convection flow past a vertical surface in a porous medium with variable viscosity. *J Appl Math* 2012;2012:1–15. [\[CrossRef\]](#)
- [11] Reddy PS, Chamkha AJ. Soret and Dufour effects on MHD convective flow of Al₂O₃-water and TiO₂-water nanofluids past a stretching sheet in porous media with heat generation/absorption. *Adv Powder Technol* 2016;27:1207–1218. [\[CrossRef\]](#)
- [12] Jha BK, Gambo YY. Soret and Dufour effects on transient free convection heat and mass transfer flow in a vertical channel with ramped wall temperature and specie concentration: an analytical approach. *Arab J Basic Appl Sci* 2020;27:344–357. [\[CrossRef\]](#)
- [13] Bilal Ashraf M, Hayat T, Alsaedi A, Shehzad SA. Soret and Dufour effects on the mixed convection flow of an Oldroyd-B fluid with convective boundary conditions. *Results Phys* 2016;6:917–924. [\[CrossRef\]](#)
- [14] Gbadeyan JA, Idowu AS, Ogunsola AW, Agboola OO, Olanrewaju PO. Heat and mass transfer for

- Soret and Dufour's effect on mixed convection boundary layer flow over a stretching vertical surface in a porous medium filled with a viscoelastic fluid in the presence of magnetic field. *Glob J Sci Frontier Res A Phys Space Sci* 2011;11:97–114.
- [15] Srinivasacharya D, Chetteti Ramreddy. Soret and Dufour Effects on Mixed Convection from an Exponentially Stretching Surface. *Int J Nonlinear Sci* 2011;12:60–68.
- [16] Siddique I, Nadeem M, Awrejcewicz J, Pawłowski W. Soret and Dufour effects on unsteady MHD second-grade nanofluid flow across an exponentially stretching surface. *Sci Rep* 2022;12:11811. [\[CrossRef\]](#)
- [17] Mahdy A. Soret and Dufour effect on double diffusion mixed convection from a vertical surface in a porous medium saturated with a non-Newtonian fluid. *J Non-Newtonian Fluid Mech* 2010;165:568–575. [\[CrossRef\]](#)
- [18] Srinivasacharya D, RamReddy C. Cross-diffusion effects on mixed convection from an exponentially stretching surface in non-Darcy porous medium. *Heat Transfer Asian Res* 2013;42:111–124. [\[CrossRef\]](#)
- [19] Shateyi S, Motsa SS, Sibanda P. The effects of thermal radiation, Hall currents, Soret, and Dufour on MHD flow by mixed convection over a vertical surface in porous media. *Math Probl Eng* 2010;2010:1–20. [\[CrossRef\]](#)
- [20] Shaheen N, Alshehri HM, Ramzan M, Shah Z, Kumam P. Soret and Dufour effects on a Casson nanofluid flow past a deformable cylinder with variable characteristics and Arrhenius activation energy. *Sci Rep* 2021;11:19282. [\[CrossRef\]](#)
- [21] Quader A, Alam MM. Soret and Dufour effects on unsteady free convection fluid flow in the presence of hall current and heat flux. *J Appl Math Phys* 2021;9:1611–1638. [\[CrossRef\]](#)
- [22] Subhakar MJ, Gangadhar K, Reddy NB. Soret and Dufour effects on MHD convective flow of heat and mass transfer over a moving non-isothermal vertical plate with heat generation/absorption. *Adv Appl Sci Res* 2012;3:3165–3184.
- [23] Balla CS, Ramesh A, Kishan N, Rashad AM. Impact of Soret and Dufour on bioconvective flow of nanofluid in porous square cavity. *Heat Transfer* 2021;50:5123–5147. [\[CrossRef\]](#)
- [24] Siva Kumar Reddy B, Surya Narayana Rao KV, Vijaya RB. Soret and Dufour effects on MHD flow of viscoelastic fluid past an infinite vertical stretching sheet. *Heat Transfer* 2020;49:2330–2343. [\[CrossRef\]](#)
- [25] Rashed GMA. Variable viscosity and thermophoresis for Soret and Dufour's effects on MHD generation in a non-Daracian porosity fluid. *Heat Transfer* 2020;49:1444–1457. [\[CrossRef\]](#)
- [26] Sheikholeslami M, Gerdroodbary MB, Shafee A, Tlili I. Hybrid nanoparticles dispersion into water inside a porous wavy tank involving magnetic force. *J Therm Anal Calorim* 2019;141:1993–1999. [\[CrossRef\]](#)
- [27] Manh TD, Bahramkhoo M, Barzegar Gerdroodbary MB, Nam ND, Tlili I. Investigation of nanomaterial flow through non-parallel plates. *J Therm Anal Calorim* 2020;143:3867–3875. [\[CrossRef\]](#)
- [28] Manh TD, Abazari AM, Barzegar Gerdroodbary MB, Nam ND, Moradi R, Babazadeh H. Computational simulation of variable magnetic force on heat characteristics of backward-facing step flow. *J Therm Anal Calorim* 2020;144:1585–1596. [\[CrossRef\]](#)
- [29] Tlili I, Moradi R, Barzegar Gerdroodbary MB. Transient nanofluid squeezing cooling process using aluminum oxide nanoparticle. *Int J Mod Phys C* 2019;30:1950078. [\[CrossRef\]](#)
- [30] Barzegar Gerdroodbary MB. Application of neural network on heat transfer enhancement of magneto-hydrodynamic nanofluid. *Heat Transfer Asian Res* 2019;49:197–212. [\[CrossRef\]](#)
- [31] Man Y, Gerdroodbary MB. Influence of Lorentz forces on forced convection of nanofluid in a porous enclosure. *J Porous Media* 2024;27:15–25. [\[CrossRef\]](#)
- [32] Barzegar Gerdroodbary MB, Jafaryar M, Sheikholeslami M, Amini Y. The efficacy of magnetic force on thermal performance of ferrofluid in a screw tube. *Case Stud Therm Eng* 2023;49:103187. [\[CrossRef\]](#)
- [33] Kumar KA, Sugunamma V, Sandeep N. Thermophoresis and Brownian motion effects on MHD micropolar nanofluid flow past a stretching surface with non-uniform heat source/sink. *Comput Therm Sci An Int J* 2020;12:55–77. [\[CrossRef\]](#)
- [34] Roja P, Sankar Reddy T, Ibrahim SM, Lorenzini G, Sidik NAC. The effect of thermophoresis on MHD stream of a micropolar liquid through a porous medium with variable heat and mass flux and thermal radiation. *CFD Lett* 2022;14:118–136. [\[CrossRef\]](#)
- [35] Zaman SU, Nauman Aslam M, Hussain A. Chemically reactive MHD fluid flow along with thermophoresis and Brownian effects. *Adv Mech Eng* 2023;15:16878132231193326. [\[CrossRef\]](#)
- [36] Sidahmed AOM, Department of Mathematics, College of Science and Arts, King Abdulaziz University, Rabigh 21911, Saudi Arabia. Numerical solution for MHD flow of an Oldroyd–B fluid over a stretching sheet in the presence of thermophoresis with chemical reaction effects. *Int J Adv Appl Sci* 2023;10:121–131. [\[CrossRef\]](#)
- [37] Hussain Z, Ayaz M, Islam S. Effects of thermophoresis and Brownian motion on radiative MHD hybrid nanofluid flow over a stretching sheet with convective boundary conditions: A homotopic approach. *Proc Inst Mech Eng Part N J Nanomater Nanoeng Nanosyst* 2024;23977914231225019. [\[CrossRef\]](#)
- [38] Chintalapudi R, Koti H, Shashidar Reddy B, Saritha K. Mechanisms of diffusion thermo and thermal diffusion on MHD mixed convection flow of Casson fluid over a vertical cone with porous material in the

- presence of thermophoresis and Brownian motion. *J Adv Res Numer Heat Transf* 2024;17:29–43. [\[CrossRef\]](#)
- [39] Sarma AK, Sarma D. MHD flow in free convection over an exponentially stretched sheet submerged in a double-stratified medium. *Int J Ambient Energy* 2024;45:1. [\[CrossRef\]](#)
- [40] Sarma AK. Unsteady radiative MHD flow over a porous stretching plate. In: *Modeling and Simulation of Fluid Flow and Heat Transfer*. 2024:96–109. [\[CrossRef\]](#)
- [41] Choudhary K, Jha AK, Mishra LN, Vandana Ms. Buoyancy and chemical reaction effects on MHD free convective slip flow of Newtonian and polar fluid through porous medium in the presence of thermal radiation and ohmic heating with Dufour effect. *Facta Univ, Ser Math Informatics* 2018;33:1. [\[CrossRef\]](#)
- [42] Sharma N, Mishra LN, Mishra VN. Fixed point theorems for expansive type mappings in multiplicative metric spaces. *Turkish J Anal Numer Theory* 2018;6:52–56.
- [43] Bhat IA, Mishra LN, Mishra VN, Tunç C, Tunç O. Precision and efficiency of an interpolation approach to weakly singular integral equations. *Int J Numer Methods Heat Fluid Flow* 2024;34:1479–1499. [\[CrossRef\]](#)
- [44] 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. [\[CrossRef\]](#)
- [45] Sharma MK, Chaudhary S, Rathour L, Mishra VN. Modified genetic algorithm with novel crossover and mutation operator in traveling salesman problem. *Sigma J Eng Nat Sci* 2024;42:6. [\[CrossRef\]](#)
- [46] 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. [\[CrossRef\]](#)
- [47] 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. [\[CrossRef\]](#)
- [48] 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. [\[CrossRef\]](#)
- [49] 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. [\[CrossRef\]](#)
- [50] 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. [\[CrossRef\]](#)
- [51] 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. [\[CrossRef\]](#)
- [52] 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. [\[CrossRef\]](#)
- [53] 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. [\[CrossRef\]](#)
- [54] 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* 2020;537:122710. [\[CrossRef\]](#)
- [55] Kataria HR, Mittal AS. Analysis of Casson nanofluid flow in presence of magnetic field and radiation. *Math Today* 2017;33:99–120.
- [56] 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. [\[CrossRef\]](#)
- [57] Kataria HR, Mittal AS. Mathematical analysis of three dimensional nanofluid flow in a rotating system considering thermal interfacial resistance and Brownian motion in suspensions through porous medium. *Math Today* 2018;34:A:7–24.
- [58] 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. [\[CrossRef\]](#)
- [59] Mittal AS. Analysis of water-based composite MHD fluid flow using HAM. *Int J Ambient Energy* 2021;42:1538–1550. [\[CrossRef\]](#)
- [60] Mittal AS. Study of radiation effects on unsteady 2D MHD Al₂O₃-water flow through parallel squeezing plates. *Int J Ambient Energy* 2022;43:653–660. [\[CrossRef\]](#)
- [61] Kataria HR, Mistry M, Mittal A. Influence of non-linear radiation on MHD micropolar fluid flow with viscous dissipation. *Heat Transfer* 2022;51:1449–1467. [\[CrossRef\]](#)
- [62] Kataria H, Mittal AS, Mistry M. Effect of nonlinear radiation on entropy optimised MHD fluid flow. *Int J Ambient Energy* 2022;43:6909–6918. [\[CrossRef\]](#)
- [63] Palani S. A comparative exploration of a magnetized power-law fluid past an inclined plate with variable physical properties. *Sigma J Eng Nat Sci* 2024;1:72. [\[CrossRef\]](#)

-
- [64] El Hattab M. MHD natural convection in a square enclosure using carbon nanotube-water nanofluid with two isothermal fins. *Sigma J Eng Nat Sci* 2024;1:87. [\[CrossRef\]](#)
- [65] Kumar Sarma A. MHD flow and effects of Hall current, Joule heating, and heat source across a chemically reacting and radiating exponentially stretching sheet in a porous medium. *Sigma J Eng Nat Sci* 2024;1:110. [\[CrossRef\]](#)
- [66] Ishak A. Unsteady MHD flow and heat transfer over a stretching plate. *J Appl Sci* 2010;10:2127–2131. [\[CrossRef\]](#)
- [67] Shampine LF, Kierzenka J, Reichelt MW. Solving boundary value problems for ordinary differential equations in MATLAB with bvp4c. *Tutorial notes* 2000;1–27.
- [68] Mwamba N, Okelo Abonyo J, Awuor KO. Effects of thermal radiation and chemical reaction on hydro-magnetic fluid flow in a cylindrical collapsible tube with an obstacle. *Int J Math Math Sci* 2023;2023:1–15. [\[CrossRef\]](#)
- [69] Elbashbeshy EMA, Aldawody DA. Heat transfer over an unsteady stretching surface with variable heat flux in the presence of a heat source or sink. *Comput Math Appl* 2010;60:2806–2811. [\[CrossRef\]](#)



Research Article

Trends and future perspectives in wind energy studies: A bibliometric analysis with time-series neural networks

Nihal GENC^{1,2,*}, Ferdi OZBILGIN³, Volkan BASER⁴

¹Graduate School of Natural and Applied Science, Karadeniz Technical University, Trabzon, 61080, Türkiye

²Department of Geomatics Engineering, Giresun University, Giresun, 28200, Türkiye

³Department of Electrical and Electronics Engineering, Giresun University, Giresun, 28200, Türkiye

⁴Department of Cartography and Cadastre, School of Vocational Education, Giresun University, Giresun, 28049, Türkiye

ARTICLE INFO

Article history

Received: 03 June 2024

Revised: 05 August 2024

Accepted: 04 October 2024

Keywords:

Bibliometric Analysis; Nonlinear Autoregressive Exogenous (Narx); Time Series; Web of Science (Wos); Wind Energy; Renewable Energy; Vosviewer

ABSTRACT

Energy, which is a fundamental resource for the universe and one of the indispensable requirements of today's modern world, plays an important role in many fields. In contrast to traditional energy sources, the utilization of renewable energy sources that offer a sustainable, clean, environmentally-friendly future is growing rapidly. Wind energy, as one of these sources, has become widespread and a focal point for energy production. This study conducted a bibliometric analysis, a data mining technique, on English publications (articles, reviews, proceedings, and book chapters) indexed in SCI-E, SSCI, and ESCI in Web of Science from 2010 to 2023. The analysis examined publication numbers, citations, authors, institutions, countries, journals, publishers, research areas, and keywords, and investigated their connections. To visualize trends, clusters, spatial distribution, and common interactions, Vosviewer, ArcGIS, and Excel were used. Temporal changes were assessed by dividing the years into three periods, showing an increasing trend in publications and citations in recent years. To forecast future changes, a Nonlinear Autoregressive Exogenous time series method predicted the number of future publications and citations. The analysis predicts continued growth in publications and citations until 2030, with a focus on energy fuels research. Finally, a multidisciplinary assessment was conducted to provide a broad perspective on wind energy and contribute to the versatility of the study. The study aims to be a comprehensive resource and guide for researchers working on wind energy and bibliometric analysis.

Cite this article as: Genc N, Ozbilgin F, Baser V. Trends and future perspectives in wind energy studies: A bibliometric analysis with time-series neural networks. Sigma J Eng Nat Sci 2025;43(5):1547–1571.

INTRODUCTION

Throughout history, human beings have questioned life to understand the universe, make progress, and sustain

existence. By observing daily life, people discovered that energy is the primary solution source [1]. Defining energy is challenging due to its dynamic structure and interdisciplinary roles. Aristotle's association of the word "entelechia,"

*Corresponding author.

*E-mail address: nihal.genc@giresun.edu.tr

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



meaning “complete reality,” with the word “energeia” in his work Metaphysics led to the derivation of the word “energy,” which modernly formulates in terms of power, force, and work [2, 3]. Energy, generally defined as the capacity to do work, represents all forms that can exist in various ways such as kinetic, potential, mechanical, heat, magnetic, electrical, chemical, and nuclear energy [4].

Energy has played an essential role in the development and evolution of the universe, encompassing both living and non-living entities throughout history [5]. It is possible to categorize energy, one of the most significant factors in the emergence of civilizations, meeting the needs of human beings, and the socio-economic-cultural development of countries, into different resource groups [6]. Energy resources can be divided into two groups: renewable and non-renewable energy resources based on whether they are exhaustible or not. Unsustainable and non-continuous energy resources are at risk of depletion because they cannot renew themselves despite their presence in nature [7]. These resources, which are divided into two as fossil and nuclear energy, are classified into subgroups as coal, oil, natural gas, uranium and thorium.

The energy that is derived from natural processes that are repetitive and permanent is called renewable energy. Solar, wind, hydroelectric, geothermal, biomass and wave are among renewable energy sources [8, 9]. These sources are also known as alternative energy sources and can meet a significant proportion of the world's total energy demand [8]. Population growth, urbanization, industrialization, economic growth, and changes in living standards all contribute to an increase in energy demand. To meet this demand, providing energy supply with renewable energy sources is a sustainable and economically viable solution compared to fossil fuels, which are exhaustible and costly. The production of renewable energy has been on the rise in recent years. In Figure 1, which is based on data from the International Renewable Energy Agency (IRENA), renewable energy production statistics between 2012–2020 are presented both in general and separately for each renewable energy type [10]. Hydroelectric energy stands out as the most produced

renewable energy type, with a large production amount that has been on an upward trend over the years. Wind energy production has also significantly increased over time, making it the second most produced renewable energy type after hydroelectric energy. Solar energy production has experienced a significant increase as well, especially after 2015, following a period of stagnation between 2012 and 2015. Biomass, geothermal, and wave energy follow solar energy in terms of production amounts. Overall, there has been a dominant increasing trend in the production of renewable energy. The figure illustrates a growing trend in the production of renewable energy sources, suggesting that their use is likely to increase in proportion.

The sun, which is the main energy source of the earth, is one of the renewable energy sources that influence the emergence of many other energy sources [11]. Solar energy that reaches the earth's atmosphere through reflection is examined in three different ways: solar photovoltaic isolated, photovoltaic grid-connected and thermal solar energy [12–14]. The sun, which causes temperature and pressure differences while heating the earth, is also the main source of winds and thus wind energy. Wind energy, which is easy to generate thanks to turbines and is both climate and environmentally friendly, is a clean resource that is rapidly developing. Solar and wind energy are preferable because they are accessible resources and can be easily produced at an affordable cost. This increases their use significantly and makes them stand out.

Hydroelectric energy is another type of energy produced by utilizing the movement caused by the sun directing the flow of water. This energy is generated from a comprehensive system consisting of a dam, reservoir and power plant. The water stored in hydroelectric power plants is directed to the turbines with a high pressure and thus the mechanical energy is converted into electrical energy by rotating the turbines. In the generation of hydroelectric energy, the sun is utilized as well as the energy generated by waves and tides [15]. Globally, about 1150 GW of hydroelectric energy generation has made it one of the largest renewable energy sources [16].

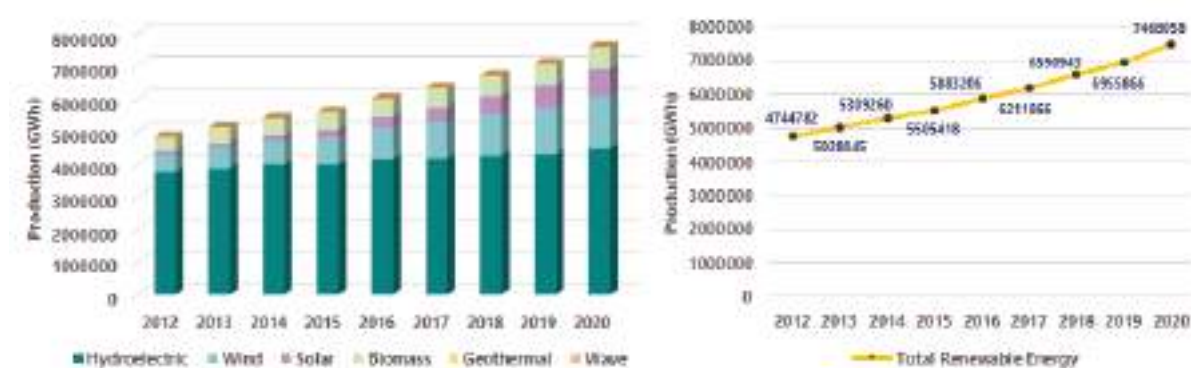


Figure 1. Production statistics for renewable energy and resources worldwide between 2012 and 2020.

Obtained by utilizing the internal thermal energy of the earth and independent of solar energy, geothermal energy is one of the renewable energy sources used effectively in electricity generation. Geothermal systems, which consist of three components: a heat source, a carrier fluid that allows the heat to be transferred, and a reservoir, enable the heat to be taken from the source and transferred to the

surface [17]. Environmentally friendly geothermal energy is used electricity generation as well as in heating and power applications.

Another renewable energy source influenced by solar energy is biomass. Obtained through plants and animal matter and consisting of carbohydrates, biomass is a special source [18]. Its basic structure includes materials left over

Table 1. Applications, advantages and capacities of renewable energy sources

Renewable Energy Source	Applications	Advantages	Worldwide Capacity (2021) [10]
 Wind	- Electricity generation - Electricity demand of factories and industrial areas - Lighting - Water pumping and storage - Grain grinding - Cooling systems - Logistics - Battery and charging stations	- Environmentally friendly - Can be used in small and large scale - Easy maintenance and long lifespan of the system - Quite economical - Abundant and freely available in the atmosphere, thus easily accessible - A clean source	824 874 MW
 Solar	- Electricity generation - Heating - Ventilation and cooling systems - Irrigation systems - Electricity for devices - Electric vehicles - Battery and charging stations	- Environmentally friendly - Unlimited - Economical - Reduces greenhouse gas emissions - Easy maintenance and long lifespan of the system - Effective for energy security - A clean source	849 473 MW
 Hydroelectric	- Electricity generation - Irrigation systems - Grain grinding - Industrial areas	- A clean energy - Reduces greenhouse gas emissions - Easy maintenance and long lifespan of the system - High energy efficiency - Does not require fuel as it benefits from water, preventing the formation of toxic gases - Benefits agriculture, fishing, tourism, and other fields.	1 360 054 MW
 Geothermal	- Electricity generation - Heating and cooling systems - Greenhouse systems - Health tourism - Drying	- A clean source - Reduces greenhouse gas emissions - Energy potential is quite large - Being a natural resource, it does not require fuel - High efficiency	15 644 MW
 Biomass	- Electricity generation - Heating and cooling systems - As fuel	- Contributes to recycling - Environmentally friendly - Reduces greenhouse gas emissions - Provides continuous energy supply - Easily stored	143 371 MW
 Wave	- Electricity generation - Industrial areas and facilities - Heating systems - Lighting	- Environmentally friendly - Provides continuous energy supply - Has no polluting effect - Helps areas like fishing and water sports - Does not harm the land.	524 MW

from agricultural and forestry practices and the residues of industrial and animal products. This makes biomass energy one of the sustainable energy sources globally. While biomass energy is generally used in heating and electricity generation applications, it is also frequently preferred in industrial studies and transportation.

Generated by winds, wave energy is a type of renewable energy whose main source is the sun, since wind is also caused by the sun [19]. The wind passing over the oceans creates waves on the surface. These waves move over long distances depending on the increase in wind speed. The waves rise as the distance increases, thus enhancing the wave energy [20]. The high energy density of waves makes wave energy advantageous compared to other types of energy. Information on the application patterns, advantages and worldwide capacities of these renewable energy sources are given in detail in Table 1.

Considering the world in general, it is seen that most of the countries are dependent on fossil fuels [21]. The use of fossil fuels to meet energy needs brings along the problems of global warming and climate change. The use of renewable energy resources is of great importance in combating these problems that negatively affect our world and all humanity. Using these resources will prevent carbon emissions, reduce environmental destruction and enable clean production, thus contributing to the sustainability of nature. Supporting this is the 7th Sustainable Development Goal, which states. With this goal, which is “Ensure access to affordable, reliable, sustainable and modern energy for all”, it is aimed to gain momentum in increasing and expanding the use of renewable energy [22]. The widespread use of green energy for a sustainable nature and life, the construction of green buildings and the development and effective use of conscious policies such as green transportation will contribute to the solution of social problems. The role of renewable energy sources is very important in accelerating all these [23].

In this study, wind energy, which is one of the renewable energy sources that is widely used today and will be more effective in our lives over time, has been detailed. Occurring due to pressure differences caused by the sun, wind is one of the natural, clean and sustainable renewable energy sources. With these features, it contributes to the prevention of problems such as air pollution, global warming and climate change. On the other hand, it makes positive contributions both locally and globally in terms of economic, political and energy efficiency by eliminating foreign dependency, being land-friendly, solving the electricity problem in rural areas, enabling agricultural and industrial practices in the regions where power plants are established, and providing employment opportunities. All these also support the Sustainable Development Goals. Wind turbines are used to generate wind energy. Electricity is generated by wind power plants established because of appropriate site selection based on wind speed, land topography, transportation facilities, socioeconomic and environmental factors. All these opportunities increase the use of wind energy

today. According to a report published by the World Wind Energy Association (WWEA), global wind energy capacity reached 874 Gigawatts in June 2022, representing a 13% growth compared to the previous year. On the other hand, it is estimated that it will surpass 955 Gigawatts by the end of 2022 and exceed 1 million Megawatts by the middle of 2023 [24]. These statistics are other indicators of the rapid development of wind energy. Wind energy makes many contributions to our world and human beings and will continue to do so in the future.

When looking at the literature, it is seen that there are thousands of studies on wind energy. These studies have been carried out in various fields of science, using various methods, by many authors at different universities. On the one hand, each study contributes to the literature and our world, as well as its scientific characteristics vary. Examining these studies that have gained a place in the literature in different aspects will provide both an evaluation of the current situation and a direction for future studies. The examinations and evaluations can be carried out effectively through bibliometric analysis.

Background of Bibliometric Analysis

The word bibliometrics is derived from two different words of Latin and Greek origin. The concept of bibliometrics, which is used today, has emerged with the combination of the words “biblion” and “metrics”, which means “book” and “science of measurement” [25]. Towards the end of the 20th century, it is possible to find many definitions of this concept in the literature. One of these definitions, which can be considered basic, was made by Alan Pritchard in 1969. Pritchard defined bibliometrics as “the application of mathematical and statistical methods to books and other media of communication” [26]. Research, analysis and evaluations based on bibliometrics are called bibliometric analysis. With bibliometric analysis, characteristics such as subject, field of science, author, citation, institution, journal, country, etc. of scientific studies within the scope of a determined discipline are evaluated numerically and statistically and the relationships between them are examined. In this respect, it can be said that bibliometric analysis is a quantitative method.

Although the concept of bibliometrics was introduced to the literature at the end of the 20th century, the beginning of bibliometric analysis applications dates to the 1870s. In 1869 and 1874, Francis Galton took the first steps in this sense with his works “Hereditary Genius” and “English Men of Science” based on the measurement of science [27, 28]. This was followed by the bibliography-themed book “Theory of the National and International Bibliography” written by Frank Cambbell in 1896 [25]. In 1917, Francis Joseph Cole and Nellie Barbara Eales wrote “The History of Comparative Anatomy: Part I- A Statistical Analysis of The Literature”, which deals with citation evaluations in the field of anatomy [29].

Today, bibliometric analysis studies have become quite widespread. With this analysis, comprehensive evaluations can be made by considering the relationship between

Table 2. Comparison of bibliometric analysis and review

Bibliometric Analysis	Review
• Evaluates the characteristics of publications quantitatively. • Examines the relationships between the characteristics of publications. • Enables understanding of trends, distributions, and clustering in a specific field. • Uses statistical methods in the analysis process and establishes network relationships between quantitative characteristics. • As it is based on statistics, the results depend on the relevance and accuracy of the data. • Evaluates the characteristics of studies conducted in a specific field from a broad perspective.	• Evaluates the topics and outputs of publications qualitatively. • Examines the results of publications and the relationships between them. • Enables understanding of the quality level of publications in a field. • The evaluation process involves reading, summarizing, and discussing the results of publications. • As it is content-based, the outputs are dependent on the author's perspective and may therefore include a subjective approach. • Evaluates the results of studies conducted in a specific field from a broad perspective.

certain characteristics of a scientific study and the outputs obtained can be presented with visualization tools such as graphs, tables and maps. Bibliometric analysis is also a data mining technique because it generates new data from data. As a data mining technique, it sources data from comprehensive databases like Web of Science (WoS), Scopus, and PubMed. These databases provide qualitative and quantitative information on authors, citations, institutions, journals, research areas, indexes, keywords, countries, and publishers. Visualization tools like Vosviewer, Gephi, Citespace, Leximancer, and SciMAT perform network analysis, producing interpretable visuals of distributions, trends, clusters, and their interrelationships. Moreover, the use of artificial intelligence (AI) in today's technology allows for future predictions. Data from bibliometric analysis can forecast future studies and trends through AI techniques. Thanks to the visuals produced, the distributions, trends, clusters and their relationships with each other can be interpreted. On the other hand, the use of artificial intelligence (AI) methods in today's developing technology is gaining importance day by day. Future predictions can be made by utilizing these methods. By using the data obtained because of bibliometric analysis, predictions can be made about future studies through AI techniques, and an idea about how future studies will be affected can be acquired by means of foresights.

The review studies commonly encountered in the literature may seem like bibliometric analysis, but in fact, they are very different from each other. While bibliometric analysis involves the evaluation of certain factors of studies conducted in a specific field using mathematical and statistical methods, review studies focus on the content of existing publications on a specific topic, examine their results, and provide recommendations for the future. Although

they have some similarities, these two types of studies are different in terms of their purpose, methods, and results obtained. Table 2 compares bibliometric analysis with review studies.

When conducting a scientific study, utilizing review articles serves as a fundamental means of developing the topic. Bibliometric analyses, on the other hand, enable the evaluation of the global perspective of the field in which the study will be conducted and facilitate the tracking of the existing network of this field worldwide. Therefore, bibliometric analyses provide a broader perspective on a scientific field. This broad perspective enables new studies to be planned in a manner consistent with their purpose, directing these plans, and thus allowing them to directly reach their target audience. It can be understood from this that bibliometric analyses serve as a guide for new studies.

Research Objective

This study aims to conduct a bibliometric analysis of wind energy studies indexed in the Web of Science database from 2010 to 2023. Reviewing the literature, it can be observed that there are few bibliometric analysis studies on wind energy. Some of these studies generally focus on the concept of renewable energy, covering other energy types and wind energy, and examining all publication types between certain years. It can be seen that these studies were conducted either in a single country [30, 31] or globally [32-34]. Another group of studies carried out bibliometric analyses on wind energy types [35-37] and components of wind energy systems [38-40]. On the other hand, other studies in the literature have examined the economic and financial aspects of both renewable energy types and wind energy alone through bibliometric analyses [41-45]. Finally, bibliometric studies focusing on specific methods for wind speed prediction can also be found in the literature [46, 47].

In general, these studies were conducted between 2013 and 2023. The differences of this study from the studies in the literature can be listed as “the comprehensive consideration of many features and situations, the use of techniques belonging to today’s artificial intelligence technology in addition to basic statistical methods, and the multidisciplinary evaluation of the obtained results.” Providing details of all these differences will help to better understand the contribution of the study to the literature. The comprehensive consideration of many features and situations includes providing detailed explanations about the main topic, having highly detailed statistical analyses, turning analysis outputs into effective and interpretable visuals to make them clear and understandable, examining a wide range of time periods, using up-to-date and hardware-strong software in the analysis and visualization processes, and conducting comparative analyses of changes through periodic evaluations.

On the other hand, the outputs obtained from the analysis process were evaluated using Nonlinear autoregressive exogenous (Narx) neural network, which is one of the AI methods. With the Narx method, predictions have been made for the future based on past and current data. Thus, some recommendations have been offered based on future trends and expectations.

Finally, for the study to be multidimensional, multidisciplinary evaluations have been made for the subject, taking into account today’s technology, living conditions and the current global situation. In this context, the concepts of environment, engineering, land management, economics,

politics and AI are discussed. By establishing the relationship between these concepts and wind energy, remarkable explanations for the current situation and the future have been made and presented with their justifications.

In this study, in Chapter 1, wind energy is discussed, examined in detail from various aspects and comprehensive information about bibliometric analysis is given. In Chapter 2, the materials and methods used in the study are explained and supported with visuals. In Chapter 3, bibliometric analysis was performed, the outputs obtained were evaluated individually and comprehensively, and graphs, figures and maps were produced and presented effectively. In Chapter 4, all outputs and the visuals produced were re-evaluated, changes were observed, comparisons were made and forward-looking predictions were made using Narx time series analysis. With the results obtained from these, future foresights were made and recommendations were presented. In the 5th and final chapter, the entire study is evaluated in general and the purpose and contributions of the study are emphasized.

MATERIALS AND METHODS

In this study, the studies published within the scope of the subject between certain years in the literature were handled for the purpose of examining them quantitatively. The studies were evaluated by bibliometric analysis, and the results were effectively visualized. Information about each step and details of the study are shown in Figure 2.

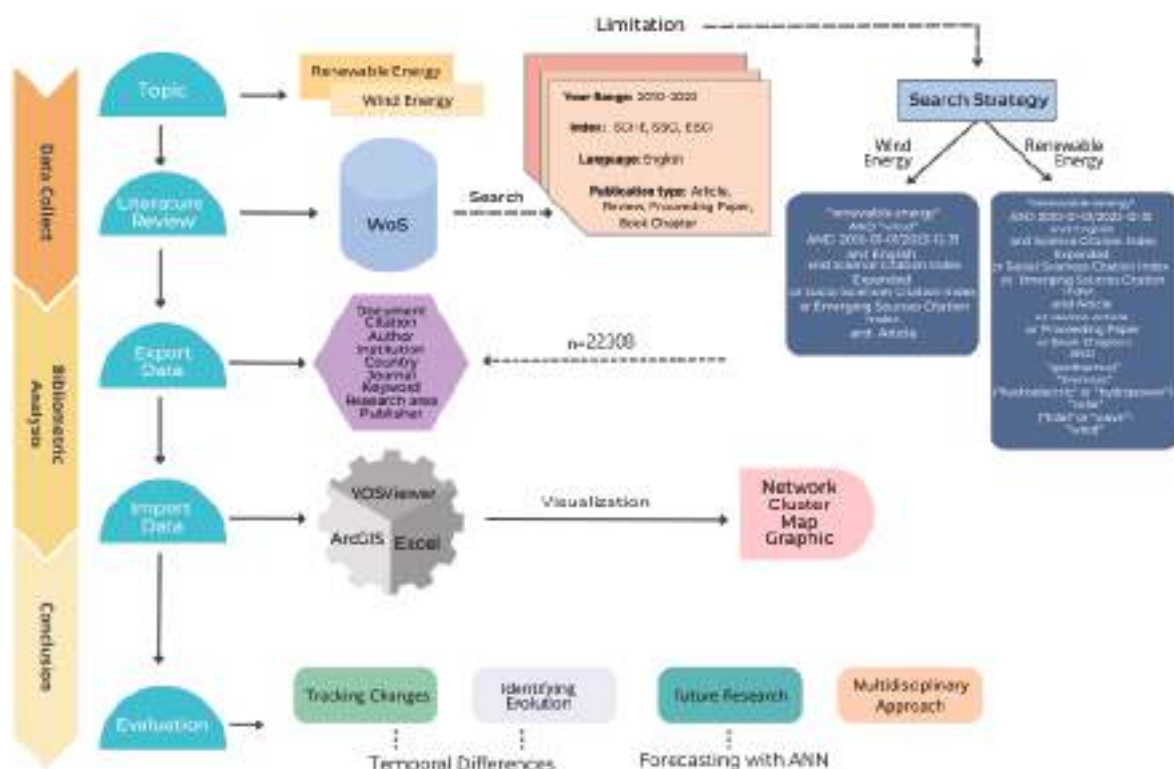


Figure 2. Multifaceted bibliometric analysis framework.

Data Source and Search Strategy

The study utilized Web of Science (WoS), an academic database that provides access to publications in various fields and offers a wide range of features. The data source for this study was the publications written in English between January 1, 2010, and December 31, 2023, in the form of articles, reviews, proceedings papers, and book chapters, available in the indexes of SCI-E, SSCI, and ESCI. Figure 2 provides a flowchart that gives a detailed overview of the filters and search strategies used in the data acquisition process.

The first part of the study analyzed publications in the five types of renewable energy sources, providing an overview and making evaluations. In the second part, articles published on wind energy were discussed, and a comprehensive analysis of these articles was conducted. The results obtained from the analysis were presented in detail.

Data Processing, Analysis and Visualization

After analyzing and processing the data, it is crucial to map and visualize the outputs to present the findings effectively. Various presentation tools were employed in this study to visualize large datasets, display data relationships, and identify trends and patterns. Specifically, Vosviewer 1.6.20 software was utilized to visualize relationships between data, ArcGIS 10.8 was used to create maps, Microsoft Excel 2021 was used for data processing and graphing, and MATLAB 2023b software was used for forecasting based on AI methods.

The main software program used in this study, Vosviewer, was developed by the Science and Technology Research Centre at Leiden University in the Netherlands. The program provides for visualization of authors, institutions, sources, countries, citations, and keywords, as well as their relationships between publications in the form of network analyses [48]. The acronym “VOS” stands for “Visualization of Similarities”. The program’s fundamental principle is to generate a similarity matrix and utilize the correlation strength to assess the similarity between item pairs in the co-occurrence data. The correlation strength is computed using the formula illustrated in the equation below:

$$S_{ij} = c_{ij}/w_i w_j \quad (1)$$

To assess the similarity between item i and item j , we use S_{ij} , which considers the number of co-occurrences (c_{ij}) of the items as well as their individual occurrence counts (w_i and w_j). The program visualizes the data by reducing the weighted sum of the squared Euclidean distance between all item pairs. The squared distance’s weight in the sum calculation increases with greater similarity between items. The expectation minimization algorithm aims to minimize the objective function shown in

formula (2) while adhering to the constraint conditions in formula (3).

$$V(x_1, x_2, \dots, x_n) = \sum_{i < j} S_{ij} \|x_i - x_j\|^2 \quad (2)$$

$$[2/n(n-1)] \sum_{i < j} \|x_i - x_j\| = 1 \quad (3)$$

The position of item i in the two-dimensional information graph is represented by the vector x_i , and the Euclidean norm is indicated by $\|\bullet\|$. ArcGIS 10.8 software was utilized to investigate the global distribution of the data based on countries and analyze them on a location-based basis. Microsoft Excel was used for storing, processing and graphing the data. Additionally, MATLAB software, known for its high capability in mathematical operations, was employed to generate future predictions and create artificial neural networks [49].

Forecasting with Artificial Neural Network

Artificial Neural Networks (ANN) are a type of machine learning techniques that draw inspiration from biological neural networks. Composed of neurons, ANN is capable of performing functions such as classification, clustering, prediction, and model building in various fields [50, 51]. The network structure comprises three fundamental components: neurons as the primary units, the network architecture formed by weighted connections between neurons, and the learning algorithm [52, 53]. One of the most significant advantages of ANN is its capacity to handle vast and complex data and generate new data. Consequently, ANN has become a popular data mining technique. The layers in ANN, formed by combining neurons, carry out various operations. ANN comprises three layers: input, hidden, and output layers and provides analytical solutions through the relationships established between them.

This study employed ANN to forecast the number of publications, citations, and research area-specific publications from 2024 to 2030, based on data from 2010 to 2023. The MATLAB programming language was used to create neural networks for future forecasting. Specifically, the Narx (Nonlinear Autoregressive with eXogenous inputs) neural network architecture was used, which is widely recognized for its effectiveness in time series analysis (Fig. 3).

The NARX neural network uses the previous values of the target variable to predict its next values. In this study, the previous three values used for prediction. Different activation functions are used to process input parameters. The hidden layer of the network contains 10 neurons that use the hyperbolic tangent function, while the output layer utilizes a linear function. The Levenberg-Marquardt algorithm was employed to train the network due to its fast-processing speed.

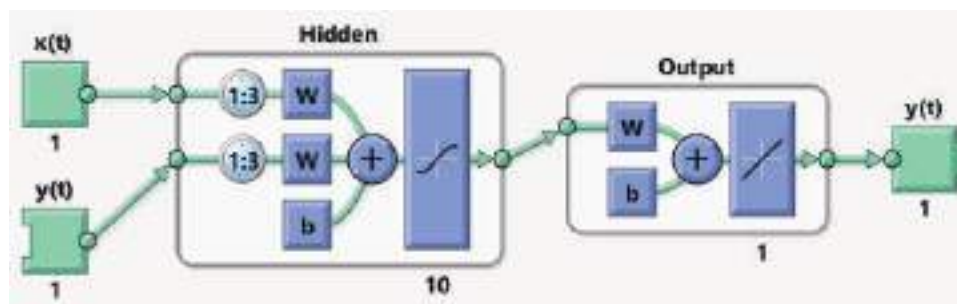


Figure 3. Narx architecture used in the study.

RESULTS AND DISCUSSION

The results of the study are presented in two main sections. The first section includes the number of publications, document types, and research areas related to renewable energy and renewable energy sources. The second section provides a detailed bibliometric analysis of studies on wind energy.

An Overview of Publications on Renewable Energy

Publications on renewable energy are found in a wide variety of fields and in different categories in the literature. In this section, renewable energy is examined from a general perspective. The information on the publications of each renewable energy type is analyzed quantitatively and evaluated.

Publication History of Renewable Energy

The graph presented as Figure 4 displays the number of renewable energy-related publications and citations by year, covering the period between 2010 and 2023. During this time, a total of 103234 publications were published, with the number of publications steadily rising from 1382 in 2010 to 17991 in 2023. The year 2023 had the highest number of publications. Over the same period, there were

a total of 2944347 citations to these publications, with the highest number of citations, 642025, occurring in 2023. As shown in the figure, both the number of publications and citations have been increasing continuously over the years, indicating that renewable energy is a popular topic of study and research. This trend suggests that studies on renewable energy will likely continue to increase in the future.

Figure 5 presents the quantitative distribution of publications for each type of renewable energy source from 2010 to 2023. The highest number of publications for “hydroelectric”, “solar”, “wind”, “biomass”, “geothermal”, and “wave” sources were made in 2023. The number of publications for each relevant year were 518, 4425, 3879, 1274, 394, and 473, respectively. As shown in the figure, there has been a general increase in the number of publications in all renewable energy types over time.

Publication Type Distribution of Renewable Energy

In WoS, all publications are classified into different document types. However, a single study can be assigned to more than one document type if it is categorized under multiple publication types. As a result, the total number of

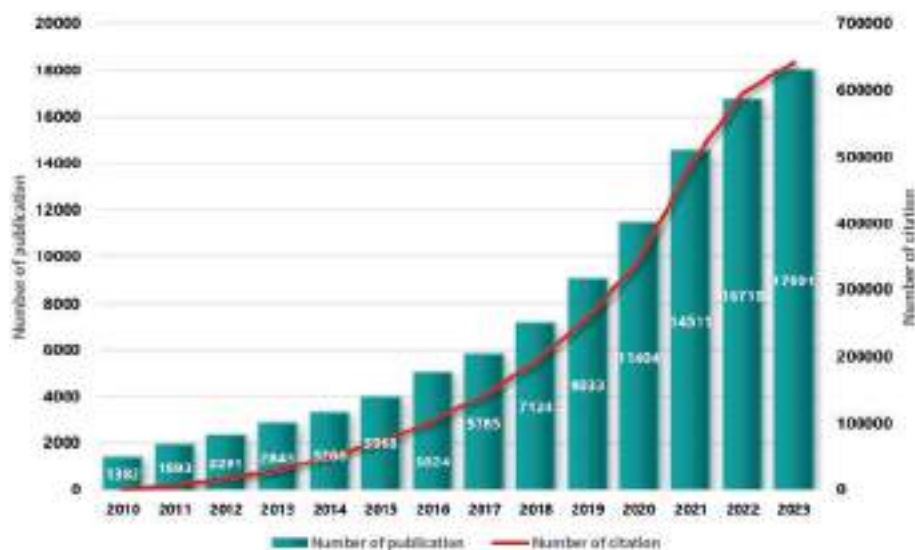


Figure 4. Number of publications and citations by year.

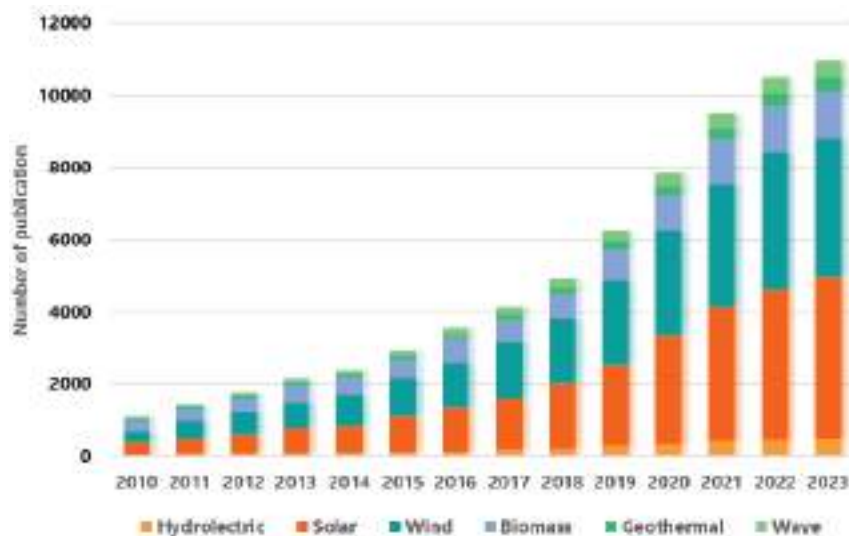


Figure 5. Number of publications of renewable energy types by year.

Table 3. Distribution of renewable energy types by document types

Document Types	Renewable Energy	Hydroelectric	Solar	Wind	Biomass	Geothermal	Wave
Article	90989	2686	21988	22308	8114	1933	2781
Review Article	12245	485	3590	2750	1844	483	452
Proceeding Paper	2328	57	608	599	251	64	85
Book Chapter	59	0	15	13	10	2	1
Other	1224	38	285	252	85	20	23

a particular document type may be affected by another document type. Table 3 shows the distribution of publications, and upon analysis, it becomes clear that articles are the most common type of publication for renewable energy and each type of renewable energy. This is followed by reviews, proceeding papers, and book chapters, in that order.

Research Area Distribution of Publications in Renewable Energy

Each publication is associated with at least one research area. However, as a publication may fall within the scope of multiple research areas, this could result in duplicate counts. Figure 6 displays numerical data regarding the top

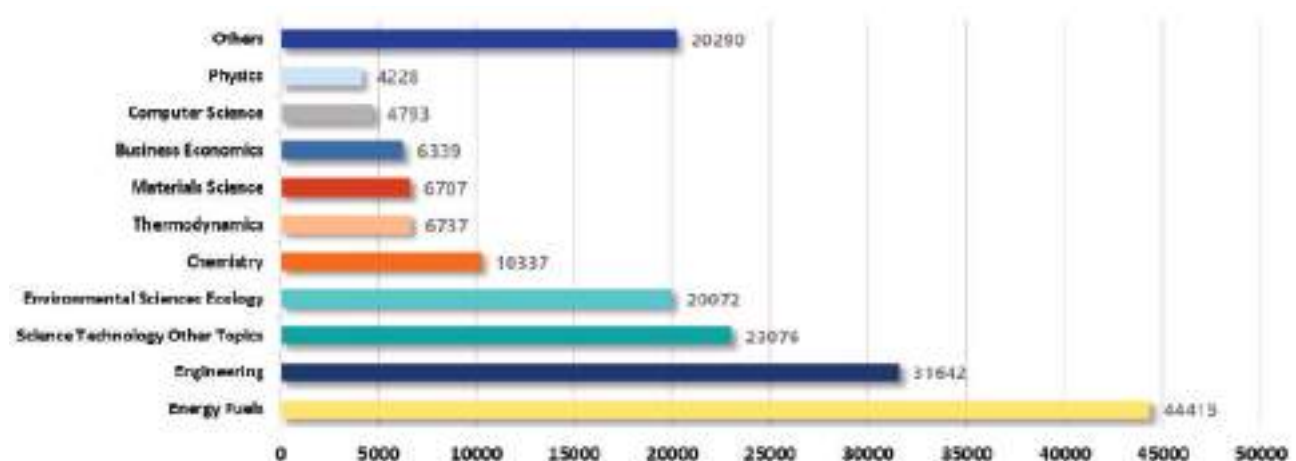


Figure 6. Top 10 research areas for renewable energy publications.

10 research areas with the highest number of publications on renewable energy studies.

The most popular research areas are Energy Fuels with 44415 publications and Engineering with 31642 publications. These areas are followed by Science Technology Other Topics with 23076 publications, Environmental Sciences Ecology with 20072 publications, and Chemistry with 10337 publications. Overall, research areas related to energy, engineering, economics, and basic sciences are stand out.

Evaluation of Wind Energy with Statistical Concepts

In this section, a comprehensive bibliometric analysis of wind energy studies in the literature is presented in a detailed and effective manner. The publications between 2010 and 2023 were evaluated using the WOS database for the analysis. Through the parameters examined, trends, clusters, networks, and frequency statistics were investigated and analyzed.

The distribution of publications over time between 2010 and 2023 was examined to evaluate the trends and changes from the past to the present. The groups formed based on the relationship between the parameters of the publications were analyzed using cluster analysis, and the characteristics of these groups were investigated. Networks played a crucial role in the formation of clusters and the relationships between them. By evaluating the networks, the bases of these relationships were investigated and inferences were

made. Frequency values were also analyzed to investigate the repetitiveness and percentage distribution of the parameters. To present all these statistical analyses, the data used and the outputs obtained were visualized and interpreted using tables, graphs, maps, network structures, and figures.

Quantitative Information and Temporal Distribution of Publications

To investigate information on wind energy publications, filtering was carried out from the WoS database, and Section 2 provides detailed information on the filtering process. Quantitative information on all publication types between 2010-2023 shows a total of 24907 publications, with 22177 being articles, 597 proceeding papers, 2730 reviews, 13 book chapters, and 5707 being other types of publications. Most of these studies are articles, followed by reviews and proceeding papers, respectively. Details of the quantitative information of each publication type are given in Figure 7.

Figure 7 clearly shows the continuous increase in the number of publications from 2010 to 2023. Notably, there has been a significant rise in the number of publications, especially after 2018.

In the literature review, it was observed that although there are numerous studies on wind energy, articles are the most used publication type. Analyzing this type of publication will provide more comprehensive information. Therefore, the analysis in the rest of the study focuses on

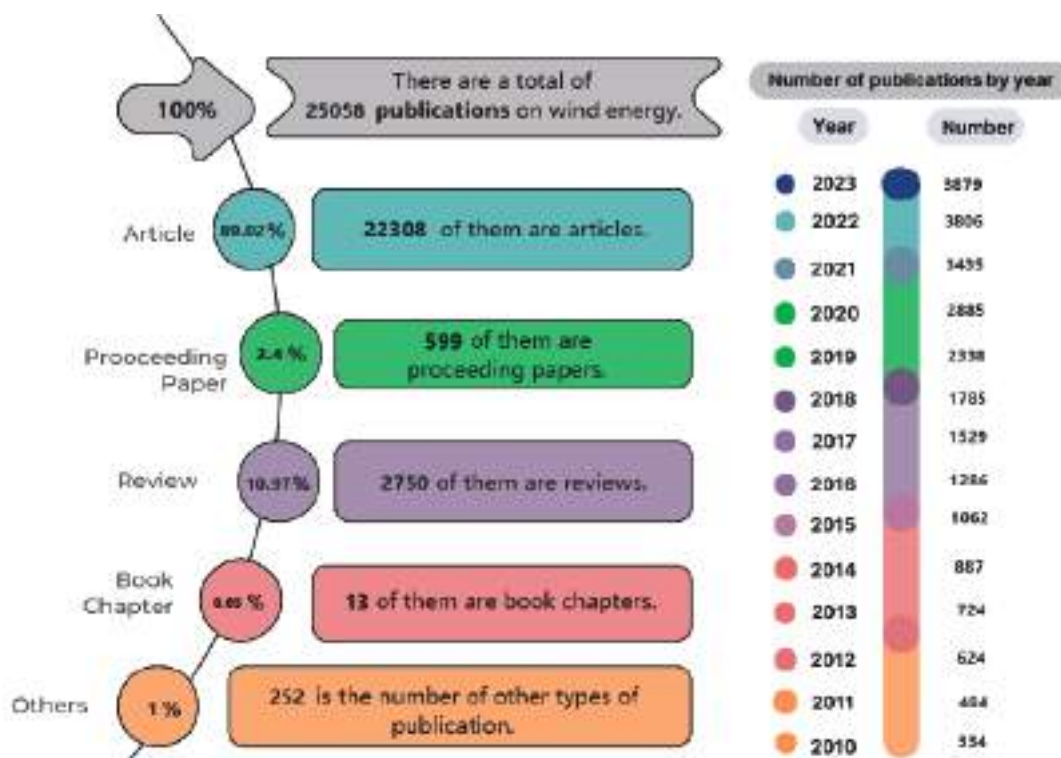


Figure 7. Number of publications published on wind energy and their distribution by year between 2010-2023.

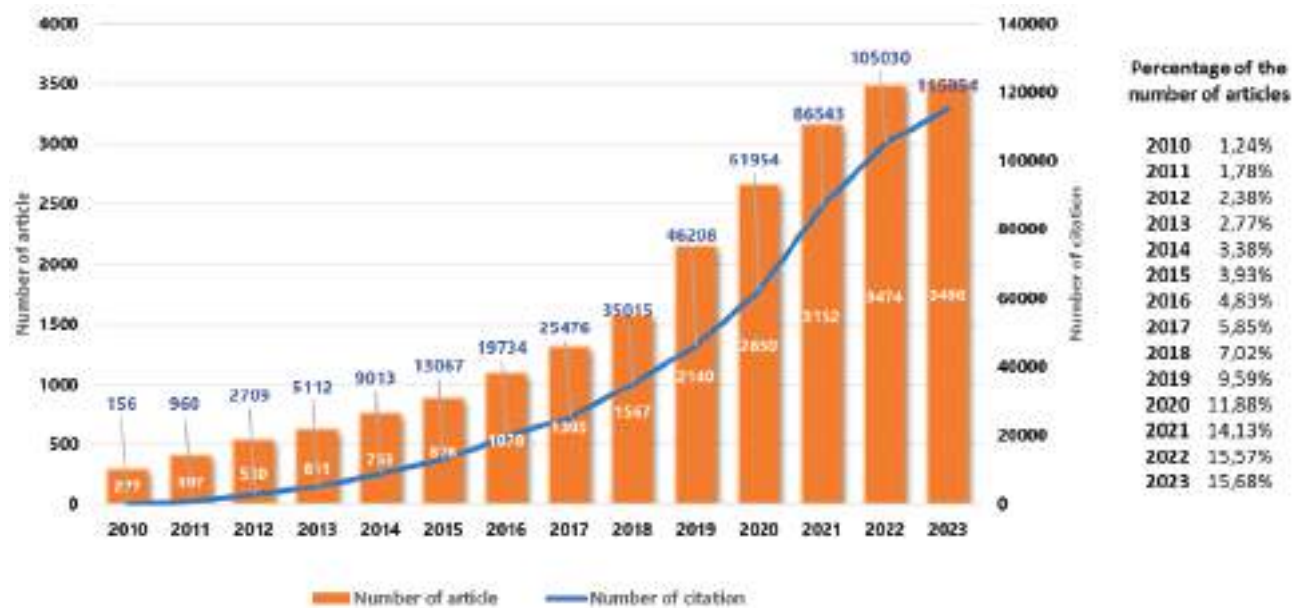


Figure 8. Number, percentage, and citations of articles published on wind energy between 2010 and 2023.

articles published between 2010 and 2023. Figure 8 illustrates the distribution of articles over time. The graph clearly indicates that the number of articles has been continuously increasing over the years. It is important to note that some of the proceeding papers are also included in the number of articles. As papers presented at conferences and symposiums are published in journals, they are reflected in the numerical information of both articles and proceeding papers. On the other hand, the number of citations is constantly increasing in proportion to the number of articles.

Most Productive Authors and Citations Analysis

The analysis of authors working in a research area or subject can provide insights into the popularity, development, and leading researchers of the area. To this end,

a total of 47030 authors were identified in the 22308 articles analyzed in the field of wind energy. The high number of authors suggests significant interest in the subject and its research. Notably, 200 authors have more than 16 articles, highlighting the prominence and importance of wind energy research in contemporary times.

Table 4 shows the most productive authors based on the number of publications on wind energy. According to this table, Liu Y (82 papers) from the China University of Petroleum, Kamel S (78 papers) from Aswan University, and Li Y (74 papers) from the Xi'an Jiaotong University are the top three authors. Their universities are located in People's Republic of China, Egypt, and People's Republic of China, respectively. When examining the h-index, which

Table 4. Top 10 most productive authors

Number of Publications	Author	h-index	Institution	Country
82	Liu Y	23	China University of Petroleum	People's Republic of China
78	Kamel S	25	Aswan University	Egypt
74	Li Y	21	Xi'an Jiaotong University	People's Republic of China
69	Chen Z	31	Aalborg University	Denmark
66	Kumar A	21	Maharishi Markandeshwar University	India
66	Senjyu T	21	Univeristy of the Ryukyus	Japan
65	Mohammadi-ivatloo B	29	University of Tabriz	Iran
63	Breyer C	37	Lappeenranta University of Technology	Finland
59	Wang Y	29	Nanjing University	People's Republic of China
56	Zhang Y	26	Hohai University	People's Republic of China

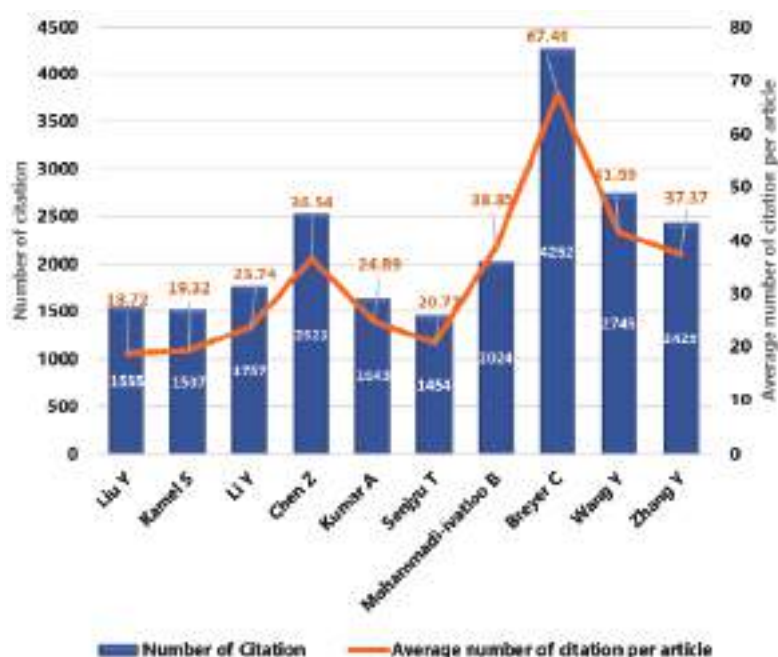


Figure 9. Citation frequencies of the top 10 most productive authors.

provides information about the productivity of the author, it is revealed that the value is 23, 25, and 21 for the top three authors, respectively. Figure 9 shows the citation information of the top 10 most productive authors' articles on wind energy.

When considering Table 4 and Figure 9 together, it becomes apparent that the authors are affiliated with various universities, with four authors from People's Republic of China and the rest from different countries. Among the authors, Breyer C stands out with an h-index value of 37. Examining the number of citations, it is evident that Breyer C has the highest citation count with 4252 citations. Breyer C, who is associated with Lappeenranta University of Technology in Finland, has published 63 articles in the wind energy literature.

Figure 10 shows a network of connections and clusters among authors and their collaborating researchers. Each cluster is represented by a different color, the prominent node of each cluster indicates the lead author of the cluster, with node size indicating the number of publications. The network is based on authors who have published more than 5 articles and received more than 5 citations on wind energy, resulting in 39 clusters. Nodes with the same color represent authors who have collaborated. The three most prominent clusters, indicated by the colors red, green, and blue, have 77, 54, and 46 authors, respectively. Furthermore, the top three authors with the most links are Kamel S, Senjyu T, and Breyer C, with 164, 166, and 135 links, respectively.

Table 5 presents information on the top 10 most cited articles on wind energy. The first three articles have 2401,

2232, and 1182 citations, respectively. In 2015, an article was published in *Applied Energy* by Luo X et al. [54], in which they reviewed and evaluated the technology of electrical energy storage systems. In 2019, Gielen D. et al. [55] conducted research on the energy transition of renewable energy technology from today to 2050, which was published in the journal *Energy Strategy Reviews*. In 2010, Whipple DT and Kenis PJA [56] presented a number of recommendations for the problem of climate change and sustainable energy production in the *Journal of Physical Chemistry Letters*. Looking at the table, the top 10 most cited articles were mostly published in 2012 and 2015. Additionally, *Applied Energy* and *Energy Strategy Reviews* are the prominent journals in this table.

Most Prominent Countries, Institutions and Their Spatial Distribution

In the wind energy literature, there are a total of 165 countries with published articles. Figure 11 presents the top 25 countries with the highest number of articles, with People's Republic of China leading at 4476, followed by the USA with 3186 and the United Kingdom with 1971. The map provides a clear view of the spatial distribution of the number of articles about countries, regions and continents, revealing that many countries are working on this topic. Notably, the North American and Australian continents, European continent, and the southern region of the Asian continent stand out in terms of the number of publications. Conversely, the African continent, the northern region of the Asian continent, and the

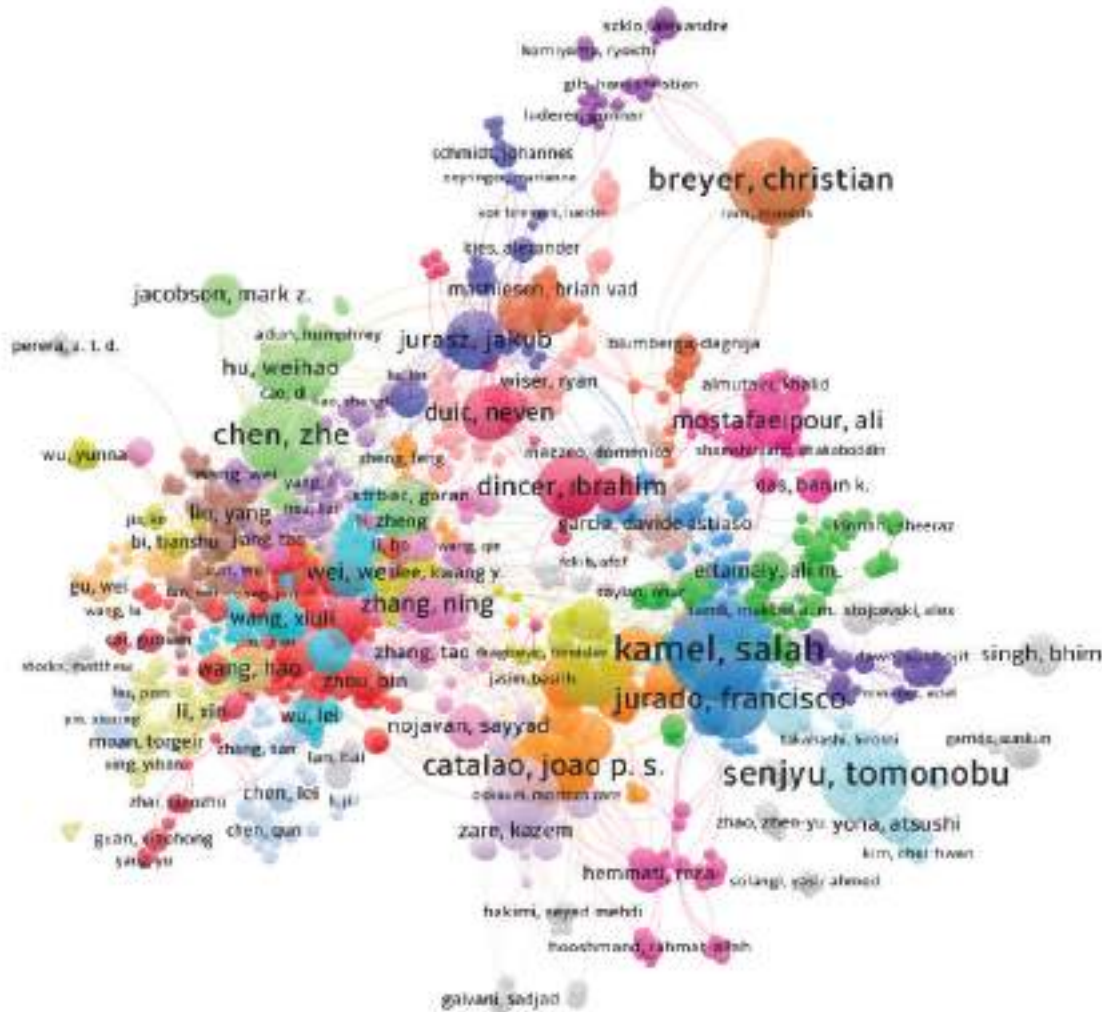


Figure 10. Network of co-authorship of authors.

southern region of the South American continent have fewer publications.

Authors can collaborate from the same country or different countries. Figure 12 depicts the links, clusters, and number of co-authorship articles between countries. To ensure visibility, countries with at least 5 articles and 5 citations were included in the network. Out of the 165 countries analyzed, 117 countries met this requirement. Clusters are represented by different colors, with a total of ten clusters. The leading country of each cluster is indicated by the prominent node, and the node size represents the number of publications. The first three clusters with the most countries are represented by the colors red, green, and blue, respectively, with 25, 21, and 14 countries in each cluster. On the other hand, People's Republic of China, USA, and England have the most links to authors in other countries with 2435, 2234, and 1649 links, respectively. The lines between nodes illustrate the co-author links between countries.

According to the literature, there are a total of 12050 institutions working on wind energy. Table 6 shows the top 10 universities with the most publications. The top three universities are North China Electric Power University in People's Republic of China with 534 articles, the National Institute of Technology in India with 409 articles, and Islamic Azad University in Iranian with 358 articles. Upon examining the table, it is seen that most of the universities in the top 10 are in People's Republic of China, Denmark, and India.

Collaborations between authors from different institutions can be analyzed. Figure 13 depicts a network that shows the links, clusters, and number of publications co-authored between institutions. This network was created using institutions with at least 10 articles and 10 citations, resulting in 940 institutions out of 12050 meeting the criteria. The network consists of 15 clusters, represented by different colors, where the prominent node of each cluster indicates the leading institution

Table 5. Top 10 most cited articles

Number of Citation	Authors	Title	Publication Year	Journal
2401	Luo, X et al. [54]	Overview of current development in electrical energy storage technologies and the application potential in power system operation	2015	Applied Energy
2232	Gielen, D. Et al. [55]	The role of renewable energy in the global energy transformation	2019	Energy Strategy Reviews
1182	Whipple, DT and Kenis PJA [56]	Prospects of co2 utilization via direct heterogeneous electrochemical reduction	2010	Journal of Physical Chemistry Letters
1010	Ursua, A; Gandia, LM and Sanchis, P [57]	Hydrogen production from water electrolysis: current status and future trends	2012	Proceedings of the IEEE
969	Atwa, YM et al. [58]	Optimal renewable resources mix for distribution system energy loss minimization	2010	IEEE Transactions on Power Systems
821	Zhong et al. [59]	Accelerated discovery of CO2 electrocatalysts using active machine learning	2020	Nature
812	Lin, KX et al. [60]	Alkaline quinone flow battery	2015	Science
802	Yan, Y et al. [61]	A survey on smart grid communication infrastructures: motivations, requirements and challenges	2013	IEEE communications surveys & tutorials
800	Mathiesen, BV et al. [62]	Smart energy systems for coherent 100% renewable energy and transport solutions	2015	Applied Energy
760	Hayashi, A et al. [63]	Superionic glass-ceramic electrolytes for room-temperature rechargeable sodium batteries	2012	Nature Communications

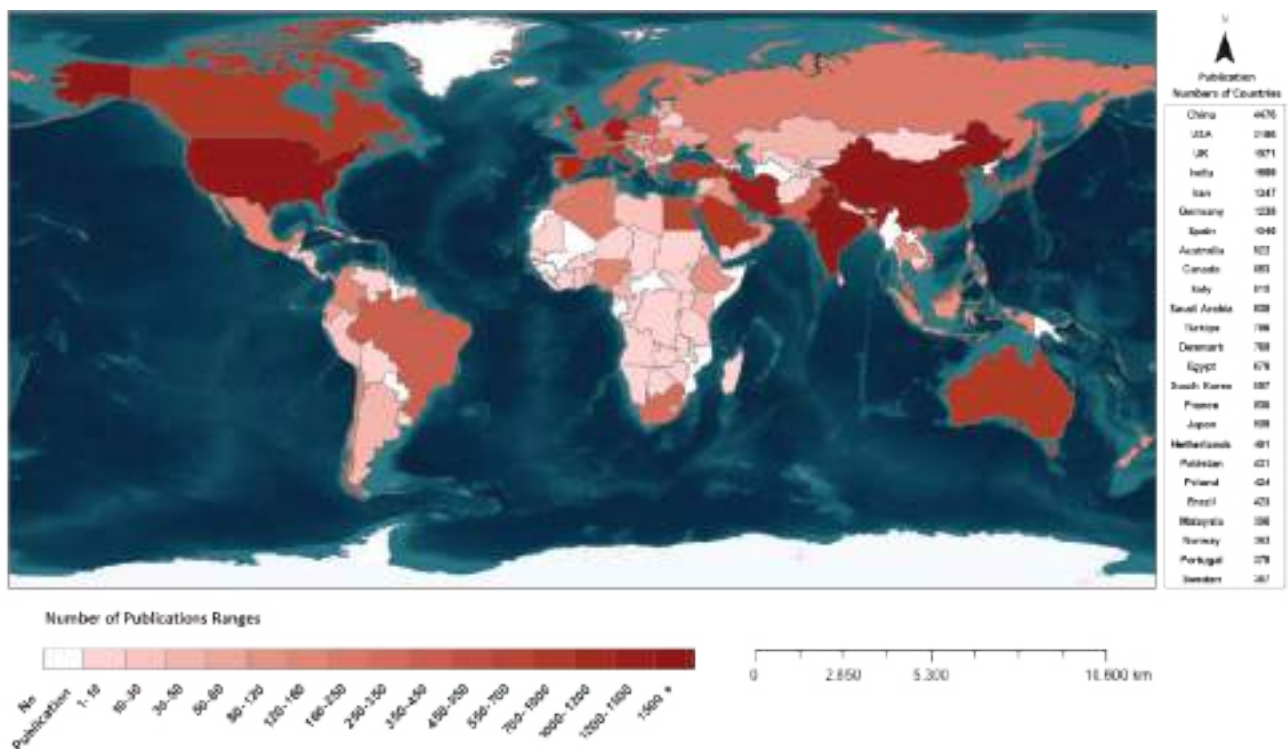


Figure 11. Spatial distribution of the number of articles by countries.

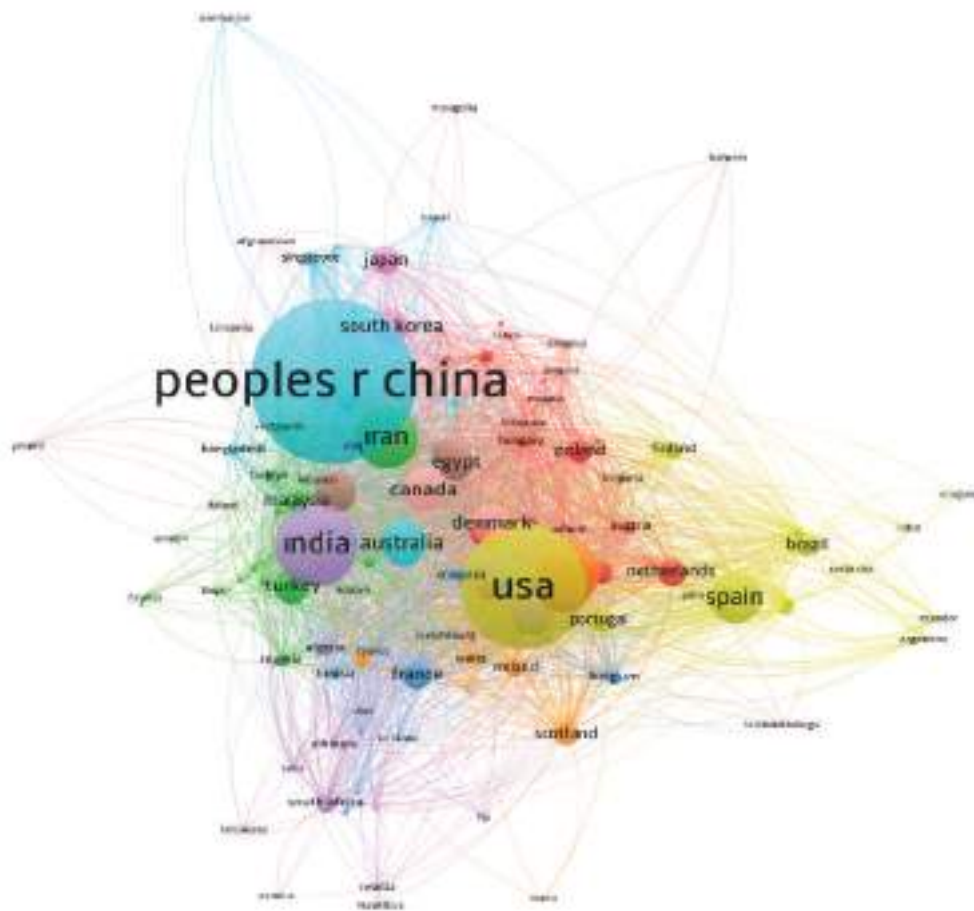


Figure 12. The co-author network of countries.

Table 6. Top 10 most prominent institutions

Institutions	Country	Number of Publications
North China Electric Power University	People's Republic of China	534
National Institute of Technology	India	409
Islamic Azad University	Iranian	358
Aalborg University	Denmark	346
Indian Institute of Technology System	India	311
Tsinghua University	People's Republic of China	285
Technical University of Denmark	Denmark	281
University of California System	USA	212
Xi'an Jiaotong University	People's Republic of China	178
Norwegian University of Science and Technology	Norway	166

of the cluster, and the size of the node symbolizes the number of publications. The first three clusters with the most institutions are red, green, and blue, respectively, consisting of 141, 137, and 110 institutions. The top three institutions with the most links to authors from other institutions are Aalborg University in Denmark

with 397 links, Islamic Azad University in Iran with 370 links, and North China Electric Power University in People's Republic of China with 336 links. The lines between the nodes indicate the co-author links between the institutions.

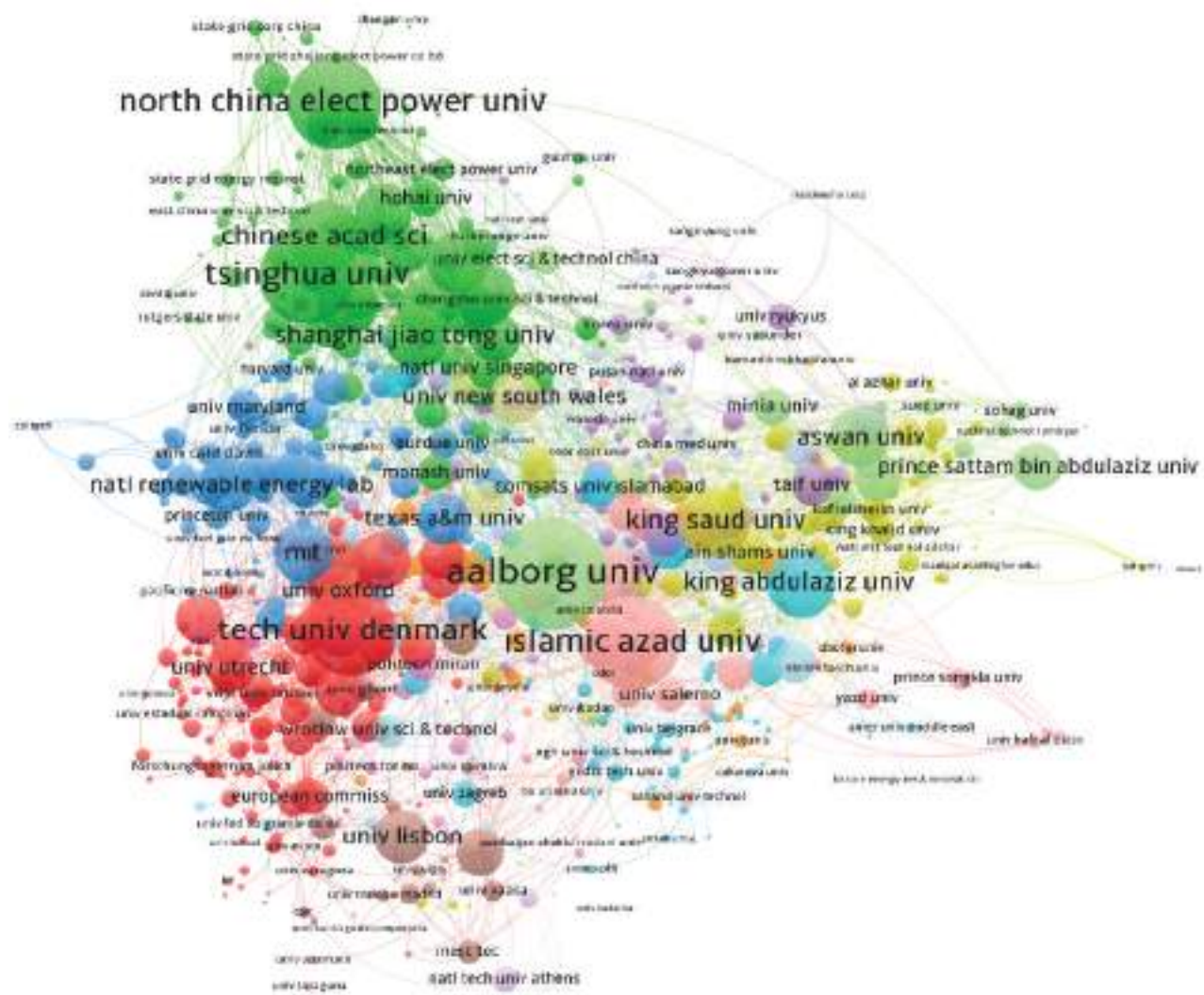


Figure 13. The co-author network of institutions.

Priority Research Areas, Leading Journals and Prominent Publisher

Wind energy has been studied in various research areas, and an examination of the WoS database shows that there are 93 research areas related to this topic. The top 10 research areas with the most studies are displayed in Figure 14. The most researched areas are Energy Fuels with 12915 articles, Engineering with 8366 articles, and Science Technology Other Topics with 6068 articles, while many studies have also been conducted in Environmental Sciences Ecology. It is evident from Figure 14 that wind energy is a multidisciplinary topic that covers various research areas.

Table 7 shows the top 10 journals that have published the most articles on wind energy. The three leading journals are Energies with 1504 articles, Renewable Energy with 1149 articles, and Energy with 1141 articles. It is evident from the establishment years of the journals that most of them have been operating for many years and are considered the

leading journals in the field. In terms of impact factors, Applied Energy, Journal of Cleaner Production, and Energy Conversion and Management hold the top three positions with 11.2, 11.1 and 10.4, respectively. This indicates that these journals are highly influential, prestigious, and successful in their respective areas of research.

Another criterion for understanding the impact of journals is examining their number of citations. Figure 15 shows the citation information for the top 10 journals. The three journals with the highest citation values are Applied Energy with 47086 citations, Renewable Energy with 44329 citations, and Energy with 43275 citations. It is evident from their high citation values that these top three journals are competent and core journals in this field. Moreover, they are also among the journals with the highest number of publications.

Table 8 shows the top 10 publishers that have published the most articles on wind energy. It can be observed that

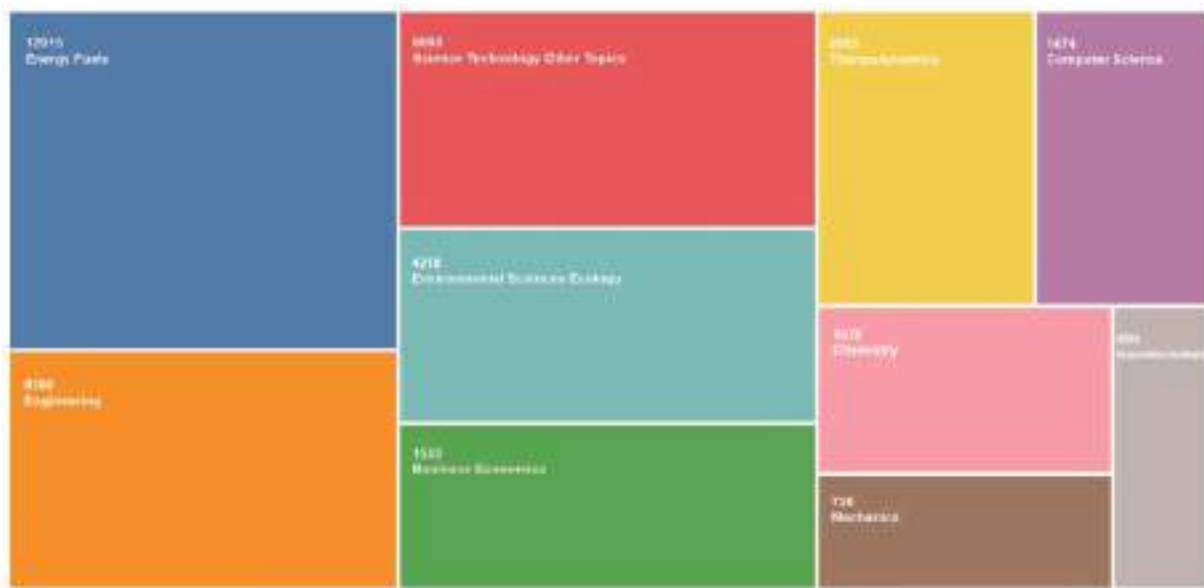


Figure 14. Top 10 priority research areas.

Table 7. Leading journals in the Top 10 by number of publications

Journal	Number of Publication	Foundation Year	IF (2023)
Energies	1504	2008	3.2
Renewable Energy	1149	1991	8.7
Energy	1141	1976	8.9
Applied Energy	879	1975	11.2
Energy Policy	719	1973	9
Sustainability	605	2009	3.9
Energy Conversion and Management	494	1961	10.4
IEEE Access	466	2013	3.9
Journal of Cleaner Production	437	1993	11.1
International Journal of Hydrogen Energy (IJHE)	356	1976	7.2

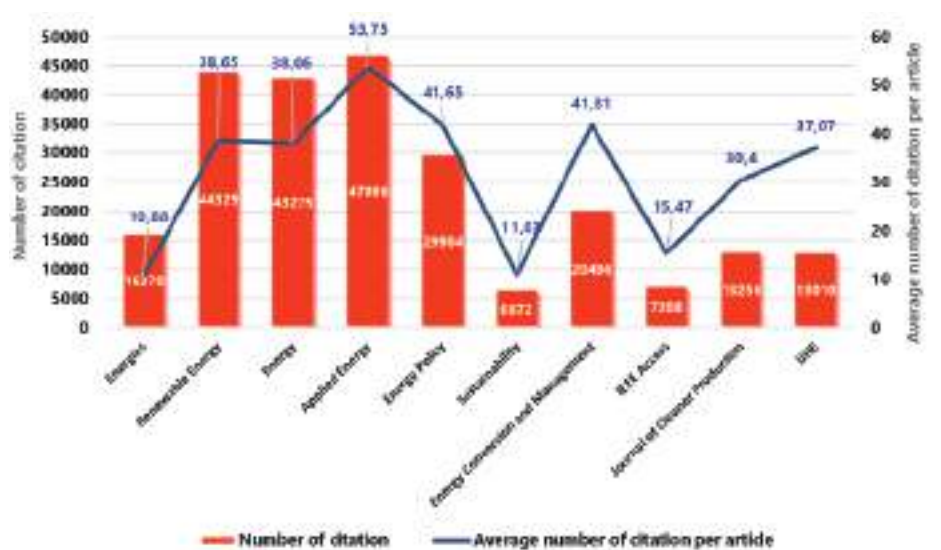


Figure 15. The citation frequencies of the leading journals in the top 10 by the number of publications.

can be inferred from this that these keywords form the basis for studies on wind energy.

To see other keywords related to wind energy, a wind rose visualization was created and presented in Figure 16. In this context, the prominence of keywords for different disciplines indicates that multidisciplinary studies are being carried out on wind energy, while at the same time providing an understanding of other fields in which the subject is directed.

To evaluate the co-occurrence of keywords, a network was created as shown in Figure 17. The examination of the literature revealed the usage of a total of 43148 keywords. To create the network, it was assumed that each keyword can co-occur with at least 5 other keywords, based on the approximate number of keywords used in an article. Accordingly, a network comprising of 3030 keywords was generated.

The network shown in Figure 17 consists of 9 clusters, each represented by a different color. The leading keyword of each cluster is highlighted in a prominent frame, while the size of the frame indicates the frequency of co-occurrence with other words. The first three clusters formed by

the most frequently co-occurring keywords are colored red, green, and blue, respectively, and contain 275, 172, and 168 keywords, respectively. The keywords that most commonly co-occur with other keywords are renewable energy with 10737 links, renewable energy sources with 5844 links, and wind energy with 4036 links. Additionally, the figure shows other prominent keywords, such as wind turbine, energy transition, energy policy, energy storage, deep learning, optimization, and power system stability.

Changes, Evaluation, Comparisons and Future Research

As a result of the bibliometric analysis conducted in the previous section, the studies on wind energy between 2010 and 2023 have been comprehensively analyzed and evaluated in detail. In this evaluation, publications were analyzed both based on each year and holistically. Thus, a general evaluation of the determined year interval was made.

This section divided the evaluated year interval into three periods and analyzed the number of publications and citations in each period to monitor the changes and development in wind energy. Additionally, predictions for the future of wind energy studies were made using artificial

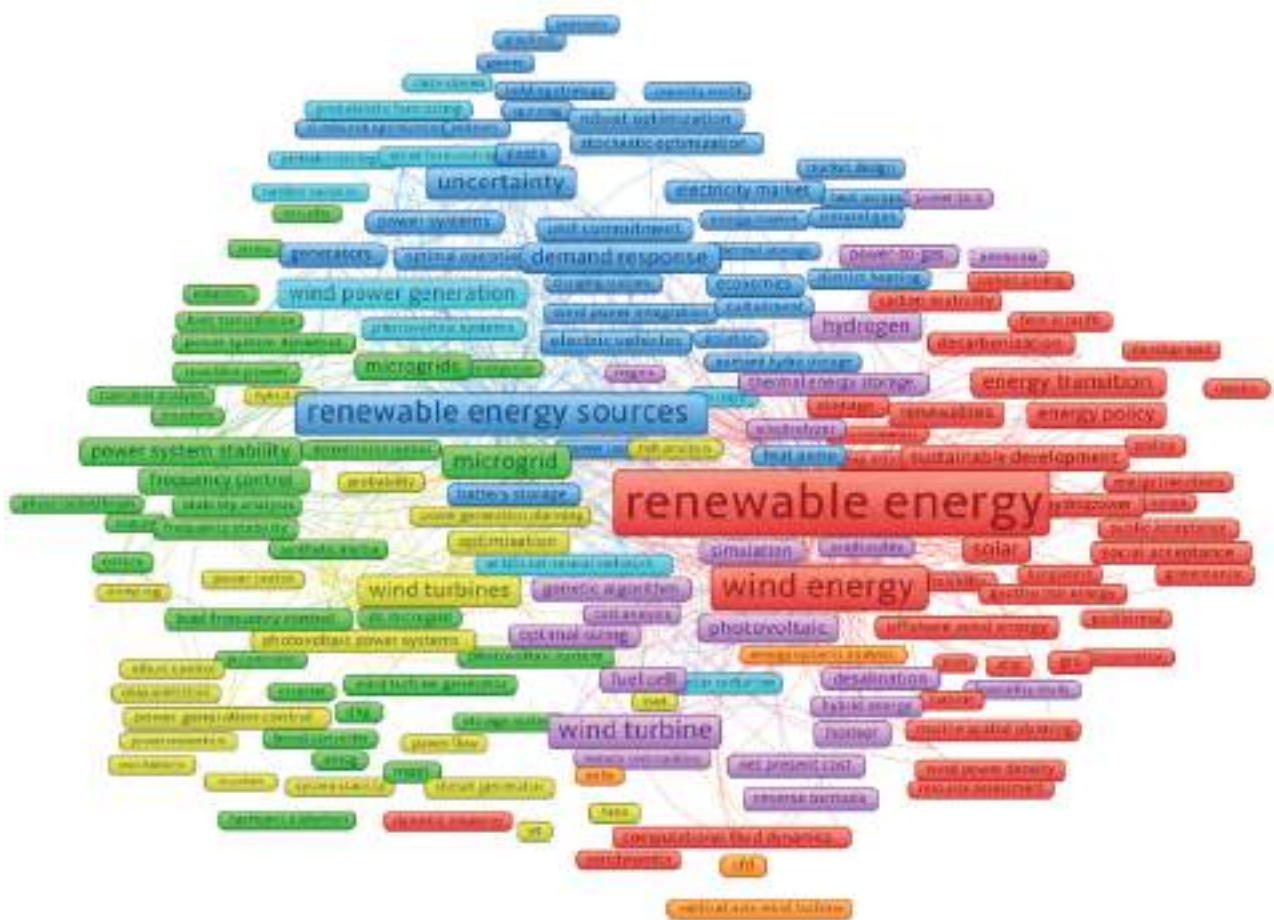


Figure 17. Network of co-occurrences of keywords.

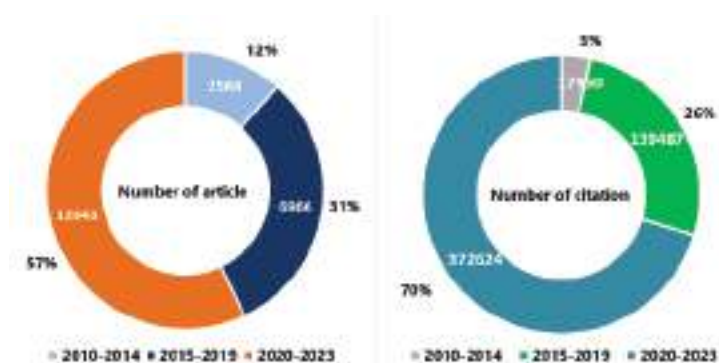


Figure 18a. Number of publications for the years 2010-2014, 2015-2019, 2020-2023,



Figure 18b. Number of citations for the years 2010-2014, 2015-2019, 2020-2023.

neural networks, a data mining method. The predictions made using ANN include the number of publications and citations between 2024 and 2030, as well as the number of studies that could potentially be published in the top 10 research areas with the most studies. Based on these predictions, evaluations were made regarding the future direction of wind energy.

Figure 18a and Figure 18b analyzes quantitative information on the number of publications and citations between 2010 and 2023, divided into three periods. Figure 18a shows that the number of publications has increased significantly over time. In the second period, the number of publications was almost three times that of the first period, while in the third period it was almost five times that of the first period and almost twice that of the second period. The trend suggests that wind energy has become an increasingly prominent topic, especially between 2020 and 2023. Figure 18b reflects the effects of the data obtained in Figure 18a, showing the number of citations between 2010 and 2023, divided into the same three periods. The data indicates that the number of citations has also increased considerably over time. In the second period, the number of citations is more than nine times higher than in the first period. In the third period, the number of citations is almost three times that of

the second period and more than twenty-three times that of the first period. These findings indicate that the number of studies has increased considerably in the last period and that publications in this field have been cited from different areas of study.

Figure 19 illustrates the prediction results obtained by the ANN for the number of publications and citations expected between 2024 and 2030. The graph indicates that both the number of publications and citations will continue to increase steadily. These findings suggest that wind energy will become increasingly popular in the years to come and will continue to evolve as a field of study with the increase in research.

Figure 20 depicts the predicted number of publications from an ANN analysis, providing insight into the top 10 research areas with the most published and prominent research from 2024 to 2030. As shown in the figure, the number of publications is expected to increase in all research areas. Notably, Energy Fuels is projected to be the leading research area during this period, followed by Engineering, Science Technology Other Topics, and Environmental Science Ecology.

Table 10 displays the predicted percentages of publications for the top 10 fields from 2024 to 2030. The table

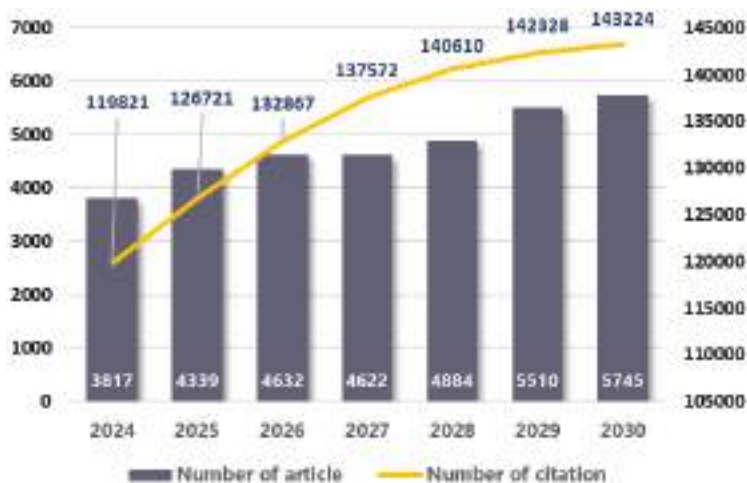


Figure 19. Forecasting of number of publication and citation between 2024-2030 with ANN.

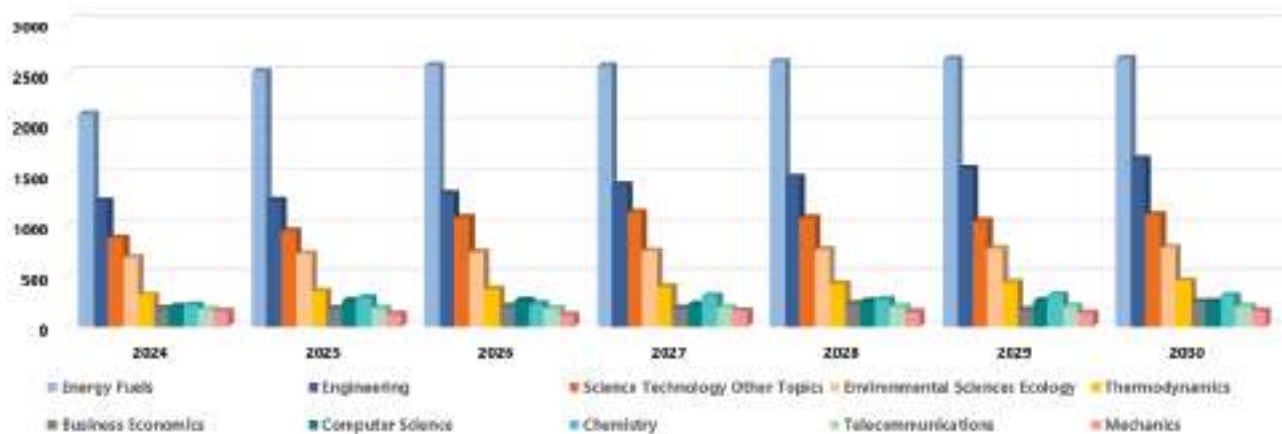


Figure 20. Forecasting of the number of publications between 2024-2030 in the top 10 research areas with ANN.

provides for a more effective observation of the differences between fields and the expected trends over the years. Based on the percentage changes, it is expected that research in Engineering, Science Technology Other Topics, Environmental Science Ecology, Chemistry, Telecommunication, and Mechanics will increase in the future, potentially becoming more prominent.

Multidisciplinary Approach to Wind Energy

Renewable energy is a versatile resource that encompasses various types of energy and interacts with many factors across multiple disciplines. The use and development of these resources depend on complex and dynamic relationships between disciplines. When considering the versatility of wind energy, various areas become prominent, including the environment, engineering, land management, economy, politics, and artificial intelligence.

The use of renewable energy sources, which are inexhaustible and have the potential to mitigate environmental

problems like global warming, climate change, and air pollution, is rapidly increasing as an alternative to fossil fuels. Wind energy, one of the green energy sources, plays an important role in combating these problems. Since wind energy systems do not require the use of fossil fuels during the electricity generation process, they help to reduce greenhouse gas emissions. This, in turn, provides a solution to the adverse effects of these problems while laying the foundation for a sustainable energy future.

The installation of wind energy systems requires collaboration with various engineering disciplines during processes such as design, planning, and utilization. These fields include environmental engineering, which analyzes environmental impacts and creates environmentally friendly designs; mechanical engineering, which focuses on the design, testing, and development of wind turbines; electrical-electronic engineering, which designs and controls the systems necessary for the generation of electricity, efficient

Table 10. Percentages of the number of publications forecasts of the research areas between 2024-2030

Year	Energy Fuels	Eng.	Sci. Tech. Other Top.	Env. Sci. Eco.	Thermo.	Business Eco.	Computer Sci.	Chemistry	Mech.	Telecom.
2024	34.10	20.19	14.18	11.18	5.14	2.98	3.29	3.53	2.88	2.53
2025	36.79	18.30	13.92	10.62	5.23	2.66	3.73	4.11	2.73	1.91
2026	36.38	18.57	15.20	10.45	5.37	2.89	3.62	3.24	2.60	1.69
2027	35.10	19.09	15.38	10.26	5.47	2.58	3.06	4.20	2.66	2.20
2028	35.08	19.85	14.36	10.24	5.71	3.06	3.40	3.58	2.74	2.00
2029	34.95	20.71	13.79	10.23	5.84	2.30	3.37	4.24	2.72	1.85
2030	33.89	21.14	14.12	10.01	5.84	3.16	3.15	3.98	2.63	2.07

storage, transportation, distribution of energy obtained from turbines, and minimization of energy losses; materials engineering, which develops lightweight and durable materials; structural engineering, which performs the structural analyses necessary for the installation and operation of the system; and finally, geomatics engineering, which selects the most suitable location for the installation of these systems to operate with maximum efficiency.

Selecting the suitable location for wind energy systems is crucial for achieving high energy production potential and maximum environmental protection [64, 65]. Land management is a discipline utilized at this stage, which enables the effective use of land resources, mainly soil, in an environmentally and economically sustainable manner by making and implementing appropriate decisions [66]. Geomatics engineers play a crucial role in determining the site selection according to the principles of land management. Today, Geographic Information Systems (GIS), which enable all these principles and information from other disciplines to be collected, queried and evaluated holistically, have become a frequently used tool for site selection studies [67-70]. GIS tools, which are used effectively by geomatics engineers, provide great easiness in site selection studies.

On the other hand, one of the principles of land management is that a land can be effectively used for more than one purpose, resulting in efficient land use. Installing wind energy systems in areas suitable for other activities such as agriculture, animal husbandry, industry, and tourism means that the lands are used to their full potential, contributing to the economy and creating new areas of employment. These factors highlight the significant impact of wind energy on the economy. Investments in wind energy not only increase economic growth but also support sustainable development [71]. Therefore, countries emphasize wind energy and other renewable energy types by offering various incentives in their energy policies [72].

Recently, artificial intelligence technologies have been increasingly utilized in the renewable energy sector, including the wind energy field. AI allows for designing facilities that can provide high potential and maximum performance, as well as making them smarter and easier to monitor. This enhances the safety and sustainability of the systems. As

AI technologies continue to develop and spread, they will play an increasingly important role in optimizing the use of energy resources in the future.

To provide a comprehensive analysis of the scientific areas related to wind energy, a bibliometric analysis was conducted, examining various fields. Today, the fields of Energy Fuels, Engineering, Science Technology Other Topics, and Environmental Science Ecology are prominent in wind energy research, and they are expected to remain the focus in the future. With the growing popularity of AI, it is believed that AI will also play a role in wind energy research and shape the field in the future.

CONCLUSION

This comprehensive bibliometric analysis highlights the rapid growth and evolving trends in wind energy research from 2010 to 2023. By employing advanced analytical tools and predictive models, this study offers valuable insights into the past, present, and future of wind energy studies. Our findings underscore the increasing prominence of wind energy in scientific research and its critical role in sustainable development. Future research should continue to explore multidisciplinary approaches, integrating technological, environmental, and economic perspectives to advance the field further.

First, the study examined the general distribution, publication and citation numbers of 24907 wind energy studies conducted between 2010-2023. Then, the focus narrowed to article types, where 22177 articles were analyzed using Vosviewer, ArcGIS, and Excel programs. The evaluation included factors such as author, citation, institution, country, journal, publisher, research area, and keywords. The analysis looked at each factor individually and their common connections. Prominent author *Liu Y*, most cited author *Breyer C*, and most cited publication “*Overview of current development in electrical energy storage technologies and the application potential in power system operation*” emerged from the analysis. *People’s Republic of China* was the most prominent country, and *North China Electric Power University* was the most cited institution. *Energy Fuels* was the most studied research area, *Energies* was the

most widely published journal, and *Elsevier* was the most used publisher. *Renewable energy* was the most used keyword. The analysis generated statistical outputs, which were presented through graphs, networks, and maps to visualize trends, clusters, spatial distributions, and connections among wind energy studies.

To examine the development process of wind energy, the dataset was evaluated in three periods based on the number of publications and citations. Notably, there was an increase in the number of publications and citations between 2019 and 2022. To forecast the future direction of wind energy, the Artificial Neural Network method was used to predict the number of publications, citations, and research areas from 2024 to 2030. The results of the forecasting study suggest a continuous increase in the number of publications and citations. In terms of research areas, the study predicts an increase in the field of Energy Fuels, while the fields of Environmental Science Ecology, Chemistry, Telecommunication, and Mechanics may also become more prominent. Finally, wind energy was evaluated from a multidisciplinary and broad perspective, considering both the present and the future.

AUTHORSHIP CONTRIBUTION

Nihal Genc: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing - original draft;

Ferdi Ozbilgin: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing - original draft;

Volkan Baser: Conceptualization; Methodology; Supervision; 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] McKeon R. The basic works of Aristotle. New York: Modern Library; 2009.
- [2] Smil V. Energy: a beginner's guide. London: Simon and Schuster; 2017.
- [3] Cleveland CJ, Morris CG. Handbook of energy: chronologies, top ten lists, and word clouds. Amsterdam: Elsevier; 2013.
- [4] Quaschnig V. Understanding renewable energy systems. London: Routledge; 2016. [\[CrossRef\]](#)
- [5] Kabak M, Taşkınöz G. Determination of the installation sites of wind power plants with spatial analysis: a model proposal. Sigma J Eng Nat Sci 2020;38:441–457.
- [6] Bhattacharyya SC. Energy economics: concepts, issues, markets and governance. Cham: Springer Nature; 2019. [\[CrossRef\]](#)
- [7] Kelly RA. Energy supply and renewable resources. New York: Infobase Publishing; 2010.
- [8] Panwar NL, Kaushik SC, Kothari S. Role of renewable energy sources in environmental protection: a review. Renew Sustain Energy Rev 2011;15:1513–1524. [\[CrossRef\]](#)
- [9] Twidell J. Renewable energy resources. London: Routledge; 2021. [\[CrossRef\]](#)
- [10] International Renewable Energy Agency (IRENA). Renewable energy statistics. Abu Dhabi: IRENA; 2022.
- [11] Sorensen B. Renewable energy: physics, engineering, environmental impacts, economics and planning. Cambridge (MA): Academic Press; 2017.
- [12] Karunathilake H, Hewage K, Mérida W, Sadiq R. Renewable energy selection for net-zero energy communities: life cycle based decision making under uncertainty. Renew Energy 2019;130:558–573. [\[CrossRef\]](#)
- [13] Bortoluzzi M, de Souza CC, Furlan M. Bibliometric analysis of renewable energy types using key performance indicators and multicriteria decision models. Renew Sustain Energy Rev 2021;143:110958. [\[CrossRef\]](#)
- [14] Pang Y, Pan L, Zhang J, Chen J, Dong Y, Sun H. Integrated sizing and scheduling of an off-grid integrated energy system for an isolated renewable energy hydrogen refueling station. Appl Energy 2022;323:119573. [\[CrossRef\]](#)
- [15] Ellabban O, Abu-Rub H, Blaabjerg F. Renewable energy resources: current status, future prospects and their enabling technology. Renew Sustain Energy Rev 2014;39:748–764. [\[CrossRef\]](#)
- [16] Rahman A, Farrok O, Haque MM. Environmental impact of renewable energy source based electrical power plants: solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. Renew Sustain Energy Rev 2022;161:112279. [\[CrossRef\]](#)

- [17] Dickson M, Fanelli M. Geothermal energy: utilization and technologies. Paris: UNESCO; 2003.
- [18] Saidur R, Abdelaziz E, Demirbas A, Hossain M, Mekhilef S. A review on biomass as a fuel for boilers. *Renew Sustain Energy Rev* 2011;15:2262–2289. [\[CrossRef\]](#)
- [19] Charlier RH, Finkl CW. Ocean energy: tide and tidal power. Berlin: Springer; 2009.
- [20] Jacobson MZ, Delucchi MA. Providing all global energy with wind, water, and solar power, Part I: technologies, energy resources, quantities and areas of infrastructure, and materials. *Energy Policy* 2011;39:1154–1169. [\[CrossRef\]](#)
- [21] International Renewable Energy Agency (IRENA). World energy transitions outlook 2022: 1.5°C pathway. Abu Dhabi: IRENA; 2022.
- [22] United Nations. The sustainable development goals report. New York: United Nations Publications; 2022.
- [23] Osman AI, Chen L, Yang M, Msigwa G, Farghali M, Fawzy S, et al. Cost, environmental impact, and resilience of renewable energy under a changing climate: a review. *Environ Chem Lett* 2022;21:741–764. [\[CrossRef\]](#)
- [24] World Wind Energy Association (WWEA). WWEA half-year report. Bonn: WWEA; 2022. [Internet]. Available at: <https://wwindea.org/wwea-annual-report-2022> Accessed on Sep 03, 2025.
- [25] Sengupta IN. Bibliometrics, informetrics, scientometrics and librametrics: an overview. *Libri* 1992;42:75–98. [\[CrossRef\]](#)
- [26] Pritchard A. Statistical bibliography or bibliometrics. *J Doc* 1969;25:348.
- [27] Godin B. From eugenics to scientometrics: Galton, Cattell, and men of science. *Soc Stud Sci* 2007;37:691–728. [\[CrossRef\]](#)
- [28] Gorraiz J, Gumpenberger C, Wieland M. Galton 2011 revisited: a bibliometric journey in the footprints of a universal genius. *Scientometrics* 2011;88:627–652. [\[CrossRef\]](#)
- [29] Danesh F, Mardani-Nejad A. A historical overview of bibliometrics. In: *Handbook bibliometrics*. Berlin: De Gruyter Saur; 2020. p. 7–18. [\[CrossRef\]](#)
- [30] Montoya FG, Montoya MG, Gomez J, Manzano-Agugliaro F, Alameda-Hernandez E. The research on energy in Spain: a scientometric approach. *Renew Sustain Energy Rev* 2014;29:173–183. [\[CrossRef\]](#)
- [31] Adedayo H, Adio S, Oboirien B. Energy research in Nigeria: a bibliometric analysis. *Energy Strateg Rev* 2021;34:100629. [\[CrossRef\]](#)
- [32] Mao G, Liu X, Du H, Zuo J, Wang L. Way forward for alternative energy research: a bibliometric analysis during 1994–2013. *Renew Sustain Energy Rev* 2015;48:276–286. [\[CrossRef\]](#)
- [33] Manzano-Agugliaro F, Alcayde A, Montoya FG, Zapata-Sierra A, Gil C. Scientific production of renewable energies worldwide: an overview. *Renew Sustain Energy Rev* 2013;18:134–143. [\[CrossRef\]](#)
- [34] Novas N, Alcayde A, Robalo I, Manzano-Agugliaro F, Montoya FG. Energies and its worldwide research. *Energies* 2020;13:6700. [\[CrossRef\]](#)
- [35] Ye P, Li Y, Zhang H, Shen H. Bibliometric analysis on the research of offshore wind power based on web of science. *Econ Res Ekon Istraz* 2020;33:887–903. [\[CrossRef\]](#)
- [36] Mendonça AKS, Vaz CR, Lezana ÁGR, Anacleto CA, Paladini EP. Comparing patent and scientific literature in airborne wind energy. *Sustainability* 2017;9:915. [\[CrossRef\]](#)
- [37] Chen CH, Su NJ. Global trends and characteristics of offshore wind farm research over the past three decades: a bibliometric analysis. *J Mar Sci Eng* 2022;10:1339. [\[CrossRef\]](#)
- [38] Eligüzel İM, Özceylan E. A bibliometric, social network and clustering analysis for a comprehensive review on end-of-life wind turbines. *J Clean Prod* 2022;135004. [\[CrossRef\]](#)
- [39] Morgan EF, Abdel-Rahim O, Megahed TF, Suehiro J, Abdelkader SM. Fault ride-through techniques for permanent magnet synchronous generator wind turbines (PMSG-WTGs): a systematic literature review. *Energies* 2022;15:9116. [\[CrossRef\]](#)
- [40] Barra P, de Carvalho W, Menezes T, Fernandes R, Courty D. A review on wind power smoothing using high-power energy storage systems. *Renew Sustain Energy Rev* 2021;137:110455. [\[CrossRef\]](#)
- [41] Gao C, Sun M, Geng Y, Wu R, Chen W. A bibliometric analysis based review on wind power price. *Appl Energy* 2016;182:602–612. [\[CrossRef\]](#)
- [42] Elie L, Granier C, Rigot S. The different types of renewable energy finance: a bibliometric analysis. *Energy Econ* 2021;93:104997. [\[CrossRef\]](#)
- [43] Proskuryakova LN, Ermolenko GV. The future of Russia's renewable energy sector: trends, scenarios and policies. *Renew Energy* 2019;143:1670–1686. [\[CrossRef\]](#)
- [44] Gan L, Jiang P, Lev B, Zhou X. Balancing of supply and demand of renewable energy power system: a review and bibliometric analysis. *Sustain Futures* 2020;2:100013. [\[CrossRef\]](#)
- [45] Marques RS, Martins LOS, Fernandes FM, Silva MS, Freires FGM. Wind power and competitiveness: a bibliometric analysis. *Inf Soc Estud* 2020;30:24. [\[CrossRef\]](#)
- [46] Ying C, Wang W, Yu J, Li Q, Yu D, Liu J. Deep learning for renewable energy forecasting: a taxonomy, and systematic literature review. *J Clean Prod* 2022;135414. [\[CrossRef\]](#)
- [47] Lagos A, Caicedo JE, Coria G, Quete AR, Martinez M, Suvire G, et al. State-of-the-art using bibliometric analysis of wind-speed and-power forecasting methods applied in power systems. *Energies* 2022;15:6545. [\[CrossRef\]](#)
- [48] Hicks D, Wouters P, Waltman L, De Rijcke S, Rafols I. Bibliometrics: the Leiden manifesto for research metrics. *Nature* 2015;520:429–431. [\[CrossRef\]](#)

- [49] Wang S, Zhang N, Wu L, Wang Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew Energy* 2016;94:629–636. [\[CrossRef\]](#)
- [50] Fausett LV. *Fundamentals of neural networks: architectures, algorithms and applications*. New Delhi: Pearson Education India; 2006.
- [51] Sumathi S, Paneerselvam S. *Computational intelligence paradigms: theory & applications using MATLAB*. Boca Raton: CRC Press; 2010. [\[CrossRef\]](#)
- [52] 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. [\[CrossRef\]](#)
- [53] Şenol H, Dereli MA, Özbilgin F. Investigation of the distribution of bovine manure-based biomethane potential using an artificial neural network in Turkey to 2030. *Renew Sustain Energy Rev* 2021;149:111338. [\[CrossRef\]](#)
- [54] Luo X, Wang J, Dooner M, Clarke J. Overview of current development in electrical energy storage technologies and the application potential in power system operation. *Appl Energy* 2015;137:511–536. [\[CrossRef\]](#)
- [55] Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. *Energy Strateg Rev* 2019;24:38–50. [\[CrossRef\]](#)
- [56] Whipple DT, Kenis PJ. Prospects of CO₂ utilization via direct heterogeneous electrochemical reduction. *J Phys Chem Lett* 2010;1:3451–3458. [\[CrossRef\]](#)
- [57] Ursua A, Gandia LM, Sanchis P. Hydrogen production from water electrolysis: current status and future trends. *Proc IEEE* 2011;100:410–426. [\[CrossRef\]](#)
- [58] Atwa Y, El-Saadany E, Salama M, Seethapathy R. Optimal renewable resources mix for distribution system energy loss minimization. *IEEE Trans Power Syst* 2009;25:360–370. [\[CrossRef\]](#)
- [59] Zhong M, Tran K, Min Y, Wang C, Wang Z, Dinh CT, et al. Accelerated discovery of CO₂ electrocatalysts using active machine learning. *Nature* 2020;581:178–183. [\[CrossRef\]](#)
- [60] Lin K, Chen Q, Gerhardt MR, Tong L, Kim SB, Eisenach L, et al. Alkaline quinone flow battery. *Science* 2015;349:1529–1532. [\[CrossRef\]](#)
- [61] Yan Y, Qian Y, Sharif H, Tipper D. A survey on smart grid communication infrastructures: motivations, requirements and challenges. *IEEE Commun Surv Tutor* 2012;15:5–20. [\[CrossRef\]](#)
- [62] Mathiesen BV, Lund H, Connolly D, Wenzel H, Ostergaard PA, Möller B, et al. Smart energy systems for coherent 100% renewable energy and transport solutions. *Appl Energy* 2015;145:139–154. [\[CrossRef\]](#)
- [63] Hayashi A, Noi K, Sakuda A, Tatsumisago M. Superionic glass-ceramic electrolytes for room-temperature rechargeable sodium batteries. *Nat Commun* 2012;3:856. [\[CrossRef\]](#)
- [64] Asadi M, Ramezanzade M, Pourhossein K. A global evaluation model applied to wind power plant site selection. *Appl Energy* 2023;336:120840. [\[CrossRef\]](#)
- [65] Sánchez-Lozano J, García-Cascales M, Lamata M. GIS-based onshore wind farm site selection using fuzzy multi-criteria decision making methods: evaluating the case of Southeastern Spain. *Appl Energy* 2016;171:86–102. [\[CrossRef\]](#)
- [66] Yomralioglu T, McLaughlin J. *Cadastre: geo-information innovations in land administration*. Cham: Springer; 2017. [\[CrossRef\]](#)
- [67] Colak HE, Memisoglu T, Gercek Y. Optimal site selection for solar photovoltaic power plants using GIS and AHP: a case study of Malatya Province, Turkey. *Renew Energy* 2020;149:565–576. [\[CrossRef\]](#)
- [68] Coruhlu YE, Solgun N, Baser V, Terzi F. Revealing the solar energy potential by integration of GIS and AHP in order to compare decisions of the land use on the environmental plans. *Land Use Policy* 2022;113:105899. [\[CrossRef\]](#)
- [69] Güner F, Başer V, Zenk H. Evaluation of offshore wind power plant sustainability: a case study of Sinop/Gerze, Turkey. *Int J Glob Warming* 2021;23:370–384. [\[CrossRef\]](#)
- [70] Çolak HE, Memişoğlu Baykal T, Genç N. Multicriteria decision and sensitivity analysis support for optimal airport site locations in Ordu Province, Turkey. *Ann GIS* 2023;1–28. [\[CrossRef\]](#)
- [71] International Climate Policy Initiative. *Global landscape of renewable energy finance*. Abu Dhabi: International Renewable Energy Agency; 2023.
- [72] Lindman Å, Söderholm P. Wind energy and green economy in Europe: measuring policy-induced innovation using patent data. *Appl Energy* 2016;179:1351–1359. [\[CrossRef\]](#)



Research Article

Real-time tomato flower estimation using deep learning models for e-agriculture applications

Bini D¹, Pamela D², Shajin PRINCE³, K Martin SAGAYAM³, Hatıra GÜNERHAN^{4,*}

¹Division of Robotics Engineering, Karunya Institute of Technology and Sciences, Coimbatore, 641114, India

²Division of Biomedical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, 641114, India

³Division of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, 641114, India

⁴Department of Mathematics, Faculty of Education, Kafkas University, Kars, 36040, Türkiye

ARTICLE INFO

Article history

Received: 25 March 2024

Revised: 12 June 2024

Accepted: 15 July 2024

Keywords:

Deep Learning; Flower Counting and Estimation; Precision Agriculture; Transfer Learning

ABSTRACT

In early spring, growers prune the extra blossoms and fruitlets off the crops and trees to enhance the yield of berries. Numerous automated machine vision techniques for estimating floral intensity have been proposed, however, their overall performance is still inadequate. The floral intensity is related with the harvest which will assist the government to frame the governance policies for business and trade. For the agricultural task of detecting the tomato blooms in crop images, the performance of six pretrained deep learning architectures was evaluated. This study presents a technique to detect tomato flowers that is reliable to occlusions, variations in illumination conditions, and orientation. The real-time crop images across the fields were acquired in daylight conditions using a 13 Mega-pixel RGB camera. The image acquisition technique would have an impact on the quality of the real-time images. One of the most important computational techniques utilized in the smart digital world of agricultural application is deep learning. The key objective of this research article is to identify the best improvement strategies for recognizing the tomato flowers and berries for real-time crop yield estimation and yield management, thereby the analysis of six deep learning architectures, AlexNet, Resnet50, VGG16, Faster R-CNN, YOLOv3 and YOLOv5, was measured. The YOLOv5 model outperformed other existing models with a 0.975 F1 score on a real-time tomato flower dataset.

Cite this article as: Bini D, Pamela D, Prince S, Sagayam KM, Günerhan H. Real-time tomato flower estimation using deep learning models for e-agriculture applications. Sigma J Eng Nat Sci 2025;43(5):1572–1579.

INTRODUCTION

The sensors placed on agricultural farms can access precise topography, temperature, climatic forecasts, and acidity of the topsoil. Precision farming collects information from

agricultural sensors, drones, and satellites to organize farming practices. In India, tomato harvesting time is based on Growers can use agricultural technologies to generate a map of their farmland, control pests, manage fertilizer, detect environmental threats, and crop yields, and plan to harvest.

*Corresponding author.

*E-mail address: gunerhanhatira@gmail.com

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



The count of blooms recorded and compared to the number of fruits counted before the harvesting phase for citrus trees allowed for crop yield forecast [1]. At the beginning of spring, the bloom intensity in an orchard has an impact on crop management. The deployment of a deep learning model to estimate the flower count will however be useful for crop production forecasting, pruning of the parts and thinning process, which influence the productivity of the crop [2]. The ability to recognize vine harvest assists the growers in preparing for transporting the yield and scheduling the trade. Deep learning-based classifiers have improved over the years where they can identify crop diseases in agricultural farms under controlled and real-world conditions [3]. The diseases that develop during the blooming and fruit development stages may have an impact on crop production predictions.

The significant fluctuation in luminance, fruit occlusion, the shape of the fruit, uniformity in color and morphological texture of the fruits and the leaves are all significant concerns in agricultural scenes for computer vision detection of fruit [4–6]. Detecting blooms accurately and early in the growing season is crucial for timely intervention strategies to mitigate potential yield losses due to factors such as nutrient deficiencies or pest infestations. Object classification in agricultural sceneries using manually determined handcrafted parameters for features like color models, shape, intensity, form, the texture and the Histogram of Oriented Gradients (HOG), Haar-like features, and the local binary patterns (LBP), has been reported numerous times over the last decade [7].

Deep convolutional networks have recently been sustaining the future upon a diverse selection of computer vision applications. [8,9] developed a (Faster R-CNN) Faster Regionbased Convolutional Neural Network at Oxford. Transfer learning is used to recognize diverse fruits using the Visual Geometry Group network (VGGNet). A pretrained model has been further trained using a limited number of images, with F1 scores of 0.848, 0.932, 0.942, 0.948, 0.938, 0.915, and 0.828 for rockmelon, avocados, mangoes, strawberries, apples, orange, and green pepper, correspondingly [10]. Table 1 highlights a few research

works that have been performed applying diverse image processing techniques.

The training of deep learning frameworks encompasses the optimization of losses, and cost function which is a unique combination of confidence, classification, precision, recall and box regression loss for its minimum error in object recognition and classification. Since deep learning approaches have a large number of parameters, large datasets like ImageNet [17], PASCAL VOC [18], and COCO [19] are commonly used to train these models. These digital datasets provide free labelled images along with annotation files for various object classes which are used for object detection model training and benchmarking. Although these datasets do not include images of tomato flowers, deep-learning models can indeed be reused in these other applications by transferring acquired key features using transfer learning [20,21]. The transfer learning technique is employed for re-training the information from a model learnt on a large dataset, such as convolutional weights, to build a new model using a smaller set of images.

A general deep learning framework for classification involves:

- (i) input image pre-processing;
- (ii) training the pre-trained model;
- (iii) classification of tomato flower;
- (iv) Performance analysis.

In this research, six pretrained models for object recognition employing classifiers and detectors, such as AlexNet, Resnet50, VGG16, Fast-er R-CNN, YOLOv3 and YOLOv5, were supervised and assessed for tomato blossom detection.

MATERIALS AND METHODS

Dataset Acquisition

The image datasets of tomato flowers were collected in a nearby field in Coimbatore, India. The images were captured using a Canon EOS 1500D digital camera, with a resolution measuring 3648 X 2056 pixels. Most of the images were captured under natural illumination with varying degrees of occlusion, lighting and alignment. Nearly 916

Table 1. Diverse deep learning methods used for tomato flower estimation

Crop	No. of images	Method	Performance Analysis	Application	Authors
Tomato flower	1069	Segmentation in HSV color space	F1 score=0.73	Detection	[11]
Tomato flower	1445	Faster RCNN	Precision=96.02% Recall=93.09%	Detect and count	[12]
Lemon	1113	YOLOv3	Accuracy=92%	Classification	[13]
Tomato flower	1400	ResNet-152	Accuracy=90.2%	Classification	[14]
Tomato Flower	1500	CNN	Accuracy = 89.6%	Disease Detection	[15]
Tomato Flower	1100	YOLOv4	Accuracy = 91.7%	Disease Detection	[16]



Figure 1. Sample images with tomato flowers.

images with tomato flowers were collected and categorized as train sets and test sets. The samples of the input tomato flower image captured are shown in Figure 1.

Image Pre-Processing

The training and testing of the datasets were implemented on a high-performance Intel Core (i5 processor) with an NVIDIA GPU GEFORCE GTX. During training, various object detection algorithms resize the input range to a predetermined network resolution. Deep learning systems models employ feature extractors that require a square input resolution. The resolution of the network can be enhanced, to handle higher input images, however, this comes at the expense of more memory and computing. The Faster R-CNN model, fine-tune the input image to 600 X 600 pixels, whereas the Alexnet, resnet50 and inceptionv3 neural networks, resize the input image to 227 X 227, 224 X 224, 299 X 299 pixels, respectively. The YOLO architecture resizes the input images to 416 X 416 pixels.

Data Augmentation

The data augmentation techniques such as cropping, resizing and scaling were used in this research work before training the model to increase the training set. The class label and coordinates of every ground truth with an annotated bounding box in the train image datasets are required for object detection models to be trained. While labelling is a time-consuming and labor-intensive task, annotating the tiles was significantly simpler than annotating the entire image due to the reduced amount of flowers on each tile compared to the entire image. For artificially increasing the dataset with label-preserving modifications, data augmentation is a typical approach to increase variability in the training data [22]. The method improves the ability of networks to generalize while reducing overfitting. To increase shape variety in the dataset, this study used vertical flip, horizontal flip, and image rotation techniques. To prevent pre-computing the broad array of random augmentations, augmentation was done by increasing the dataset during each training phase.

Data Annotation

The annotation technique was effective since it identifies and labels a rectangular bounding box on the specified part of the crop in the image along with location

information and the labelled class. It is indeed essential to identify between the concepts of image classification as well as to object detection at this phase. Object detection detects the labelled class and position of each object from an image, whereas classification determines, whether an image possesses the object class. Detection seems to be more challenging than the classification algorithm, and the number of samples required for object detection is greater than the number of images required for image classification. A single frame in real-time agricultural images contains numerous artifacts of various categories, which should have been assessed by using crop class probability and its orientation. The annotation method used in this research is time-consuming and comparable to that of the PASCAL Visual Object Classes dataset [18]. The following are the rules for labelling the datasets:

- (i) Each occurrence of an object should indeed be labelled out when an image contains many such objects.
- (ii) When two objects in an image overlap, the occluded regions should be drawn with a box all across the visible exposed regions. It is necessary to completely wrap and label each object in the image.
- (iii) If there are multiple objects of the same class in an image, all the individual objects are boxed, even if the bounding boxes overlap. Only one bounding box is drawn around the occurrences if they are physically interconnected.

Bloom Counting Deep Learning Models

Faster R-CNN, like most other neural network models, is hampered by its reliance on a vast number of higher-quality, well-labelled data. Researchers might need transfer learning to get around the necessity for a huge amount of new data. A CNN that has already been trained on a task with a large amount of labelled training data will be able to tackle a new task with significantly fewer labelled instances. As a reference point, the ImageNet dataset, which comprises nearly 1000 object classes and 1.2 million images, is frequently employed. Using the pre-trained CNN features on the ImageNet dataset, state-of-the-art actual outcomes have been produced on a variety of image processing techniques, including image classification and image labelling. Figure 2 depicts the overview of the tomato flower estimation using deep learning models. The classification and

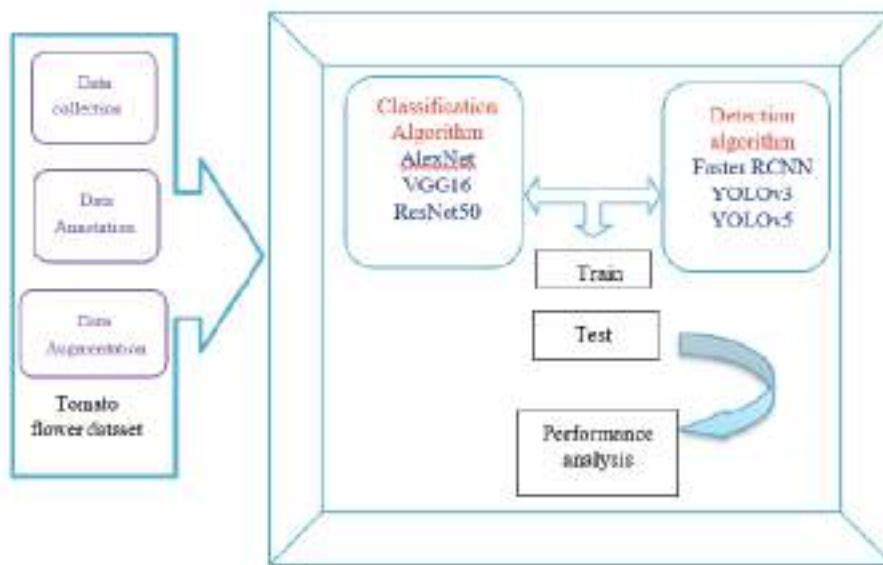


Figure 2. A framework of the Flower estimation deep learning model.

detection algorithms are trained with the real-time tomato flower datasets for the estimation/counting of tomato flowers. The early recognition of tomato flowers can predict yield management in precision agriculture.

Alexnet

The Alexnet architecture is a deep CNN model developed to classify nearly 1.2 million images, with five convolution layers, a Rectified Linear Unit layers (ReLU), max-pooling layers and three fully connected layers [17]. The convolutional kernels for the AlexNet architecture are computed during the back-propagation optimization phase, which involves optimizing the entire cost structure with the stochastic gradient descent (SGD) technique. In general, the convolutional layers employ sliding convolutional kernels for constructing feature maps, while the pooling layers generate a max-pooling operation on these feature maps to consolidate the data within the provided neighborhood region. The training is made faster using non-saturating neurons with an efficient GPU execution. The Alexnet framework utilizes a recently developed normalization technique called “dropout” which proved to be particularly effective in reducing the overfitting in fully-connected layers.

Resnet50

Deep neural networks are difficult to train due to the well-known diminishing gradient problem: since the gradient being back-propagated towards the previous levels, continuous multiplication can reduce the gradient to infinity [13]. Besides implementing a short link to transfer input from one layer towards another layer without even any change in the input, ResNet may have a very deep CNN model with approximately 152 layers. In the implementation of ResNet, there are two types of auxiliary modules.

The first is an identity block that does not have a convolution layer of any kind. In this instance, input and output have the same dimensions [23]. The other one is the convolution frame, which has a shortcut for the convolution layer. In this instance, the input measure is far less than the resultant dimensions. In each block, 1 X 1 convolution layers are connected to the beginning and at the end of the network structure. The bottleneck approach is a process for decreasing the quantity of parameters and retaining the network performance.

VGG16

VGG includes 3 fully - connected layers and 13 convolutional layers. Max-pooling layers connect consecutive convolution layers. Every group has three convolution layers, ranging in size from 64 at the beginning of the implementation to 512 at the completion [10]. This network was constructed in preparation for the ImageNet 1000 category classification task. The most distinguishing aspect of VGG16 seems to be that, instead of only reaching a substantial amount of hyper-parameters, the developers designed to generate 3x3 convolution layers of stride 1, and also a comparable padding and max-pool layer featuring 2x2 filter of stride 2. The convolution, as well as the max-pool layers, are grouped similarly throughout the architecture. Finally, it has two completely connected layers and a soft-max for output. VGG16 alludes to 16 layers with different weights. This network is quite huge, with over 138 million parameters.

Faster R-CNN

A third refinement of the network design and region proposal technique is the Faster R-CNN. The early R-CNN used the region proposal approach for recognizing

Table 2. Hyperparameters of the deep learning models

Parameter	Alexnet	Resnet 50	VGG 16	Faster RCNN	YOLOv3	YOLOv5
Learning Rate	0.01	0.1	0.001	0.001	0.001	0.01
Batch size	128	256	64	256	64	16
Momentum	0.9	0.9	0.9	0.9	0.9	0.937

regions of interest, including Selective Search, as a prospective area comprising the target item. The ROIs are the inputs to a deep CNN, which would extract features for classification by a support vector machine, resulting in a visual representation of the process. External ROI extraction methods are typically computationally intensive processes that slow down the entire algorithm pipeline. The later version employs a new method for classifying ROIs [8]. Rather than applying a CNN to each of the area proposals, the image is fully inserted into the CNN, with the output inserted into a spatial pyramid pooling (SPP) layer. In the SPP layer, only ROIs acquired using a region proposal approach are employed. This method eliminates the need for a full forward run through the CNN for each ROI, reducing execution time by 10 to 100 times at test time and 3 times at training time. However, it still involves a computing bottleneck in the form of an external region proposal algorithm.

YOLOV3

YOLOv3 is an upgrade over YOLOv2 and is built on Darknet-53. Darknet is a C and CUDA-based open-source deep learning framework. The Darknet-53 is indeed a major feature extractor which improves upon Darknet-19 through adding 3X3 and 1X1 convolutional layers including residual blocks, similar with ResNet. Similar to SSD, YOLOv3 employs feature pyramids to automatically extract the features across three levels for box prediction. Lower-level feature maps are integrated using up-sampled extracted feature maps from the upper layers and then processed to collect more relevant and fine-grained data. Similar to Faster R-CNN, YOLOv3 employs logistic regression model to estimate the objectness score for the bounding boxes, however, each ground truth object only has one bounding box anchor. YOLOv3 employs 75 convolution layers as well as 3 detection layers. [3,13,19,24].

YOLOV5

The network model is a CNN architecture that uses multiple convolution layers and max-pooling layers to retrieve the feature maps of various sizes from the input images. The backbone network generates four layers of feature maps, with the following dimensions: 152X152 pixels, 76X76 pixels, 38X38 pixels, and 19X19 pixels [25]. The neck network integrates the feature maps from multiple levels of different sizes to gain additional contextual information to reduce information loss. YOLOv5

employs CSPDarknet53 and also, however the neck system employs the Feature Pyramid Network (FPN) and Pixel Aggregation Network (PAN) architecture. It is comparable to YOLOv4 in order to improve accuracy however outperforms it in order of speed due to its small model size. The hyperparameters of the diverse deep learning models for the image classification and object detection are depicted in table 2.

DEEP LEARNING MODEL PERFORMANCE

Classification And Detection Results

The tomato flower datasets are split into three categories: 80% training and 20% testing. The training is completed on the training dataset images, followed by an evaluation on the test set to reduce overfitting. After the training of the deep learning model, parameter selection has been completed, and the performance is evaluated using the testing set. Non-Maximum Suppression reduces the loss among the ground truth and predicted results, and the occurrence of false positives in the final results, by selecting only candidates with an Intersection-over-Unions (IoU > 0.5) relative to the primary annotated ground truth. The sample results of the tomato flower estimation models are summarized in Figure 3. To boost the robustness of every model, images with a specific instance were combined with images including many objects present and many occurrences.

Performance Metrics

The pretrained deep learning architectures were trained, relative to the labelling annotated images using the train sets and the test sets. The trained neural networks were validated for tomato flower datasets with the performance metrics such as F1 score, average precision, inference time and prediction rate. Also, with training data and testing data, the network is trained using SGD momentum including a learning rate of 0.01. Every epoch, the images in the datasets are randomized. The randomized real-time images are used for the testing phase, whereas the output is classified according to the given target specified in the FC layer. The accuracy and performance analysis in terms of F1-score, average precision, Inference time are listed in Table 3.

The average of training images per class for each class is reflected by the 'average precision. The results show that YOLOv5 model outperformed the other classification and



Figure 3. Sample results of tomato flower detection.

Table 3. Performance analysis of the deep learning models

Models	Average Precision	F1 Score	Inference time (ms)
Alexnet	0.918	0.932	52
Resnet50	0.953	0.935	39
VGG16	0.924	0.892	46
Faster RCNN	0.962	0.939	67
YOLOv3	0.969	0.947	25
YOLOv5	0.975	0.962	20

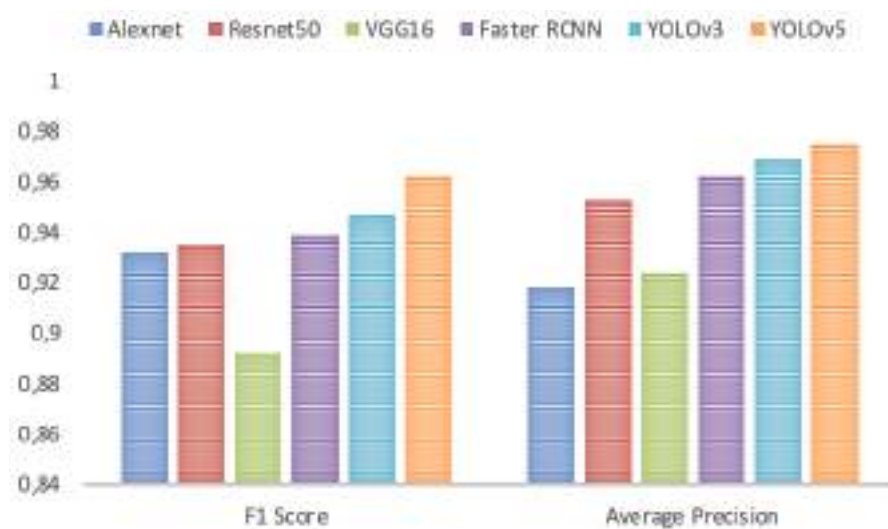


Figure 4. Performance of tomato-estimation deep learning models.

detection models for tomato flower estimation. The Faster RCNN model was too slow to reach the preferred frames per second. The evaluation of the models for the metrics is shown in Figure 4.

DISCUSSIONS

The models were evaluated based on how well they could manage various scenarios that arise in the field, including occlusions, illumination variations, and different

orientations. Based on its precision and resilience in real-time recognition tasks, YOLOv5 showed the best performance among these designs, with an F1 score of 0.975. The YOLOv5 framework's efficiency in processing images fast while retaining good accuracy is one of its standout features. In agricultural applications, where rapid identification of plant health concerns, such as deficiencies in nutrients indicated by bloom patterns, can greatly effect crop output and quality, this skill is essential. Notwithstanding its merits, the research highlights a number of shortcomings that require further development. The accuracy of detection is still greatly influenced by picture quality since differences in lighting and image resolution might have an impact on the performance of the model.

The implementation of deep learning approaches signifies a paradigm change from conventional methods, including inspecting manually or basic image processing techniques, in terms of agricultural contributions. Improved accuracy and flexibility provided by deep learning models allow automated evaluation of large-scale datasets that are difficult for standard methods to handle effectively. With proactive pest and disease control and optimum utilization of resources, this capacity not only lowers labor costs but also improves decision-making, promoting sustainable farming practices. Deep learning has implications that go beyond its ability to detect. The ability of these models to adjust and learn from fresh data enables ongoing development and adaptation to changing agricultural requirements. Through the integration of cutting-edge technology such as YOLOv5 into commercial process flow, stakeholders may take use of precision cultivation practices and predictive modeling to attain increased production while demonstrating resilience against environmental obstacles.

CONCLUSION

A domain-specific image dataset of tomato flower images for precision agriculture is discussed in this paper. The goal of this research is to compare the deep-learning-based classification and detection models for real-time agricultural tasks. For the agricultural task of detecting tomato blooms in crop images, the performance of six deep neural network architectures was evaluated. These studies revealed a system for detecting tomato blooms that is robust to occlusions, lighting fluctuations, and orientation. The real-time image quality would be affected by the image acquisition technique. However, the deep learning approach is used in diverse agricultural tasks for smart farming. On a real-time tomato flower dataset, the performance of six existing deep learning architectures, AlexNet, Res-net50, VGG16, Faster R-CNN, YOLOv3 and YOLOv5, was analyzed and found that the YOLOv5 framework outscored other conceptual methods with an F1 score of 0.975.

The detection of tomato blooming and berry development phase assists in the early detection of macronutrient imbalances. New varieties of crop species images can be

integrated and annotated to the datasets in the future to train the architecture for tasks such as harvesting and agricultural yield management.

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] Dorj U, Lee M, Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing. *Comput Electron Agric* 2017;140:103–112. [\[CrossRef\]](#)
- [2] Dias PA, Tabb A, Medeiros H. Multispecies fruit flower detection using a refined semantic segmentation network. *IEEE Robot Autom Lett* 2018;3:3003–3010. [\[CrossRef\]](#)
- [3] Liu J, Wang X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front Plant Sci* 2020;11:898. [\[CrossRef\]](#)
- [4] Font D, Tresanchez M, Martínez D, Moreno J, Clotet E, Palacín J. Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night. *Sensors* 2015;15:8284–8301. [\[CrossRef\]](#)
- [5] Gong A, Yu J, He Y, Qiu Z. Citrus yield estimation based on images processed by an Android mobile phone. *Biosyst Eng* 2013;115:162–170. [\[CrossRef\]](#)
- [6] Bargouti S, Underwood JP. Image segmentation for fruit detection and yield estimation in apple orchards. *J Field Robot* 2017;34:1039–1060. [\[CrossRef\]](#)

- [7] Gongal A, Amatya S, Karkee M, Zhang Q, Lewis K. Sensors and systems for fruit detection and localization: a review. *Comput Electron Agric* 2015;116:8–19. [\[CrossRef\]](#)
- [8] Ren RSJ, He S, Girshick K. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2016;39:1137–1149. [\[CrossRef\]](#)
- [9] Bini D, Pamela D, Prince S. Machine vision and machine learning for intelligent agrobots: a review. In: *Proceedings of the 5th International Conference on Devices, Circuits and Systems (ICDCS)*. 2020. p. 12–16. [\[CrossRef\]](#)
- [10] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Am J Health Syst Pharm* 2018;75:398–406.
- [11] Oppenheim D, Edan Y, Shani G. Detecting tomato flowers in greenhouses using computer vision. *Int J Comput Inf Eng* 2017;11:104–109.
- [12] Rahim UF, Mineno H. Tomato flower detection and counting in greenhouses using faster region-based convolutional neural network. *J Image Graph* 2020;8:107–113. [\[CrossRef\]](#)
- [13] Zheng YY, Kong JL, Jin XB, Wang XY, Su TL, Zuo M. Cropdeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors* 2019;19:1058. [\[CrossRef\]](#)
- [14] Rajasree R, Latha CBC, Paul S. Application of transfer learning with a fine-tuned ResNet-152 for evaluation of disease severity in tomato plants. *Mob Comput Sustain Inform* 2022;:695–710. [\[CrossRef\]](#)
- [15] Abdulridha J, Ehsani R, Abd-Elrahman A, Ampatzidis Y. A remote sensing technique for detecting laurel wilt disease in avocado in presence of other biotic and abiotic stresses. *Comput Electron Agric* 2020;156:549–557. [\[CrossRef\]](#)
- [16] Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. PlantDoc: a dataset for visual plant disease detection. In: *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 2020. p. 249–253. [\[CrossRef\]](#)
- [17] Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2012;60:84–90. [\[CrossRef\]](#)
- [18] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *Int J Comput Vis* 2010;88:303–338. [\[CrossRef\]](#)
- [19] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018;321:321–331. [\[CrossRef\]](#)
- [20] Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. Deepfruits: a fruit detection system using deep neural networks. *Sensors* 2016;16:1222. [\[CrossRef\]](#)
- [21] Coulibaly S, Kamsu-Foguem B, Kamissoko D, Traore D. Deep neural networks with transfer learning in millet crop images. *Comput Ind* 2019;108:115–120. [\[CrossRef\]](#)
- [22] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018;321:321–331. [\[CrossRef\]](#)
- [23] Sharma R, Sungheetha A. An efficient dimension reduction based fusion of CNN and SVM model for detection of abnormal incident in video surveillance. *J Soft Comput Paradigm* 2021;3:55–69. [\[CrossRef\]](#)
- [24] Lawal MO. Tomato detection based on modified YOLOv3 framework. *Sci Rep* 2021;11:1–11. [\[CrossRef\]](#)
- [25] Bini D, Pamela D, Hemanth DJ. Evaluating YOLOv5 deep learning models for the detection of tomato berries for crop yield estimation. *Front Artif Intell Appl* 2022;347:429–436. [\[CrossRef\]](#)



Research Article

Classification of ancient coins in archaeology using a novel deep learning approach: Bayesian convolutional neural network

Ebru PEKEL ÖZMEN^{1,*}, Soner ÖZMEN², Mehmet Emre SERTKAYA³, Tuncay ÖZCAN⁴,
Vedat KELEŞ⁵

¹Department of Industrial Engineering, Samsun University, Samsun, 55000, Türkiye

²Department of Archaeology, Ondokuz Mayıs University, Samsun, 55270, Türkiye

³Distance Education Application and Research Center, Samsun University, 55000, Samsun, Türkiye

⁴Department of Management Engineering Department, Istanbul Technical University, Istanbul, 34467, Türkiye

⁵Department of Archaeology, Ondokuz Mayıs University, Samsun, 55270, Türkiye

ARTICLE INFO

Article history

Received: 04 April 2024

Revised: 03 July 2024

Accepted: 24 September 2024

Keywords:

Archaeology; Bayesian
Optimization; Convolutional
Neural Network; Dating; Deep
Learning

ABSTRACT

This study looks at classifying and dating old coins. Coins are important in archaeology because they tell us about history, culture, and economy. Knowing the right date of coins helps to understand excavation sites and also helps studies in art, politics, and social life. Normally, numismatics experts do this work, but it takes a lot of time and their judgment can be different. In this research, we used some deep learning models like DenseNet-201, GoogLeNet, InceptionV3, MobileNetV2, and Xception. We also tested a new model called Bayesian Convolutional Neural Network (B-CNN). This model uses Bayesian optimization to choose parameters. The B-CNN reached about 97% accuracy, which is better than the other models. The results show that B-CNN can be a good tool for archaeologists, especially for dating coins. It gives more clear and correct results and reduces the need for special experts. The new part of this study is mixing Bayesian optimization with CNNs. This makes the model stronger than older methods. The work connects archaeology and computer science and shows better sensitivity and performance, but it also needs more training time.

Cite this article as: Pekel Özmen E, Özmen S, Sertkaya ME, Özcan T, Keleş V. Classification of ancient coins in archaeology using a novel deep learning approach: Bayesian convolutional neural network. Sigma J Eng Nat Sci 2025;43(4):1580–1591.

INTRODUCTION

One of the fundamental working concepts of archaeology is the requirement to appropriately date the material uncovered. The ancient coins, which were discovered in vast quantities during the excavations, give vital information for

the identification and date of the region where they were discovered, as well as other archaeological material in the area [1]. Many archaeological works are interpreted with a wide range of dates rather than a specific period, because their building technique and aesthetic qualities have been

*Corresponding author.

*E-mail address: ebru.pekel@samsun.edu.tr

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



maintained for years. On the other hand, the ancient coins can be dated in time intervals, sometimes up to several years, by interpreting the figures, inscriptions, and monograms on them as a result of the political, military, religious, economic, and social events that the politically dominant people or the ruler experienced with the society in a short period of time [2, 3]. Experts (numismatists) do coin recognition and dating because they give thorough and trustworthy information on the location and/or geography. The identification and dating of coins are important since it directly helps sub-studies such as pottery, art, architecture, politics, religion, geography, social and economic life, and notably archaeology.

Traditionally, experts in numismatics (coin specialists) have been responsible for the recognition and dating of these coins, as they provide thorough and reliable information about the geography and time period of their origin. The identification and dating of coins are crucial because they contribute directly to sub-disciplines such as pottery, art, architecture, politics, religion, geography, social and economic life, and, most importantly, archaeology.

Classifying and dating coins is important, but it usually takes a long time and often depends on what experts think. Doing this work by hand also has problems. It needs a lot of historical knowledge and very close study of small details on the coins.

In the literature, there are only a few studies that use deep learning for coin classification and dating. Earlier studies mostly depended on traditional techniques or very basic machine learning models. To reach better accuracy and reliability, there is now a need to explore stronger deep learning methods, such as Convolutional Neural Networks (CNNs).

In this work, we tested different CNN models, including DenseNet-201, GoogLeNet, InceptionV3, MobileNetV2, and Xception. We also proposed a Bayesian Convolutional Neural Network (B-CNN), which uses Bayesian optimization to fine-tune its parameters. This work represents the first known use of Bayesian optimization in CNN-based ancient coin classification. The results suggest that the method reduces subjectivity, shortens the analysis process, and provides a stronger basis for interpretation. It can be a

helpful support tool for archaeologists and numismatists. In this way, the study fills a gap in the literature and shows an interdisciplinary method that brings deep learning into archaeological research.

The remainder of this paper is organized as follows: In the second section, the literature review is presented. In the third section, the proposed approaches for the classification of ancient coins are introduced. The fourth section explained information which performance measures are selected for interpretation. The fifth part shows experimental results to compare the performances of the proposed approaches. Finally, the discussion and conclusions are presented.

LITERATURE REVIEW

In this study, the literature review was carried out in two different stages. In the first stage, a general review of the Convolutional Neural Networks method was conducted. Afterwards, Deep Learning studies in the archaeological area were reviewed. There is a few studies in the literature that applies deep learning algorithms to the classification problem of ancient coins.

Convolutional neural networks are now employed successfully in a variety of applications such as classification, pattern recognition and so on. There is no standard neural network model since the problem types differ. The difficulty of the problem and the structure of the neural network model are affected by factors such as the inputs, predicted outputs, and output form of the problem.

[4] conducted a study on the automatic recognition of ancient coin types using a semi-supervised learning method called Graph Transduction Games (GTG). This method deals with the difficulty of classifying ancient coins. The problem is hard because different coin types can look very similar, and coins from the same type can look very different. Another problem is that there are not many large labeled datasets.

The authors introduced a new dataset, RRC-60, which is an extension of a previous dataset, both in quantity and diversity. In practice, when dealing with old coins, researchers rarely trust a single glance. Most of the time, the coin

Table 1. Summarizes several research from the literature

Paper	Dataset	Method	Accuracy
(4)	Roman Empire	Graph Transduction Games (GTG)	97.4%
(5)	Roman Republican	Feature Fusion and Attention Mechanisms	98.5%
(6)	Roman Empire	MobileNetV3-L	95.2%
(7)	Roman Empire, Ancient Greece	Transfer Learning	98.32%
(8)	Turkish Republic	CNN, GAN	97.71%
(9)	Roman Empire	Graph Transduction Games (GTG)	87.2%
(10)	Roman Republican	Local Feature Matching, Geometric Consistency	83.3%
(11)	Roman Republican	radial-polar tiling scheme	91.68%

is turned over, examined again, and even compared with similar samples. This is a very ordinary step for a human expert, but when transferred into a model, it creates a rather clever mechanism. In one recent approach, the front and back sides of a coin were not treated separately, but instead as two sources of evidence that could “talk” to each other. The idea sounds almost obvious—if one side is blurred or worn, the other side may still carry useful details—but building this logic into a formal system is not trivial. By letting both sides support each other, the model manages to imitate the reasoning process of a person who is trying not to miss any clue. That is why the method has been described as game-theoretic, although in reality it simply reflects the back-and-forth nature of how experts already work.

The experiments showed that this method boosts performance in this difficult task, even when only a small number of labeled images are available.

Roman Republican coins were studied with a model called CoinNet, which effectively handled challenges caused by very old and damaged coins, especially in recognizing reverse side motifs such as objects, faces, animals, and buildings. A dataset of more than 18,000 images and 228 motif types was used, and the architecture—combining bilinear pooling, residual blocks, and attention layers—achieved 98.5% accuracy, significantly outperforming Bag-of-Visual-Words and traditional CNN models, thus demonstrating the advancement of modern deep learning methods [5].

Roman period coins were analyzed with deep learning models using the RRC-60 dataset, where Xception, MobileNetV3-L, EfficientNetB0, and DenseNet201 were tested. The best result was obtained by MobileNetV3-L with 95.2% accuracy, 98.2% Precision, and 96.8% Recall. The study also emphasized that considering both sides of a coin provides richer context and improves classification, much like how historians examine both sides before making judgments [6].

Ancient coins were also examined with local feature extraction and geometric checks combined with CNNs, which improved accuracy, particularly for visually similar coins [7].

A dataset named TurCoins, consisting of 11,080 images of Turkish Republic coins, was introduced, where ResNet50 with transfer learning was used and synthetic image generation for rare classes increased the accuracy to 97.71% [8]. Another approach, Graph Transduction Games (GTG), treated coin classification as a game where coins adopted labels based on peers, achieving 73.6% accuracy with one image per class and 87.2% with two on 180 Roman coins across 60 classes [9].

The ILAC project focused on Roman Republican coins with methods such as image matching, legend recognition, and coin recognition, reaching up to 83.3% accuracy on 4,100 coins, proving effective even for worn or damaged samples [10].

Motif recognition on the reverse side of Roman Republican coins was also conducted with a Bag of Visual Words (BoVWs) model, where radial-polar tiling provided the best performance with 91.68% accuracy [11].

Research Aim

When the literature studies are examined, it is seen that the classification of ancient coins using deep learning algorithms can be an innovative and important research topic. This study allows that the deep learning algorithms will be used for the first time to perform dating and classification tasks that have hitherto only been carried out by humans. As a result, this approach helps overcome limitations such as avoiding or overlooking areas that are difficult to access or hard to study with the naked eye for humanitarian reasons.

Moreover, this study marks the first attempt to analyze coins from two of the most significant periods of antiquity—the Greek and Roman eras—using deep learning methods. Compared with many other archaeological materials, coins stand out as especially valuable artifacts, since they preserve highly detailed and comprehensive information uncovered during excavations. Only experts (numismatists) can evaluate and interpret coins, which can only be identified after mechanical and chemical cleaning procedures, in addition to going through the phases of high-resolution photography and microscope analysis. While the tests conducted, it is possible to run into human factors like ignoring, neglecting, different interpretation, and quick-wrong conclusion. A specialist numismatist is not always present in excavations or projects for scientific archaeological investigations that are completed in a short amount of time, under challenging circumstances, and with few possibilities. At this point, this study presents a decision support system that can be used by non-technical experts for the classification of ancient coins.

In this study, a different perspective was taken by introducing the B-CNN model. The motivation behind developing this model was to see whether combining a standard CNN with Bayesian-based optimization could make the training process smarter and more efficient. Instead of relying only on fixed parameters, the model was allowed to “learn” better settings during the process. This way, the B-CNN does not simply copy what traditional CNNs already do, but rather tries to improve them by adapting to the data in a more flexible manner. The study shows that such an approach can open new directions for deep learning applications, especially in areas where accuracy and reliability are critical. Notably, the uniqueness of this study extends beyond archaeology; it also introduces novel contributions to image processing technologies.

MATERIALS AND METHODS

Pure Convolutional Neural Networks Models

Deep learning is a helpful technology that came about because regular neural networks got better. Deep learning

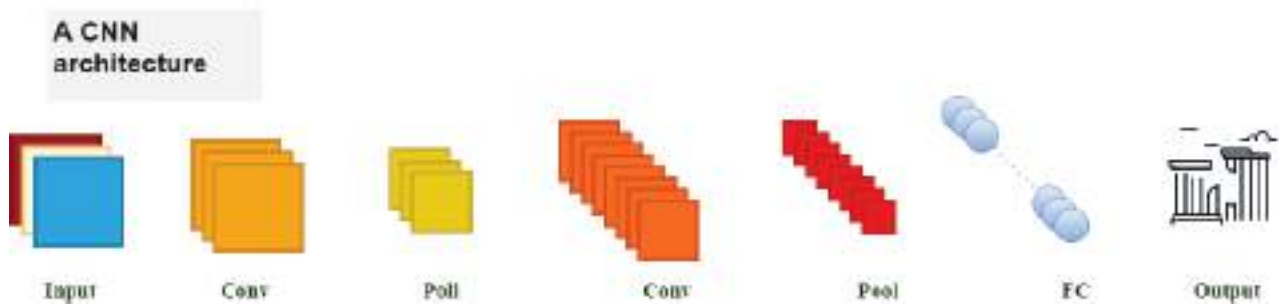


Figure 1. CNN architecture

is widely used in classification, image segmentation, identification, prediction, and detection. Deep learning algorithms can predict complex connections between inputs and outputs and manage large amounts of data. CNN is one of the most effective and extensively used deep learning algorithms. With the convolution process, the attributes of the input data are automatically retrieved and transferred to the following layers with these properties. The convolution method reveals the image's most significant features. Data is transferred through the next layer of the CNN's multilayered structure by performing a different operation on each layer as seen in Figure 1. Each layer has a distinct purpose.

The layers of CNN can be summarized as follows:

Convolution: In the convolution process, a matrix of any size, such as 3×3 , 5×5 , 7×7 so on passes along the image matrix. The previously described tiny size matrix cycles the entire image matrix and concretizes the image's attributes. After that, a new image matrix is created.

Pooling: After the convolution process, the pooling layer is frequently applied. A matrix of size $n \times n$ is generated with n determined by the original picture size. This empty matrix slides along the image matrix to obtain the biggest value in the relevant region of the image matrix and assigns a new image matrix's value. The volume of data handled in the network is also lowered in this way.

Flattening: This layer's goal is to prepare the data for the fully connected layer's entrance. Neural networks are typically fed data from a one-dimensional (1D) array. When it comes to the flattening layer, however, the data is in the form of two-dimensional (2D) matrices.

Fully connected: This layer's data is one-dimensional. In this layer, all neuron connections are complete. Each neuron is connected to the neuron in front of it. As a result, it is referred to as the fully connected layer. In ReLU layer, an activation procedure is carried out. The data is sent to the activation function, which returns a value. Many activation functions may be utilized in this layer, including hyperbolic tangent, sinus, and sigmoid.

Softmax: It calculates the classification process's probability value and allocates the class with the highest probability.

The CNN models used in this study are summarized in the following subsections.

DENSENET-201

Deep learning models started to show good results in image classification around 2012. At that time, shallow networks were replaced by deeper ones with many layers. Figure 1 shows the overall CNN model and its layers. The recently created Deep Learning networks have varied architectures and have found solutions to issues like gradient fading and numerous shortcut connections [12]. Contrary to other techniques, DENSENET-201 has improved the efficiency of all the characteristics created by the earlier layers by allowing them to enter the subsequent layer. By allowing access to all feature maps created by the previous layers, the deep layers in the networks can reuse the features [13]. The Densenet-201 structure is shown in Figure 2. The current layer can draw conclusions from earlier layers' features since all previous levels are related to one another [14].

GoogleNet

GoogleNet is a convolutional neural network (CNN) model with 22 layers. The model showed its success by

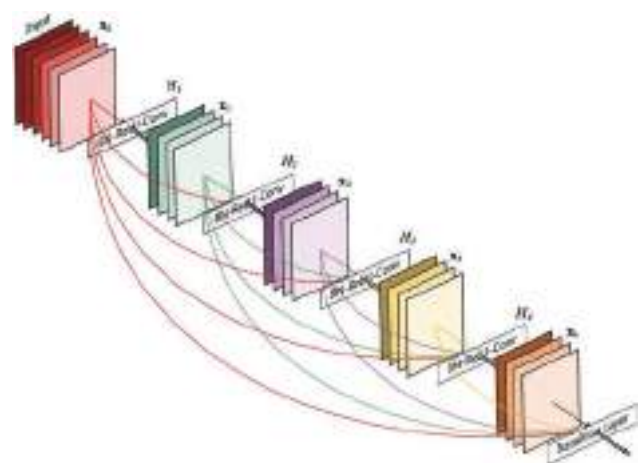


Figure 2. DENSENET-201 architecture [14] [created by the authors]

winning the ILSVRC competition held in 2014. Unlike Densenet-201 model in the GoogleNet architecture, multiple layers work in parallel at the same time. Parallel layers consist of variable sized convolution kernels, 1×1 , 3×3 and 5×5 . The purpose of using 1×1 filters used in the GoogleNet model is to reduce binary dimensionality and summarize the contents in the previous layer [15]. The initial branch of GoogleNet's branching structure is the 1×1 convolution kernel. A 1×1 convolution kernel and a 3×3 convolution kernel are combined in the second branch. A 1×1 convolution kernel and a 5×5 convolution kernel are combined in the third branch. The fourth branch consists of a convolution kernel and a 3×3 maximum pooling kernel. Four outputs are produced by the model's four branches, which process the input data. The final output, the feature map, is produced by combining these outputs with the channel dimensions [16].

Inception V3

The Inception V3 model, which entered the top 5 in ILSVRC-2015, is a 48-layer model. Since the model is initially sparse, it can produce few convolutions [17]. Inception V3 suggests a method to optimize the network and has a bigger mesh than Inception-V1 and V2 models [18]. Increasing computational efficiency in application instances and minimizing the impact of low parameter values on the model are two of the model's primary objectives. These characteristics make it quick to react and simple to use. Convolutional cores of various sizes are used by Inception V3 to enable it to have various receiver areas. Batch normalization (BN) is a layer that is used in Inception V3. Between the fully connected layer and the classifier, this layer serves as a moderator [19].

MobileNet V2

A popular deep learning model for classifying images is MobileNet V2. It is so widespread because it functions on both mobile devices and computers with less processing power. The model has a controllable structure that allows it to regulate the parameters accuracy and latency [20]. The MobileNet V2 mesh consists of convolution layers each 3

x 3 deep, followed by a 1×1 point convolution layer that combines this filter output. Each of the 13 blocks in the previously stated structure contains a point convolution layer. The point convolution layer that MobileNet V1 utilizes allows it to reflect data with many channels to tensors with much fewer channels. Residual connectivity is another benefit of MobileNet V2, since it facilitates gradient flow across the network [21]. The network does not use the usual residual connections used to connect two extended units. Instead, inverted residual links provide the connection between thin blocks. The model deals with the possibility of representing the versatility obtained from a pixel in less dimensional space through channels by expanding the width factor [22].

Xception

Xception is an enhanced version of Inception V3. This model has the advantage of using channel correlation and spatial correlation separately. Therefore, Xception provides a deeply separable network model. There are two distinct components to the deeply separable convolution structure. Each input channel is subjected to independent spatial convolution after which point-to-point convolution is applied using 1×1 cores for point-to-point wrapping. The Xception structure is shown in Figure 3 [23]. There are 71 layers, 36 convolution layers, and 14 blocks in the Xception model. All layers, except for the first and last, are connected linearly. To preserve the sequential features within the network, a linear residual link is introduced as a shortcut. This approach helps to overcome both bottleneck and vanishing gradient problems. Instead of simply concatenating outputs, the method uses an aggregation process to pass information from one layer to the next, which improves stability and learning [24].

Bayesian Convolutional Neural Networks Approach

Bayesian optimization

Bayesian Optimization (BO) has been tried in many different areas, like physics, chemistry, and engineering, and it has shown good results. One reason it works well is that it can keep a balance: it explores new options while still using

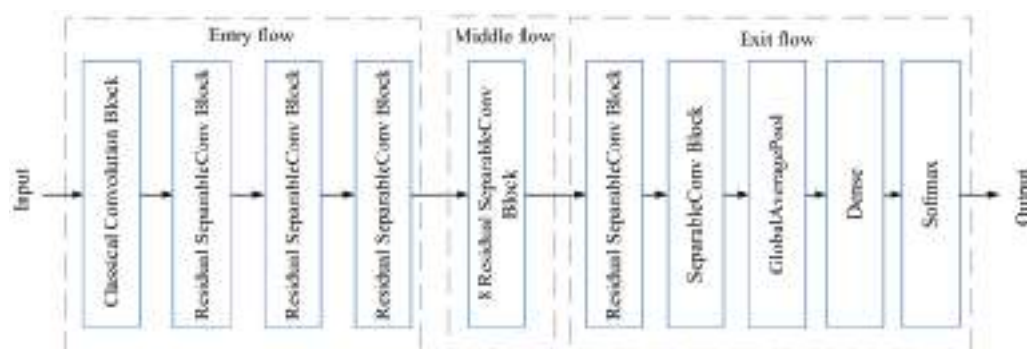


Figure 3. Xception architecture [23] [created by the authors]

the best choices it already knows. This balance is very helpful when decisions need to be made under uncertainty.

Another factor behind its popularity is the growth of open-source tools. These platforms made BO easier to use, and they also opened the door for applying it to multi-objective optimization and experimental design.

Overall, BO is both efficient with data and dependable in practice. Today it is seen as a useful method across many fields, and it has become an especially important tool in machine learning and artificial intelligence.

Bayesian Optimization typically relies on a probabilistic model, often modelled using a Gaussian Process (GP). The GP is a distribution over functions, defined by its mean function $\mu(x)$ and covariance function $k(x, x')$. For a set of observed data $D = \{(x_i, y_i)\}$ where x_i represents input parameters and y_i represents the corresponding function values, the predictive distribution of the GP at a new point x_* is given by Eq.(1).

$$P(f_* | x_*, D) = N(\mu_*, \sigma_*^2) \quad (1)$$

Here, μ_* is the predictive mean, and σ_*^2 is the predictive variance.

The acquisition function guides the optimization process by balancing exploration and exploitation. One common acquisition function is the Expected Improvement (EI), defined as in Eq.(2).

$$EI(x) = E[\max(0, f(x) - f(x^+))] \quad (2)$$

where $f(x)$ is the mean of the GP, $f(x^+)$ is the best observed value so far, and the expectation is taken over the predictive distribution.

The steps of the BO are given in the following:

1. **Set up the surrogate model:** Start with some initial data D and build a Gaussian Process model that will act as an approximation of the real function.
2. **Choose the next point:** Use the acquisition function to decide where to sample next (x_{next}).
3. **Check the real outcome:** Evaluate the actual objective function at that point to get the result (y_{next}).
4. **Update and repeat:** Add this new information into the surrogate model. Keep repeating steps 2–4 until the method either converges or a chosen stopping rule says it's done.

The Gaussian Process is defined by Eq.(3-5).

$$(x) \sim gp(\mu(x), k(x, x')) \quad (3)$$

$$\mu(x) = K(X, x)^T [K(X, X) + \sigma_n^2 I]^{-1} y \quad (4)$$

$$K(x, x') = k_0 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d + x'_d)^2}{l_d^2}\right) \quad (5)$$

where $K(X, x)$ is the vector of covariances between x and all points in the training set X , $K(X, X)$ is the covariance matrix for all pairs of points in X , σ_n^2 is the noise variance, y is the vector of function values in the training set, k_0 is a constant, and l_d are length scale parameters.

Bayesian Optimization leverages the mathematical foundation of Gaussian Processes and acquisition functions to efficiently explore and exploit the parameter space, making it a valuable tool for optimizing complex and computationally expensive objective functions.

A novel deep learning approach by bayesian convolutional neural network

This study initially establishes a simple Convolutional Neural Network (CNN) structure designed for image classification. In this work, we developed a CNN model with eight layers for image classification. The network takes RGB images at 227×227 resolution as input. This size was selected intentionally to shorten the training time and to make the model more practical on different computer systems.

The first convolutional stage applies small filters and is followed by normalization and ReLU activation, after which a pooling step reduces the feature dimensions. This same idea is repeated with more filters in the following layers so that the network can capture increasingly detailed visual patterns.

Once the convolutional part is complete, the extracted features are passed to a fully connected layer with two outputs. Finally, a softmax layer converts the outputs into probabilities, and the classification layer assigns each image to its class. In addition, we applied parameter optimization during training to improve both efficiency and accuracy.

The Bayesian CNN model incorporates Bayesian optimization to enhance the learning processes of traditional Convolutional Neural Network (CNN) architectures. This model is designed to identify the most effective values for important hyperparameters, including MaxEpochs (the maximum number of epochs), InitialLearnRate (the initial learning rate), and MiniBatchSize (the size of the mini-batches).

Bayesian optimization is commonly applied when tuning hyperparameters. Rather than testing values at random, the method evaluates how the model performs and keeps track of the outcomes. Using this information, it proposes the next set of values to examine. Through this gradual adjustment, the model is guided toward better performance in a more practical and less wasteful manner.

When settings like MaxEpochs, InitialLearnRate, and MiniBatchSize are adjusted, the Bayesian CNN can train faster. It also adapts well when evaluated on unseen data. In practice, this not only increases accuracy but also reduces the overall computational cost. As a result, the training process becomes both easier and more efficient.

We can describe the core math behind the B-CNN model as follows:

Algorithm 1. The pseudocode for B-CNN

```

datastore ← (datasetPath, 'IncludeSubfolders', true, 'LabelSource', 'foldernames')
[trainingSet, testSet] ← splitdata(datastore, 0.7, 'randomized')
inputSize ← [227 227 3]
trainingSet_Resized ← arrange_image_size(inputSize, trainingSet)
testSet_Resized ← arrange_image_size(inputSize, testSet)
objectiveFunction = @(params) cnnObjectiveFunction(params, trainingSet_Resized, testSet_Resized, testSet)
Define the parameter ranges for optimization:
param1 = optimizableVariable('MaxEpochs', [5, 50], 'Type', 'integer')
param2 = optimizableVariable('InitialLearnRate', [0.001, 0.1], 'Transform', 'log')
param3 = optimizableVariable('MiniBatchSize', [16, 64], 'Type', 'integer')
Create a Bayesian optimization object:
bayesOpt = bayesopt(objectiveFunction, [param1, param2, param3], 'maximum_iterations', 100)
Get the optimized parameters:
optimizedParams ← bestPoint(bayesOpt)
Train the model using the best parameters:
Layers ← Assign the layer to be optimized
Options ← trainingOptions('sgdm', ...
'MaxEpochs', optimizedParams(MaxEpochs), ...
'InitialLearnRate', optimizedParams(InitialLearnRate), ...
'MiniBatchSize', optimizedParams(MiniBatchSize))
Make predictions using the optimized model:
YPredOptimized = classify testSet_Aug by optimizedNet
YTest ← get the Labels from testSet
Create confusion_matrix (YTest, YPredOptimized);
Get the accuracy value

```

$f(x)$, represents an objective function measuring the performance of the CNN model, such as accuracy or loss function value (accuracy for this study). x , represents hyperparameter values. For example, $x = (\text{MaxEpochs}, \text{InitialLearnRate}, \text{MiniBatchSize})$.

Gaussian Process (GP) Model is constructed as seen in Eq. (6).

$$f(x) \sim GP(m(x), k(x, x')) \quad (6)$$

$m(x)$ is the mean function and $k(x, x')$ is the kernel function. On the other hand, the acquisition function is established as given in Eq. (7)

$$\text{Acquisition Function: } x_{\text{next}} = \arg \max EI(x) \quad (7)$$

x_{next} is the next hyperparameter suggestion in Eq. (7).

$$EI(x) = E[0, f(x_{\text{best}}) - f(x)] \quad (8)$$

$f(x_{\text{best}})$ is the best performance value observed so far. A new hyperparameter set is chosen with x_{next} . This hyperparameter set is given to the CNN model, and the model

is trained. The performance of the model is measured, and the objective function value is obtained. This value is added to the Gaussian Process model, updating the model. This process is repeated until a certain criterion is met (e.g., a specific number of iterations).

This process mathematically expresses the goal of Bayesian optimization to find the best hyperparameter set for improving the performance of a CNN model.

PERFORMANCE MEASURES

In a classification problem, classifiers can be compared using performance measures such as accuracy, precision, precision, and F-measure. The model's success is proportional to the number of samples assigned to the correct class and the number of samples assigned to the incorrect class in the test set. The confusion matrix is used to calculate the performance measures of the classifiers. In the confusion matrix shown in Table 2, the rows represent the actual values of the samples and the columns represent the predicted values [25].

Table 2. Confusion matrix

Actual	Predicted	
	C0	Not-C0
	C0	Not-C0
	(TP)	(FN)
	(FP)	(TN)

The accuracy rate is defined as the ratio of true categorized samples ($TP + TN$) to total samples ($TP + TN + FP + FN$). Eq.(9) can be used to calculate the accuracy rate.

$$AR = (TP + TN)/(TP + TN + FN + FP) \quad (9)$$

Precision is defined as the ratio of True Positives (TP) samples predicted as class 1 to the total number of samples projected as class 1 ($TP + FP$). Eq.(10) can be used to calculate the precision.

$$P = TP/(TP + FP) \quad (10)$$

Sensitivity is defined as the ratio of properly categorized positive samples (TP) to total positive samples ($TP + FN$). Eq.(11) can be used to calculate the sensitivity.

$$S = TP/(TP + FN) \quad (11)$$

Another metric used to assess the success of classification algorithms is the F-measure. The F-measure is the harmonic mean of precision (P) and sensitivity (S), and it can be determined using Eq.(12).

$$F\_measure = (2 * P * S)/(P + S) \quad (12)$$

In practice, these metrics tell slightly different stories about the same model. Accuracy gives a quick overall picture, but precision and sensitivity (recall) point to opposite

kinds of mistakes: precision punishes false positives, while sensitivity punishes false negatives. When we inspected a few borderline cases, the coins with heavy wear tended to hurt sensitivity more than precision, which fits the intuition that faint motifs are easier to miss than to hallucinate. For that reason, we report F-measure alongside accuracy to keep both effects in view.

APPLICATION

Parion is one of the most prominent port cities in the Troas region, which was colonized and built around 709 BC. It is situated in northwest Turkey, in the province of Çanakkale, in the Biga district, close to the village of Kemer. A large number of coins have been uncovered in the ancient city of Parion, where archaeological excavations and surveys have continued for more than two decades. These coins are mainly linked to the Roman and Greek periods, which were therefore chosen as the class labels in this study. Out of the 112 samples, 56 belong to the Roman class, while the remaining 56 are associated with the Greek class. For each coin, the width and height were recorded, and Figures 4–5 illustrate representative examples from both categories.

While training the models, 70% of the dataset was used for training and the rest for testing. When training models in deep learning, a considerable quantity of data is typically necessary. In this way, the model is aimed to increase performance by detecting more features. However, in our study, since the ancient coins belonging to 2 different eras are in a uniform structure, methods such as image reproduction led to overfitting.

Matlab R2021B 64 bit software was employed for analysis. The computer's characteristics include an AMD Ryzen 7 3800X 8-core CPU 3.90 GHz as the processor. RAM has a storage capacity of 32 GB. Matlab trained the models

**Figure 4.** Greek coins**Figure 5.** Roman coins

using the graphics card processor. The graphics card on the machine is an NVIDIA GeForce RTX 3060 12 GB.

Figure 4 presents the performance of the pure CNN models for the training data. In this figure, the x-axis shows the number of iterations, and the y-axis shows the accuracy rate. As can be seen in Figure 5, Densenet-201, Googlenet and InceptionV3 models gave the best results, respectively. The accuracy of the models increased with the number of iterations.

The performance analysis of the pure CNN models for the test data is presented in Figure 6. As can be seen in Figure 7, when the models were tested, the Densenet-201 model performed the best. The models' validation baselines grew by 40–60% on average and varied linearly across 110 iterations. The accuracy rate of the Densenet-201 model for the test data is 95.54%. When compared to other models, Googlenet produced the second-best result, with a accuracy rate of 93.75%. The main factor in the Densenet-201 model's performance is that it eliminates fading gradients, which reduces feature loss. To do this, each layer creates a feature map that incorporates all of the preceding levels' characteristics. However, it is believed that the parallel layer design is the major component in producing good results, such as the Googlenet model, which ranks second. The

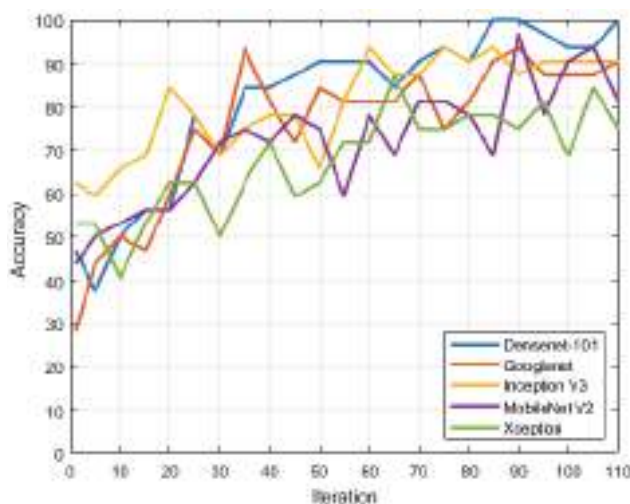


Figure 6. Performance analysis of the pure CNN models for the training data

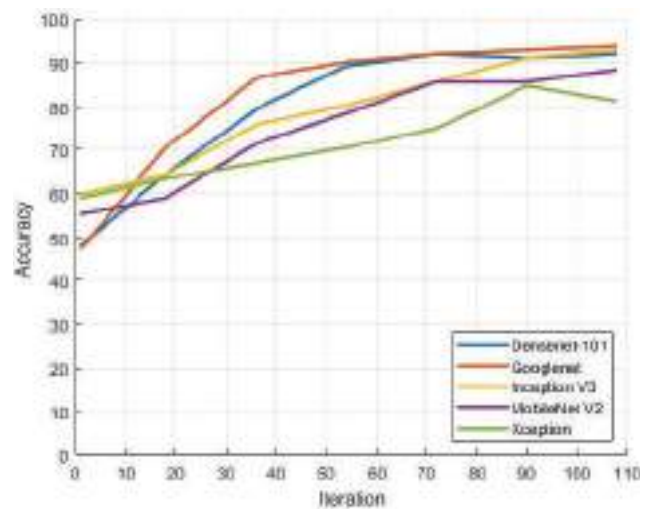


Figure 7. Performance analysis of the pure CNN models for the test data

reason that these two models outperform other models is that they strive to minimize feature loss. When the graphs of both models are reviewed, the gradient loss is kept to a minimal, allowing them to draw a linear trend.

The performance measures and run times of the pure models are presented in Table 3.

The proposed model within the scope of the study is the B-CNN model, and performance metric values associated with the best objective function obtained after a 100-iteration optimization process have been calculated.

The hyperparameter ranges utilized for the B-CNN approach are provided in the table below. Additionally, the best parameter values obtained at the end of the optimization process are also presented in this table.

Table 3. Hyperparameter to be optimized in B-CNN

Hyperparameter	Solution space	Optimum value
MaxEpoch	(5, 50)	8
InitialLearnRate	(0.001, 0.1)	0.0011289
MiniBatchSize	(16, 64)	35

Table 4. Performance analysis of the pure CNN models

MODEL	Se. (%)	Sp. (%)	Pre. (%)	F-Scr.	Acc (%)	Collapse Time
DENSENET-201	98.11	93.22	92.86	0.954	95.54	6 min 17 sec
GoogleNet	92.86	94.64	94.55	0.936	93.75	2 min 38 sec
InceptionV3	92.98	91.07	91.38	0.921	92.86	3 min 56 sec
MobilenetV2	85.71	91.07	90.57	0.880	88.39	5 min 8 sec
Xception	84.91	83.93	83.33	0.841	82.14	3 min 4 sec
B-CNN	94.44	100	100	0.981	97.06	15 min 3 sec

The optimal hyperparameter values obtained in Table 3 are provided. These values were obtained on the training set and applied to the test dataset as well.

Table 4 compares the results of different deep learning models for classifying ancient coins. The table shows Sensitivity (Se.), Specificity (Sp.), Precision (Pre.), F-Score, Accuracy (Acc.), and Collapse Time.

DenseNet-201 gave good results, with 98.11% sensitivity and 95.54% accuracy. It took around six minutes to train, which isn't the fastest, but it works. GoogleNet performed well, with 93.75% accuracy and 94.64% specificity, and it also trained very quickly—only 2 minutes and 38 seconds. InceptionV3 gave balanced results, showing 92.86% accuracy and 91.07% specificity.

MobileNetV2 and Xception performed less strongly. MobileNetV2 reached 88.39% accuracy, while Xception achieved only 82.14%. Although these models are faster, their accuracy is lower.

The Bayesian Convolutional Neural Network (B-CNN) produced the best outcomes, reaching 100% specificity and precision with about 97% accuracy. However, training took much longer, nearly 15 minutes. This indicates that B-CNN is highly dependable but requires more computational power. For situations where training speed is important, DenseNet-201 and GoogleNet are more practical options.

Overall, B-CNN delivered the highest accuracy and reliability, making it valuable for sensitive fields such as archaeology. Still, the longer training time needs to be considered when selecting the model.

If we take a closer look at the methods listed in Table 4, it quickly becomes apparent that no single approach can be considered flawless. Each has certain points where it performs well and others where it lags behind. For instance, DENSENET-201 often shows a strong outcome in terms of sensitivity, which makes it appealing when that metric is of central importance. However, this strength should not be interpreted as a guarantee of equally strong results in every other category. In fact, in some cases it may fall short, and this highlights the need to evaluate the model with a balanced view. On the other hand, GoogleNet is usually mentioned for its shorter training time. This can be an undeniable advantage when rapid experimentation or limited computing resources are part of the research setting. Its speed does not always guarantee the highest level of accuracy, and this kind of trade-off is something that scholars and practitioners alike need to weigh with care. After all, in most real research settings there is rarely a single “best” option that works under every condition. Much depends on what the study is trying to achieve and the constraints under which it is carried out. For some projects, efficiency and shorter training times may be the decisive factor, while in others the focus may shift toward maximizing precision, even if that requires more time and resources. Archaeological applications provide a good example of this dilemma: a misclassified artifact is not just a numerical error but could potentially alter the

historical interpretation of an entire site. For this reason, accuracy often receives special emphasis in such contexts, though in time-sensitive studies a more balanced or pragmatic solution may still be considered acceptable. For some projects, efficiency and shorter training times may be the decisive factor, while in others the focus may shift toward maximizing precision, even if that requires more time and resources. Archaeological applications provide a good example of this dilemma: a misclassified artifact is not just a numerical error but could potentially alter the historical interpretation of an entire site. For this reason, accuracy often receives special emphasis in such contexts, though in time-sensitive studies a more balanced or pragmatic solution may still be considered acceptable. In other studies with stricter time constraints, however, the faster option could be more practical.

CONCLUSION

This study shows a good example of cooperation between archaeology and computer science. It proves that an interdisciplinary approach can help solve complex problems. Dating and classifying old coins, especially from Greek and Roman times, has always been difficult in archaeology and needs special experts. This research uses deep learning to make the process faster and less dependent on humans.

Different Convolutional Neural Networks were tested. Among them, DenseNet-201 gave the best result with 95.54% accuracy, so it was the strongest of the classical models. The main focus, however, was on the Bayesian Convolutional Neural Network (B-CNN). This model uses Bayesian optimization to improve CNN performance. After 100 iterations of optimization, the model reached 97.06% accuracy, with better sensitivity, specificity, and precision than the other models.

The B-CNN improved performance a lot, but training took longer. The collapse time was about 15 minutes, which can be a problem when time is limited. For this reason, the choice of model should depend on the needs of the archaeological work.

In conclusion, this study shows how deep learning can be applied in archaeology. The use of Bayesian optimization makes the method different from earlier studies and gives high accuracy and reliability. This interdisciplinary work creates a base for future research and opens new ways to study historical artifacts with artificial intelligence.

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] Thonemann P. The Hellenistic World: using coins as sources. Cambridge: Cambridge University Press; 2015. [\[CrossRef\]](#)
- [2] Atlan S. Grek sikkeleri. Istanbul: Arkeoloji ve Sanat Yayınları; 1993.
- [3] Klawans ZH. Handbook of ancient Greek and Roman coins: Whitman Coin Pub; 1995.
- [4] Aslan S, Vascon S, Pelillo M. Two sides of the same coin: Improved ancient coin classification using graph transduction games. Pattern Recogn Letters 2020;131:158-165. [\[CrossRef\]](#)
- [5] Anwar H, Anwar S, Zambanini S, Porikli F. Deep ancient Roman Republican coin classification via feature fusion and attention. Pattern Recogn 2021;114:107871. [\[CrossRef\]](#)
- [6] Kayaalp K, Özkaner F. Bi-directional classification of roman period coins by deep learning methods. Int J Eng Innov Res 2023;5:161-169. [\[CrossRef\]](#)
- [7] Zambanini S, Kavelar A, Kampel M, editors. Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation. 2014 22nd International Conference on Pattern Recognition; 2014: IEEE. [\[CrossRef\]](#)
- [8] Temiz H, Gökberk B, Akarun L, editors. TurCoins: Turkish republic coin dataset. 2021 29th Signal Processing and Communications Applications Conference (SIU); 2021: IEEE. [\[CrossRef\]](#)
- [9] Aslan S, Vascon S, Pelillo M, editors. Ancient coin classification using graph transduction games. 2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo); 2018: IEEE. [\[CrossRef\]](#)
- [10] Kavelar A, Zambanini S, Kampel M, Vondrovec K, Siegl K. The ILAC-project: Supporting ancient coin classification by means of image analysis. Int Arch Photogram Remote Sens Spat Inform Sci 2013;40:373-378. [\[CrossRef\]](#)
- [11] Anwar H, Zambanini S, Kampel M. Coarse-grained ancient coin classification using image-based reverse side motif recognition. Machine Vision Appl 2015;26:295-304. [\[CrossRef\]](#)
- [12] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al, (editors). Going deeper with convolutions. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2015. [\[CrossRef\]](#)
- [13] Yu X, Zeng N, Liu S, Zhang Y-DJMV, Applications. Utilization of DenseNet201 for diagnosis of breast abnormality. Machine Vision Appl 2019;30:1135-1144. [\[CrossRef\]](#)
- [14] Adam K, Mohamed II, Ibrahim YJIA. A selective mitigation technique of soft errors for dnn models used in healthcare applications: Densenet201 case study. IEEE Access 2021;9:65803-65823. [\[CrossRef\]](#)
- [15] Balagourouchetty L, Pragatheeswaran JK, Pottakkat B, Ramkumar GJljob, informatics h. GoogLeNet-based ensemble FCNet classifier for focal liver lesion diagnosis. IEEE J Biomed Health Inform 2019;24:1686-1694. [\[CrossRef\]](#)
- [16] Yang N, Zhang Z, Yang J, Hong Z, Shi JJNRR. A convolutional neural network of GoogLeNet applied in mineral prospectivity prediction based on multi-source geoinformation. Nat Resour Res 2021;30:3905-3923. [\[CrossRef\]](#)
- [17] Al Husaini MAS, Habaebi MH, Gunawan TS, Islam MR, Elsheikh EA, Suliman FJNC, et al. Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4. Neural Comput Appl 2022;34:333-348. [\[CrossRef\]](#)
- [18] Mujahid M, Rustam F, Álvarez R, Luis Vidal Mazón J, Díez IdlT, Ashraf IJD. Pneumonia classification from X-ray images with inception-V3 and convolutional neural network. Diagnostics (Basel) 2022;12:1280. [\[CrossRef\]](#)
- [19] Cao J, Yan M, Jia Y, Tian X, Zhang Z. Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals. EURASIP J Adv Signal Process 2021;2021:1-25. [\[CrossRef\]](#)
- [20] Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM. Sensors (Basel) 2021;21:2852. [\[CrossRef\]](#)
- [21] Michele A, Colin V, Santika DD. Mobilenet convolutional neural networks and support vector machines for palmprint recognition. Proced Comput Sci 2019;157:110-117. [\[CrossRef\]](#)
- [22] Shen Y, Sun H, Xu X, Zhou J, (editors). Detection and positioning of surface defects on galvanized sheet based on improved MobileNet v2. 2019 Chinese Control Conference (CCC); 2019: IEEE. [\[CrossRef\]](#)
- [23] Chen B, Liu X, Zheng Y, Zhao G, Shi Y, Technology SfV. A robust GAN-generated face detection method based on dual-color spaces and an improved Xception. IEEE 2021;32:3527-3538. [\[CrossRef\]](#)

-
- [24] Kassani SH, Kassani PH, Khazaeinezhad R, Wesolowski MJ, Schneider KA, Deters R, (editors). Diabetic retinopathy classification using a modified xception architecture. 2019 IEEE international symposium on signal processing and information technology (ISSPIT); 2019: IEEE. [\[CrossRef\]](#)
- [25] Pekel Özmen E, Özcan T. Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm. J Forecast 2020;39:661-670. [\[CrossRef\]](#)



Research Article

Cat swarm optimized tumor segmentation and an ensemble model for glioblastoma patient survival prediction

Bhagyalaxmi K¹, Dwarakanath B²

¹Matrusri Engineering College Computer Science and Engineering Department, Hyderabad, Telangana, India

²SRM Institute of Science and Technology Department of Information Technology, Ramapuram, Tamilnadu, India

ARTICLE INFO

Article history

Received: 18 May 2024

Revised: 24 June 2024

Accepted: 18 September 2024

Keywords:

Cat Swarm Optimization, Glioblastoma, Stereotactic Radio surgery, Survival Prediction, Deep Learning, ResNet, U-Net, Ensemble Prediction, Random Forest, Tumor Segmentation

ABSTRACT

Glioblastoma is characterized as the most common and lethal primary brain tumor among adults, glioblastoma is associated with enormous clinical management challenges due to its high rate of recurrence, poor prognosis and underlying complexity. Here we propose a study to improve the prediction of overall survival in GBM patients treated with stereotactic radio-surgery using advanced image segmentation and machine learning techniques. In this paper, we propose a novel fusion of Cat Swarm Optimization based hybrid ResNet and U-Net models to achieve an accurate segmentation of the tumor as well as an ensemble of machine learning algorithms for survival prediction enabling us to overcome limitations of conventional techniques. We demonstrate that on the BraTS2020 related regions we are able to produce almost perfect segmentations with metrics, like 99.2% segmentation accuracy, loss of 0.023, recall of 0.986, a mean intersection over union (IOU) of 0.991, a dice coefficient of 0.96, a precision of 0.991, a sensitivity of 0.991, and a specificity of 0.997. For in the field of survival prediction we looked at many machine learning models which we found out that the Random Forest did an outstanding job at handling the complex issues presented by the segmented images. Also we saw that the Ensemble method did very well in this area which we report to have achieved 60.01% accuracy. We present that which is the use of in depth image segmentation in combination with machine learning greatly improves results in Glioblastoma survival prediction. Also this approach does not only improve the prognostic accuracy but also brings to the table what may be game changing elements in clinical management and personalized treatment.

Cite this article as: Bhagyalaxmi K, Dwarakanath B. Cat swarm optimized tumor segmentation and an ensemble model for glioblastoma patient survival prediction. Sigma J Eng Nat Sci 2025;43(5):1592–1606.

INTRODUCTION

The High degree of recurrence, poor prognosis, tumor heterogeneity, and limited treatment options are the main

issues in Glioblastoma. Glioblastomas' complex and heterogeneous nature contribute to poor overall survival and also to treatment issues. While we see an increase in the use of Stereotactic Radiosurgery (SRS) which is very precise in

*Corresponding author.

*E-mail address: bhagyareddy20380@gmail.com, bk8019@srmist.edu.in

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



its delivery of radiation therapy to the tumor thus sparing the surrounding healthy tissue we are still have issues in accurate patient outcome prediction. In the area of artificial intelligence (AI) which includes deep learning (DL) we are seeing very promising results in the identification and management of glioblastomas' but still accurate prognosis is impaired by tumor complexity.

Our research has put forth a new approach which is a combination of Cat Swarm Optimization-assisted hybrid ResNet and U-Net architectures for accurate tumor segmentation with machine learning algorithms for very robust survival prediction. This we have done to outperform traditional models which we do so by improving segmentation accuracy and at the same time we are able to present more reliable survival predictions. We are into this with the aim to improve individualized treatment plans for GBM patients that are undergoing stereotactic radiosurgery. What we have done not only improves the care we give to glioblastoma patients but also we put forth the wide scale promise of AI driven methods in oncology and personal medicine. Our work gives out that which may be applied across many cancer types which in turn bridges the gap between very advanced AI techniques and clinical practice.

Related Work

The inroads that machine learning (ML) has made in brain tumor image analysis which includes the use of MPMRI scans [1] is a result of initiatives like the International Brain Tumor Segmentation (BraTS) challenge. In this field Semantic segmentation is a key element which it enables the tell between target objects and adjacent tissues. As for the methods used, Fully Convolutional Networks (FCN) which use up sampling methods to do accurate pixel level labeling have proven to be very effective.

To better predict overall survival (OS) we have seen the introduction of more complex approaches. For example a hybrid model which puts together a DenseNet based 3D neural network with a position encoding convolutional layer (PECL) [2] has been put forth. The model described above applies semi-supervised learning to minimize overfitting while extracting essential features from T1 contrast MRIs, T2 MRIs, and pre-segmented sub-regions. Furthermore, as previously mentioned, texture features retrieved from the MRI are useful for biologically characterizing subtypes in a noninvasive manner. Numerous researchers have been active in the area and have created various models and approaches. Thuong-Cang Phan [3] used MobileNet-V2,

Table 1. Summary of literature review

Author Name and Reference No	Segmentation methods	Prediction methods	Dataset	Results
Raza et al., [10]	3D deep residual U-Net	hybrid of the deep residual network and U-Net model	BraTS	dice score for the tumor core (TC), whole tumor (WT), and enhancing tumor (ET) on the BraTS 2020 dataset of 0.8357, 0.8660, and 0.8004, respectively.
Vijay, et al.,[11]	3D U-Net	residual Spatial Pyramid Pooling-powered	BraTS	average dice score of 0.883 and a Hausdorff distance of 7.84 on Brats 2021 cross validation.
Yang D et al.,[12]	tumor segmented Manually	Ensemble classifier - Random Forest	Cancer Genome Atlas (TCGA)	Area Under Curve: 0.62: features with aerial scan 0.61 features with coronal scan 0.62: features with sagittal scan
Montaha et al., [13]	2D U-net architecture	CNN	BraTS	acc of 99.41% and dice similarity coeff (DSC) of 93%
Sangui, S et al., [14]	modified U-Net architecture	U-Net	BraTS	Test acc of 99.4%
Mohammadreza et al.,[15]	Res-Net Segmentation method	Random Forest	BraTS	Segmentation acc:0.80 Survival validation accuracy:0.45 Testing acc: 0.52
Fabian, et al.,[16]	U-Net Segmentation Method	NA	BraTS	Segmentation Acc Core: 98% Whole:96%
Muhammad Junaid Ali, et al.,[17]	2D and 3D U-Net Segmentation method	Random Forest Repressors	BraTS	Segmentation Acc Dice Score: 78% Prediction Acc:48%

ResNet-101, and DenseNet-121 for tumor segmentation in detail and used 3D slicer tools for MRI imaging. Eyad Gad [4] explored an attention-based 3D UNet model for segmentation and claimed to have improved the segmentation by concentrating on useful portions of the image. IKE Purnama [5] used a modified U-Net and claimed to have improved detail and accuracy in the segmented images. Zhao J [6] proposed Uncertainty-aware Multi-dimensional Mutual Learning.

Gayathri Ramasamy [7] constructed a multi-modal modified Link-Net model utilizing Squeeze and Excitation ResNet152 as a segmentation backbone for improved segmentation performance. Cihan M [8] applied convolutional neural networks for fusion of phase MRI images: pre-contrast, venous, arterial, and delayed phases. With all these advancements, there are still limitations on the applications of research results. Some studies, for instance, report an OS prediction accuracy of merely 44.8%, illustrating ongoing difficulties in achieving dependable clinical results. A. Durgut [9] studies image processing using parallel and serial programming methods and finds parallel programming reduces the time required for processing both whole and segmented images. Whole image processing outperformed segmented image processing. The research highlights the efficiency of parallel programming in improving image processing tasks. Table 1 provides further insights into related research contributions and their respective metrics.

Using nature-inspired optimization methods such as genetic algorithms as well as hybrid filters can provide solutions to the issues of brain tumor image analysis[18]. The improvement in the use of these approaches provide methods for optimizing hyperparameter parameters of a model, reduce the effects of over fitting and improve the original feature extraction to enhance model accuracy and generalizability. Through mimicking natural behavior and incorporating heterogeneous filtering methods, these approaches present a valuable strategy to enhance productivity and applicability of machine learning models for use in clinical settings.

MATERIALS AND METHODS

The issue we address in this work is that of accurate prediction of overall survival (OS) in glioblastoma multiform (GBM) patients that are undergoing stereotactic radio surgery (SRS). GBM is known for its aggressive nature and poor prognosis which in turn sees the rise of SRS as a key treatment modality for precise tumor cell targeting. That said accurately prognosticating OS in GBM patients post SRS is still a challenge which we put down to the heterogeneous nature of the disease and also to the variable patient responses to treatment. Also included are the goals of our work.

- Preprocessing the Dataset for feature extraction and classification by applying a Hybrid Bilateral_Gaussian

Filter, and normalization to reduce noise, and improve performance.

- Segmentation for tumor by Cat Swarm Optimization for Medical Image Segmentation (CSO-MIS) assisted hybrid ResNet and U-Net deep learning model.
- Ensemble of popular machine learning algorithms—Random Forest, KNN models, and SVM—as well as the deep learning model ANN are employed to predict the OS.
- To evaluate the performance. Compare accuracy, sensitivity, specificity, and area under the ROC curve metrics

Implementation

Figure 1 showcases the remarkable process of Glioma sub region segmentation and predicting overall survival in glioblastoma patients using deep learning architecture. First out of the gate we do in depth pre processing of the BRATS 2020 dataset which we do to that very meticulous degree which we do for the sake of accuracy. Then we do segmenting which in turn does what it can to improve the accuracy of the results. Also we put the data into train, test, and validate sets which we do to make our analysis as thorough as possible. We see this approach to be very effective in that it puts out very high in terms of accuracy and reliability in what we predict for overall survival in patients with glioblastoma.

Dataset

As part of an annual challenge related to the segmentation of brain tumors using multimodal MRI scans, a set of datasets called brain tumor segmentation (BraTS) 2020 has been made available. The BraTS 2020 dataset contains glioma patient magnetic resonance imaging (MRI) scans from different medical facilities. The participating institutions are focused on glioma scans. To ensure accuracy and reproducibility of results, standardization of protocols has been done. The dataset also contains expert annotated delineations for various tumor subregions, which are essential for training supervised learning models. The delineations are demonstrated through:

- Label 1: Non-Enhancing Tumor Core (NCR/NET)
- Label 2: Peritumoral Edema (ED)
- Label 3: GD-enhancing Tumor (ET)

For each case, clinical information such as the age of the patient is provided. Survival data, i.e., the number of days the patient survived post-initial diagnosis, might also be available for a subset of patients.

Data Preprocessing

To preprocess BraTS2020 MRI dataset, performed a sequence of normalization followed by application of hybrid filter for noise reduction. This approach ensures that the data is scaled properly and noise is reduced, which are crucial steps for improving the performance of downstream tasks like segmentation and classification. A crucial first step in data analysis, particularly in the domains of machine learning and image processing, is normalization.

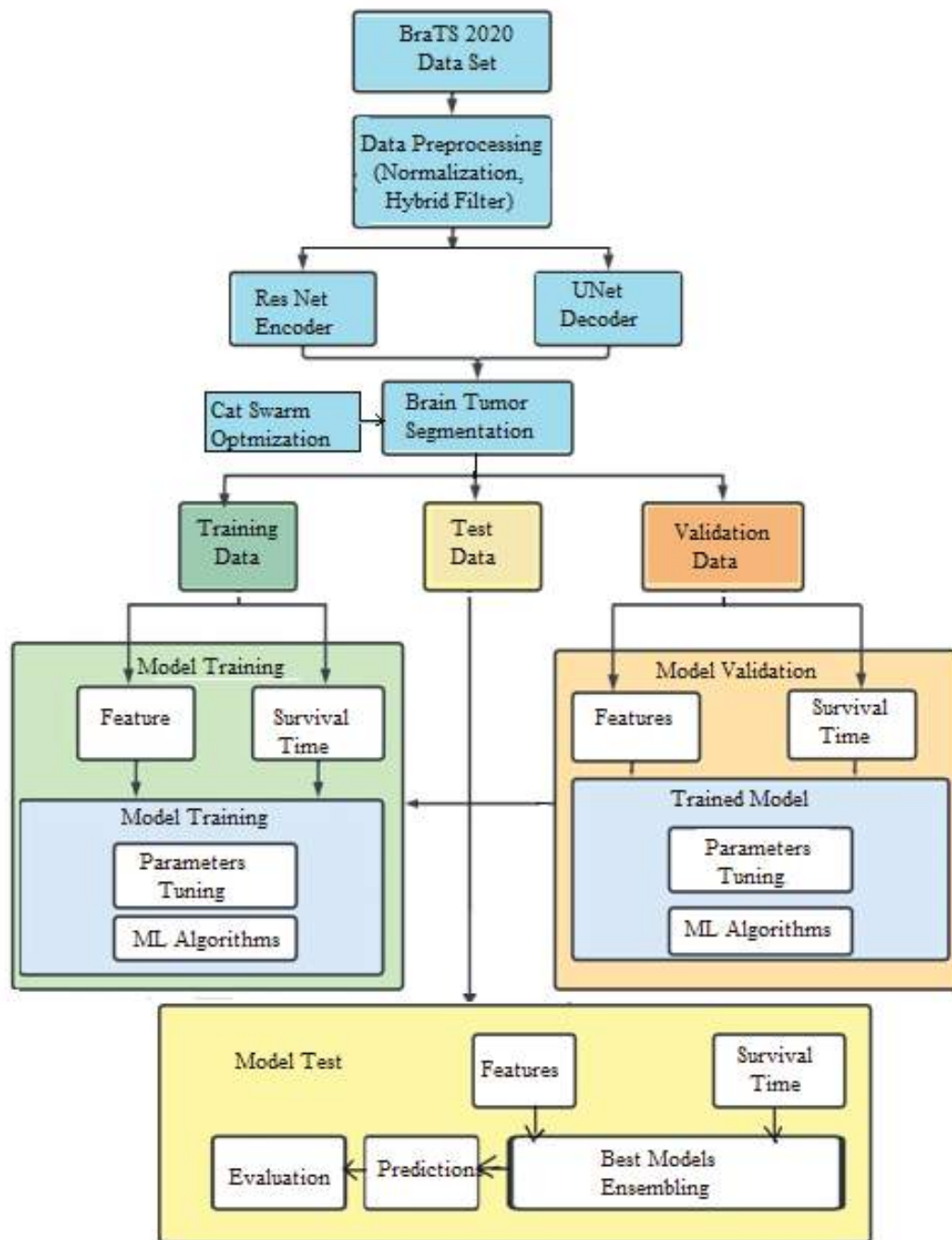


Figure 1. Architecture diagram.

This includes setting the data values onto a common scale so we can stabilize training processes and improve algorithm performance. Normalization can account for variability in intensity from medical images (e.g., MRI scans) that might occur due to the use of different scanning protocols, or machine calibration during scans from other institutions.

This process generally rescales image intensity values of MRI scans to a standard range of $[0, 1]$ or $[-1, 1]$, which enhances image comparison and analysis. Perform

the normalization and then apply a Hybrid filter as given in Algorithm 1 to each slice to reduce noise. As given, the Hybrid filter reduces noise within the MRI images while improving the overall image quality with local analysis of pixel neighborhoods. This normalization and noise reduction with the Hybrid filter prepares the BraTS2020 MRI dataset for more effective analysis in segmentation models or other machine learning tasks by fostering higher quality images and greater consistency throughout the dataset.

HybridBilateral_Gaussian Filter

Algorithm1: Hybrid Filtering (Bilateral_Gaussian Filter)

```

1   Input:
2       Original image  $org(i,j)$ 
3       Gaussian sigma value  $\sigma$  for Gaussian filtering

4   Bilateral Filtering
5       Apply bilateral filtering to the org image to result a Denoised image
6       Denoised img:  $org\_denoised(i,j)=denoise\_bilateral(org(i,j))$ 

7   Gaussian Smoothing
8       Apply Gaussian smoothing to the org image with the specified sigma value to obtain a smoothed version.
9       Smoothed image:  $org\_smoothed(i,j)=gaussian\_filter(org(i,j),\sigma)$ 

10  Hybrid Filtering
11      For each pixel  $(i,j)$  in the image:
12          Calculate the hybrid filtered pixel value as the sum of the Denoised pixel value
              and the difference between the org pixel value and the smoothed pixel value.

13          Hybrid filtered pixel value:
14           $hybrid\_filtered(i,j)=org\_denoised(i,j)+(org(i,j)-org\_smoothed(i,j))$ 

15  Output:
16  Return Hybrid filtered image:  $hybrid\_filtered(i,j)$ 

```

Model

The suggested work uses a ResNet encoder and a U-Net decoder which captures the advantages of both networks. This fused method uses the efficient feature extraction of ResNet and the segmentation characteristics of U-Net to achieve feature extraction and spatial information preservation concurrently. The deep layers of ResNet improve the fine details retrieval, and U-Net's encoder-decoder configuration accurately captures the details in precise segments. The residual connections of ResNet also address overfitting to improve generalization. This enhances the model's strength on the variability and complexity of the medical images, which make it difficult to analyze. This fusion has demonstrated state-of-the-art performance across segmentation tasks, offering a powerful and robust solution for accurate tumor segmentation.

Resnet as Encoder U-Net As Decoder

To sufficiently segment brain tumors within the BraTS2020 dataset using the strengths of both ResNet and U-Net architectures, it is important to comprehend how ResNet may serve as an encoder in a U-Net framework and how this bestows certain advantages, along with its influence on the overall functionality of the network. The ResNet-Unet Architecture is represented in Figure 2.

The encoder portion of U-Net, harnesses ResNet's capacities for deep learning while the U-Net architecture allows for ample amount of local context with its skip connections. This allows for ResNets reliable feature extraction via residual blocks, particularly with deeper architectures such as ResNet-50 or ResNet-101. Each residual block is fundamentally composed of two parts: several convolutional layers that are viewable as executing the main operation; and a skip connection, which sums the block input with the block output. This particular arrangement solves the issue of vanishing gradients, expediting the ability to train deeper network architectures without degradation in performance because of sufficient gradient flow. U-Net in both its fundamental approach and framework for model development as well.

The methods given above are highly specialized for brain tumor segmentation and their development needed expertise and extensive experimentation [19]. The network gains from the ability to encode more abstract and complicated information from the input images when ResNet is used as the encoder in U-Net. This is especially useful for medical imaging jobs where it can be critical to discern minute distinctions between tumor types and normal tissues, such as tumor segmentation in MRI scans. ResNet's improved depth and feature extraction capabilities enable the network to collect

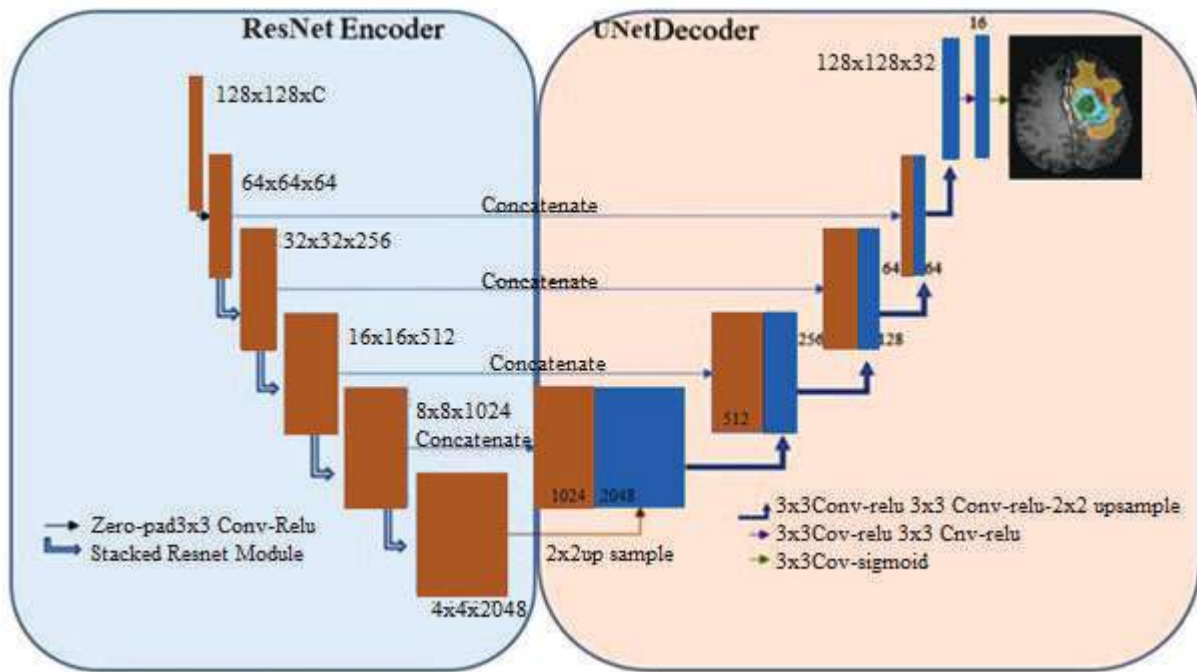


Figure 2. ResNet-Unet architecture.

the finer features and the larger context required for precise segmentation.

Mathematically, a residual block in ResNet can be represented as in equation 1:

$$\text{Output} = \text{ReLU}(\text{BatchNorm}(\text{Conv}(x))) + x \quad (1)$$

Where: x is the input to the block, Conv denotes a convolutional layer, Batch Norm represents batch normalization.

ReLU is the rectified linear unit activation function and $+x$ represents the skip connection. Several residual blocks stacked together make up the encoder part of the network. The encoded representation is the output of the last residual block. In the U-Net decoder, the segmentation map is generated from the encoded features through the sequential processes of up convolution and concatenation. Up convolution (also termed transpose convolution) is a process in the decoder that enlarges the spatial dimensions of the feature maps. These up sample maps are then concatenated with the same connected feature maps from the encoder, via the skip connection. This procedure preserves spatial hierarchies, which aids in correct localization, and enhances the feature maps with fine-grained information required for exact segmentation. Mathematically, the up sampling operation can be represented as in equation 2:

$$\text{Upsampled} = \text{Conv}(\text{Upsample}(x)) \quad (2)$$

Where x is the input to be up sampled, up sample denotes the up sampling operation, Conv represents a

convolutional layer. Then, to incorporate skip connections, we concatenated the corresponding feature maps from the encoder with the up sampled feature maps in the decoder. The combined feature maps are then processed through additional convolutional layers to refine the segmentation mask.

Optimization

In this study we present our put forth of Novel Cat Swarm Optimization for Medical Image Segmentation (CSO-MIS). This Cat Swarm Optimization which we have used improves medical image segmentation by fine tuning the parameters of the deep learning models we use for tumor segmentation. We apply CSO to optimize the hybrid ResNet and U Net architectures which in turn improves their performance in accurate tumor detection. By tuning hyperparameters and network settings we see an improvement in segmentation accuracy and efficiency which in turn leads to better identification of key tumor areas. In this algorithm cats we see exploration of promising solutions and at the same time we have elements of random exploration which together do a good job at the balance of exploration and exploitation to find high quality segmentations. Our algorithm has cats which over time play in and out of the solution space to do what it takes to find the best segmentations of the medical image. CSO is inspired by the hunting behavior of cats and their ability to cooperate to solve complex tasks. Here's a simplified version of how CSO can be applies for this implementation:

Algorithm 2: Cat Swarm Optimization for Medical Image Segmentation (CSO-MIS)

```

1  Input: Medical image img, Number of cats Nm, Maximum iterations T, Learning rate  $\eta$ , Exploration rate  $\epsilon$ 
2  Output: Segmented image Seg*
3  Procedure:
4  Initialization:
5      Randomly initialize Nm cats in the solution space representing potential segmentations of the image.
6      Set best solution found so far Seg*=None and best fitness best_fitness= $\infty$ .
7  Main Loop:
8      for  $t=1$  to T do:
9          Update each cat:
10             Move each cat towards the best solution found so far with a probability of  $\eta$ .
11             else, randomly explore the solution space with a probability of  $\epsilon$ .
12             Evaluate the fitness of each solution (segmentation).
13             Update best solution:
14                 if any solution has a better fitness than best_fitness then:
15                     Update best_fitness and Seg* with the best-performing solution.
16 Segmentation Inference:
17     Use the best solution Seg* obtained from CSO to segment new medical images.
19 Output:
20     Segmented image Seg*.
End

```

Randomly initialize a cats population representing different sets of model parameters. for each iteration, Evaluate the fitness of each model (cat) using the fitness function and the training dataset. The fitness function is as given in equation 3.

$$Fitness(S) = \frac{2X|S \cap G|}{|S| + |G|} \quad (3)$$

Where: S is the predicted segmentation mask,

G is the ground truth segmentation mask,

$|S \cap G|$ is the number of overlapping pixels between S and G,

$|S|$ and $|G|$ are the total number of pixels in S and G respectively.

Update the present best solution based on fitness values. Update the position of each cat using the rules of CSO which include both exploitation (movement toward the best solution) and exploration (random search). The new position $x_i(t+1)$ of the i -th cat at time $t+1$ is given by the equation. 4:

$$x_i(t+1) = x_i(t) + \eta \cdot (global\_best - x_i(t)) + \epsilon \cdot rand() \cdot (upr\_bound - lwr\_bound) \quad (4)$$

Where: At time t the position of the i -th cat is x_i , and η is the learning rate.

Global_best is the best solution to date, ϵ is the exploration rate.

rand () returns a random number between 0 and 1.

Upr bound and lwr bound define the solution space.

Update the best solution found so far based on the fitness of the cats' positions. If any Cat finds which set of

parameters performs better for improved segmentation results, report that as the best solution at present. Do this for a fixed number of iterations or until we see that the model has converged. Once the optimization process is done, train the segmentation model with the best set of parameters found by the CSO on the full training dataset. Then put the trained model to the test on a separate validation set to see how it does. This step also makes sure the model does well on data it has not seen before.

RESULTS AND DISCUSSION

The in Figure 3 we present a ROC curve which reports on a classification model's performance across three different classes. We plot True Positive Rate (sensitivity) against False Positive Rate which in turn gives us info into how the model does at different threshold levels. A curve that reaches towards the top left hand corner of the plot indicates better performance which in turn means high sensitivity and low false positives. By looking at the three curves we are able to see which class the model does best at identifying as well as the models' overall performance which in turn may point out areas for improvement. Also if we see large deviations from the ideal curve that may indicate issues with the models' predictions in certain areas which in turn may require more refinement. The dice coefficient for both the training and validation datasets over 50 epochs is presented in Figure 4.

The plot represents an overall upward trend, indicating that segmentation accuracy improves as training progresses. This suggests that the model is learning effectively and becoming more accurate with each epoch for both the training and validation data.

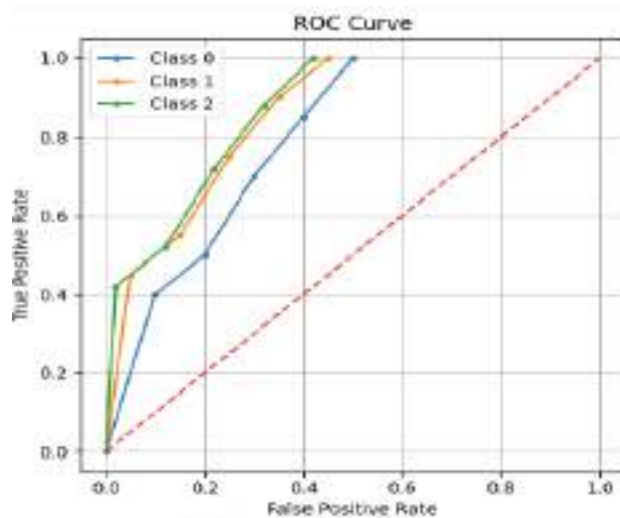


Figure 3. ROC curve.

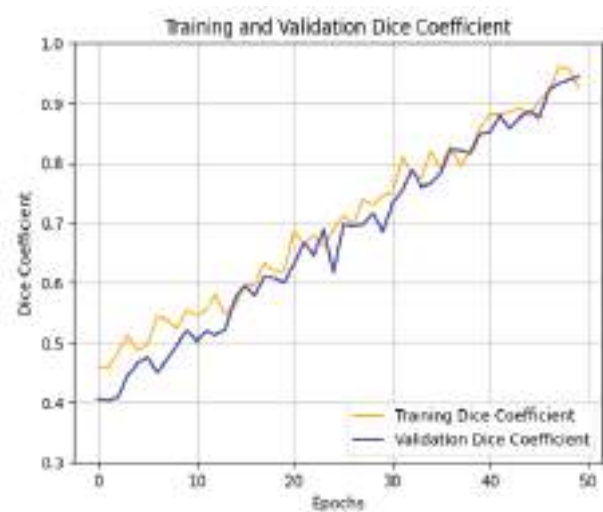


Figure 4. Dice coefficient.

Figure 5 shows the training and validation accuracy (left) and loss (right) over 50 epochs. From the accuracy curves showing the training and validation group data, we see that both are shifting in a consistent upward direction, suggesting that the model is improving in making accurate predictions over time. We also see that the loss curves are decreasing in a consistent manner, indicating that, as training continues, the model's predictions are making fewer mistakes (errors) with each epoch. The general stability in the training and validation accuracy and loss curves suggests that the model is learning to generalize from the training data to the validation data with no signs of overfitting, which improves overall model performance for segmentation tasks. Figure 6 shows the Training and Validation Mean Intersection over Union (IOU) over 50 epochs, where

the X-axis represents the epochs and the Y-axis shows the Mean IOU score. Both the Training and Validation Mean IOU scores show a steady, nearly identical increase, indicating consistent performance and minimal overfitting. The values approach high accuracy in image segmentation.

The final stage of the U-Net structure assigns each pixel a distinct class label (an example is shown in Figure 7 with labels including “NOT tumor,” “NECROTIC/CORE,” “EDEMA,” and “ENHANCING”). While it is the U-Net decoder that is providing precise localization of the features, ResNet encoder is providing distinct feature extraction providing a segmentation (i.e., class) map that preserves high resolution similar to the training dataset and test ground truth. The segmented brain tumors additionally classified into separate classes: i.e., core, edema, and enhancing

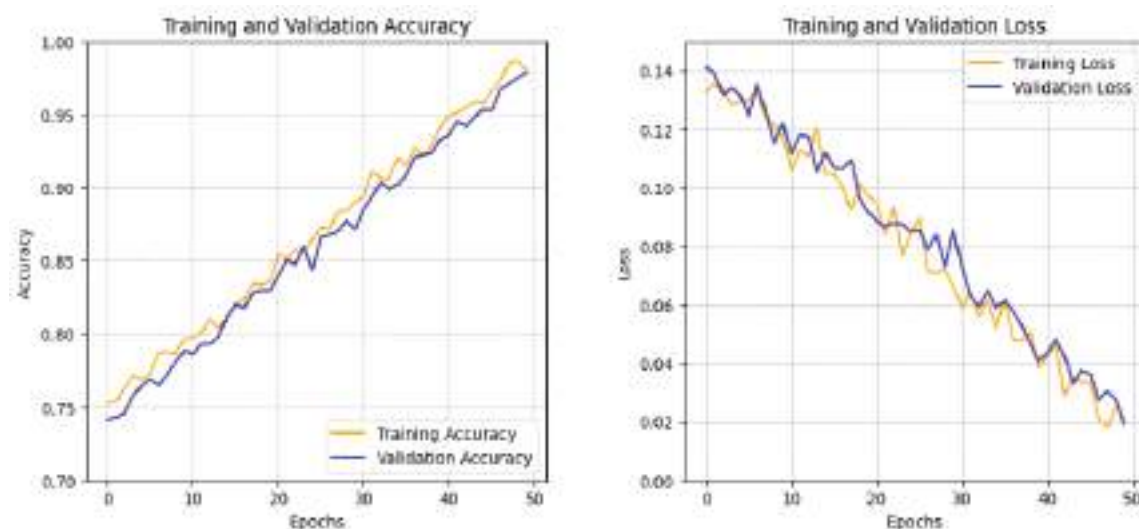


Figure 5. Training and Validation Accuracy and loss plot.

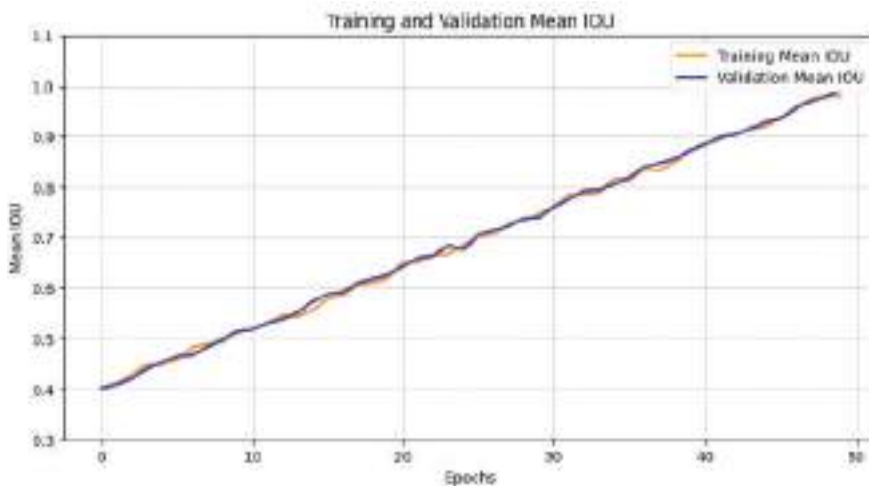


Figure 6. Mean IOU plot.

tumor were represented collectively as the ground truth, and represent a comprehensive view of the tumor classes. This way of representing tumor classes will ultimately assist in data accuracy and reliability concerning deep learning based brain tumor segmentation.

The 99.2% segmentation accuracy in this study shows a significant improvement over common Glioblastoma segmentation methods due to the tumor's complexity. This improvement is a function of the hybrid ResNet-U-Net that leverages ResNet's feature extraction and U-Net's good spatial accuracy, which leads to the segmenting of the glioma accurately. The Hybrid Bilateral-Gaussian Filter is also instrumental in helping reduce noise and at the same time maintain the edges that are important for the model not to receive garbage inputs into the network. Also, CSO helps the model with hyperparameter tuning, which optimized ResNet and U-Net for better performance. Overall, these techniques (ResNet-U-Net, CSO, Hybrid Bilateral-Gaussian Filter) are better than existing approaches which means greater precision in Glioblastoma segmentation and better clinical decision-making.

Survival Prediction

The BraTS2020 challenge requires the use of a dataset that includes features obtained from segmented MRI brain images, patient age, and surviving days for survival analysis [20]. The aim is to predict overall survival (OS) throughout intervals like short-, mid-, and long-term. To predict the Overall Survival time, radiomic and image-based features were extracted. These features have been shown to be successful in different studies for OS tasks [21]. In order to approach the task, we used a process of feature extraction and engineering combined clinical data and the derived imaging features that describe the tumor size and region [22]. After which we trained a number of machine learning models e.g., frequently used methods-support vector

machine (SVM), random forest, k-nearest neighbors, and artificial neural networks in the combined dataset. [23].

The first step in the procedure is gathering data, which includes patient demographics such as age, specific survival durations, and MRI images that are divided into different tumor locations (necrotic/core, edema, enhancing). We combined the volumetric and location features with the patient's age to train a model using linear regression [24]. Quantitative properties of the tumor areas, including their size, shape, and intensity, are extracted from these MRI scans and are thought to have an impact on survival rates. Features can generate prognostic imaging signatures for OS prediction and patient stratification for GBM [25]. Which are very effective with a small number of training data? Random Forest is an ensemble algorithm that is made up of multiple decision trees and operates on them as an ensemble. It is one of the most frequent [26].

Machine learning models like Random Forests, Support Vector Machines, or Neural Networks are then trained on this integrated dataset, which combines clinical data with imaging-derived features. To estimate how long patients with similar profiles will live after diagnosis, the models are trained to correlate the features with survival durations. All models were fit with knowledge that the models were finally tuned and evaluated with methods such as cross-validation to provide a safeguard against overfitting and optimize hyperparameters. All models were measured with accuracy, precision, recall, F1-score, and area AUC-ROC. These statistical methods provided a capacity for predicting survival with a compelling analysis that used advanced analytic techniques to improve health outcomes on a patient level. To visualize, Figure 8 provides a pictorial of survival rates. The x-axis is days survived, and the y-axis is people. This relationship with features may allow some associations to be developed in survival rates that may assist in developing

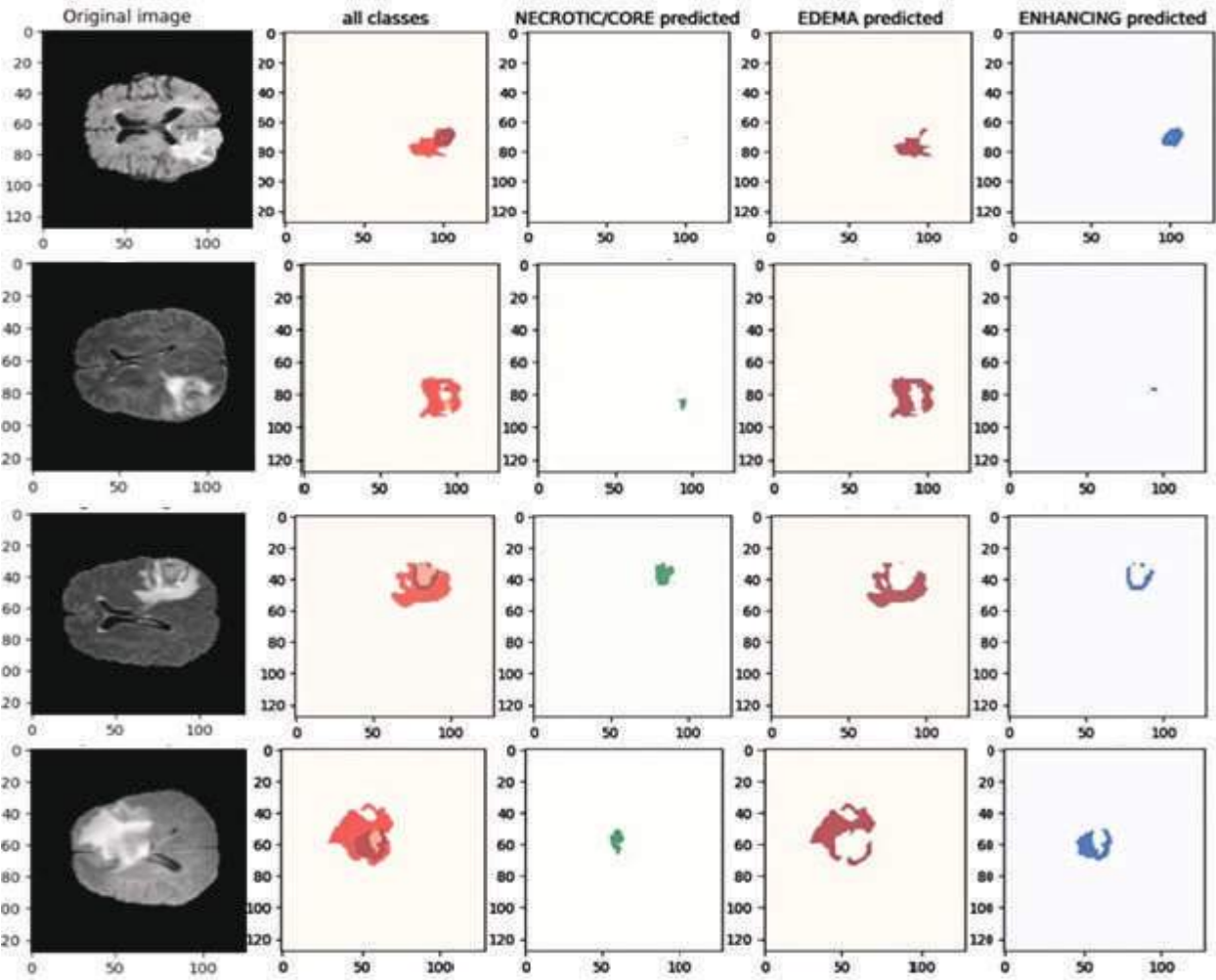


Figure 7. Segmentation results.

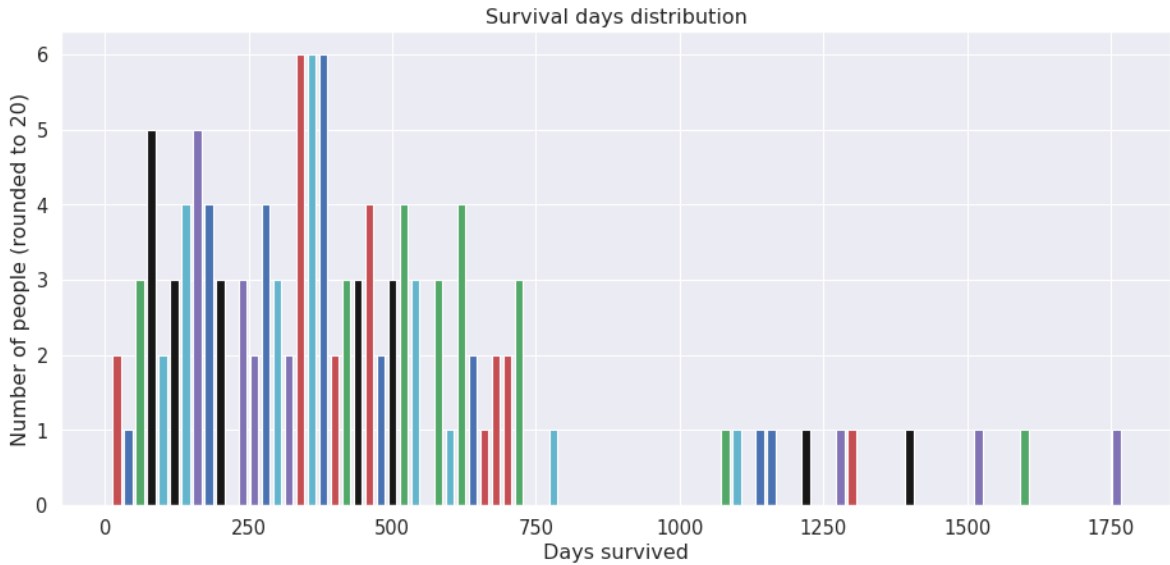


Figure 8. Survival data analysis.

beneficial treatment for variations in patients with various medical conditions.

Figure 9 provides insight into the dependency of survival data and segmentation, determining whether the survival rate is more dependent on the core region, edema, or enhancing tumor. Based on the image, it appears that the survival rate depends largely on the edema region. This information can be useful in guiding the development of targeted treatment strategies for patients with glioblastoma.

As we evaluated predictive survival models, we explored a range of supervised machine learning models, which included random forest classifier, SVM, KNN, and ANN. Random forest produced the highest testing accuracy (0.54) and precision (0.8) but had a relatively low training score (0.37), which can be an indicator of a model that is not learning from the training data as we want. The other models we explored (i.e. the SVM, KNN, and ANN models) had similarly low performance, especially SVM and KNN, and overall had lower accuracy and precision than the random forest model for testing. We used the three best models, Random Forest (RF), Artificial Neural Network (ANN),

and support vector machine (SVM) to then enhance the overall prediction accuracy of survival using an ensemble strategy see Figure 10.

The ensemble model is here to be defined:

- **Random Forest (RF):** Random Forest is an ensemble model that builds multiple decision trees and combines the model outputs, so it can model complex interactions between variables that single models may be unable to decode. Random Forest also provides feature importance scores that allow for the evaluation of which factors contribute most to the survival outcomes, which is important for clinical interpretation and personalized treatment models.
- **Artificial Neural Network (ANN):** ANN builds on layers of nodes so that identifier complex, non-linear patterns are possible in large datasets.
- **Support Vector Machine (SVM):** SVM finds the optimal hyperplane for dividing classes while performing especially well in higher dimensional spaces.

The Ensemble method (or approach) takes advantage of the unique strengths of each model (where each model has

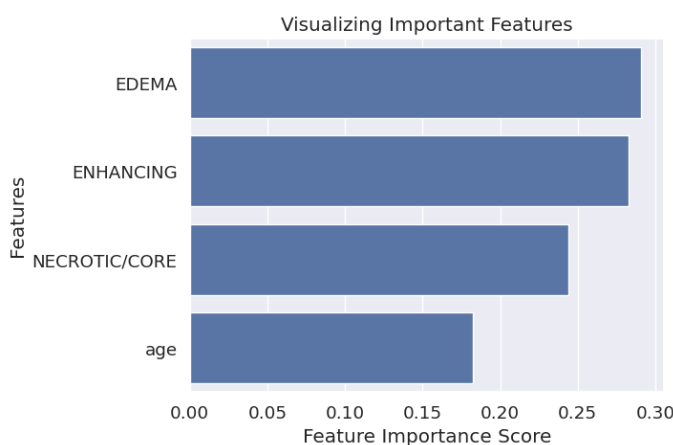


Figure 9. Feature visualization.

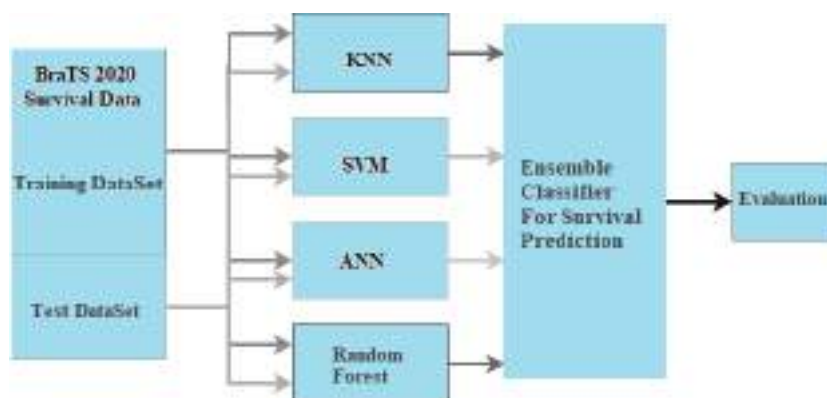


Figure 10. Ensemble model.

unique strengths: Random Forest makes robust, low-variance predictions; ANN can model highly complex non-linear relationships; SVM can handle high-dimensional spaces). The power of the Random Forest algorithm is its ability to capture the complex relationships between the array of different features, which is particularly useful when working with high-dimensional medical data (for example, the imaging-derived features from GBM). However, the Random Forest and other models could not combine and capture all of the patterns in the data, which is what the ensemble method was intended for. Each model captures its own features of the available data and reduces the overall errors in prediction because the weaknesses of each individual model are reduced by other models. However, by ensemble methods, predictions are based on a combined model with the strength of each individual model, providing more stability and robustness by dampening individual model biases and variability. The ensemble method provides improved prediction accuracy and reliability and could be particularly advantageous for the difficult survival

prediction for the cohort of Glioblastoma patients. The overall predicted value for an input given a model can be represented as follows

$$\text{Ensemble Prediction} = (W_{RF} \cdot P_{RF}) + (W_{ANN} \cdot P_{ANN}) + (W_{SVM} \cdot P_{SVM}) \quad (5)$$

Where: W_{RF} , W_{ANN} and W_{SVM} are the weights assigned to the RF, ANN, and SVM models, and P_{RF} , P_{ANN} and P_{SVM} are prediction probabilities of RF, ANN, and SVM models respectively. The confusion matrix for the survival prediction is represented in Figure 11. Table 2 displays several different models for prediction of survival outcomes with metrics, including accuracy, precision, training score and testing score. The Ensemble model, which attained the highest accuracy of 60.01% out of all models presented, is the best model for this type of data. Moreover, the Ensemble model achieved a precision score of 0.85 meaning that it is better at correctly predicting positive survival outcomes when present. Its model training score of 0.63 and model

Table 2. Evaluation metrics

Model	Accuracy	Precision	Training score	Testing score
Random forest	0.54	0.8	0.37	0.54
SVM	0.41	0.6	0.45	0.41
KNN	0.37	0.6	0.43	0.37
ANN	0.41	0.7	0.32	0.42
Ensemble	0.60	0.85	0.63	0.60

SVM: Support vector machine; KNN: K-nearest neighbors; ANN: Artificial neural networks.

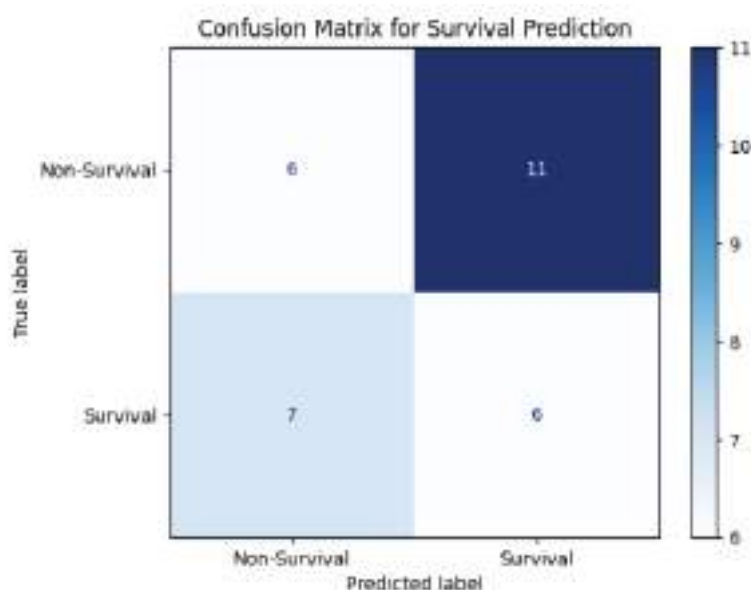


Figure 11. Confusion matrix of survival prediction.

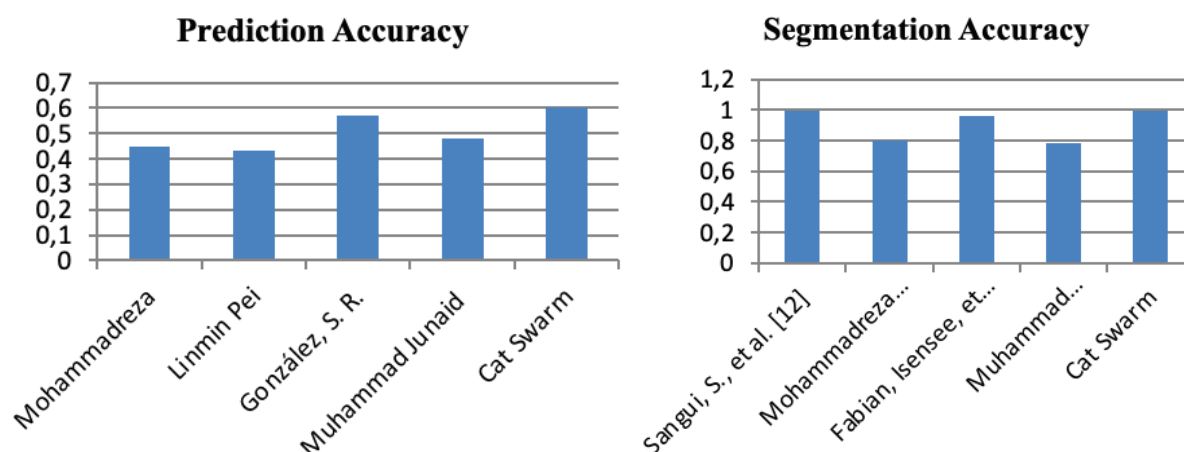


Figure 12. Comparision of results.

testing score of 0.60 indicate that the model generalizes sufficiently to new or unseen data making it an ideal model for survival prediction tasks.

The Figure 12 compares the accuracies of segmentation and survival predictions made by different models. Each bar represents a different model. The survival prediction accuracy figure indicates that our models have different levels of correspondence with survival outcomes, with attainment of the highest accuracy from predicted survival, while other models performed on a moderate level indicating that some potential exists for improving survival predictions. In the segmentation prediction accuracy figure, our models were shown to be superior, attaining the highest accuracy, while the other models had lower accuracy. Thus, the other models may be less able to perform well with precise segmentation tasks.

This study differs from typical clinical management methods of survival prediction in Glioblastoma patients using BraTS2020 dataset in that, all we did was create an ensemble of advanced machine learning models (i.e., Random Forest, ANN, and SVM) to analyze all of the complex imaging data in a more sophisticated manner than just clinical features and subjective manual interpretation. Typical clinical management of glioblastoma patients often relies on clinical features from prior research and clinical decision support systems to inform their suggested management. This approach may be subjective and the reliability is somewhat less refined. The study's approach used an advanced ensemble of models which captured complex data patterns found within the dataset and produced a more objective and precise prediction system. The added benefits of this approach include the ensemble produced a much better accuracy (60.01%), it generalized better to new data, and has the potential to enhance clinical decision making by having a more reliable prediction system for survival during clinical management of Glioblastoma patients.

CONCLUSION

Glioblastoma is the most common and aggressive form of primary brain tumor in adults that poses significant clinical management difficulties including high rates of recurrence, poor prognosis, and complexity and heterogeneity. This study aims to improve predictions of overall survival for Glioblastoma patients treated with stereotactic radiosurgery using advanced image segmentation and machine learning. We used Cat Swarm Optimization assisted hybrid ResNet and U-Net architectures to perform precise tumor segmentation and additional machine learning algorithms as an ensemble to make strong predictions of survival time. Utilizing the BraTS2020 dataset, we accurately delineated three critical tumor regions: core, edema, and enhancing tumor. The segmentation model achieved remarkable metrics, including a segmentation accuracy of 99.2%, a loss of 0.023, a recall of 0.986, a mean intersection over union (IOU) of 0.991, a dice coefficient of 0.96, a precision of 0.991, a sensitivity of 0.991, and a specificity of 0.997. In this study, we evaluated a number of machine learning models for survival prediction, and prioritized using different machine learning models for survival prediction while the Random Forest algorithm was ranked the best tool for survival prediction post deriving the features from segmented images due to its ability to handle the complexity of the features. Although the ensembling approach for survival prediction at 60.01% improved accuracy, these results demonstrated that there is a lot of potential to improve survival prediction for Glioblastoma patients in terms of clinical management and prognosis by utilizing advanced image segmentation methods with machine learning models. In future work using the BraTS dataset, advanced machine learning models, particularly Transformers to process spatial data should be employed. In addition, we can extend the use of multi-modal data, which may include genomic profiles and comprehensive medical history, to improve prediction accuracy and provide personalized treatment strategies. Improving

methods for data handling that can ensure privacy could result in a more broaden data demographic for a training dataset dataset. Performance and randomized controll trails in real-time clinical decision support systems could foster patient outcomes helpful toward personal medicine management.

ACKNOWLEDGMENTS

We extend our sincere thanks to all researchers and contributors in the field of glioma research. Their pioneering work has greatly inspired and guided this 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.

REFERENCES

- [1] Rajput S, Kapdi RA, Raval MS, Roy M, Bhalodiya JM. Deriving and interpreting robust features for survival prediction of brain tumor patients. *Int J Imaging Syst Technol* 2024;34:e23105. [\[CrossRef\]](#)
- [2] Hong S, Wu J, Zhu L, Chen W. Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP. *PLoS One* 2024;19:e0298102. [\[CrossRef\]](#)
- [3] Phan TC, Phan AC, Nguyen KT, Tran HD. Detection and segmentation of brain tumors on 3D MR images using 3D U-Net. In: *Future Data and Security Engineering*. Singapore: Springer Nature; 2023. p. 528–541. [\[CrossRef\]](#)
- [4] Gad E, Soliman S, Darweesh MS. Advancing brain tumor segmentation via attention-based 3D U-Net architecture and digital image processing. In: *Model and Data Engineering*. Cham: Springer Nature; 2023. p. 245–258. [\[CrossRef\]](#)
- [5] Purnama IKE, Rachmadi RF. Modified 3D U-Net for brain tumor segmentation. In: *14th Int Conf Information & Communication Technology and System (ICTS)*. Piscataway (NJ): IEEE; 2023. p. 88–93. [\[CrossRef\]](#)
- [6] Zhao J, Xing Z, Chen Z, Wan L, Han T, Fu H, Zhu L. Uncertainty-aware multi-dimensional mutual learning for brain and brain tumor segmentation. *IEEE J Biomed Health Inform* 2023;27:4362–4372. [\[CrossRef\]](#)
- [7] Ramasamy G, Singh T, Yuan X. Multi-modal semantic segmentation model using encoder-based Link-Net architecture for BraTS 2020 challenge. *Procedia Comput Sci* 2023;218:732–740. [\[CrossRef\]](#)
- [8] 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;41:119–122. [\[CrossRef\]](#)
- [9] Durgut A, Biroğul S, Güvenç U. Comparison of the effects of image segmentation on image processing performance with parallel programming. *Sigma J Eng Nat Sci* 2016;8:145–154.
- [10] Raza R, Bajwa UI, Mehmood Y, Anwar MW, Jamal MH. dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI. *Biomed Signal Process Control* 2023;79:103861. [\[CrossRef\]](#)
- [11] Vijay S, Guhan T, Srinivasan K, Vincent PDR, Chang CY. MRI brain tumor segmentation using residual spatial pyramid pooling-powered 3D U-Net. *Front Public Health* 2023;11:1091850. [\[CrossRef\]](#)
- [12] Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. *Med Phys* 2015;42:4328–4338. [\[CrossRef\]](#)
- [13] Montaha S, Azam S, Rafid AKMRH, Hasan MZ, Karim A. Brain tumor segmentation from 3D MRI scans using U-Net. *SN Comput Sci* 2023;4:386. [\[CrossRef\]](#)
- [14] Sangui S, Iqbal T, Chandra PC, Ghosh SK, Ghosh A. 3D MRI segmentation using U-Net architecture for the detection of brain tumor. *Procedia Comput Sci* 2023;218:542–553. [\[CrossRef\]](#)
- [15] Soltaninejad M, Pridmore T, Pound M. Efficient MRI brain tumor segmentation using multi-resolution encoder-decoder networks. In: *Med Image Comput Comput Assist Interv*. Cham: Springer; 2020. p.31–39. [\[CrossRef\]](#)
- [16] Isensee F, Jäger PE, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for brain tumor segmentation. *arxiv.org/abs/2011.00848* doi: 10.48550/arXiv.2011.00848

- [17] Ali MJ, Akram MT, Saleem H, Raza B, Shahid AR. Glioma segmentation using ensemble of 2D/3D U-Nets and survival prediction using multiple features fusion. In: *Med Image Comput Comput Assist Interv*. Cham: Springer; 2020. p.179–188. [\[CrossRef\]](#)
- [18] Das S, Bose S, Nayak GK, Satapathy SC, Saxena S. Brain tumor segmentation and overall survival prediction in glioblastoma multiforme using radiomic features. *Concurr Comput Pract Exp* 2020;32:e6501. [\[CrossRef\]](#)
- [19] Wang S, Dai C, Mo Y, Angelini E, Guo Y, Bai W. Automatic brain tumour segmentation and biophysics-guided survival prediction. In: Crimi A, Bakas S, editors. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. *Lect Notes Comput Sci* 11993. Cham: Springer; 2020. p. 29–41. [\[CrossRef\]](#)
- [20] Das S, Swain MK, Nayak GK, Saxena S, Satapathy SC. Effect of learning parameters on the performance of U-Net model in segmentation of brain tumor. *Multimed Tools Appl* 2022;81:34717–34735. [\[CrossRef\]](#)
- [21] Hajianfar G, Haddadi Avval A, Hosseini SA, Nazari M, Oveisi M, Shiri I, et al. Time-to-event overall survival prediction in glioblastoma multiforme patients using magnetic resonance imaging radiomics. *Radiol Med* 2023;128:1521–1534. [\[CrossRef\]](#)
- [22] Bhagyalaxmi K, Dwarakanath B, Reddy PVP. Deep learning for multi-grade brain tumor detection and classification: a prospective survey. *Multimed Tools Appl* 2024;83:65889–65911. [\[CrossRef\]](#)
- [23] Saxena S, Agrawal A, Dash P, Jena B, Khanna NN, Paul S, et al. Prediction of O-6-methylguanine-DNA methyltransferase and overall survival of glioblastoma patients using MRI-based hybrid radiomics signatures in machine and deep learning framework. *Neural Comput Appl* 2023;35:13647–13663. [\[CrossRef\]](#)
- [24] Babaei R, Bonakdar A, Shakourifar N, Soltani M, Raahemifar K. Evaluating maximum diameters of tumor sub-regions for survival prediction in glioblastoma patients via machine learning considering resection status. *J Electron Inf Syst* 2024;6:22–38. [\[CrossRef\]](#)
- [25] Peng B, Wang K, Xu R, Guo C, Lu T, Li Y, et al. Preoperative computed tomography-based tumoral radiomic features prediction for overall survival in resectable non-small cell lung cancer. *Front Oncol* 2023;13:1131816. [\[CrossRef\]](#)
- [26] Bathla G, Soni N, Ward C, Maheshwarappa RP, Agarwal A, Priya S. Clinical and magnetic resonance imaging radiomics-based survival prediction in glioblastoma using multiparametric magnetic resonance imaging. *J Comput Assist Tomogr* 2023;47:919–923. [\[CrossRef\]](#)



Research Article

Examining dimensionality reduction effect of principal component analysis via hierarchical clustering techniques

Yağmur ŞAN^{1,*}, Tolunay GÖÇKEN²

¹Department of Industrial Engineering, Faculty of Engineering, Erciyes University, Kayseri, 38039, Türkiye

²Department of Industrial Engineering, Faculty of Engineering, Adana Alparslan Türkeş Science and Technology University, Adana, 01250, Türkiye

ARTICLE INFO

Article history

Received: 27 June 2024

Revised: 09 August 2024

Accepted: 03 October 2024

Keywords:

Baker's Gamma Correlation Coefficient; Cophenetic Correlation Coefficient
Dimension Reduction; F_M Index; Hierarchical Cluster Analysis, Principal Component Analysis

ABSTRACT

This study purposes examine the effect of the Principal Component Analysis method on Hierarchical Clustering techniques in terms of dimension reduction in high-dimensional data sets. The study was carried out using Principal Component Analysis and Hierarchical Clustering methods on the data sets with 22, 38, and 46 variables, created with 2020 data from the United Nations data platform. The variables of the dataset1 consist of the general information (GI) variables and the economic indicator (EI) variables of the countries and the objects of the data set consist of Africa countries. The variables of the dataset2 consist of the general information (GI) variables, the economic indicator (EI) variables and the social indicator (SI) variables of the countries and the objects of the data set consist of Europe countries. The variables of the dataset3 consist of the general information (GI) variables, the economic indicator (EI) variables, the social indicator (SI) variables and the environmental and infrastructural indicator (EII) variables of the countries and the objects of the data set consist of Asia countries. For dataset1, the mean absolute correlation value is 0.2426, and dimension reduction with PCA is decreased 22 variables to 8 variables. For dataset2, the mean absolute correlation value is 0.2346, and dimension reduction with PCA is decreased 38 variables to 10 variables. For dataset3 the mean absolute correlation value is 0.2265 and dimension reduction with PCA is decreased 46 variables to 11 variables. The results obtained from the analysis were compared and interpreted using tanglegrams and some similarity coefficients. The results of the study showed that the Principal Component Analysis method had positive effects on hierarchical clustering results and dendrograms despite low correlation and outliers. In this study, despite the outlier and noise problems of high-dimensional datasets, the facilitating role of PCA in clustering analysis is investigated.

Cite this article as: Şan Y, Göçken T. Examining dimensionality reduction effect of Principal component analysis via hierarchical clustering techniques. Sigma J Eng Nat Sci 2025;43(5):1607–1627.

*Corresponding author.

*E-mail address: yagmursan01@gmail.com

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



INTRODUCTION

In the digitalized world, due to scientific and technological developments, every transaction we do in our daily life and even every step we take appears as data. According to Statista's research and forecasts, by 2025, the volume of data/information created, captured, copied, and consumed will be more than 180 zettabytes [1]. This vast amount and variety of data consist of bits of information that are meaningless on their own. The scientific discipline that includes the processes and methods of transforming data into information is data mining. As a general definition, "Data mining is the process of discovering interesting patterns and knowledge from large amounts of data." [2]. The main purpose of cluster analysis is to discover the natural groups in the data set by collecting the closest observations in the same cluster. However, sometimes these natural groups or clusters can be difficult to observe due to the noise of outliers and unrelated features in the dataset. This is especially true for high-dimensional datasets. Dimension reduction techniques are used to reduce the number of unnecessary features in the data set and to perform operations with fewer variables more easily in matters such as noisy data and outliers. Principal Component Analysis (PCA) is the most widely used dimension reduction method, especially in cluster analysis. PCA reveals the basic pattern and characteristic information in the dataset by compressing the data according to the variance values. PCA uses three different approaches to determine the number of components and it is possible to obtain different results with these approaches.

PCA and clustering have been used in some agricultural studies to assess species variation or product quality [3-7]. Some studies used clustering and PCA methods to classify and evaluate data from different sectors [8-11,12]. Carried out customer segmentation using PCA, hierarchical clustering, and k-means clustering in his study. On the other hand, [13] compared the dimension reduction effect of PCA and non-negative matrix factorization methods on clustering. In another study that proposes a new method for graph-based dimensionality reduction, a hybrid method that is a combination of NPE and PCA linear dimensionality reduction methods is presented. The presented hybrid method produces a transformation matrix for the generalized eigenvalue problem. According to the analysis results of the study, the hybrid method showed the best performance among PCA, NPE and the presented hybrid method (HDR) [14]. In another study conducted by the same authors, according to the results obtained through empirical analysis using graphic data sets, in linear methods, the principal component analysis, singular value decomposition, and neighborhood preserving embedding methods have been showed better performance than other methods of the statistical information category, dictionary methods, and embedding methods, respectively [15]. In a study, the effectiveness of PCA dimension reduction and SVM

classification techniques was examined for anomaly detection over network data. In the study, the positive results of the dimensionality reduction effect of PCA on classification quality and processing time are shown [16]. In another study, PCA and LTP methods and the BAT algorithm were used to reduce the difficulties and processing time in face recognition systems. The dimensionality reduction effects of PCA and LTP were comparatively examined by integrating them with the feature selection function of the Bat algorithm [17]. In another study, the prediction performance of machine learning methods was investigated using multidimensional data obtained in cancer cases. In the research, the effect of PCA and Kernel PCA's dimensionality reduction function on the prediction performance was evaluated comparatively [18]. In the study, which focuses on improving automatic intrusion detection with the aim of minimum fault and correct classification, the system performance was evaluated with the dimensionality reduction effect of random projection and PCA techniques. The effect of the two methods on the results and their accuracy rates were compared [19]. In a study conducted in the field of fluid mechanics, the performances of linear and nonlinear dimension reduction techniques were compared. The methods examined in the study; PCA, Independent Component Analysis, Isometric Mapping and Local Linear Embedding dimensionality reduction techniques. The performance of each method and their suitability according to the characteristics of the flow fields were evaluated separately [20]. PCA is a very effective technique that can be used for many different purposes such as clustering, data reduction and dimensionality reduction. We encounter the use of PCA in different ways in the literature, but the dimensionality reduction ability of PCA has never been examined in detail before. In this study, the dimensionality reduction ability of PCA is investigated and demonstrated in detail with graphics and numerical analysis using high-dimensional data sets. Despite the outlier and noise problems of high-dimensional datasets, the facilitating role of PCA in clustering analysis is searched.

The rest of the study; in section 2, hierarchical clustering methods and their properties are examined in detail. In section 3, PCA and the approaches used to determine the number of components are handled. In section 4, some coefficients used for comparing hierarchical clustering results are given with formulas. In section 5, the datasets and the methodology used in the study are explained in detail. Section 6 consists of the evaluations of the results for each dataset separately. In the conclusion section, the outputs of the study are interpreted in a general framework.

HIERARCHICAL CLUSTERING

Hierarchical clustering methods are the methods that perform clustering operations by following a hierarchical structure. These methods form clusters by grouping the objects in the data set according to their similarities, and as

a result of the operations, a tree-like hierarchical structure form emerges [2]. Hierarchical methods reveal natural clusters in the dataset without the need to specify the number of cluster. This is known as one of the important advantages of hierarchical methods [21]. One of the weaknesses of the hierarchical methods is that they have high computational complexity. The time complexity of hierarchical methods is expressed by $O(n^2)$ and they are not suitable for large datasets [22]. In hierarchical methods, the tree-shaped diagram that presents the clustering results of the data in a hierarchical structure is called a dendrogram [23]. A dendrogram is a clear representation form that is easy to interpret and understand. The vertical axis of the dendrogram shows the distance values between clusters and data points, while the horizontal axis shows the data points [24]. In hierarchical clustering, the cutting point of the dendrogram has a significant role in determining the clusters. A different number of various cluster combinations can be obtained by cutting the dendrogram at different points [22]. Hierarchical clustering methods are divided into two groups in terms of the strategic way they follow. These two groups operate in opposite directions, bottom-up (agglomerative) and top-down (divisive) [25]. The method in which clusters are brought together using the lines and distance matrix based on attribute vectors and distance criteria is called the linkage technique [26]. There are many different linkage methods according to different distance preferences such as minimum, maximum, average, median, centroid and minimum variance [27].

Single Linkage

In the single linkage method, also called the *nearest neighbour*, the minimum distance between two clusters is accepted as the criterion for merging [21,25,28]. In each iteration, the two closest clusters are merged and the operations are repeated until the clustering is complete [29]. This method is faster than other hierarchical techniques but, because of its local behaviour, it is sensitive to noises and outliers [30, 31]. For example, the clusters which are not closest in real can be merged for the closest objects in the clusters, and this fault also affects the following iterations of the process [26,30]. This case is known as the *chaining effect*.

Complete Linkage

In the complete linkage method, unlike the single linkage method, the merging criterion is the furthest distance between two clusters [27]. Due to its non-local behaviour, smaller and tighter clusters are obtained [33]. However, also in this method, the presence of outliers can have negative effects on clustering because of its sensitivity to noise and outliers [26,30,31].

Average Linkage

In the average linkage method, the merging criterion is the average distances of all the pairs of points of the clusters [21-23]. In each iteration, the procedure is repeated

according to this criterion until the clustering process is completed. This method has the advantage of outlier insensitivity and is difficult to use for categorical data, but also a very effective method for numerical data [2].

Centroid Linkage

This method performs the merging operation according to the proximity between the centroids of the clusters [21, 32, 33]. Here centroid denotes the center point of the cluster.

Median Linkage

In the merge of two clusters with different sizes, the median value is used to prevent the centroid of the newly formed cluster from shifting predominantly towards the larger cluster [25, 23]. The median value is the midpoint of the distance between the two centroids. It is designed to overcome the disadvantage of the centroid method.

Ward's Criterion Linkage

The main purpose of Ward's criterion linkage method is to obtain homogeneous clusters by keeping the sum of squares of the error of the distances within the cluster minimum [22,25,34]. Contrary to some methods, cluster size is effective on this method. There are two versions of this method; ward1 and ward2. The main difference between these two versions is; Squared Euclidean distance is used in ward1, while Euclidean distance is used in ward2 [35,36]. Significant formulas used for Hierarchical clustering are shown in Table 1 and Table 2.

Table 1. The time complexity functions of the hierarchical methods for dataset with n objects

<i>Divisive</i>	$O(2^n)$	<i>Average linkage</i>	$O(n^2 \log n)$
<i>Agglomerative</i>	$O(n^2)$	<i>Centroid linkage</i>	$O(n^2 \log n)$
<i>Single linkage</i>	$O(n^2)$	<i>Median linkage</i>	$O(n^2 \log n)$
<i>Complete linkage</i>	$O(n^2 \log n)$	<i>Ward's linkage</i>	$O(n^2)$

Principal Component Analysis

PCA, also known as Karhunen Loeve expansion or the Hotelling transformation, is one of the oldest multivariate methods frequently used in many scientific fields. Although its history back to Pearson (1901), the form used today is defined by Hotelling in 1933 [37]. PCA is the process of creating new variables, called the principal components, to obtain important information by looking at the spread of observation values over variables. These new orthogonal and uncorrelated variables are the linear combinations of the original variables in the data set. PCA is a fast running and computationally easy method. The areas and applications where PCA is used are; pattern recognition, dimension reduction, computer vision, image compression, signal processing, video surveillance, face recognition,

Table 2. The merging criteria formulas of the hierarchical linkage methods

<i>Single linkage</i>	$D(A, B) = \min\{d(y_i, y_j), \text{ for } y_i \text{ in } A \text{ and } y_j \text{ in } B\}$
<i>Complete linkage</i>	$D(A, B) = \max\{d(y_i, y_j), \text{ for } y_i \text{ in } A \text{ and } y_j \text{ in } B\}$
<i>Average linkage</i>	$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j)$
<i>Centroid linkage</i>	$D(A, B) = d(\bar{y}_A, \bar{y}_B)$
<i>Median linkage</i>	$M_{AB} = \frac{1}{2}(\bar{y}_A + \bar{y}_B)$
<i>Ward's linkage</i>	$SSE_A = \sum_{i=1}^{n_A} (y_i - \bar{y}_A)'(y_i - \bar{y}_A), SSE_B = \sum_{i=1}^{n_B} (y_i - \bar{y}_B)'(y_i - \bar{y}_B),$ $SSE_{AB} = \sum_{i=1}^{n_{AB}} (y_i - \bar{y}_{AB})'(y_i - \bar{y}_{AB}); I_{AB} = SSE_{AB} - (SSE_A + SSE_B)$

latent semantic indexing, ranking, and collaborative filtering [38,39]. In PCA, the first principal component (PC1) is the linear combination of the observations which have maximum variance and the second principal component (PC2) is the linear combination of the observations which have maximum variance as orthogonal to the first principal component (PC1). $y_1, y_2, \dots, y_n$ are the observation vectors in p dimensional space, and $\bar{y}$ is the mean vector of the observation vectors. $y_1, y_2, \dots, y_p$ are the swarm of points for the variables. The first operation in PCA is the transformation. In the transformation operation, the origin of each y_i are translated to $\bar{y}$ and in certain cases the transformation of $y_i - \bar{y}$ is applied. But it is generally assumed that each y_i is centralized. The main purpose of this procedure is finding the optimal axes for the points. The second important procedure of the PCA is rotation. In the rotation phase, each of y_i is multiplied with the orthogonal matrix A and the new variables (PCs) $z_1, z_2, \dots, z_p$ are obtained. The orthogonal matrix transforms each point y_i into a point z_i that is the same distance from the origin, and these new variables are uncorrelated.

$$A = \begin{pmatrix} a'_{11} \\ \vdots \\ a'_{1p} \end{pmatrix}; \quad z_i = A * y_i, \quad i = 1, \dots, p$$

$z_1 = a'_{11}y_1, z_2 = a'_{12}y_2, \dots, z_p = a'_{1p}y_p \rightarrow$ Principal Components

$z_1 = a_{11}y_1 + a_{12}y_2 + \dots + a_{1p}y_p \rightarrow$ PC1.

$z_2 = a_{21}y_1 + a_{22}y_2 + \dots + a_{2p}y_p \rightarrow$ PC2.

$\vdots$

$z_p = a_{p1}y_1 + a_{p2}y_2 + \dots + a_{pp}y_p \rightarrow$ PCp.

If we handle the other details of PCA; the sample covariance matrix of z (S_z) is calculated by using the orthogonal matrix (A) and the sample covariance matrix of $y_1, y_2, \dots, y_n$ (S).

$$S_z = ASA' = \begin{pmatrix} S_{z_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & S_{z_p}^2 \end{pmatrix}$$

Here $S_{z_i}^2 = \lambda_i$ and $\lambda_i, i = 1, \dots, p$ are the eigenvalues of S which equals to variance values of the principal components. In PCA, the variance values of the components have an arrangement as $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Therefore, PC1. has the largest variance value and PCp. has the smallest variance value respectively. The fact that the variance value of the first principal component is the largest is since it represents most of the variables proportionally. How to decide the number of the principal components?

1. Starting from the first component, the components whose sum of the explained variance percentages exceed %80 are selected [40].
2. Starting from the first component, the number of components whose λ value is greater than the average $\bar{\lambda} = \sum_{i=1}^p \lambda_i / p$ value is selected.
3. In the graph in which the λ values of the components are included (Scree graph), the number of components can be decided by looking at the point where the natural break (Elbow test) between the large and small values occurs [25, 37].
4. If the correlation between variables in the data set is high, the k value (number of components selected) will be much smaller than the p value. But otherwise, if the correlation between variables is low, the value of k will be close to p . This case will reduce the effectiveness of the principal component analysis in terms of dimension reduction [25].

THE INDEXES FOR COMPARING HIERARCHICAL CLUSTERING

CPCC (Cophenetic Correlation Coefficient)

CPCC is expressed as the correlation value between the cophenetic matrix created according to the height values in

dendrogram and similarity matrix [41-43]. CPCC takes the value in the range of [-1,1]. The high values of the coefficient express the high similarities between the cophenetic matrix and the distance matrix, while the low values close to zero express the low similarities. CPCC aims to measure how well hierarchical clustering is performed [44-46].

FM-Index (Fowlkes and Mallows Index)

FM_index (Fowlkes and Mallows index) measures the similarity between the clustering results [43]. It uses the number of cluster as the parameter in the calculations. Therefore, different results are obtained with different cluster number input. FM_index takes the value between the range of [0,1] [47]. The values close to 1 express the high similarity and the values close to 0 express the low similarity between the clustering results.

Baker's Gamma Coefficient (Goodman and Kruskal's Gamma Coefficient)

The Goodman and Kruskal's Gamma Coefficient was proposed in 1954 by Goodman & Kruskal to measure the relationship depending on the probabilities π_c and π_d [48]. In generally γ coefficient is used for ordinal variables and it takes the values between the range of [-1,1] [48, 49]. Here is the values close to 0 express the low relationship and the values close to -1 and 1 express the negative and positive relationship respectively. The formulations of the indexes are given in Table 3.

MATERIALS AND METHODS

In this study, hierarchical linkage techniques (Single, Complete, Average, Centroid, Median and Ward's Criterion) are used as clustering methods and clustering processes are applied to the high dimensional data sets separately for each method. The most important factor in

choosing hierarchical clustering methods in this study is that natural groups in the dataset spontaneously emerge with the application of the technique. That is, there is no need to determine the number of clusters with this technique. Clustering operations are performed both with PCA and without PCA. The main purpose here is to evaluate the effects of PCA on the hierarchical clustering process and its results. FM_index, CPCC, and Baker's gamma coefficient metrics are used to measure similarities between dendrograms and to evaluate the results. In addition, to observe the change in dendrograms, *tanglegrams* are used for each method. Tanglegrams are plots that show two hierarchical clustering or two dendrograms comparatively. The data sets used in the analysis are compiled from the United Nations data platform [50]. The missing observation values in the data sets are completed from the relevant international data platforms [51, 52, 53, 54]. While the objects of the datasets are created from the relevant countries, the variables in the datasets are created from the relevant indicator variables. The indicators and the variables used for the datasets are given in the appendices in the relevant tables. All the methods and stages used in the analysis are applied to each dataset in the same way. All the phases of the analysis are carried out in R programming, and the Cluster package and Denextend package are used for the clustering analysis procedure.

Datasets

Dataset1 consists of 22 variables (dimension), 54 objects and a total of 1188 observations. The variables of the data set consist of the general information (GI) as Pop. density (per km², 2020), sex ratio(male/female) and surface area variables and the economic indicator (EI) as growth rates, sectoral employment rates and international trade data variables of the countries and the objects of the data set consist of Africa countries. To evaluate the correlation level

Table 3. The indexes for comparing hierarchical clustering methods

CPCC (Cophenetic Correlation Coefficient)	$\frac{\sum_{i<j} (d_{ij} - \bar{d})(d_{ij}^* - \bar{d}^*)}{\sqrt{\sum_{i<j} (d_{ij} - \bar{d})^2 \sum_{i<j} (d_{ij}^* - \bar{d}^*)^2}}$	d_{ij} : distance between the pairs (i,j) d_{ij}^* : cophenetic distance between the pairs (i,j) $\bar{d}$: average distance for similarity matrix $\bar{d}^*$: average distance for cophenetic matrix
FM-INDEX (Fowlkes and Mallows Index)	$\sqrt{\frac{TP}{TP + FP} * \frac{TP}{TP + FN}}$	TP : The count of pairs which are in the same cluster both in C_1 and C_2 . FP : The count of pairs which are in the same cluster in C_1 but not in C_2 . FN : The count of pairs which are in the same cluster in C_2 but not in C_1 .
Baker's Gamma Coefficient (Goodman and Kruskal's Gamma Coefficient)	$\gamma = \frac{\pi_c - \pi_d}{\pi_c + \pi_d}$	π_c : The probability of concordant pairs of observations. π_d : The probability of discordant pairs of observations.

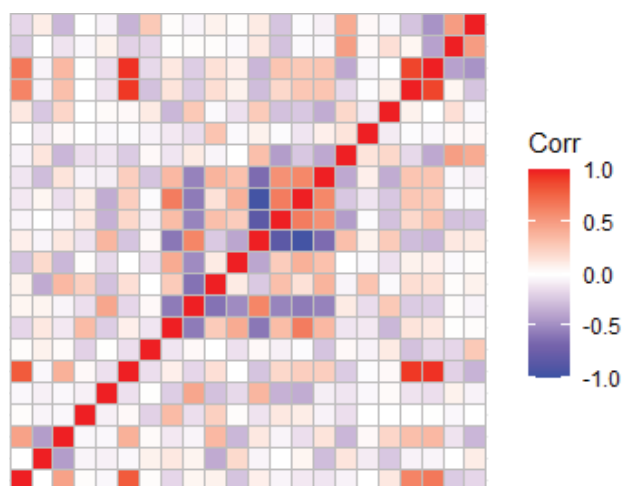


Figure 1. Colored correlation matrix of the variables for dataset1. between the variables in the data set, the colored correlation matrix for the variables of the dataset1 is given below in Figure 1.

The mean absolute correlation between variables for the dataset1 is calculated as 0.2426. Due to the low level of correlation between the variables in the data set, it is seen that light colors are dominant in the colored correlation matrix. Below is a scatterplot showing how African countries are distributed according to the economic indicator variables.

When the distribution of the normalized data in the graph is examined, it can be seen that there are so many outliers in the data set. Some of these outliers are labeled on the scatterplot in Figure 2.

Outliers in the data set, depending on their number, directly affect the correlation between variables negatively. Correlation shows the strength and direction of the relationship between two variables. More deterministic methods other than scatterplot should be used to express the high values of the objects on the variables as outliers. One of the most robust ways to identify outliers in a data set is to plot a Box and Whisker plot for each variable. Potential outlier values are determined based on the quartiles and median values in the boxplot in Figure 3.

Dataset2 consists of 38 variables (dimension), 39 objects and a total of 1482 observations. The variables of the data set consist of the general information (GI) variables, the economic indicator (EI) variables and the social indicator (SI) as population growth rate, international migration stock, education and health data variables of the countries and the objects of the data set consist of Europe countries.

The mean absolute correlation between variables for the dataset2 is calculated as 0.2346. Due to the low level of correlation between the variables in the data set, it is seen that light colors are dominant in the colored correlation matrix in Figure 4.

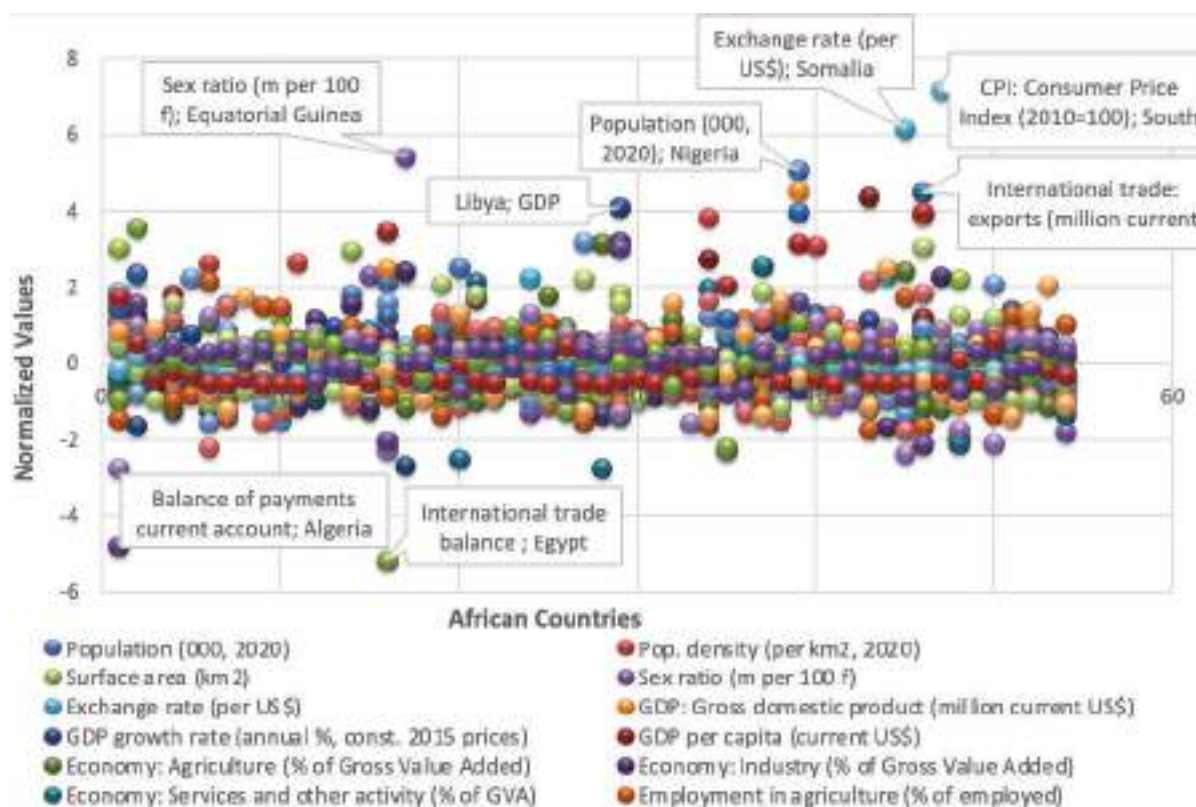


Figure 2. Scatterplot of the normalized dataset1 for the economic indicator variables.

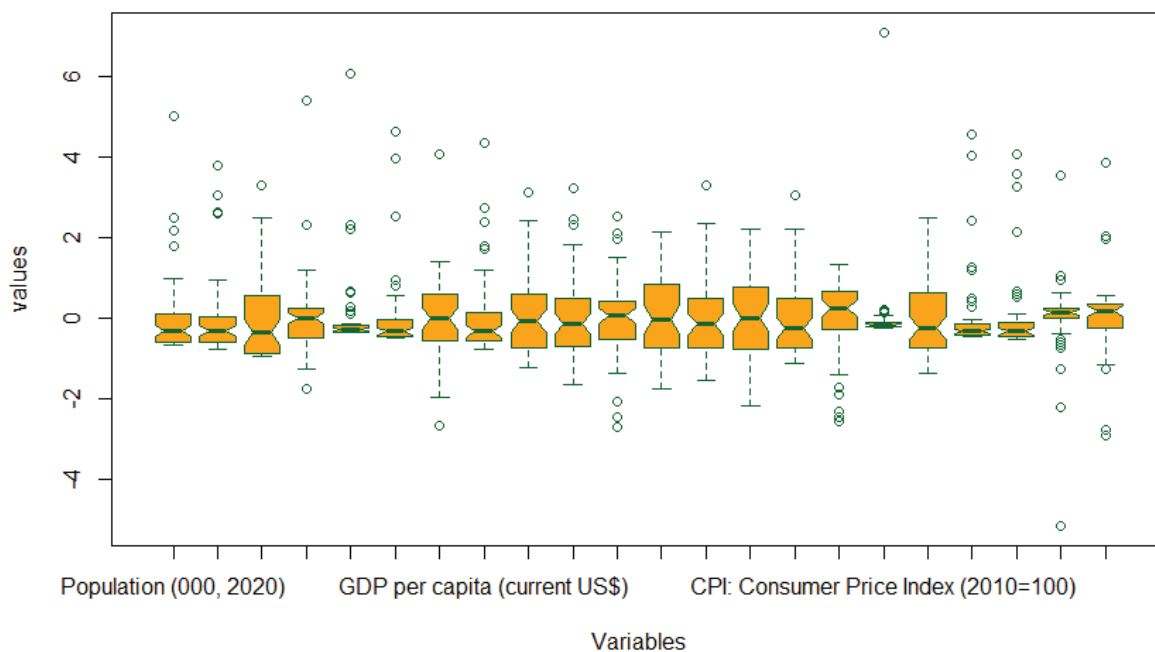


Figure 3. Boxplots for dataset1.

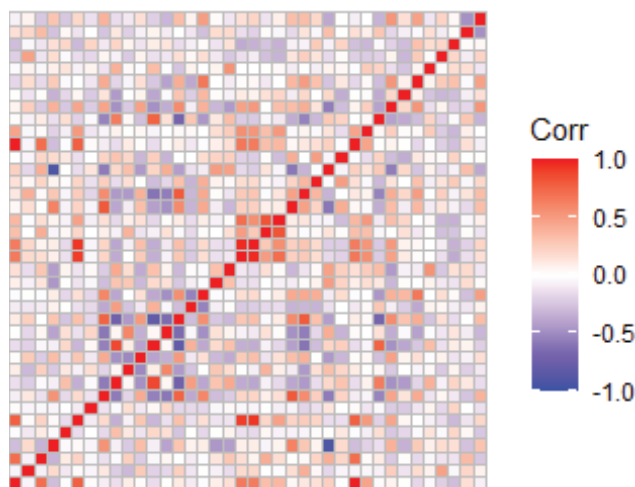


Figure 4. Colored correlation matrix of the variables for dataset2.

Figure 5 is a scatterplot showing how Europe countries are distributed according to the economic indicator and the social indicator variables. It seems to be so many outliers here as in the first dataset. The existence of many outliers can be seen clearly in the boxplots for dataset2 in Figure 6. Several situations cause outliers to exist; measurement errors, data entry errors, or the actual value of the data. Since the observation values in the data set are real values, it is not possible and accurate to clear outliers. However, it is a fact that this situation will negatively affect the analysis process and results.

Dataset3 consists of 46 variables (dimension), 48 objects and a total of 2208 observations. The variables of the data set consist of the general information (GI) variables, the economic indicator (EI) variables, the social indicator (SI) variables and the environmental and infrastructural indicator (EII) as individual internet use, CO2 emission estimates and energy production amounts variables of the countries and the objects of the data set consist of Asia countries. Colored correlation matrix for Dataset3 is shown in Figure 7 and the scatterplot is shown in Figure 8.

Figure 9 is a scatterplot showing how Asia countries are distributed according to economic indicator, social indicator and, environment and infrastructure indicator variables. When the distribution of the normalized data in the graph is examined, it can be seen that there are so many outliers in the data set as the first and second datasets. Object China, for many variables, has extreme values compared to other countries. The boxplots plotted with normalized values show that there are many outlier values in almost all variables in a way that supports the scatterplot.

RESULTS AND DISCUSSION

In this section, before clustering analysis results and dendrogram comparisons, principal component analysis results are evaluated in terms of dimension reduction. Afterward, clustering analysis results are comparatively evaluated for each method.

Results for Dataset 1

In principal component analysis, three approaches can be used to decide the number of components. For the elbow

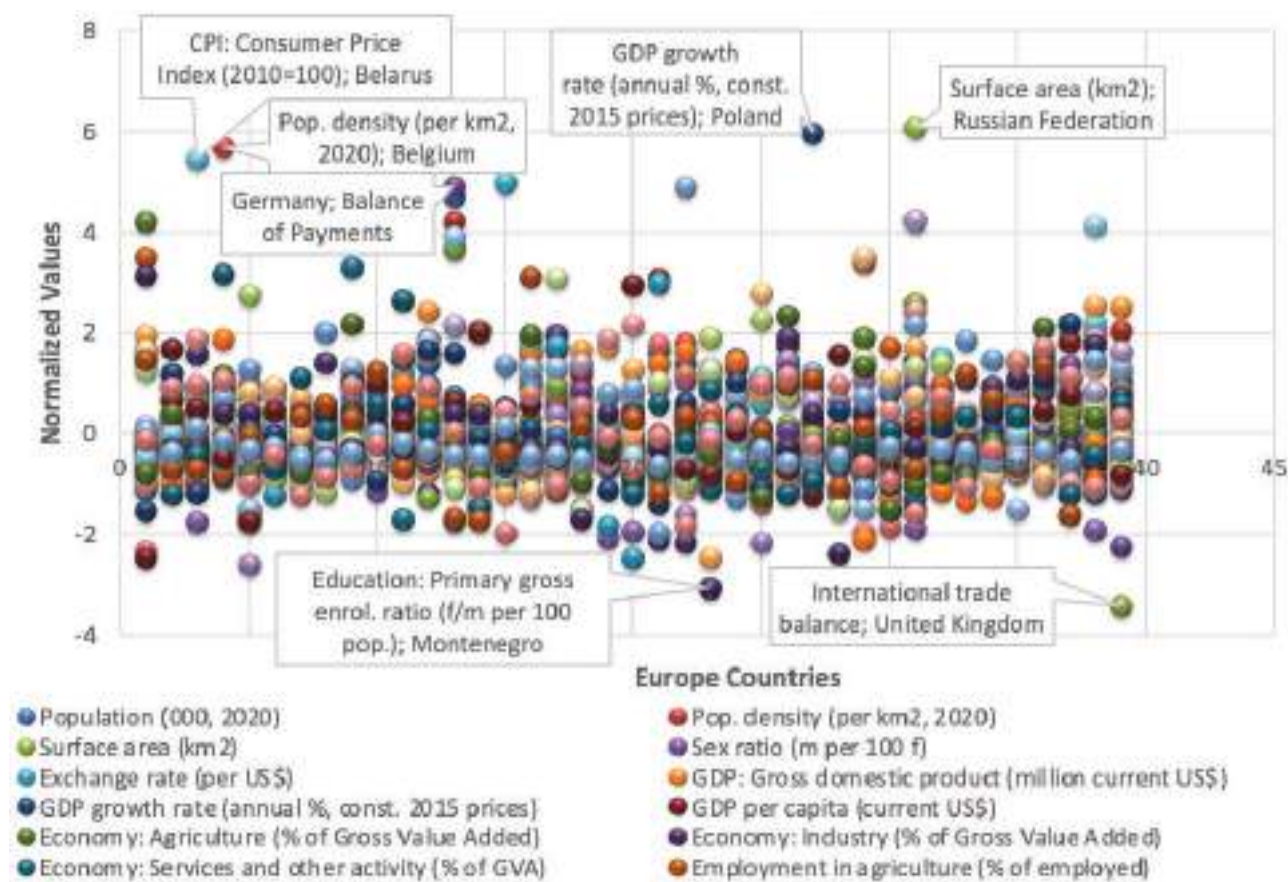


Figure 5. Scatterplot of the dataset2 for the economic indicator and social indicator variables.

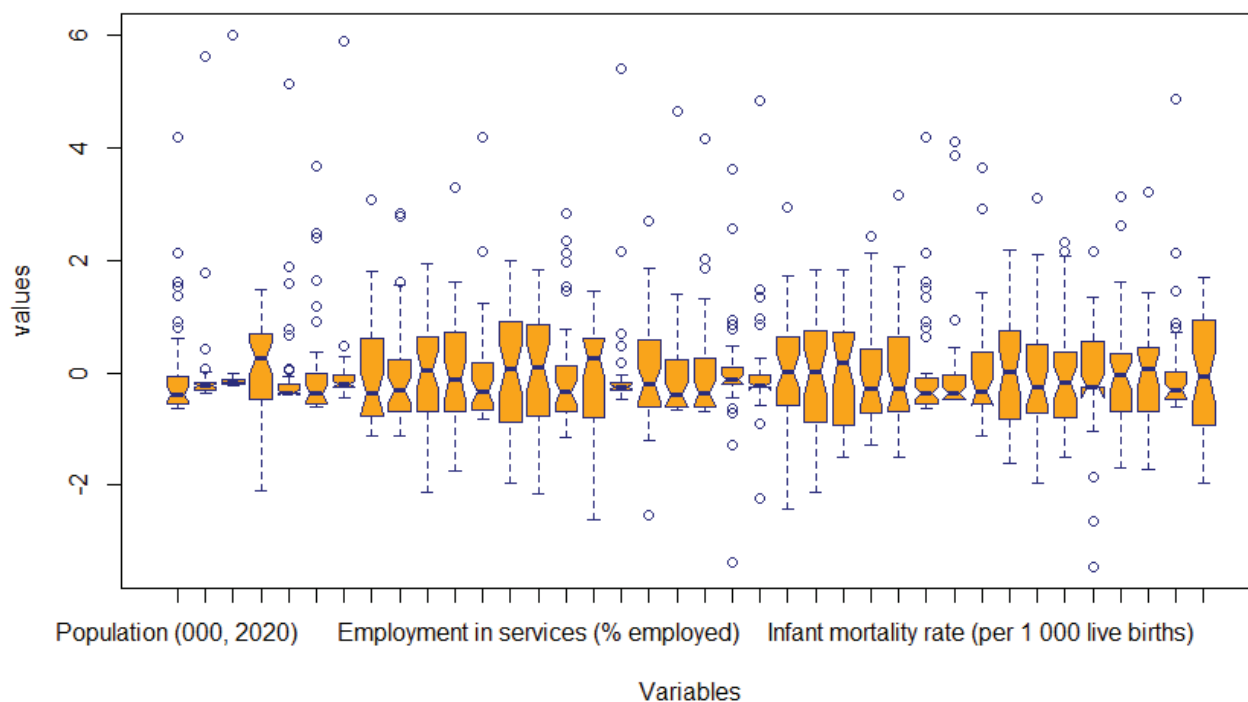


Figure 6. Boxplots for dataset2.

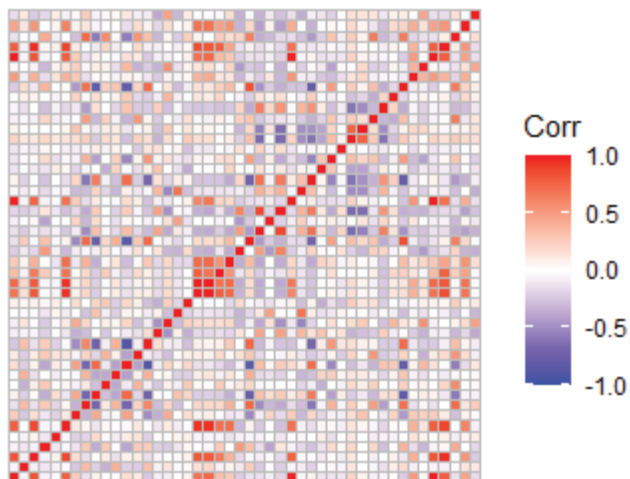


Figure 7. Colored correlation matrix of the variables for dataset3.

test, which is one of these approaches, when the scree plot which includes the variance values of the principal components given in Figure 10 is examined, it is seen that the natural breakpoint in the graph coincides with the 3rd principal component.

According to Table 4 and Table 5, it is seen that the first 3 components can explain only %51.11 of the total variance in the dataset. This ratio is not sufficient to represent the entire data set.

Another approach that can be used to decide on the number of components is to select components with

variance above the mean-variance value. For dataset1, the mean-variance value ($\bar{\lambda}$) of the principal components is calculated as 0.9815 and the variance values of the first 7 principal components are above this value. The first 7 components can explain %76.35 of the total variance in the data set, as seen in table 4.

The third approach is to select components whose sum of variance explained percentages exceeds the 80% specific ratio. The first 8 principal components should be selected to achieve this specific ratio because the first 8 components can explain %80.81 of the total variance in the data set. As a result, the number of components for a robust analysis process is determined as 8.

The CPCC (dend1) values show the fit and the similarity ratio of the cophenetic matrix and the distance matrix of the dendrogram obtained without the PCA method. Also, the CPCC (dend2) values show the fit and the similarity ratio of the cophenetic matrix and the distance matrix of the dendrogram obtained after applying the PCA method. According to the CPCC values in the Table 6, it is obvious that the results of PCA applied clustering analysis have higher similarity with the distance matrix in almost all methods. For CPCC(dend1) and CPCC(dend2), while the best clustering results belong to PCA+average and PCA+centroid methods, the worst clustering belongs to the variance-based Ward's method due to the outliers and low correlation in the dataset. Unlike the CPCC (dend1 and dend2), the CPCC (tanglegram) shows the fit and the similarity ratio between the cophenetic matrices of these two dendrograms. Consistent with the tanglegrams, the highest similarity rate belongs to average linkage and centroid

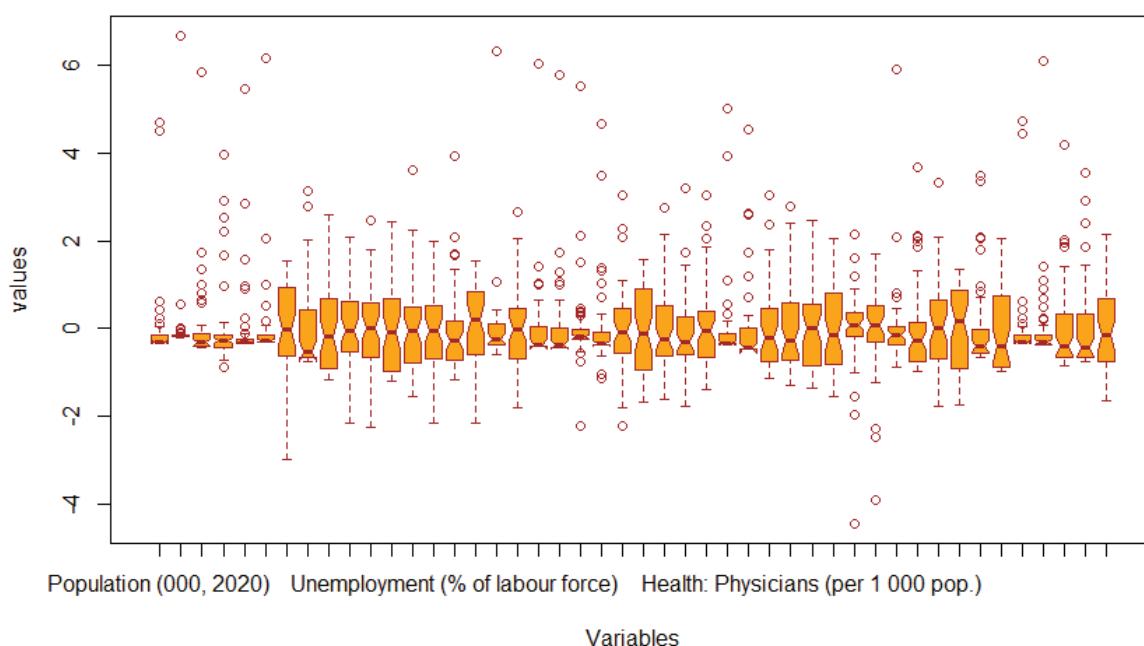


Figure 8. Boxplots for dataset3.

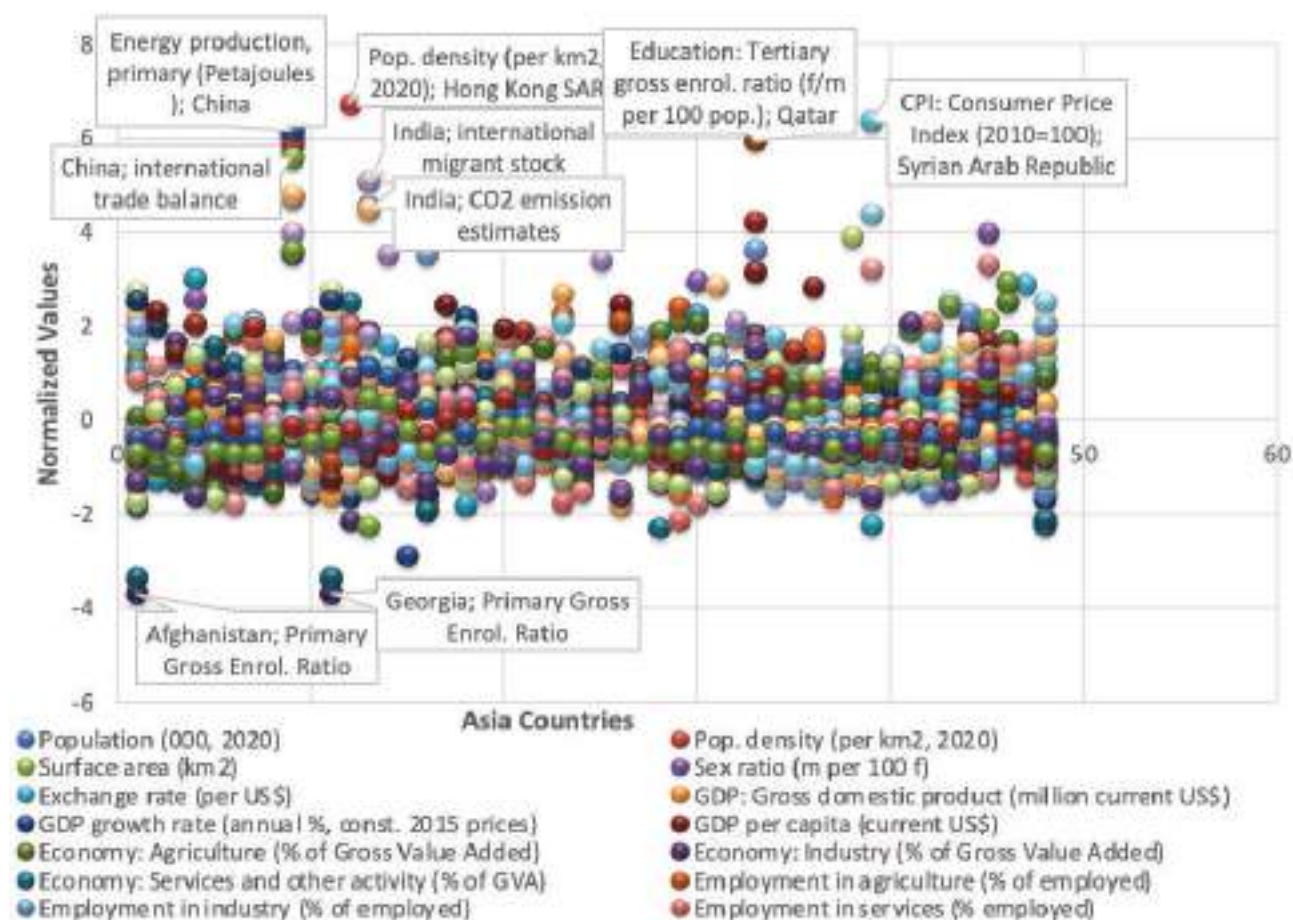


Figure 9. Scatterplot of the dataset3 for the economic indicator, social indicator and environmental and infrastructural variables.

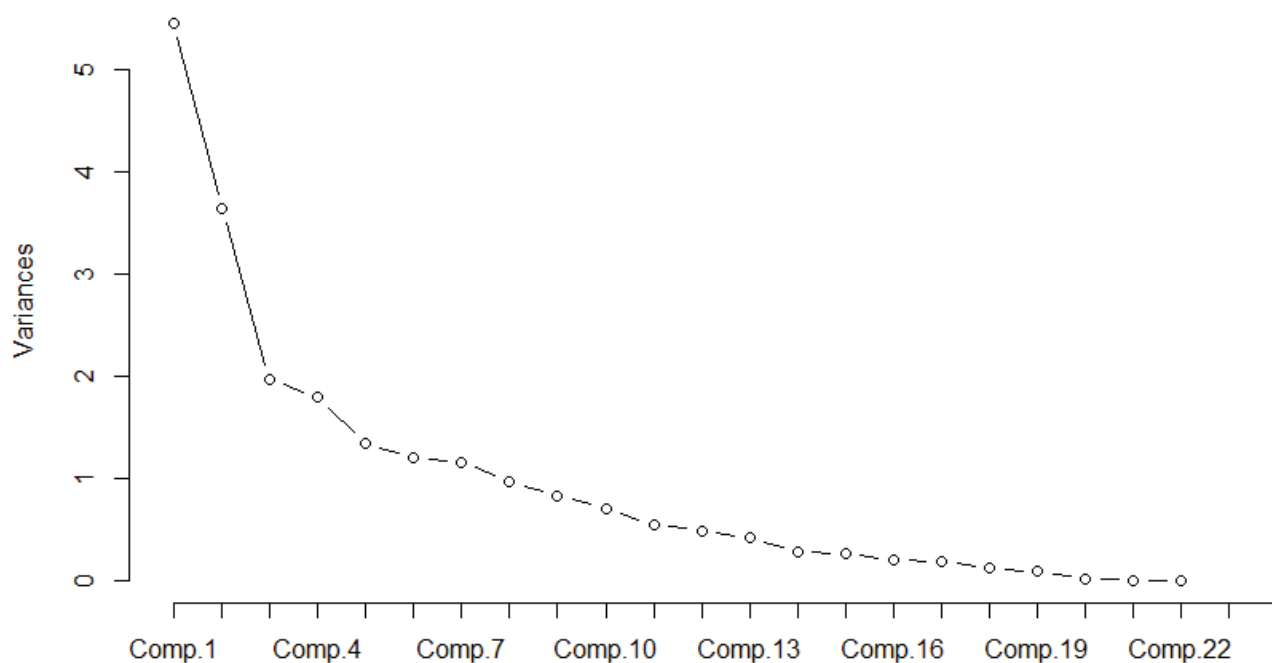


Figure 10. Scree plot of the variance values of the principal components for dataset1.

Table 4. The variance values of the principal components for dataset1

comp.1	comp.2	comp.3	comp.4	comp.5	comp.6	comp.7	comp.8	comp.9
5,44E+0	3,64E+0	1,97E+0	1,79E+00	1,34E+00	1,20E+00	1,15E+00	9,63E-01	8,28E-01

Table 5. The variance explained percentages of the principal components for dataset1

comp.1	comp.2	comp.3	comp.4	comp.5	comp.6	comp.7	comp.8	comp.9
2,52E+01	1,68E+01	9,11E+0	8,28E+0	6,22E+0	5,55E+0	5,34E+00	4,46E+00	3,83E+00

Table 6. Coefficient and index results used to measure the similarity of dendrograms for dataset1

	Methods	CPCC (dend 1)	CPCC (dend 2)	CPCC (tanglegram)	BGCC	FM_index
Data set 1.	single	0.8519	0.8704	0.7465474	0.9856606	0.9425261
	complete	0.7724	0.7261	0.9655519	0.9867451	1
	average	0.9089	0.9211	0.9710973	0.9809861	0.9201198
	centroid	0.8977	0.9102	0.994414	0.3989163	1
	median	0.7948	0.8365	0.9269909	0.428965	1
	ward.D2	0.5129	0.4961	0.921376	0.95822	0.95849

linkage methods. BGCC (Baker's gamma correlation coefficient) measures the similarity of dendrograms using the odds ratios of concordant and discordant pairs of objects. However, when evaluated together with the tanglegrams in Figure 11, it is seen that this coefficient gives inconsistent results for the centroid and median methods. The FM_index determines the similarity ratio of the dendrograms by comparing the cluster contents. Unlike the other coefficients, FM_index uses the number of clusters as a parameter in the calculations. In this study, the number of clusters parameter is determined as 3 for all data sets. If the cluster contents of the two compared dendrograms are the same, the FM_index value will be equal to 1. By evaluating the graphics and all the calculated coefficients together, it is observed that the best clustering result belonged to the PCA+Centroid method, and the cluster contents for the 3 clusters are given below.

Cluster 1: Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cabo Verde, Cameroon, Central African Republic, Chad, Comoros, Congo, Cote d'Ivoire, Dem. Rep. Of the Congo, Djibouti, Egypt, Equatorial Guinea, Eritrea, Eswatini, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Libya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Sudan, Sudan, Togo, Tunisia, Uganda, United Rep.

Of Tanzania, Zambia, Zimbabwe Cluster 2: Nigeria Cluster 3: South Africa

Results for Dataset 2

The variance values of the principal components for dataset2 are given in Figure 12 and Table 7. For the elbow test, the natural breakpoint in the graph coincides with the 3rd principal component. But the first 3 components could explain only %46.44 of the total variance in the dataset according to Table 8.

For the second approach, the mean-variance value ($\bar{\lambda}$) of the principal components is calculated as 0.9743 for dataset2. It is observed that the variance values of the first 10 principal components are above this value, and can explain %80.50 of the total variance in the data set.

For the 80% specific ratio, the first 10 principal components should be selected to achieve this specific ratio, and can explain %80.50 of the total variance in the data set.

According to the CPCC (dend1 and dend2) values in Table 9, it is obvious that the results of PCA applied clustering analysis show higher similarity with the distance matrix in almost all methods. For CPCC(dend1) and CPCC(dend2), while the best clustering results belong to PCA+average and PCA+centroid methods, the worst clustering belongs to the variance-based Ward's method due to the outliers and low correlation in the dataset. Consistent with the tanglegrams, the highest similarity rates for the CPCC (tanglegram) belong to the single linkage, centroid linkage and ward's criterion linkage methods. When evaluated together with

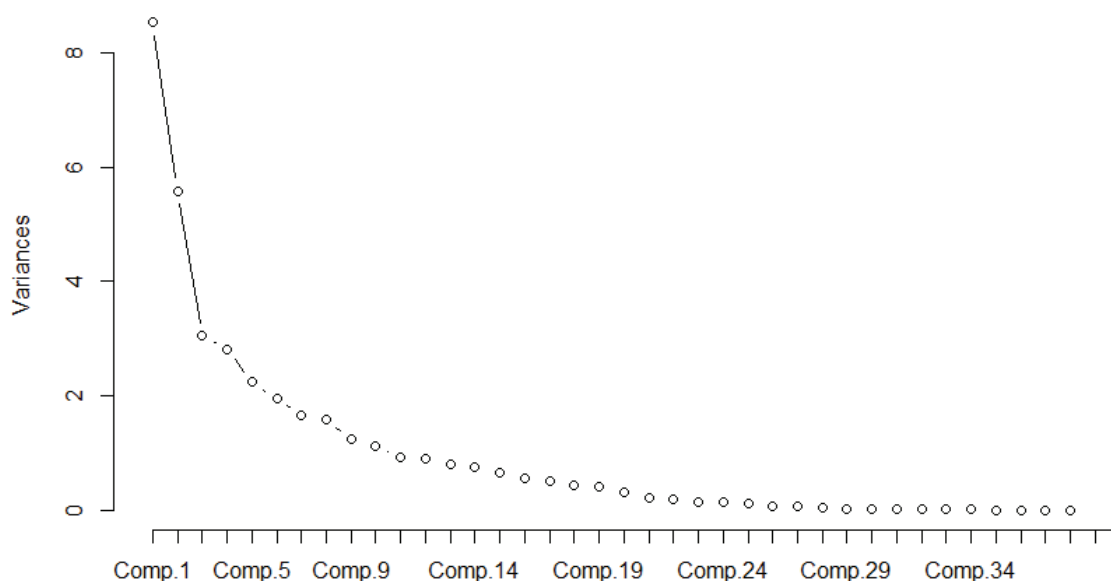


Figure 11. Scree plot of the variance values of the principal components for dataset2.

the tanglegrams in Figure 13, it is seen that BGCC (Baker's gamma correlation coefficient) gives inconsistent results for the centroid and median methods as well in dataset2. The FM_indexes in the table are equal to 1 for all methods. This means that in each method, the contents of the clusters formed as a result of clustering analysis with and without PCA are mutually identical. By evaluating the graphics and all the calculated coefficients together, it is observed that the best clustering result belonged to the PCA+Centroid method, and the cluster contents for the 3 number of clusters are given below.

Cluster 1: Albania, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro,

Netherlands, North Macedonia, Norway, Poland, Portugal, Republic of Moldova, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom
Cluster 2: Germany
Cluster 3: Russian Federation

Results for Dataset 3

The variance values of the principal components for dataset3 are given in Table 10 and Table 11. For the elbow test, when the scree plot which includes the variance values of the principal components given in Figure 12 is examined, it is seen that the natural breakpoint in the graph coincides with the 6th principal component. But the first 6 components can explain only %66.68 of the total variance in the dataset. This ratio is not sufficient to represent the entire data set.

Table 7. The variance values of the principal components for dataset2

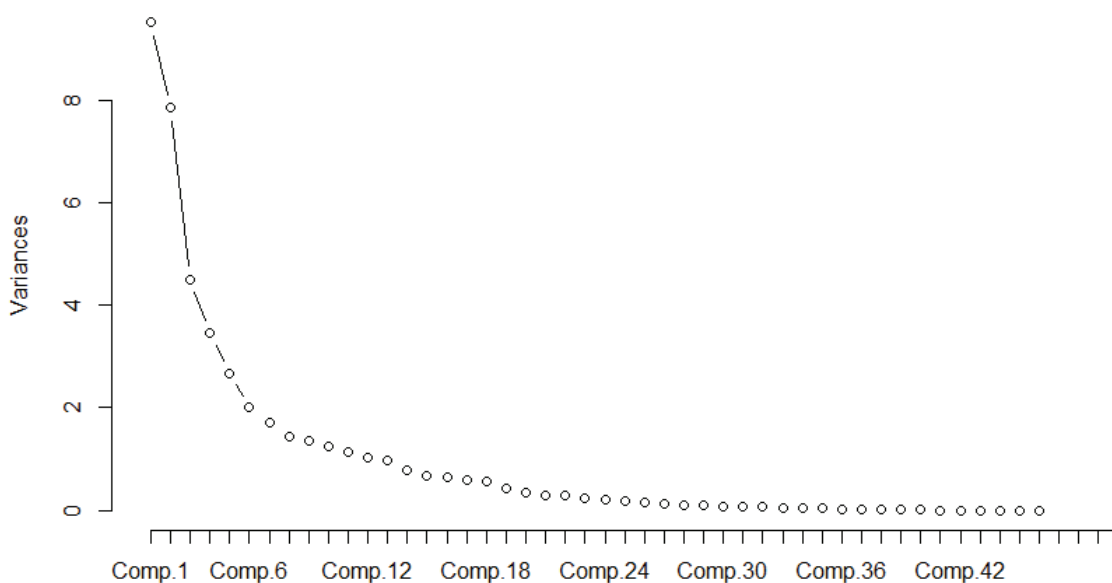
comp.1	comp.2	comp.3	comp.4	comp.5	comp.6
8,54E+00	5,58E+00	3,05E+00	2,82E+00	2,24E+00	1,96E+00
comp.7	comp.8	comp.9	comp.10	comp.11	comp.12
1,66E+00	1,59E+00	1,23E+00	1,11E+00	9,27E-01	8,87E-01

Table 8. The variance explained percentages of the principal components for dataset2

comp.1	comp.2	comp.3	comp.4	comp.5	comp.6
2,31E+01	1,51E+01	8,24E+00	7,61E+00	6,04E+00	5,29E+00
comp.7	comp.8	comp.9	comp.10	comp.11	comp.12
4,49E+00	4,31E+00	3,32E+00	3,00E+00	2,50E+00	2,40E+00

Table 9. Coefficient and index results used to measure the similarity of dendrograms for dataset2

	Methods	CPCC (dend 1)	CPCC (dend 2)	CPCC (tanglegram)	BGCC	FM_index
Data set 2.	single	0.8013	0.8092	0.9959106	0.9813049	1
	complete	0.7480	0.7351	0.8459703	0.8403557	1
	average	0.8588	0.8662	0.509475	0.8847822	1
	centroid	0.8370	0.8491	0.9947938	0.3375059	1
	median	0.7814	0.8022	0.9027062	0.4272787	1
	ward.D2	0.4782	0.4646	0.9947947	0.9996194	1

**Figure 12.** Scree plot of the variance values of the principal components for dataset3.

For the second approach, the mean-variance value ($\bar{\lambda}$) of the principal components is calculated as 0.9791 for dataset3, and it is observed that the variance values of the first 12 principal components are above this value. The first 12 components can explain %84.29 of the total variance in the data set.

For the 80% specific ratio, the first 11 principal components should be selected to achieve this specific ratio. The first 11 components can explain %82.00 of the total variance in the data set.

If the results of these three approaches are evaluated, it is clear that the mean-variance and 80% specific ratio approaches give very close results. As a result of all these evaluations, the number of components to be selected for a robust analysis process is determined as 11.

According to the CPCC (dend1 and dend2) values in Table 12, different from the results of dataset1 and dataset2, for dataset3, the results of PCA applied clustering analysis have close or lower similarity rates with the distance matrix in almost all methods except centroid linkage method. For

Table 10. The variance values of the principal components for dataset3

comp.1	comp.2	comp.3	comp.4	comp.5	comp.6
9,53E+00	7,86E+00	4,51E+00	3,45E+00	2,68E+00	2,02E+00
comp.7	comp.8	comp.9	comp.10	comp.11	comp.12
1,72E+00	1,43E+00	1,35E+00	1,25E+00	1,15E+00	1,03E+00

Table 11. The variance explained percentages of the principal components for dataset3

comp.1	comp.2	comp.3	comp.4	comp.5	comp.6
2,12E+01	1,74E+01	1,00E+01	7,66E+00	5,94E+00	4,48E+00
comp.7	comp.8	comp.9	comp.10	comp.11	comp.12
3,82E+00	3,18E+00	3,00E+00	2,78E+00	2,54E+00	2,29E+00

Table 12. Coefficient and index results used to measure the similarity of dendrograms for dataset3

	Methods	CPCC (dend 1)	CPCC (dend 2)	CPCC (tanglegram)	BGCC	FM_index
Data set 3	single	0.8718	0.8681	0.9959265	0.9962016	1
	complete	0.8458	0.7562	0.5993731	0.7446837	0.7299865
	average	0.9129	0.9108	0.9060994	0.9764306	1
	centroid	0.8714	0.8709	0.9939731	0.5969392	1
	median	0.8325	0.8174	0.9428428	0.430804	1
	ward. D2	0.5453	0.5148	0.8058044	0.7859016	0.7952145

CPCC(dend1) and CPCC(dend2), while the best clustering results belong to PCA+centroid and PCA+average methods, the worst clustering belongs to the variance-based Ward's method due to the outliers and low correlation in the dataset similar to dataset1 and dataset2. Consistent with the tanglegrams, the highest similarity rate belongs to single linkage and centroid linkage methods for the CPCC (tanglegram). When evaluated together with the tanglegrams above, it is seen that BGCC (Baker's gamma correlation coefficient) gives inconsistent results for the centroid and median methods as well in dataset3, and the highest value belongs to single linkage method. The FM_indexes in table 12 are equal to 1 for all methods except complete linkage and ward's criterion linkage method. This means that in each method except for complete and ward's linkage, the contents of the clusters formed as a result of clustering analysis with and without PCA are mutually identical. By evaluating the graphics and all the calculated coefficients together, it is observed that the best clustering result belonged to the PCA+Centroid method and the cluster contents for the 3 number of clusters are given below.

Cluster 1: Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Brunei Darussalam, Cambodia, Cyprus, Georgia, Hong Kong SAR, Indonesia, Iran, Iraq, Israel, Japan, Jordan, Kazakhstan, Kuwait, Kyrgyzstan, Lao People's Dem. Rep., Lebanon, Malaysia, Maldives, Mongolia, Myanmar, Nepal, Oman, Pakistan, Philippines, Qatar, Republic of Korea, Saudi Arabia, Singapore, Sri Lanka, State of Palestine, Syrian Arab Republic, Tajikistan, Thailand, Timor-Leste, Turkey, Turkmenistan, United Arab Emirates, Uzbekistan, Viet Nam, Yemen Cluster 2: India Cluster 3: China

CONCLUSION

PCA is one of the most important and frequently used methods to perform dimension reduction and to cope with the difficulties such as the increase in the amount and dimension of the data in the data pre-processing stages. The main purpose of this study is to examine the effect of PCA on hierarchical methods in dimension reduction for high-dimensional datasets. In the light of the observed results, it is seen that, while the Elbow method, which is one of the approaches used in PCA, gave misleading and insufficient results, the mean-variance and 80% specific rate approach gave consistent and reliable results. Low correlation between variables reduces the dimension reduction efficiency of PCA [26]. Despite the outliers and low correlation in the datasets, effective results are obtained in terms of dimension reduction performance, especially in the 2nd and 3rd datasets. For dataset1, the mean absolute correlation value is 0.2426, and dimension reduction with PCA is decreased 22 variables to 8 variables. For dataset2, the mean absolute correlation value is 0.2346, and dimension reduction with PCA is decreased 38 variables to 10 variables. For dataset3 the mean absolute correlation value is 0.2265 and dimension reduction with PCA is decreased 46 variables to 11 variables. When the CPCC values of the dendrograms are examined, it is noteworthy that the values of the Ward method are quite low for three datasets. The main reason explaining this situation is that the Ward method includes a completely variance-oriented merging process and the datasets in our study contain many outliers affecting the total variance. Considering the coefficient values, the best clustering results for all datasets belong to Average linkage and Centroid linkage methods in terms of CPCC (dend1)

and CPCC (dend2) values. Both methods have low outlier sensitivity. For these methods, it is seen that the FM_index values are equal to 1. Therefore, the clustering results are the same and the number of clusters is three except for average and average+PCA in dataset1. Baker's Gamma Correlation Coefficients of centroid linkage and median linkage methods, shown in red in the table 6, table 9 and, table 12 are given misleading results, incompatible with tanglegrams and other coefficients. For each dataset, all tanglegrams are carefully examined and it is observed that almost all the dendrograms obtained after the use of PCA formed a much more ordered hierarchical structure. The reason for this more ordered hierarchical structure can be thought of as the reduction of the negative effects of outliers in the dataset when PCA is used. When the CPCC (tanglegram) values, which measure the similarity of the cophenetic matrices of the dendrograms, are examined, it is clearly observed that the highest for all three datasets belong to Centroid and centroid+PCA, in concordance with the graphical results. This situation can be considered as clear proof that the Centroid linkage method is the most compatible with PCA and gives the most robust results. The coefficients and graphs used in the study enabled us to consider the results from three different perspectives; the compatibility probability of the merged object pairs, the branch heights changing according to the merging criteria (cophenetic matrix), and the contents of the clusters created. As an evaluation of all coefficients and graphs, for all three datasets, the Centroid+PCA method with the highest coefficients and congruent graphical results is more robust and reliable compared to other methods. Despite the low correlation and outlier disadvantages of datasets, it is observed that PCA allows hierarchical methods to work more comfortably in less dimensional space and with less negative variance effect. High-dimensional data are data types used in analysis in many fields or sectors today. Working with high-dimensional data is quite difficult and has disadvantages such as longer processing times and lower quality of results. However, the general result of this study showed that the use of the PCA dimensionality reduction method together with clustering has a positive effect on the analysis process and results. As stated in the introduction of the study, there are many studies in the literature that support the results obtained. For example, it has been shown that PCA positively affects classification time and performance in the Support vector machine method [16]. Likewise, the positive effects of different dimension reduction techniques along with PCA have been shown in the analysis processes carried out in facial recognition systems and automatic intrusion detection systems [17, 19]. In addition, the positive effect of PCA in prediction with high-dimensional cancer data has been stated, and similarly, the positive effects of linear and non-linear dimension reduction methods have been evaluated in the study on fluid mechanics [18, 20]. In this study, the dimensionality reduction ability of PCA was evaluated only through hierarchical clustering techniques.

Three different high-dimensional datasets were used in the study, but all three are low-correlation datasets. Positive and beneficial results were obtained in the study carried out under these restrictions. As suggestions for future studies, the dimensionality reduction performance of PCA can be investigated by using different clustering techniques or classification techniques. Additionally, the dimensionality reduction effect of PCA can also be investigated on deep learning methods. Another study that may be useful is to evaluate the performance of PCA on different correlated data sets. All parameters contained in the datasets and tanglegrams comparing the tanglegrams are included in the appendix section as Table A1, Figure A1, Figure A2, Figure A3.

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] Statista. Worldwide data created. 2021. Available at: <https://www.statista.com/statistics/871513/worldwide-data-created/> Accessed on September 10, 2025.
- [2] Han J, Kamber M, Pei J. Data mining concepts and techniques. 3rd ed. San Francisco (CA): Morgan Kaufmann; 2012.
- [3] Koj FS, Saba J. Using cluster analysis and principal component analysis to group lines and determine important traits in white bean. *Procedia Environ Sci* 2015;29:38–40. [CrossRef]
- [4] Öten M, Albayrak S. Determination of Variation Between Some Alfalfa (*Medicago sativa* L.) Genotypes by Principal Component and Clustering Analysis. *Türk Tarımsal Arastırma Derg* 2018;5:222–228. [Turkish] [CrossRef]

- [5] Shoba D, Vijayan R, Robin S, Manivannan N, Iyanar K, Arunachalam P, Nadarajan N, Arumugam Pillai M, Geetha S. Electronic J Plant Breed 2019;10:1095–1104. [\[CrossRef\]](#)
- [6] Ioele G, De Luca M, Dinç E, Oliverio F, Ragno G. Artificial neural network combined with principal component analysis for resolution of complex pharmaceutical formulations. Chem Pharm Bull (Tokyo) 2011;59:35–40. [\[CrossRef\]](#)
- [7] Kim YI, Boogert ST, Honda Y, Lyapin A, Park H, Terunuma N, Tauchi T, Urakawa J. Principal component analysis of cavity beam position monitor signals. J Instrum 2014;9:P02007. [\[CrossRef\]](#)
- [8] Gaitani N, Lehmann C, Santamouris M, Mihalakakou G, Patargias P. Using principal component and cluster analysis in the heating evaluation of the school building sector. Appl Energy 2010;87:2079–2086. [\[CrossRef\]](#)
- [9] Nwangburuka CC, Kehinde OB, Ojo DK, Denton OA, Popoola AR. Morphological classification of genetic diversity in cultivated okra, *Abelmoschus esculentus* (L.) Moench using principal component analysis and single linkage cluster analysis. Afr J Biotechnol 2011;10:11165–11172. [\[CrossRef\]](#)
- [10] Margaritis A, Soenen H, Fransen E, Pipintakos G, Jacobs G, Blom J, Van den Bergh W. Identification of ageing state clusters of reclaimed asphalt binders using principal component analysis and hierarchical cluster analysis based on chemo-rheological parameters. Constr Build Mater 2020;244:118276. [\[CrossRef\]](#)
- [11] Penkova T. Principal component analysis and cluster analysis for evaluating the natural and anthropogenic territory safety. Procedia Comput Sci 2017;112:99–108. [\[CrossRef\]](#)
- [12] Abdulhafedh A. Incorporating k-means, hierarchical clustering and PCA in customer segmentation. J City Dev 2021;3:12–30.
- [13] Yıldız K, Çamurcu AY, Doğan B. A comparative analysis of principal component analysis and non-negative matrix factorization techniques in data mining. In: Akademik Bilişim'10; Turkey. 2010. p. 12.
- [14] Al-Omairi L, Abawajy J, Chowdhury MU, Al-Quraishi T. High-dimensionality graph data reduction based on a proposed new algorithm. EPIc Ser Comput 2019;63:1–10. [\[CrossRef\]](#)
- [15] Al-Omairi L, Abawajy J, Chowdhury MU, Al-Quraishi T. An empirical analysis of graph-based linear dimensionality reduction techniques. Concurr Comput 2021;33:e5990. [\[CrossRef\]](#)
- [16] Bharadiya JP. A tutorial on principal component analysis for dimensionality reduction in machine learning. Int J Innov Sci Res Technol 2023;8:2028–2032.
- [17] Mousavi A, Arefanjazi H, Sadeghi M, Ghahfarokhi AM, Beheshtinejad F, Masouleh MM. Comparison of feature extraction with PCA and LTP methods and investigating the effect of dimensionality reduction in the bat algorithm for face recognition. Int J Robot Control Syst 2023;3:500–509. [\[CrossRef\]](#)
- [18] Kabir MF, Chen T, Ludwig SA. A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction. Healthc Anal 2023;3:100125. [\[CrossRef\]](#)
- [19] Nabi F, Zhou X. Enhancing intrusion detection systems through dimensionality reduction: a comparative study of machine learning techniques for cyber security. Cyber Secur Appl 2024;2:100033. [\[CrossRef\]](#)
- [20] Wang Z, Zhang G, Xing X, Xu X, Sun T. Comparison of dimensionality reduction techniques for multi-variable spatiotemporal flow fields. Ocean Eng 2024;291:116421. [\[CrossRef\]](#)
- [21] Reddy CK, Vinzamuri B. A survey of partitional and hierarchical clustering algorithms. In: Data clustering algorithms and applications. Boca Raton (FL): CRC Press; 2014. p. 87–110. [\[CrossRef\]](#)
- [22] Xu R, Wunsch DC. Clustering. Hoboken (NJ): John Wiley & Sons; 2009.
- [23] Everitt BS, Landau S, Leese M, Stahl D. Cluster analysis. 5th ed. Hoboken (NJ): John Wiley & Sons; 2011. [\[CrossRef\]](#)
- [24] Forina M, Armanino C, Raggio V. Clustering with dendrograms on interpretation variables. Anal Chim Acta 2002;454:13–19. [\[CrossRef\]](#)
- [25] Rencher AC. Methods of multivariate analysis. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2002. [\[CrossRef\]](#)
- [26] Yim O, Ramdeen KT. Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. Quant Methods Psychol 2015;11:8–21. [\[CrossRef\]](#)
- [27] Tokuda EK, Comin CH, Costa LD. Revisiting agglomerative clustering. Physica A 2021;574:126433. [\[CrossRef\]](#)
- [28] Mohbey KK, Thakur GS. An experimental survey on single linkage clustering. Int J Comput Appl 2013;76:1–6. [\[CrossRef\]](#)
- [29] El-Hamdouchi A, Willett P. Techniques for the measurement of clustering tendency in document retrieval systems. J Inf Sci 1987;13:361–366. [\[CrossRef\]](#)
- [30] Jarman AM. Hierarchical cluster analysis: comparison of single linkage, complete linkage, average linkage and centroid linkage method. 2020. Preprint. doi: 10.13140/RG.2.2.11388.90240
- [31] Emmendorfer LR, Canuto AMP. A generalized average linkage criterion for hierarchical agglomerative clustering. Appl Soft Comput 2021;101:106990. [\[CrossRef\]](#)
- [32] Bu J, Liu W, Pan Z, Ling K. Comparative study of hydrochemical classification based on different hierarchical cluster analysis methods. Int J Environ Res Public Health 2020;17:9515. [\[CrossRef\]](#)
- [33] Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Hoboken (NJ): John Wiley & Sons; 2005.

- [34] Majerova I, Nevima J. The measurement of human development using the Ward method of cluster analysis. *J Int Stud* 2017;10:239–257. [\[CrossRef\]](#)
- [35] Szekely GJ, Rizzo ML. Hierarchical clustering via joint between-within distances: extending Ward's minimum variance method. *J Classif* 2005;22:151–183. [\[CrossRef\]](#)
- [36] Murtagh F, Legendre P. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *J Classif* 2014;31:274–295. [\[CrossRef\]](#)
- [37] Abdi H, Williams LJ. Principal component analysis. *WIREs Comput Stat* 2010;2:433–459. [\[CrossRef\]](#)
- [38] Candes EJ, Li X, Ma Y, Wright J. Robust principal component analysis. *J ACM* 2011;58:11. [\[CrossRef\]](#)
- [39] Mudrova M, Prochazka A. Principal component analysis in image processing. *Res Gate Publ* 2005. Preprint. Available at: <https://www.semanticscholar.org/paper/PRINCIPAL-COMPONENT-ANALYSIS-IN-IMAGE-PROCESSING-Mudrov%C3%A1-Proch%C3%A1zka/76a7fc9d87736c8383576865cf50403e53e74848> Accessed on September 10, 2025.
- [40] Gemperline PJ. Principal component analysis. In: *Practical guide to chemometrics*. 2nd ed. Boca Raton (FL): CRC Press; 2006. p. 211–235. [\[CrossRef\]](#)
- [41] Triayudi A, Fitri I. Comparison of parameter-free agglomerative hierarchical clustering methods. *ICIC Express Lett* 2018;12:973–980.
- [42] Silva AR, Dias CTS. A cophenetic correlation coefficient for Tocher's method. *Pesqui Agropecu Bras* 2013;48:589–596. [\[CrossRef\]](#)
- [43] Nino JO, Berzal F. Evaluation metrics for unsupervised learning algorithms. *arXiv:1905.05667*. Preprint 2019. doi: 10.48550/arXiv.1905.05667
- [44] Kumar S, Toshniwal D. Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient. *J Big Data* 2016;3:1–11. [\[CrossRef\]](#)
- [45] Rao AR, Srinivas VV. Regionalization of watersheds by hybrid-cluster analysis. *J Hydrol* 2006;318:37–56. [\[CrossRef\]](#)
- [46] Saraçlı S, Doğan N, Doğan I. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J Inequal Appl* 2013;2013:203. [\[CrossRef\]](#)
- [47] Fowlkes EB, Mallows CL. A method for comparing two hierarchical clusterings. *J Am Stat Assoc* 1983;78:553–569. [\[CrossRef\]](#)
- [48] Hryniewicz O. Goodman–Kruskal measure of dependence for fuzzy ordered categorical data. *Comput Stat Data Anal* 2006;51:323–334. [\[CrossRef\]](#)
- [49] Kılıç AF. Kategorik veride faktör analizi için kullanılabilir alternatif bir korelasyon matrisi: Goodman–Kruskal gamma. *Marmara Univ Atatürk Eğitim Fak Eğitim Bilim Derg* 2021;54:151–168. [\[CrossRef\]](#)
- [50] United Nations. UN Data. 2021. Available at: <https://data.un.org/> Accessed on September 10, 2025.
- [51] UNICEF. UNICEF data. 2021. Available at: <https://data.unicef.org/> Accessed on September 10, 2025.
- [52] World Bank. World Bank open data. 2021. Available at: <https://data.worldbank.org/> Accessed on September 10, 2025.
- [53] Our World in Data. Our World in Data. 2021. Available at: <https://ourworldindata.org/> Accessed on September 10, 2025.
- [54] Worldometer. Real time world statistics. 2021. Available at: <https://www.worldometers.info/> Accessed on September 10, 2025.

APPENDICES

Table A1. The variables used in datasets

General Information = (GI)	Social Indicators = (SI)
1. Population (000, 2020)	23. Population growth rate (average annual %)
2. Pop. density (per km ² , 2020)	24. Urban population (% of total population)
3. Surface area (km ²)	25. Fertility rate, total (live births per woman)
4. Sex ratio (m per 100 f)	26. Life expectancy at birth (females/males, years)
5. Exchange rate (per US\$)	27. Population age distribution (0-14/60+ years old, %)
Economic Indicators = (EI)	28. International migrant stock (000/% of total pop.)
6. GDP: Gross domestic product (million current US\$)	29. Refugees and others of concern to UNHCR (000)
7. GDP growth rate (annual %, const. 2015 prices)	30. Infant mortality rate (per 1 000 live births)
8. GDP per capita (current US\$)	31. Health: Current expenditure (% of GDP)
9. Economy: Agriculture (% of Gross Value Added)	32. Health: Physicians (per 1 000 pop.)
10. Economy: Industry (% of Gross Value Added)	33. Education: Government expenditure (% of GDP)
11. Economy: Services and other activity (% of GVA)	34. Education: Primary gross enrol. ratio (f/m per 100 pop.)
12. Employment in agriculture (% of employed)	35. Education: Secondary gross enrol. ratio (f/m per 100 pop.)
13. Employment in industry (% of employed)	36. Education: Tertiary gross enrol. ratio (f/m per 100 pop.)
14. Employment in services (% employed)	37. Intentional homicide rate (per 100 000 pop.)
15. Unemployment (% of labour force)	38. Seats held by women in national parliaments (%)
16. Labour force participation rate (female/male pop. %)	Environment and Infrastructure Indicators = (ENI)
17. CPI: Consumer Price Index (2010=100)	39. Individuals using the Internet (per 100 inhabitants)
18. Agricultural production index (2004-2006=100)	40. Threatened species (number)
19. International trade: exports (million current US\$)	41. Forested area (% of land area)
20. International trade: imports (million current US\$)	42. CO ₂ emission estimates (million tons/tons per capita)
21. International trade: balance (million current US\$)	43. Energy production, primary (Petajoules)
22. Balance of payments, current account (million US\$)	44. Energy supply per capita (Gigajoules)
	45. Tourist/visitor arrivals at national borders (000)
	46. Important sites for terrestrial biodiversity protected (%)

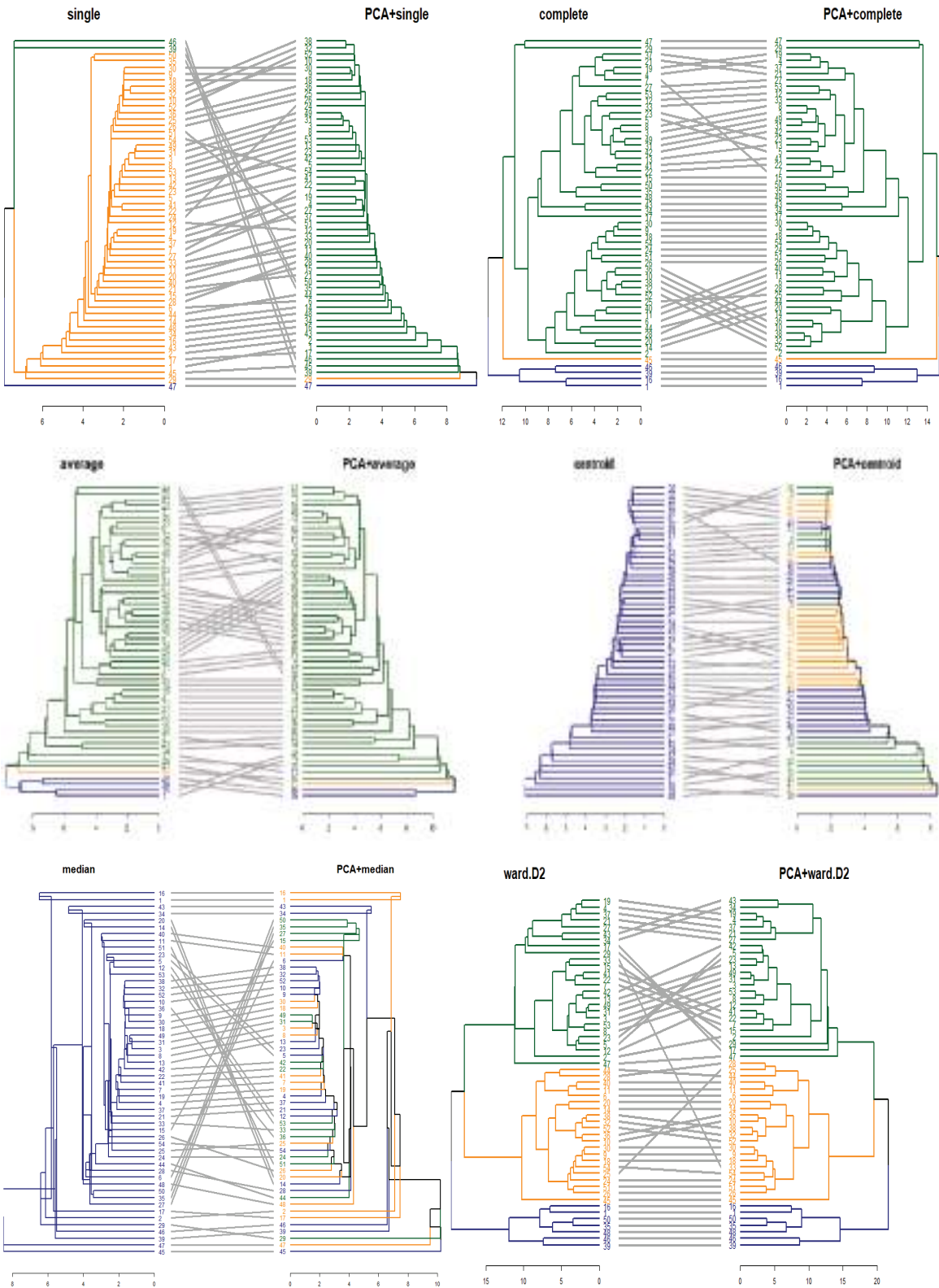


Figure A1. The tanglegrams for dataset1.

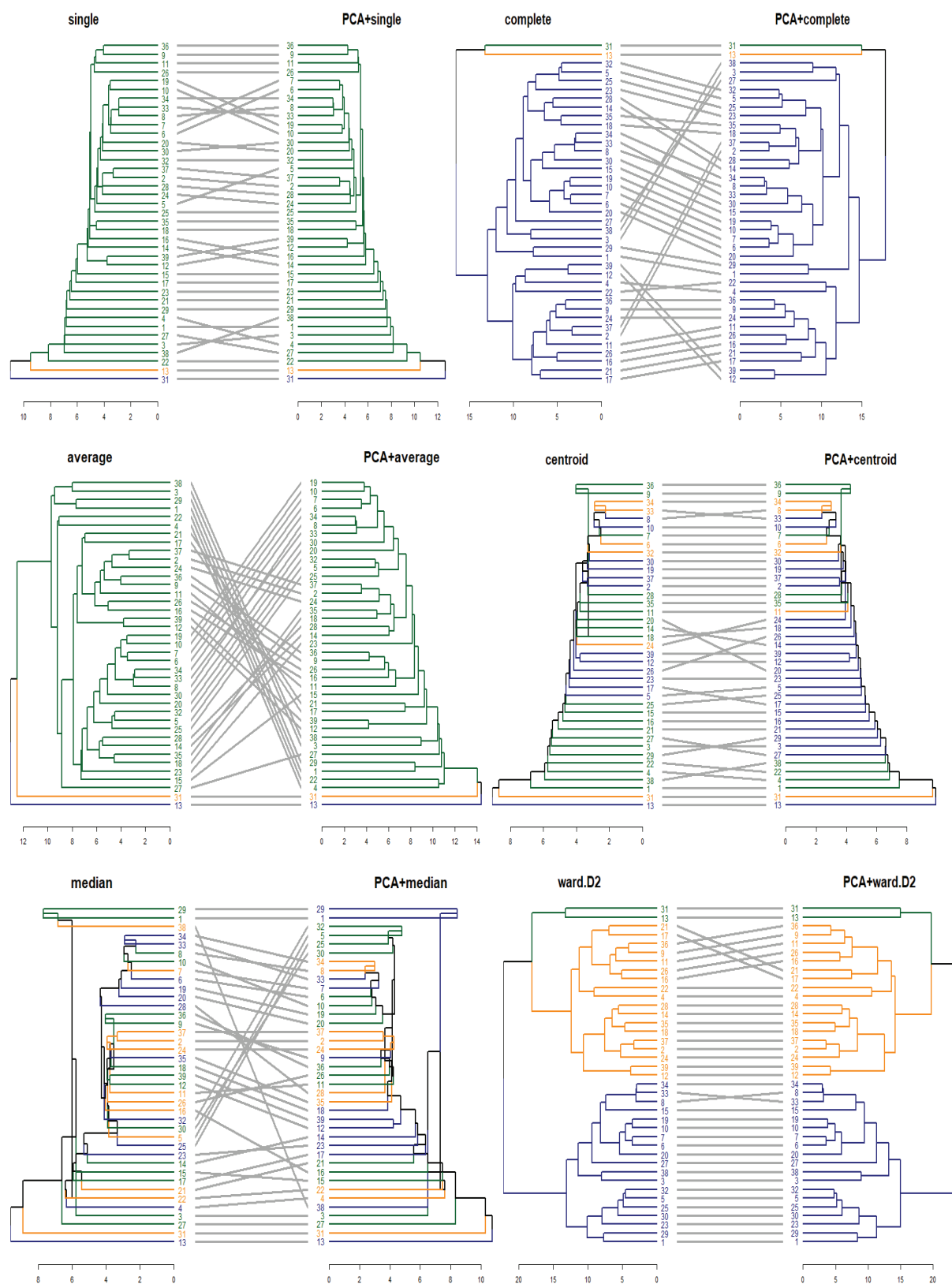


Figure A2. The tanglegrams for dataset2.

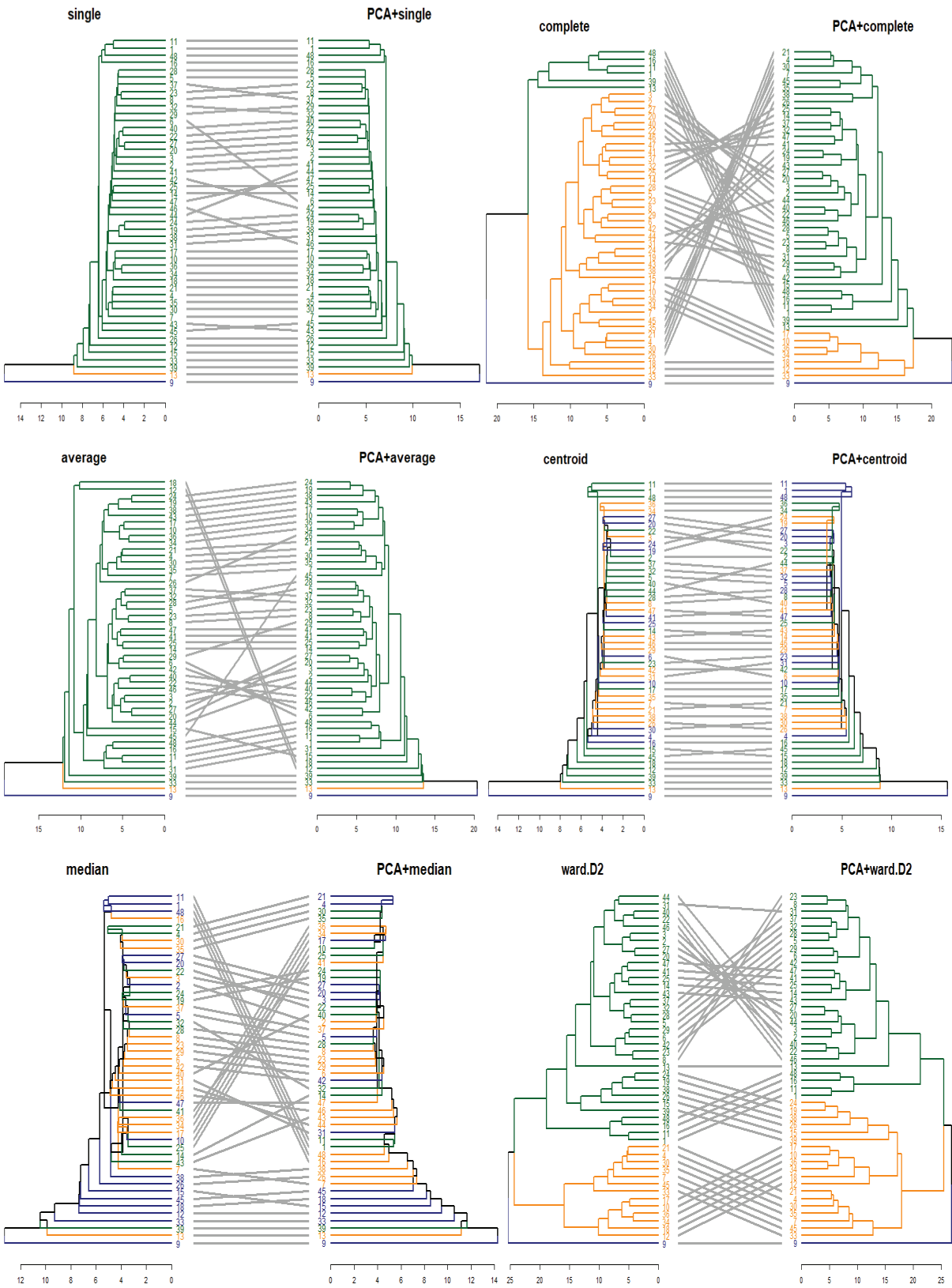


Figure A3. The tanglegrams for dataset3.



Research Article

A creative combination of time series models for accurate forecasting of monthly rainfall: A case study in Northeast India

D KARTHIKA¹, K KARTHIKEYAN^{1,*}

¹Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, India

ARTICLE INFO

Article history

Received: 08 May 2024

Revised: 24 June 2024

Accepted: 03 November 2024

Keywords:

ANN; Combined Forecasts;
Rainfall Forecasting; SARIMA;
Time Series Analysis

ABSTRACT

Rainfall forecasting is a complex and critical problem faced by many meteorologists. Classical forecasting models are struggle to capture the seasonal variations and long-term trends in rainfall data. So, it is essential to develop a more robust method to rainfall forecasting. The Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters Additive (HWA) models are used for parallel hybridization, with optimal weights determined by the variance-covariance matrix method. We evaluated the proposed model using monthly rainfall data from Jan 1990 to Dec 2017 of Northeast (NE) India. This region is divided into five divisions based on rainfall pattern that are West Bengal and Sikkim (WBS), Arunachal Pradesh (AP), Assam and Meghalaya (AM), Gangetic West Bengal (GWB), and Nagaland, Manipur, Mizoram, and Tripura (NMMT). The developed model is performed better than the classical models like SARIMA, HWA, Exponential Smoothing (ETS), Holt model, and FeedForward Neural Network (FFNN) across all regions. In the WBS region, it achieved an RMSE of 0.0798, an MAE of 0.0453, an MSE of 0.0063, an sMAPE of 0.3939, a correlation of 0.9414 between actual and predicted values, and an NSE of 0.8855. These results have significant implications for flood and drought management, climate change adaptation, and agricultural planning, particularly in the context of increasing climate variability.

Cite this article as: Karthika D, Karthikeyan K. A creative combination of time series models for accurate forecasting of monthly rainfall: A case study in Northeast India. Sigma J Eng Nat Sci 2025;43(5):1628–1642.

INTRODUCTION

Predicting rainfall precipitation is the most challenging research because of its time and spatial variance. So, precipitation prediction plays a significant role in society. The rainfall precipitation variance affects agricultural production, industry, and hydroelectric power generation, causing severe strain on the economy. The scarceness of monsoon

rainfall rigorously affects large parts of the country [1,2]. Rainfall forecasting is important for minimizing drought and flood damages, improving farming production, and establishing appropriate irrigation plans [3].

Classical forecasting models estimate rainfall based solely on previous records, resulting in insufficient accuracy. To mitigate this issue, statistical approaches have been proposed. Statistical modeling is a significant approach

*Corresponding author.

*E-mail address: k.karthikeyan@vit.ac.in

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



for testing, predicting, and deciding about hydrological cycle components. Over the past few decades, most of the academics have employed statistical techniques to address hydrological challenges [4–7].

The main factors influencing rainfall forecasts include monthly, annual, and seasonal cycles. In this article, we chose monthly rainfall data. Basically, in time series forecasting two types of models are used to predict: univariate model, and multivariate model. The multivariate time series models [6,8] are based on multiple parameters, including historical data, temperature, wind, and other influencing factors. Similarly, the univariate time series models [5,9–11] are based on single time series data for making short and long-term forecasts. To investigate the impact of univariate and multivariate time series models on the forecasting of rainfall, several forecasting techniques have been developed by many researchers. For accurate forecasting, some researchers concentrated on multivariate models. But it has some limitations: the availability of the data, and the complex structures of the model. So, our primary aim and objective of this study are to forecast rainfall precipitation using univariate time series models.

Considering the numerous monthly and seasonal patterns in model parameters and estimates, univariate time series models can be extremely beneficial for demand forecasting. In our research, we analyze various univariate forecasting models used by researchers worldwide to address hydrological issues and select suitable models for hybridization to improve forecast accuracy.

The ARIMA is a univariate time series model that is commonly used for forecasting. Box and Jenkins introduced the ARIMA model in the early 1970s. This model is also useful for identifying patterns in quasi time series. The ARIMA model is used to build the pure Seasonal ARIMA (p, d, q) (P, D, Q)s model [12]. In SARIMA model, the components P, D, and Q represent the suitable “seasonal autoregressive, integrated, and moving average components and p, d, and q represent autoregressive, integrated, and moving average components” [13,14]. The SARIMA model is commonly utilized in climatology, technology, financial statistics, and manufacturing prediction. Some authors have studied hydrological data prediction and used the SARIMA model for analyzing precipitation patterns in different regions [15–17].

Holt [18] modified exponentially weighted moving averages to provide the trend and seasonal variation. exponential smoothing is a common method for forecasting seasonal data. A Modified version of the exponential model is called the Holt-Winters model. it is a statistical forecasting model in univariate time series techniques. This model is useful for seasonal time series data, and it is predicted well [5,7].

The Artificial Neural Network (ANN) model is inspired by the human brain which is connected by neurons. It uses artificial neurons connected in layers. This model helps in predicting the nonlinear part of rainfall data. So, many

researchers used ANN models to forecast rainfall data [19–21].

Similarly, A univariate time series model for examining the level, trend, and seasonal elements of time series data is the exponential smoothing model. A modified form of the weighted moving average that takes trend and seasonal fluctuations into consideration is the Holt model [18].

Accordingly, we can integrate forecasts produced from the most precise forecasting systems for various forecasting sources and ranges to reduce their susceptibility to weather changes as well as other seasonal patterns. Bates and Granger, [22] introduced the concept of the combined forecast by adding weights to attach the individual forecasting method. Newbold and Granger,[23] compared the performance of the combined forecasts in terms of the ratio of average squared forecast errors. Averaging distinct forecasts, and combining forecasts helps reduce errors [24]. This is especially helpful when we are unsure about the best forecasting technique to apply. Winkler and Makridakis, [25] and others have made some significant contributions to combined forecasts of univariate time series models [26,27]. The combined forecast is a very productive way to achieve greater accuracy with minimum effort and time. We employ a novel strategy for this type of research. To employ the strategy, we used 28 years of monthly rainfall data (Jan 1990-Dec 2017) in NE India. NE India has a peculiar climate which means that the region receives rainfall during summer compared to other parts of India. We propose a novel parallel hybrid model using SARIMA and HWA models to forecast the rainfall precipitation in NE India. This approach assembles the unique modeling strengths of SARIMA and HWA. In terms of popular error metrics, the proposed model is compared with SARIMA, HWA, ETS, ANN, and Holt methods predicting accuracy in NE India.

Many existing forecasting models have high computational complexity for large scale applications and the scalability is still another significant problem. These are the limitations of existing models.

Our developed model overcomes this limitation by parallel processing, considering univariate dataset, improved forecasting accuracy.

In addition to addressing the significant limitations of the existing models, enhancements in our methodology pave the way for future study and implementation in the metrological department.

Objectives of the Study

The primary objective of this study is to develop a parallel hybrid model combining Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters Additive (HWA) methods for improved rainfall forecasting in Northeast India.

The secondary objective is to compare the efficacy of the presented parallel hybrid model with existing forecasting models to demonstrate its effectiveness and advantages.

The remaining article is organized as follows: The next section discusses the study area. Section 3 describes the data and methodologies of the SARIMA, HWA, ANN, and our proposed hybrid approach. Section 4 presents the empirical research evidence, which is followed by conclusions in Section 5.

STUDY AREA

Northeast (NE) India is the easternmost region of the country. In this study, we collected monthly rainfall data for the period 1990 to 2017 for the homogeneous region of NE India (Figure 1). Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, Sikkim, West Bengal, and Gangetic West Bengal were present in the Northeast Homogeneous Region. Based on the homogenous properties of rainfall patterns, these states are combined as NE India. Further, the subdivisions are made to analyze the monthly rainfall data series. The NE region of India shares its borders with various countries. This region shares an international border with several neighbouring countries. It is bounded by China in the north, Myanmar in the east, Nepal in the west and Tibet in the northeast, Bangladesh in the southwest, and Bhutan in the northwest. It has a total area of 262,230 square kilometres which is 8% of the total area of India [2]. This region has a humid subtropical climate.

That is, it has mild summers, strong monsoons, and extreme cold. It has the largest area of rainforest in the rest of India. Some parts of Northeast India receive an annual rainfall of 2000 mm. Arunachal Pradesh and Sikkim have an alpine climate with cold, frigid winters and warm summers [28]. Cherrapunji in Meghalaya receives rainfall throughout the year. The average annual rainfall of this place is 11,777mm. It is one of the rainiest regions of the world. 90% of the total rainfall of this region falls during the southwest monsoon. April to October is the rainy season and June and July are the heaviest rainy months [6].

DATA AND METHODS

Data

The monthly average rainfall precipitation of NE India, data from January 1990 to December 2017, have been collected from the Open Government Data Platform India (OGD) <https://data.gov.in/rainfall-india>. The 28 years of time series data were analyzed using the software R. To picture data, elementary statistics like Mean, Standard Deviation (SD), Minimum, and Maximum has calculated. Rainfall time series data were separated into training and testing datasets. Seventy percent of the data, from January 1990 to June 2009, were used as the training dataset, while

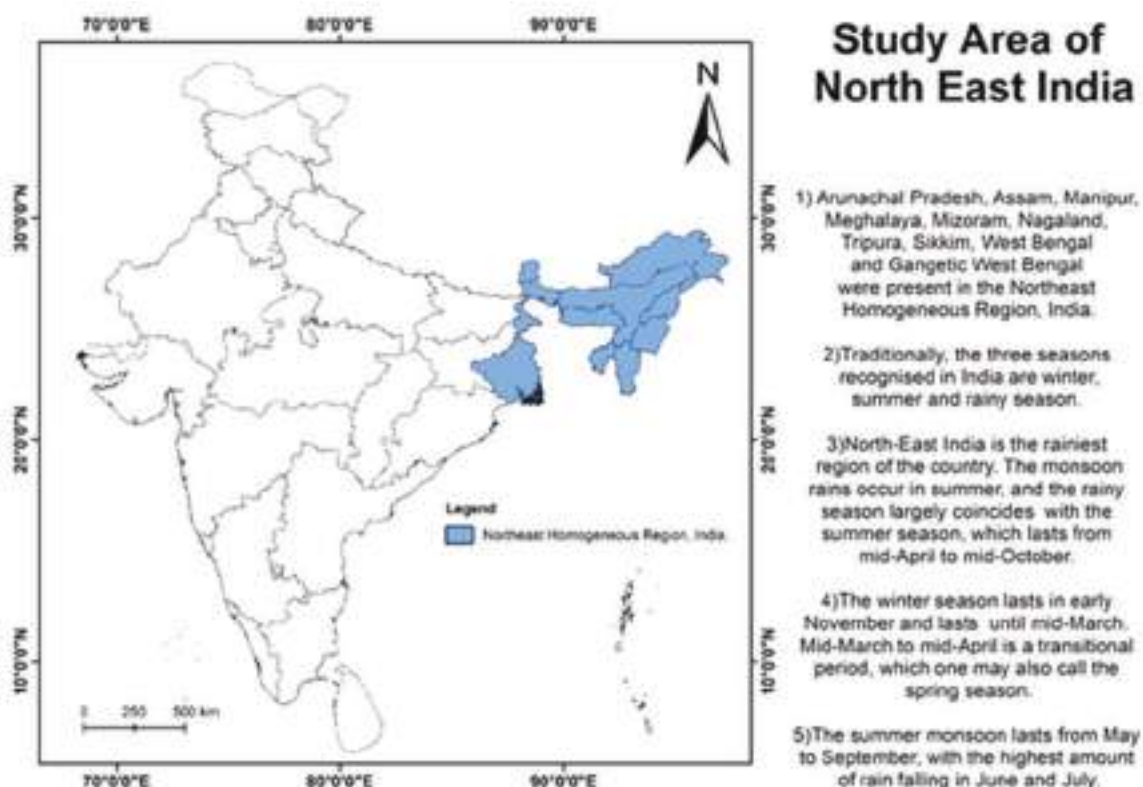


Figure 1. Study Area of Northeast India.

Table 1. Statistical Analysis for Northeast India rainfall data

a)WBS					
Rainfall Data	Minimum	Maximum	Mean	SD	CV
Actual Data	0	916.6	227.5658	233.8672	102.769
Training Data	0	916.6	235.6372	240.7008	102.148
Testing Data	0	777.9	209.049	217.3862	103.988
b)AP					
Actual Data	0.6	857.2	220.1321	193.7594	88.019
Training Data	1.4	857.2	221.4679	196.3434	88.655
Testing Data	0.6	724.9	217.0676	188.6143	86.891
c)AM					
Actual Data	0.2	848.4	203.9932	189.0011	92.650
Training Data	0.2	848.4	208.1372	190.8406	91.689
Testing Data	0.4	729.7	194.4863	185.2894	95.271
d)GWB					
Actual Data	0	633.1	132.3313	141.1562	106.668
Training Data	0	610.1	137.2491	142.7586	104.014
Testing Data	0	633.1	121.049	137.4342	113.536
e)NMMT					
Actual Data	0	619.4	179.8923	158.1471	88.26
Training Data	0	619.4	186.4244	161.7294	86.753
Testing Data	0.1	543.3	164.9069	149.2949	90.532

a) WBS- West Bengal and Sikkim, b) AP- Arunachal Pradesh, c) AM- Assam and Meghalaya, d) GWB- Gangetic West Bengal, e) NMMT- Nagaland, Manipur, Mizoram, and Tripura. SD is short form of standard deviation and CV is coefficient of variation.

thirty percent of the data, from July 2009 to December 2017, were used as the testing dataset. NE India has separated into five regions based on similarity characterization of rainfall precipitation. They are West Bengal and Sikkim (WBS), Arunachal Pradesh (AP), Assam and Meghalaya (AM), Gangetic West Bengal (GWB), Nagaland, Manipur, Mizoram, and Tripura (NMMT). The statistical analysis of the NE India rainfall data region-wise is in Table 1.

Decomposing Time Series

It is used to find the trend and seasonal components in time series data. It indicates basic components like an irregular, trend, and seasonal components. In Figure 2, The first row indicates the observed time series, the second row indicates the estimated irregular or residual component, the third row indicates the estimated seasonal component, and the final row indicates the estimated trend component [29]. The observed time series shows up and down pattern that indicates the rainfall data series remains seasonal. According to Figure 2, NE India's rainfall time series data have a seasonal component, with regular increasing and decreasing trends occurring annually. By Figure 2, WB, and Sikkim, from 2000 to 2015, almost the same trend in rainfall precipitation. Arunachal Pradesh met a downward trend from the year 2014 to 2017. GWB and AP met

an almost downscale trend in 2010. In 2014, Nagaland, Manipur, Mizoram, Tripura, Assam, and Meghalaya had an increasing trend in rainfall precipitation.

MATERTIALS AND METHODS

Normalizing the Data

The data must be pre-processed before implementing the forecasting models. In data-driven modeling methodologies, data pre-processing is often employed to reduce any anomalies, incomplete data, or inaccurate data [8]. For the whole rainfall dataset, the equation as follows is applied to normalize the time series data:

$$Norm(X) = \frac{X - Minimum(X)}{Maximum(x) - Minimum(X)} \quad (1)$$

Here, X is time series data, and Norm(X) is normalized X. we further use this value of rainfall data to compute the models and predict.

SARIMA Model

The SARIMA model is the best linear model for univariate data analysis and forecasting. it has been used to

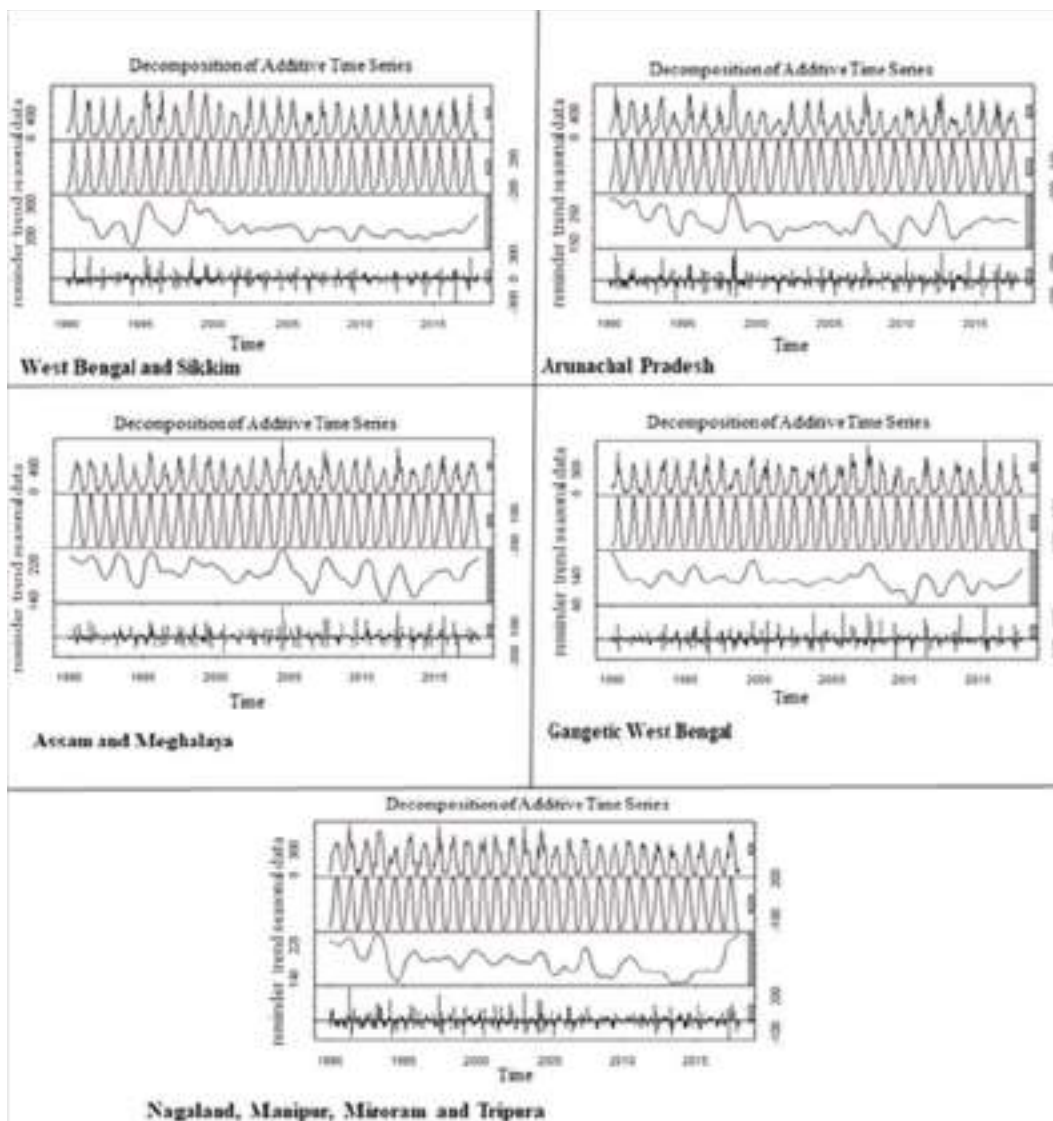


Figure 2. Decomposition of Time Series.

forecast and simulate the behaviour of univariate time series data [4,30]. The SARIMA modeling is based on correlational approaches, which must be utilized to represent components that are not obvious in the available data. The steps involved in predicting rainfall precipitation in NE using SARIMA are stationarity check, identification, and selection. The stationarity of time series data has been checked by Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values. Also, the statistical test, Augmented Dickey–Fuller (ADF), has been applied.

its general form of ARIMA (p, d, q) (P, D, Q) s is as follows [29].

$$\varphi(B)\varphi(B^s)(1-B)^d(1-B^s)^D(Z_{t-\mu}) = \theta(B)\theta(B^s)\epsilon_t \quad (2)$$

$$\text{Here, } \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

(The p order of the AR term),

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

(The q order of the MA term),

$$\varphi(B^s) = 1 - \varphi_1 B^s - \varphi_2 B^{2s} - \dots - \varphi_p B^{ps}$$

(The P order of the seasonal AR term),

$$\theta(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs}$$

(The Q order of the seasonal MA term)

and $\epsilon(t) \sim WN(0, \sigma^2)$, the difference d: quasi number, s is the absolute value which is always higher than one. Here μ is 0, if d or D is greater than 0. The ACF and PACF functions are used to estimate the model's order. The parameterization process is computed by maximum likelihood approaches after the model order has been determined by the Akaike information criteria, Bayesian information criteria, and other model selection criteria. The best fitted

SARIMA models for these regions are selected by the minimum value of RMSE, MAE, sMAPE, and MSE.

Holt-Winter Additive Method (HWA)

This approach uses a weighted moving average. It is used to evaluate seasonality, trend, and level [31]. The equations for Holt-Winters' additive approach for a series $\{X_t\}$ with period m are:

$$A_t = \gamma(X_t - S_{t-m}) + (1 - \gamma)(A_{t-1} + B_{t-1}) \quad (3)$$

$$B_t = \alpha(A_t - A_{t-1}) + (1 - \alpha)B_{t-1} \quad (4)$$

$$C_t = \beta(X_t - A_t) + (1 - \beta)C_{t-m} \quad (5)$$

$$F_n(h) = A_n + hB_n + C_{n+h-m}. \quad (6)$$

In this case, the smoothing estimates of level, trend, and seasonality at time t are represented by the variables A_t , B_t , and C_t , respectively. The smoothing parameters are γ , α , and β . They are employed to distinguish between the effects of recent and old observational data. In Eq. (6), The time series data's level at time n is represented by A_n and its trend at time n is represented by B_n . The trend is represented by hB_n , and the seasonal effect is represented by C_{n+h-m} (for yearly data $m = 12$, C_{n+h-12} is the expected seasonal in the corresponding month of the previous year). For point forecasting and model evaluation, we utilize the R software's "forecast" package.

Feed Forward Neural Networks (FFNN)

It is a type of ANN which is very good at stimulating the nonlinear patterns generally present in time series data from the real world [19]. Layers of synthetic neurons that mimic the neuronal connections seen in the human brain compose these models. The standard form of this model is NNAR (p , k); here ' p ' stands for the number of lag inputs and ' k ' for the number of hidden layers. The seasonal variant of FFNN model is indicated as NNAR (p , P , k). its working flow is single forward direction there is no recurrent or backward connections. There are no connections between neurons in the same layer [20]. A single hidden layer is typically included in NNAR models to reduce overfitting and make training and interpretation easier. This model is trained using "neuralnet" package in R. Based on the data it will optimizes the model parameters for better predicting by automatically choosing the sizes of the lag and hidden layer.

Proposed Methodology

The idea of integrating forecasts was initially proposed by Bates and Granger [22], who added weights to each forecasting technique. By integrating the predicted values from various models, they developed a linear hybrid model. A hybrid model's output is a linear combination of the results

that each model predicted, with the proper weights for each model determined by different kinds of techniques. Simple summing of predicted values, proportional MSE, and proportional forecast squared errors are a few ways to compute these weights [23, 26]. The time series models were recently merged by Najafabadipour et al. [27] utilizing particular weights that were obtained from the least squares approach.

Existing univariate rainfall forecasting methods have various drawbacks, as was previously mentioned. Whenever rainfall estimates are created with these techniques, such errors may compromise the reliability of distribution systems. We introduce a hybrid forecasting model designed especially for Northeast India to increase the precision of rainfall predictions. The parallel hybrid structure's fundamental structure is as follows:

$$f_{combined,t} = \varphi(w_1\widehat{f}_{1,t}, w_2\widehat{f}_{2,t}, \dots, w_n\widehat{f}_{n,t}) \quad t=1,2,\dots,T \quad (7)$$

Here, The entire combination variable is denoted by φ , and the adjusted predicted value of every single model at time t is indicated by $w_i\widehat{f}_{i,t}$ ($i = 1, 2, \dots, n$) [11,32].

Each predicted value is multiplied by its appropriate weight to determine the final forecast, which is obtained after applying the original data to each individual model. The parallel hybridization method, which combines a number of linear models, is used in this work to merge several forecasting models. The weighted forecasts from each separate model are added together to produce the final forecasts.

Holt-Winters Additive (HWA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are two univariate models that we unified in this study. Both the HWA and SARIMA models were used to generate forecasts for rainfall data in five different regions, and the precision of each model was evaluated. SARIMA outperformed HWA in most cases.

Thus, we merged the models as SARIMA-HWA or, in certain situations, as HWA-SARIMA. Eq. 8 yields the combined prediction model, and Eq. 9 computes the forecasting error. The variance-covariance matrix approach is used to calculate the weights. Figure 3 represents the flow chart of the proposed methodology. The following equation is then used to combine the results of the two prediction models:

$$\begin{aligned} Y_t &= \sum_{i=1}^2 w_i f_{it} \quad (t = 1, 2, \dots, n) \quad (w_1 + w_2 = 1) \\ &= w_1 * \text{forecasts from first forecast model} + w_2 * \\ &\quad \text{forecasts from second forecast model} \end{aligned} \quad (8)$$

$$E_{combined,t} = X_t - Y_t$$

$$\begin{aligned} &= \sum_{i=1}^n w_i X_t - (w_1 f_{1,t} + w_2 f_{2,t}) \\ &= w_1(X_t - w_1 f_{1,t}) + w_2(X_t - w_2 f_{2,t}) \end{aligned} \quad (9)$$

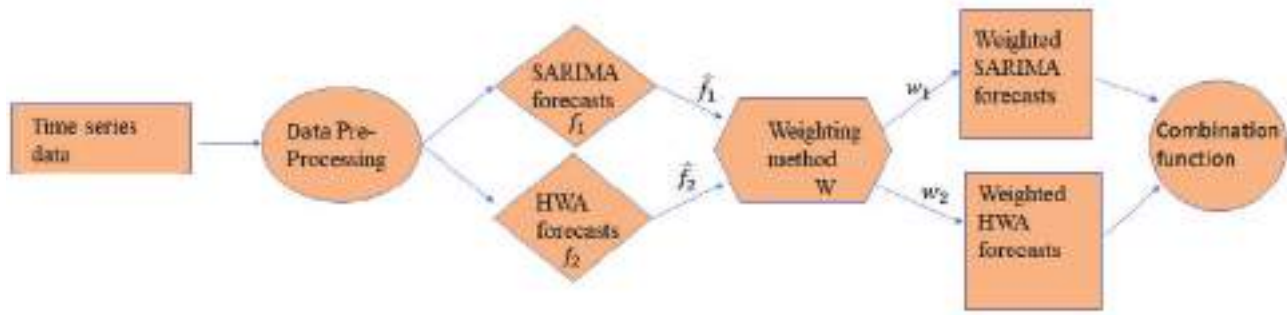


Figure 3. Flow chart of the Proposed Methodology.

Here w_1 and w_2 are the weights of first model's forecast, and weights of second model's forecast, respectively. The predicted value for the i th model is $\hat{f}_{it}$. The following is how the weights are determined:

$$w_1 = \frac{\text{var}(e_2) - \text{cov}(e_1, e_2)}{\text{var}(e_1) + \text{var}(e_2) - 2\text{cov}(e_1, e_2)} \quad (10)$$

$$w_2 = \frac{\text{var}(e_1) - \text{cov}(e_1, e_2)}{\text{var}(e_1) + \text{var}(e_2) - 2\text{cov}(e_1, e_2)} \quad (11)$$

SARIMA and HWA error values are e_1 and e_2 , respectively. We calculate $\text{var}(e_1)$, $\text{var}(e_2)$, and $\text{cov}(e_1, e_2)$, to determine w_1 and w_2 . Then, we derive w_1 and w_2 using Eq.10 and 11.

Performance Statistics for Model Evaluation

This research examines the efficiency of the fitted models in terms of the statistical indicators, like RMSE, MAE, MSE, sMAPE, NSE and Correlation Coefficient (R)

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (12)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i| \quad (13)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2 \quad (14)$$

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|X_i - Y_i|}{|X_i| + |Y_i|/2} * 100 \quad (15)$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \quad (16)$$

$$\text{Correlation Coefficient}(R) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (17)$$

Here, X_i and Y_i stand for real and predicted values, respectively. The set of data obtained is indicated by n . Prediction accuracy is indicated by RMSE, MSE values close to zero, and R and NSE values around one. When using real time series data to analyze and validate a hydrological model, NSE is the most widely utilized efficiency criteria [33].

RESULTS AND DISCUSSION

As discussed earlier, the normalized rainfall data has been used to model the proposed methodology. This methodology aims to combine univariate models to accurately forecast rainfall. The univariate models used for forecasting rainfall in Northeast India include SARIMA, Holt-Winters Additive (HWA), Feedforward Neural Networks (FFNN), Exponential Smoothing State Space Model (ETS), and the Holt Model. When it comes to rainfall forecasting, each of these models is important. The performance of each forecasting model determines whether it should be hybridized.

SARIMA Model

The steps involved in forecasting rainfall precipitation using the SARIMA model include checking for stationarity, model identification, and selection. We conducted forecasts for each region of Northeast India following these steps. To determine the stationarity of each region's rainfall data, we used the Augmented Dickey-Fuller (ADF) test with a p-value threshold of 0.05 and a lag order of 6. All regions of Northeast India had p-values less than 0.05, indicating that the normalized time series data for each region is stationary and the detailed results are presented in Table 2. The normalized time series data of each region of NE India is stationary. The SARIMA model was discovered and calculated using R software. The p , q and P , Q of the model is derived using the maximum likelihood approach. The Akaike information criterion (AIC) utilized to evaluate the model. The model with the lowest AIC, RMSE, MAE, MSE, and sMAPE is among the models chosen as the best-fitted model.

From Table 3, the ARIMA (1,0,1)(2,1,1)₁₂ model was selected as the best fit for the data in the WBS region of

Northeast India, with a minimum AIC value of -367.4906. This model has been used to forecast rainfall precipitation for the next 60 months, accounting for seasonal variance. Furthermore, ARIMA (1,0,1) (1,1,1)12 was determined as the most suitable model for rainfall forecasting in AP, with an AIC = -320.87. The ARIMA (1,0,1) (1,0,1)12 model was determined for AM, with an AIC = -415.86, and ARIMA (1,0,0) (1,1,0) 12 was chosen for GWB, with an AIC = -306.16. For NMMT, the ARIMA (1,0,1) (1,0,1) 12 model was effective, with an AIC value of -300.55. In Northeast India, these models were shown to be the most effective for predicting rainfall.

Table 4 presents the prediction accuracy of the selected SARIMA models for each region. The results of the SARIMA model for the WBS region are as follows: RMSE = 0.0855, MAE = 0.0539, MSE = 0.0073, and sMAPE = 0.4611. The true and projected values had a 0.9322 correlation, with a hydrological value of 0.8686 for the model. In the AP region, the selected model provides: RMSE = 0.1248, MAE = 0.0879, MSE = 0.0155, sMAPE = 0.6727, a correlation = 0.8327, and a hydrological model value of 0.6894. For AM, the SARIMA model gives RMSE = 0.1052, MAE = 0.0748, MSE = 0.0110, sMAPE = 0.5606, a correlation = 0.8784, and a hydrological model value of 0.7655. In GWB, the selected model results : RMSE = 0.1257, MAE = 0.0758, MSE = 0.0158, sMAPE = 0.7533, a correlation = 0.8187, and a hydrological model value of 0.6549. For NMMT, the SARIMA model yields RMSE = 0.0992, MAE = 0.0737, MSE = 0.0098, sMAPE = 0.6144, a correlation = 0.7753, and a hydrological model value of 0.9135.

The forecasting of monthly average rainfall for all regions of Northeast India was conducted after creating the appropriate time series models. The monthly data from

January 1990 to June 2009 were used for model validation, while the data from July 2009 to December 2017 were used for testing the forecasts. Using the chosen SARIMA models, Figure 4 shows the rainfall forecasting values for all of Northeast India's areas from July 2009 to December 2022. The blue line represents the predicted values, while the black line represents the actual values.

HWA Model

The Holt-Winters Additive (HWA) model was described in the technique section. The steps required in making predictions using the HWA model and analyzing the outcomes will be covered in detail in this section. Given that this model is a parametric approach, its parameters must be assigned weights. A higher weight means that the model gives the most recent observed data a higher priority when making predictions.

Using the best-fitting HWA models for all of Northeast India's areas, Figure 5 displays the rainfall predicted values from July 2009 to December 2022. The blue line represents the predicted values, while the black line represents the actual values. It is evident that the HWA model predictions in Northeast India follow the same seasonality as the historical data. To evaluate the performance of these prediction models, statistical measures such as RMSE, MAE, MSE, sMAPE, correlation coefficient R, and Nash-Sutcliffe Efficiency (NSE) are summarized in Table 4.

In the WBS region, the HWA model produces an RMSE value of 0.0983, an MAE value of 0.0792, an MSE value of 0.0096, an sMAPE value of 0.9981, a correlation coefficient of 0.9360, and a hydrological model value of 0.8262. For AP, the selected model yields an RMSE value of 0.1224, an MAE value of 0.0845, an MSE value of 0.0149, an sMAPE value

Table 2. ADF test for monthly rainfall precipitation data

Rainfall Data	Test	Calculated Value	Lag order	P value	Comment
WBS	ADF	-17.024	6	0.01	Stationary
AP	ADF	-12.88	6	0.01	Stationary
AM	ADF	-15.011	6	0.01	Stationary
GWB	ADF	-15.248	6	0.01	Stationary
NMMT	ADF	-13.314	6	0.01	Stationary

Table 3. Selected SARIMA model for monthly rainfall of each region of NE India

Region	SARIMA model	AIC	RMSE
WBS	ARIMA (1,0,1) (2,1,1) 12	-367.4906	0.0855
AP	ARIMA (1,0,1) (1,1,1) 12	-320.87	0.1248
AM	ARIMA (1,0,1) (1,0,1) 12	- 415.86	0.1052
GWB	ARIMA (1,0,0) (1,1,0) 12	-306.16.	0.1257
NMMT	ARIMA (1,0,1) (1,0,1)12	-300.55	0.0992

Table 4. Comparison of error measures for Best fitted models of monthly rainfall data series

West Bengal and Sikkim (WBS)						
Rainfall Data	RMSE	MAE	MSE	sMAPE	NSE	R
HWA-SARIMA	0.0798	0.0453	0.0063	0.3939	0.8855	0.9414
HWA	0.0983	0.0792	0.0096	0.9981	0.8262	0.9360
SARIMA	0.0855	0.0539	0.0073	0.4611	0.8686	0.9322
FFNN	0.0944	0.0688	0.0089	0.6474	0.8397	0.9196
ETS	0.0861	0.0613	0.0074	0.8972	0.8666	0.9370
Holt	0.5373	0.4461	0.2887	1.6844	-4.4490	0.0240
Arunachal Pradesh (AP)						
HWA-SARIMA	0.1206	0.0791	0.0145	0.3765	0.6986	0.8399
SARIMA	0.1248	0.0879	0.0155	0.6727	0.6894	0.8327
HWA	0.1224	0.0845	0.0149	0.4539	0.6770	0.8329
FFNN	0.1710	0.1239	0.0292	0.6104	0.3825	0.6481
ETS	0.1628	0.1060	0.0265	0.5230	0.4402	0.6866
Holt Model	0.2263	0.1909	0.0512	0.8546	-0.0817	0.1213
Assam and Meghalaya (AM)						
SARIMA-HWA	0.0961	0.0644	0.0092	0.4284	0.8046	0.8977
SARIMA	0.1052	0.0748	0.0110	0.5606	0.7655	0.8784
HWA	0.0983	0.0732	0.0096	0.8173	0.7955	0.8978
FFNN	0.1058	0.0789	0.0112	0.6411	0.7627	0.8688
ETS	0.1017	0.0746	0.0103	0.5441	0.7843	0.8934
Holt Model	0.2281	0.0520	0.1959	1.0569	-0.0139	0.0452
Gangetic WestBengal (GWB)						
SARIMA-HWA	0.0169	0.0572	0.0114	0.6630	0.7503	0.8681
SARIMA	0.1257	0.0758	0.0158	0.7533	0.6549	0.8187
HWA	0.1122	0.1933	0.0126	1.1094	0.7247	0.8583
FFNN	0.1299	0.0907	0.0168	0.7851	0.6314	0.8075
ETS	0.1186	0.0798	0.0140	0.7479	0.6164	0.8469
Holt	0.3635	0.3330	0.1321	1.2120	-2.6012	-0.0575
Nagaland, Manipur, Mizoram, and Tripura (NMMT)						
HWA-SARIMA	0.0834	0.0496	0.0069	0.4926	0.9382	0.8793
HWA	0.1137	0.0864	0.0129	0.8889	0.9186	0.8287
SARIMA	0.0992	0.0737	0.0098	0.6144	0.9135	0.7753
FFNN	0.1420	0.0956	0.0201	0.6802	0.6493	0.8156
ETS	0.1016	0.0757	0.0103	0.7368	0.8203	0.9094
Holt	0.4333	0.3655	0.1878	1.0205	-2.4605	-0.0839

of 0.4539, a correlation coefficient of 0.8329, and an NSE value of 0.6770. In AM, the HWA model gives an RMSE value of 0.0983, an MAE value of 0.0732, an MSE value of 0.0096, an sMAPE value of 0.8173, a correlation coefficient of 0.8978, and a hydrological model value of 0.7955. In GWB, the HWA model produces an RMSE value of 0.1122, an MAE value of 0.1933, an MSE value of 0.0126, an sMAPE value of 1.1094, a correlation coefficient of 0.8583, and an NSE value of 0.7247. For the NMMT region, the

HWA model results in an RMSE value of 0.1137, an MAE value of 0.0864, an MSE value of 0.0129, an sMAPE value of 0.8889, a correlation coefficient of 0.8287, and an NSE value of 0.9186.

FFNN Model

In the methodology section, the Feedforward Neural Network (FFNN) methodology was explained. This part details the processing steps for prediction using the FFNN

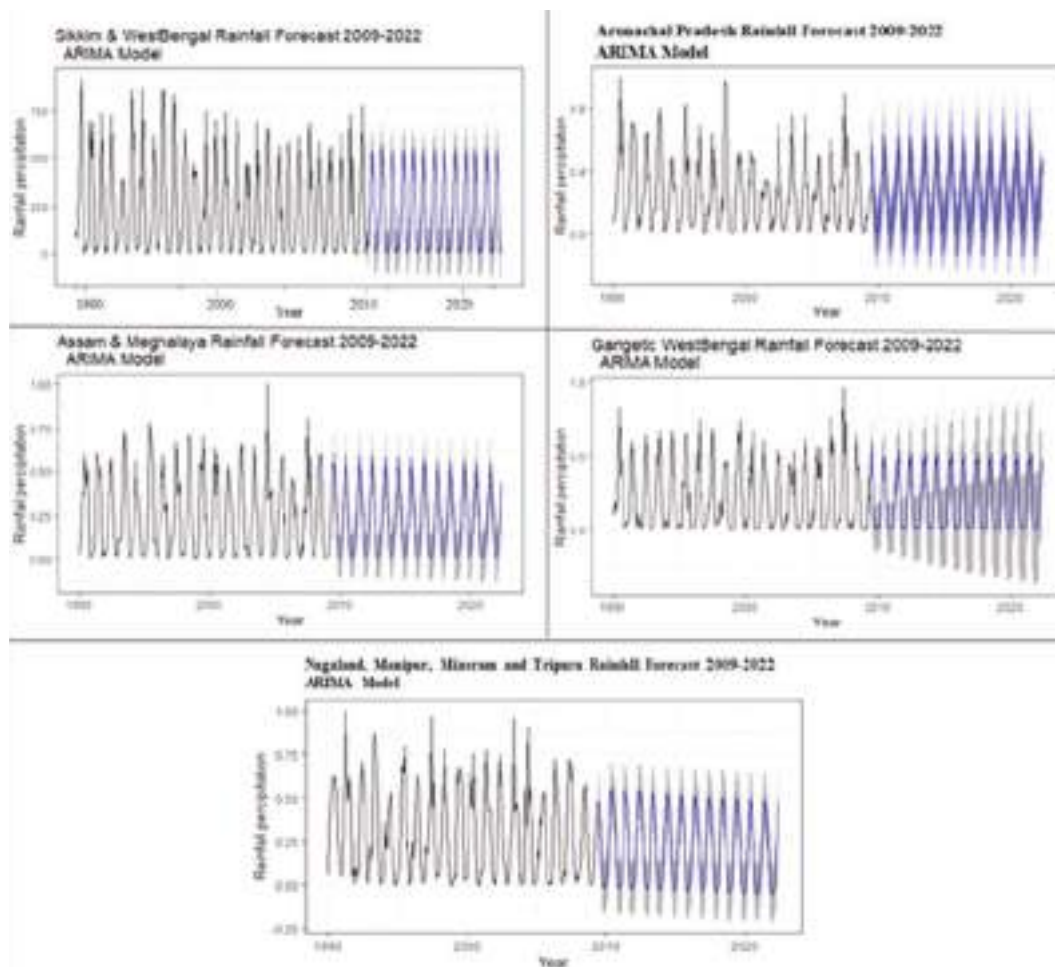


Figure 4. Forecast from best fitted SARIMA model for rainfall data series of Northeast India.

model, as well as the analysis of the results produced [31]. The artificial neural network model serves as an alternative to the SARIMA model for time series forecasting, capturing non-linear patterns in the data. In this study, we used a single hidden layer FFNN with one output node. Figure 6 displays the rainfall forecasting values for all regions of Northeast India. From Figure 6, it is evident that the FFNN model's predictions for each region follow the same seasonality as the historical data. The RMSE, MAE, MSE, sMAPE, correlation coefficient R , and Nash-Sutcliffe Efficiency (NSE) values for each region are summarized in Table 4.

The ETS and Holt models were forecasted using R software. Performance statistics for each model are shown in Table 4. The Holt model struggled with seasonal volatility when applied to this rainfall time series data. The HWA model outperformed the FFNN model in terms of error metrics, making it the best option. In comparison to the other models in the table, the SARIMA and HWA models exhibit superior performance. Although FFNN is also suitable for rainfall forecasting, it is less effective than SARIMA and HWA. Therefore, the ETS and Holt models were

disregarded. We will proceed with combining the SARIMA and HWA models.

Proposed Methodology

To combine the selected forecasting models, we used the variance-covariance matrix method to calculate the weights for each model. The weights for each forecasting model were determined using Eq. 10 and Eq. 11. Subsequently, forecasts and error values for the proposed model were computed using Eq. 8 and Eq. 9. Figure 7 shows the forecasted values of all regions of NE India using the proposed method. Blue line denotes the predicted values and the red line denotes the actual values.

The HWA- SARIMA model was chosen as effective model for WBS region. Since, RMSE value of 0.0798, MAE of 0.0453, MSE of 0.0063, sMAPE of 0.3939, a correlation between actual and predicted values of 0.9414, and an NSE value of 0.8855. For AP, the HWA-SARIMA model produced an RMSE of 0.1206, MAE of 0.0791, MSE of 0.0145, sMAPE of 0.3765, NSE of 0.6986, and a correlation value of 0.8399. In AM, the SARIMA-HWA model resulted in an RMSE of 0.0961, MAE of 0.0644, MSE of 0.0092, sMAPE of

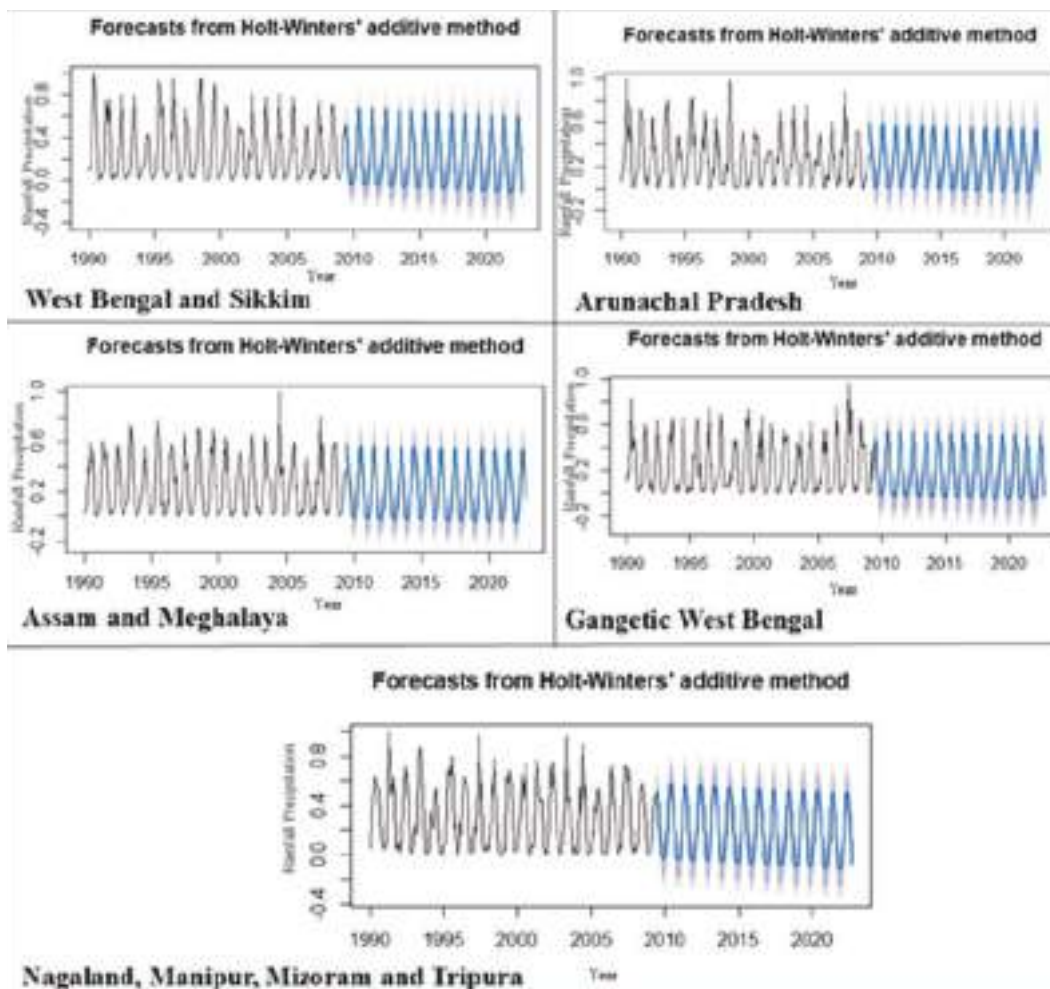


Figure 5. Forecast from best fitted Holt Winter's Additive Model for rainfall data series of Northeast India.

0.4284, NSE of 0.8046, and a correlation value of 0.8977. For GWB, the SARIMA-HWA model gave an RMSE of 0.0169, MAE of 0.0572, MSE of 0.0114, sMAPE of 0.6630, NSE of 0.7503, and a correlation value of 0.8681. In NMMT, the HWA-SARIMA model achieved an RMSE of 0.0834, MAE of 0.0496, MSE of 0.0069, sMAPE of 0.4926, NSE of 0.9382, and a correlation value of 0.8793.

These results indicate that the proposed methodology is more accurate than the existing forecasting models. The proposed approach is highly effective for rainfall forecasting. Table 4 demonstrates that the combined model outperforms the individual models and benchmark methods such as the ETS and Holt models, confirming that combining the best individual forecasting models yields superior performance.

Using monthly rainfall precipitation data from 1990 to 2017, we employed R software to forecast rainfall from 2018 to 2022. We proposed a hybrid statistical model for each region of Northeast India and compared it with individual models, including the ANN model. The results indicate that the proposed hybrid model outperformed the individual

forecasting models. Consequently, we used this model to forecast rainfall precipitation in Northeast India for the next 60 months (see Figure 7).

Rainfall patterns in Northeast India are becoming more unpredictable and intense. It is a sign of larger shifts brought on by global climate change. Recent research indicates that climate change is causing extreme rainfall and other weather events. From a 28-year study period of rainfall data, the state of Assam receives most of its rainfall in July, with a 21% variation when looking at historical data. Meghalaya receives 71% of its rainfall during the monsoon period. This area has decreasing trend in rainfall precipitation. Arunachal Pradesh had the rainfall in July, with a 31 percent variation. The state of Nagaland receives most of the rains in the month of July with 30% variation in monsoon rainfall when looking at historical data. Twenty-five percent of the total rainfall has varied [34]. Manipur has a consistent pattern. Mizoram has shown a consistent trend for overall rainfall. But while analysing the monthly trend, it shows little variation with a significant decrease in the number of rainy days. Tripura has a decreasing trend,

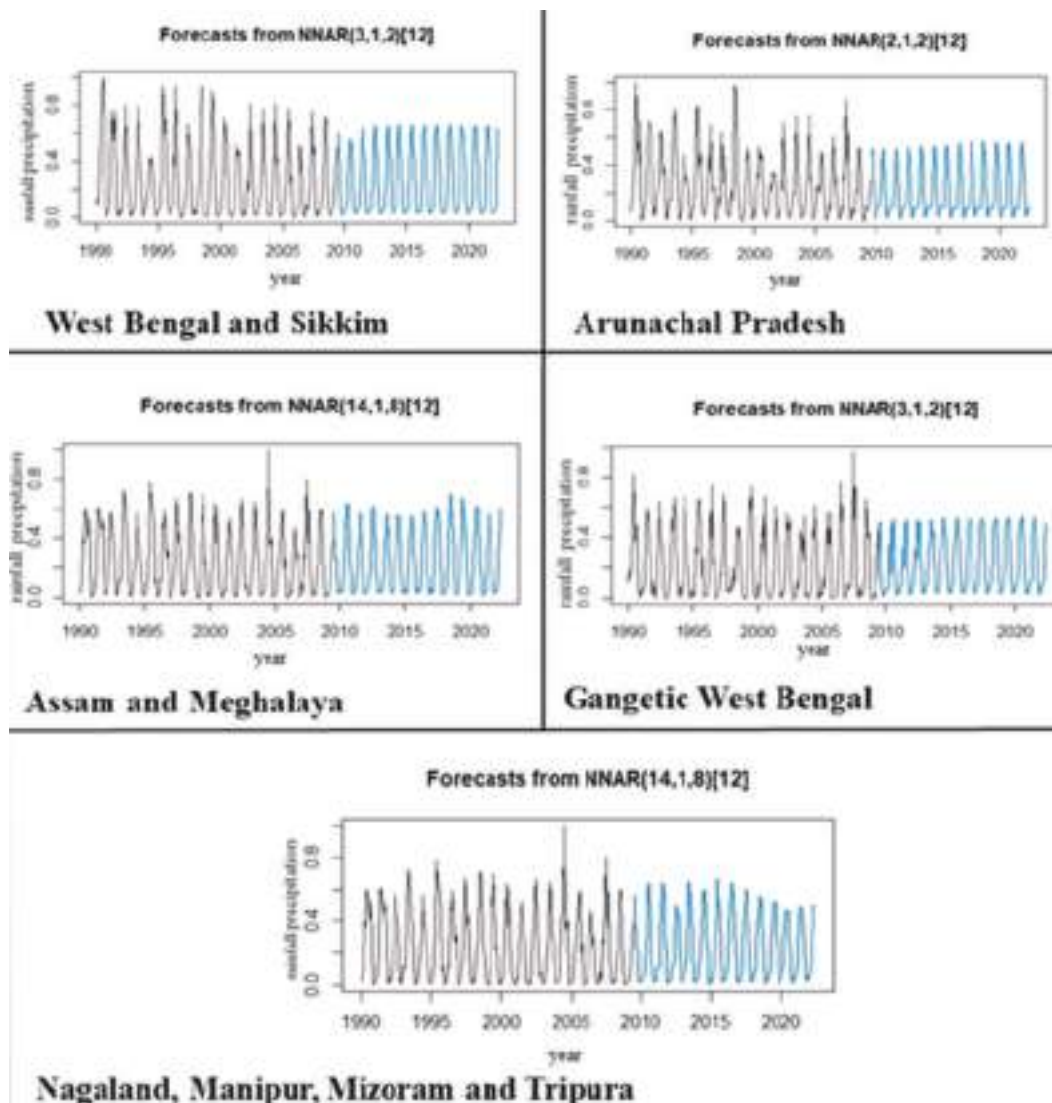


Figure 6. Forecast from Feed Forward Neural Network forecasting model rainfall data series of Northeast India.

even though it is not meaningful. Sikkim and West Bengal Monsoon rainfall increased slightly with a 23% variation. The month of July received the most rain. Gangetic West Bengal receives its highest rainfall in the month of June [35].

The choice of the preferred model based on univariate models may vary with different data sets. Therefore, it is crucial to evaluate all time series models for any location and hydrological factor to select the most suitable model for our needs. Our results indicate that statistical models are the most successful methods for rainfall prediction. They are very successful in identifying changes in rainfall time series components.

Applications of proposed model:

Flood Control: our proposed model is greatly beneficial in flood risk assessment and management. Experts can lessen the impact of floods on communities that are already

at risk by implementing early warning systems. They can taking preventative action when periods of heavy rainfall are anticipated.

Drought Preparedness: On the other hand, precise forecasts of decreased precipitation can help with drought resilience.

Infrastructure and Urban Planning: Designing and building sturdy structures requires an understanding of shifting rainfall patterns.

Policy and Decision-Making: The results highlight how crucial it is to incorporate climate change issues into national and regional policies. The model's insights can be used by policymakers to create more comprehensive and proactive plans to tackle the long-term problems caused by climate change, especially in regions that are vulnerable to drought and flooding.

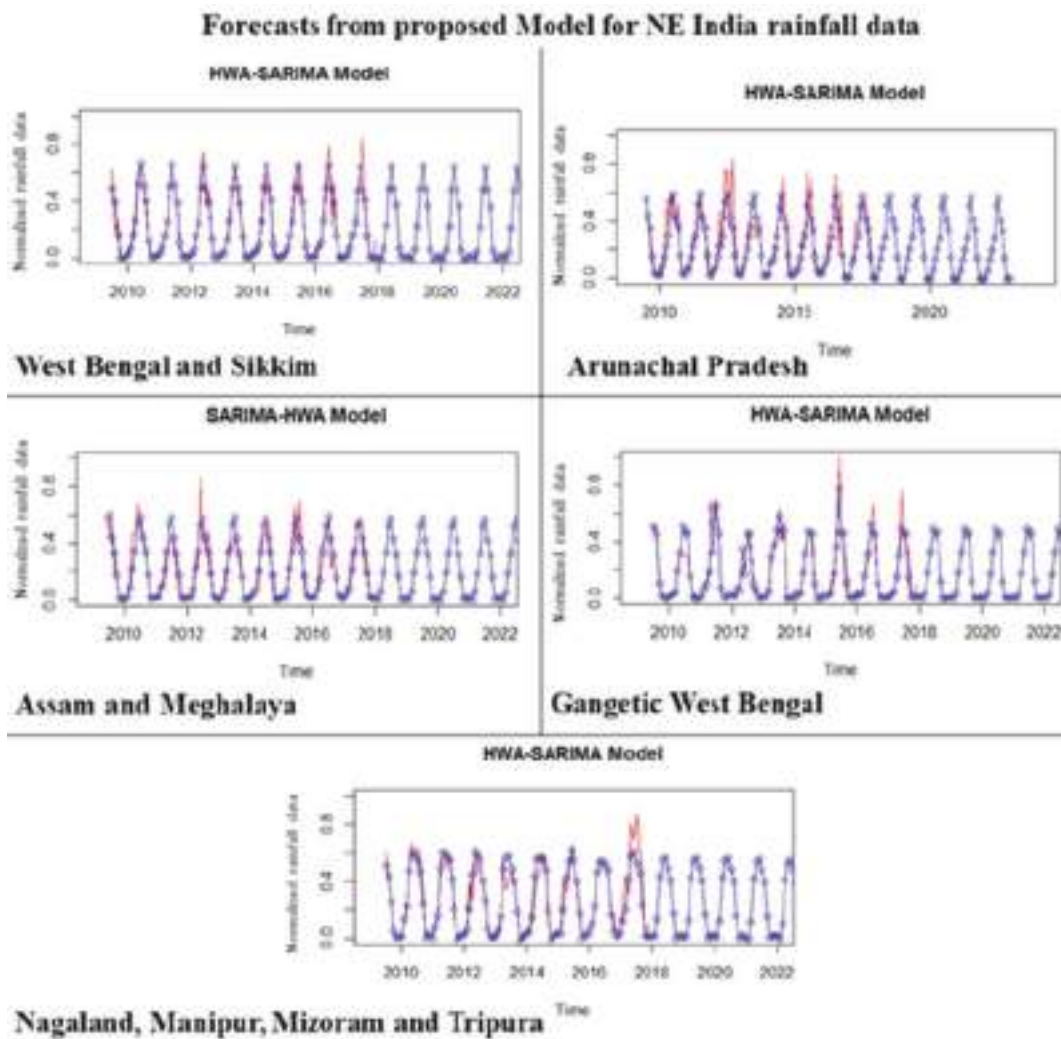


Figure 7. Forecast from combined statistical forecasting model rainfall data series of Northeast India.

In conclusion, the urgent necessity for adaptation measures is highlighted by the evidence of climate change represented in our rainfall projections. Our work supports sustainable development in the region by contributing to larger efforts to safeguard against the growing dangers connected with climate change by offering more accurate projections.

CONCLUSION

The fundamental purpose of this research is the design of a cohesive forecasting model. In order to enhance the accuracy of precipitation forecasts, our proposed model integrates the benefits of both the HWA and SARIMA frameworks. In this time series data, linear trends captured by the SARIMA model and the HWA model is proficient in detecting seasonal fluctuations. Finally, we have developed a combination method that optimizes overall forecasting efficacy by combining these models.

For rainfall forecasting, earlier studies, like [7], only used univariate models like HWA and ETS. But These models are not suitable for time series data that have seasonal variability, like Northeast India. According to our results, the HWA performed poorly in areas such as NMMT, with sMAPE values considerably greater than those derived from the suggested hybrid model.

Our suggested model is appropriate for application in the meteorological field due to its high accuracy. It makes easier to analyze and predict a range of hydrological data, such as groundwater levels and rainfall precipitation. Our model's accuracy and resilience make it appropriate for a variety of forecasting scenarios and useful in a wide range of areas. Hybrid model performs better than individual models and has applications in resource management, agricultural planning, and catastrophe preparedness. Its use could help reduce the hazards associated with extreme weather events like droughts and floods brought on by climate change throughout different parts of the world.

Future studies may investigate applying the parallel hybrid model to different places with diverse climatic circumstances to assess its wider application. Furthermore, introducing other factors such as temperature, wind speed, and humidity into the forecasting models could boost prediction accuracy by capturing more complicated interactions within the climate system.

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] Terzioğlu ZÖ, Kankal M, Yüksek Ö, Nemli MÖ, Akçay F. Analysis of the precipitation intensity values of various durations in Trabzon Province of Turkey by Şen's innovative trend method. *Sigma J Eng Nat Sci* 2019;37:241-250.
- [2] Borah P, Hazarika S, Prakash A. Assessing the state of homogeneity, variability and trends in the rainfall time series from 1969 to 2017 and its significance for groundwater in north-east India. *Nat Hazards* 2022;111:585-617. [\[CrossRef\]](#)
- [3] Nyatuame M, Agodzo SK. Stochastic ARIMA model for annual rainfall and maximum temperature forecasting over Tordzie watershed in Ghana. *J Water Land Dev* 2018;37:127-140. [\[CrossRef\]](#)
- [4] Soltani S, Modarres R, Eslamian SS. The use of time series modeling for the determination of rainfall climates of Iran. *Int J Climatol* 2007;27:819-829. [\[CrossRef\]](#)
- [5] Alonso Brito GR, Rivero Villaverde A, Lau Quan A, Ruiz Pérez ME. Comparison between SARIMA and Holt-Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Appl Sci* 2021;3:671. [\[CrossRef\]](#)
- [6] Lama A, Singh KN, Singh H, Shekhawat R, Mishra P, Gurung B. Forecasting monthly rainfall of Sub-Himalayan region of India using parametric and non-parametric modelling approaches. *Model Earth Syst Environ* 2022;8:837-845. [\[CrossRef\]](#)
- [7] Puah YJ, Huang YF, Chua KC, Lee TS. River catchment rainfall series analysis using additive Holt-Winters method. *J Earth Syst Sci* 2016;125. [\[CrossRef\]](#)
- [8] Mehdizadeh S, Fathian F, Safari MJS, Adamowski JF. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach. *J Hydrol (Amst)* 2019;579. [\[CrossRef\]](#)
- [9] Karthika D, Karthikeyan K. Estimation of electrical energy consumption in Tamil Nadu using univariate time-series analysis. *Ann Optim Theory Pract* 2021;4.
- [10] Karthika D, Karthikeyan K. Analysis of mathematical models for rainfall prediction using seasonal rainfall data: A case study for Tamil Nadu, India. In: 1st International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT 2022); 2022. [\[CrossRef\]](#)
- [11] Karthika D, Karthikeyan K. Performance of combined forecasting model for monthly rainfall precipitation. *Adv Appl Stat* 2023;90. [\[CrossRef\]](#)
- [12] Narasimha Murthy KV, Saravana R, Vijaya Kumar K. Modeling and forecasting rainfall patterns of south-west monsoons in North-East India as a SARIMA process. *Meteorol Atmos Phys* 2017;130:99-106. [\[CrossRef\]](#)
- [13] Eni D, Adeyeye FJ. Seasonal ARIMA modeling and forecasting of rainfall in Warri Town, Nigeria. *J Geosci Environ Prot* 2015;3:91-98. [\[CrossRef\]](#)
- [14] Dastorani M, Mirzavand M, Dastorani MT, Sadatinejad SJ. Comparative study among different time series models applied to monthly rainfall forecasting in semi-arid climate condition. *Nat Hazards* 2016;81:1811-1827. [\[CrossRef\]](#)
- [15] Mirzavand M, Ghazavi R. A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. *Water Resour Manag* 2015;29:1315-1328. [\[CrossRef\]](#)
- [16] Papacharalampous G, Tyralis H, Koutsoyiannis D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stoch Environ Res Risk Assess* 2019;33. [\[CrossRef\]](#)
- [17] AlSubih M, Kumari M, Mallick J, Ramakrishnan R, Islam S, Singh CK. Time series trend analysis of rainfall in last five decades and its quantification in Aseer Region of Saudi Arabia. *Arab J Geosci* 2021;14. [\[CrossRef\]](#)

- [18] Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast* 2004;20. [\[CrossRef\]](#)
- [19] Luk KC, Ball JE, Sharma A. An application of artificial neural networks for rainfall forecasting. *Math Comput Model* 2001;33. [\[CrossRef\]](#)
- [20] Chattopadhyay S. Feed forward artificial neural network model to predict the average summer-monsoon rainfall in India. *Acta Geophys* 2007;55. [\[CrossRef\]](#)
- [21] Guhathakurta P. Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. *Meteorol Atmos Phys* 2008;101. [\[CrossRef\]](#)
- [22] Bates JM, Granger CWJ. The combination of forecasts. *J Oper Res Soc* 1969;20:451–468. [\[CrossRef\]](#)
- [23] Newbold P, Granger CWJ. Experience with forecasting univariate time series and the combination of forecasts. *J R Stat Soc Ser A* 1974;137. [\[CrossRef\]](#)
- [24] Clemen RT. Combining forecasts: A review and annotated bibliography. *Int J Forecast* 1989;5. [\[CrossRef\]](#)
- [25] Winkler RL, Makridakis S. The combination of forecasts. *J R Stat Soc Ser A* 1983;146:150. [\[CrossRef\]](#)
- [26] Caiado J. Performance of combined double seasonal univariate time series models for forecasting water demand. *J Hydrol Eng* 2010;15. [\[CrossRef\]](#)
- [27] Najafabadipour A, Kamali G, Nezamabadi-pour H. The innovative combination of time series analysis methods for the forecasting of groundwater fluctuations. *Water Resour* 2022;49. [\[CrossRef\]](#)
- [28] Dikshit KR, Dikshit JK. North-East India: land, people and economy. Dordrecht: Springer Netherlands; 2014. [\[CrossRef\]](#)
- [29] Ray S, Das SS, Mishra P, Al Khatib AMG. Time series SARIMA modelling and forecasting of monthly rainfall and temperature in the South Asian countries. *Earth Syst Environ* 2021;5. [\[CrossRef\]](#)
- [30] Valipour M. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorol Appl* 2015;22. [\[CrossRef\]](#)
- [31] Winters PR. Forecasting sales by exponentially weighted moving averages. *Manage Sci* 1960;6. [\[CrossRef\]](#)
- [32] Hajirahimi Z, Khashei M. Hybrid structures in time series modeling and forecasting: A review. *Eng Appl Artif Intell* 2019;86. [\[CrossRef\]](#)
- [33] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models. Part I – A discussion of principles. *J Hydrol (Amst)* 1970;10. [\[CrossRef\]](#)
- [34] Das S, Tomar CS, Saha D, Shaw SO, Singh C. Trends in rainfall patterns over North-East India during 1961–2010. *Int J Earth Atmos Sci* 2015;2.
- [35] Ravindranath NH, Rao S, Sharma N, Nair M, Gopalakrishnan R, Rao AS, et al. Climate change vulnerability profiles for North East India. *Curr Sci* 2011;101.



Research Article

Exploration of Ulam-Hyers stability for a system of fractional integro-coupled differential equations with integral boundary conditions

Brijendra Kumar CHAURASIYA¹, Avadhesh KUMAR^{1,*}

¹Department of Mathematics, School of Advanced Sciences, VIT-AP University, Amaravati, 522237, India

ARTICLE INFO

Article history

Received: 16 May 2024

Revised: 03 July 2024

Accepted: 04 October 2024

Keywords:

Coupled System; Fractional
Integro-Differential Equation,
Green's Function; Integral
Boundary Conditions; Ulam
Stability

ABSTRACT

This study examines existence, uniqueness, and Ulam-Hyers stability for solutions of non-linear coupled fractional integro-differential equations with integral boundary conditions. Fractional systems incorporating memory and hereditary effects serve as effective models for complex processes in science and engineering applications. This study proves solution existence and uniqueness by applying the Banach fixed-point theorem within carefully constructed function spaces, then extends our analysis to investigate Ulam-Hyers stability, a framework that reveals how solutions behave when initial conditions contain small errors. Our stability analysis demonstrates that minor perturbations in starting data translate to bounded solution variations, keeping the system stable within predictable limits, which we verify through a computational example showing how controlled initial changes produce correspondingly controlled solution deviations. These results advance stability theory for coupled fractional integro-differential systems, particularly where memory effects influence system behavior, mathematical models that appear frequently in applications ranging from viscoelastic materials to population dynamics, where past states influence current evolution. By establishing rigorous stability bounds, our work provides a theoretical foundation for implementing these models in real-world scenarios where measurement uncertainties and modeling approximations are inevitable.

Cite this article as: Chaurasiya BK, Kumar A. Exploration of Ulam-Hyers stability for a system of fractional integro-coupled differential equations with integral boundary conditions. Sigma J Eng Nat Sci 2025;43(5):1643–1652.

INTRODUCTION

Fractional calculus has proven remarkably useful across mathematics and its applications. What makes it fascinating is how we can generalize traditional calculus beyond integer orders, taking derivatives of order 0.5 or 1.7, for instance. This flexibility reveals new mathematical structures that

better model phenomena where history matters, opening doors to more accurate descriptions of complex real-world systems. And it began when L'Hospital and Leibniz wrote a letter to each other in 1695. At present, there are a lot of theories and numerical findings on this topic in the scientific literature. The flow of groundwater in a stream can be described through a fractional differential equation

*Corresponding author.

*E-mail address: soni.iitkgp@gmail.com

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



that accounts for the non-integer order derivatives. The fractional derivative models the anomalous behavior of flow and transport in porous media. It may also be used in several fields, such as the theory of control, aerodynamics, communication and image filtering, and biology [1-4]. Manuals by Kilbas et al. [5], Miller and Ross [6], and Halfer were used to clarify the fundamental concepts of fractional calculus. Multiple writers have researched the existence and uniqueness of approaches to fractional order differential systems in finite and infinite-dimensional domains. Ahmad et al. discovered results involving non-linear fractional integro-differential equations utilizing integral boundary conditions [7-10].

The results obtained in this research demonstrate practical utility across multiple domains where complex differential systems with integral boundary conditions govern physical phenomena. Hydrogeological researchers have found fractional integro-differential equations particularly effective for modeling subsurface water flow dynamics [10], while chemical engineers apply these systems to understand reaction-diffusion mechanisms [11,12]. When studying population dynamics or circulatory systems, we need mathematical frameworks that can reliably predict how these systems evolve over time, making solution stability absolutely essential for any practical model [13]. Our Ulam-Hyers stability analysis tackles this challenge by establishing clear bounds on how solutions change when initial parameters vary slightly. This becomes especially important in hydrology and biomedical engineering, where professionals rely on model predictions to make decisions that can have serious consequences for public health or environmental protection. The stability criteria we've developed offer mathematical confidence that these models will behave consistently under real-world conditions, which is exactly what practitioners need when applying these tools to solve actual problems in environmental science and engineering.

Integral boundary conditions play a crucial role in hydrogeological modeling as they capture the dynamic interactions between aquifer systems and surrounding environments. These conditions can represent various physical processes, including groundwater recharge rates or discharge flows between aquifers and surface water bodies such as rivers and lakes. Additional details regarding integral boundary conditions can be found in [13-15]. To illustrate the practical significance of integral boundary conditions, consider the following real-world example

$$-\varpi'' = f(\phi)g(\phi, \varpi), \quad \varpi(a)=0, \quad \eta\varpi'(b) = \varpi(\beta)$$

where $\phi \in (a, b)$, $\beta \in (0,1]$ and η are positive constants, and a, b are constants. This represents a thermoregulator model. The problem admits well-defined solutions for the one-dimensional heat equation describing a heated bar equipped with a control system that modulates heat input based on temperature measurements from a sensor positioned at β .

This formulation extends to more general cases as the heat equation, incorporating nonlinear gradient terms and time-varying source components. The heated bar now features a controller positioned at β that regulates heat output according to temperature data collected from multiple sensors distributed along the bar [16-20]. This problem, therefore, may be stated as follows

$$\varpi'' = g(\phi, \varpi, \varpi'), \quad \varpi(a)=0, \quad \varpi'(b) = \int_a^b \varpi(s)dh(s)(\beta).$$

Ulam's type stability is a mathematical notion that describes how solutions to differential equations behave when tiny perturbations are introduced into the beginning conditions or equations themselves. It quantifies how stable or responsive a system is to disturbances. In 1940, Ulam [21] was the first to introduce the Ulam kind of stability. Hyers [22] subsequently refined this concept. Ulam and Hyers investigated stability properties across various integer-order differential equations, leading to what is now termed Hyers-Ulam stability for Cauchy functional equations. Their work produced stronger results concerning polynomials, isometric mappings, and convex functions [23-34]. Numerous researchers have since extended and generalized Hyers' framework for integer-order differential equations. The literature contains extensive investigations into additional stability concepts of Ulam type, documented in [35-44] and related works [45-51].

Motivated by the previously mentioned work related to integro-systems equipped with fractional order derivatives, we aim to extend the findings of [1, 51]. Integro-differential equations are important because they combine the properties of both differential and integral equations, allowing them to model a wide range of complex systems where current states depend on both rates of change and accumulated effects. Despite their significance, few studies have focused on investigating Ulam-type stability for integrodifferential equations involving fractional derivatives. Despite extensive work on fractional differential equations, Ulam's stability for coupled systems with integral boundary conditions remains unexplored. The proposed system is as follows

$$\begin{cases} {}^c D^m x(t) = k(t, y(t), \chi^y(t)), & m \in (1,2], t \in \mathfrak{N} \\ {}^c D^n y(t) = l(t, x(t), \chi^x(t)), & n \in (1,2], t \in \mathfrak{N} \\ \theta x(0) + \vartheta x'(0) = \int_0^1 b_1(x(s))ds, \theta x(1) + \vartheta x'(1) = \int_0^1 b_2(x(s))ds \\ \bar{\theta} y(0) + \bar{\vartheta} y'(0) = \int_0^1 \bar{b}_1(y(s))ds, \bar{\theta} y(1) + \bar{\vartheta} y'(1) = \int_0^1 \bar{b}_2(y(s))ds, \end{cases} \quad (1.1)$$

where ${}^c D$ is the Caputo fractional derivative, m, n represents the order of the derivative, $\mathfrak{N} = [0,1]$ and $k, l: [0,1] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ for $\chi^y, \chi^x: [0,1] \times [0,1] \rightarrow [0, \infty)$. And

$$\chi^y(t) = \int_0^t \hat{y}_1(t, s)y(s)ds \quad \text{and} \quad \chi^x(t) = \int_0^t \hat{y}_2(t, s)x(s)ds.$$

Here $b_1, b_2, \bar{b}_1, \bar{b}_2: \mathbb{R} \rightarrow \mathbb{R}$ and $\theta, \bar{\theta} > 0, \vartheta, \bar{\vartheta} \geq 0$ are real numbers.

We organize this work as follows: Section 2 presents essential definitions, notation, and preliminary results needed for our main analysis. In Section 3, we establish the existence and uniqueness of solutions for system (1.1). Section 4 develops our main stability results, focusing on Ulam-type stability criteria. Finally, Section 5 demonstrates these theoretical findings through a computational example that illustrates the practical implications of our stability bounds.

PRELIMINARIES

We start with preliminary results using a few helpful concepts, notations, and lemmas, which will be utilized in the subsequent parts.

Definition 1.[5] The Caputo derivative of a given function $y \in ((0,1), \mathbb{R})$ for a fractional order $m \in \mathbb{R}^+$ is presented as

$${}^c D^m y(t) = \frac{1}{\Gamma(\zeta - m)} \int_0^t (t-s)^{\zeta-m-1} y^{(\zeta)}(s) ds, \zeta = [m] + 1,$$

where $[m]$ denote the integer part of m and $\Gamma(\cdot)$ is the gamma function.

Definition 2. [5] The Riemann- Liouville fractional derivative of a function $y(t)$ of order m is presented as

$$J^m y(t) = \frac{1}{\Gamma(\zeta - m)} \left(\frac{d}{dt} \right)^\zeta \int_0^t \frac{y(s)}{(t-s)^{m-\zeta+1}} ds, \zeta = [m] + 1,$$

provided the integral exists.

Lemma 1.[5] For any real number $m > 0$, the differential equation

$${}^c D^m y(t) = 0$$

has a solution which is presented by

$$y(t) = d_0 + d_1 t + d_2 t^2 + \cdots \dots + d_{\zeta-1} t^{\zeta-1},$$

$$d_i \in \mathbb{R}, i = 0, 1, 2, \dots, \zeta - 1,$$

where $\zeta = [m] + 1$.

Lemma 2. [5] The solution for any real number $m > 0$, of the differential equation

$${}^c D^m y(t) = u(t)$$

will be presented as

$$J^m [{}^c D^m y(t)] = J^m y(t) + d_0 + d_1 t + d_2 t^2 + \cdots \dots + d_{\zeta-1} t^{\zeta-1},$$

$$d_i \in \mathbb{R}, i = 0, 1, 2, \dots, \zeta - 1,$$

where $\zeta = [m] + 1$.

Lemma 3. [9] For any function $u, \gamma_1, \gamma_2 \in C([0,1], \mathbb{R})$, the following boundary value problem has a unique solution of

$$\begin{cases} {}^c D^m y(t) = u(t), t \in [0,1] \\ \theta y(0) + \vartheta y'(0) = \int_0^1 \gamma_1(s) ds, \theta y(1) + \vartheta y'(1) = \int_0^1 \gamma_2(s) ds \end{cases} \quad (2.1)$$

is presented as

$$y(t) = \int_0^1 G_m(t,s) u(s) ds + \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 \gamma_1(s) ds + (\vartheta + \theta t) \int_0^1 \gamma_2(s) ds \right], \quad (2.2)$$

where $G_m(t,s)$ is the Green's function presented by

$$G_m(t,s) = \begin{cases} \frac{\theta(t-s)^{m-1} + (\vartheta - \theta t)(1-s)^{m-1}}{\theta \Gamma(m)} + \frac{\vartheta(\vartheta - \theta t)(1-s)^{m-2}}{\theta^2 \Gamma(m-1)}, & s \leq t \\ \frac{(\vartheta - \theta t)(1-s)^{m-1}}{\theta \Gamma(m)} + \frac{\vartheta(\vartheta - \theta t)(1-s)^{m-2}}{\theta^2 \Gamma(m-1)}, & t \leq s \end{cases} \quad (2.3)$$

Proof: To prove lemma (3) for some real constants $s_1, s_2 \in \mathbb{R}$, we have

$$y(t) = J^m u(t) - s_1 - s_2 t = \int_0^t \frac{(t-s)^{m-1}}{\Gamma(m)} u(s) ds - s_1 - s_2 t$$

Given the relation ${}^c D^m J^m y(t) = y(t)$ and $J^m J^n = J^{m+n} y(t)$ for $m, n > 0, y \in L(0,1)$, we obtain

$$y'(t) = \int_0^t \frac{(t-s)^{m-2}}{\Gamma(m-1)} u(s) ds - s_2$$

Applying the boundary conditions for (2.1), we get the following

$$s_1 = \frac{1}{\theta^2} \left[\vartheta \int_0^1 \gamma_2(s) ds - (\vartheta + \theta) \int_0^1 \gamma_1(s) ds \right] - \frac{\vartheta}{\theta \Gamma(m)} \int_0^1 (1-s)^{m-1} u(s) ds$$

$$- \frac{\vartheta^2}{\theta^2 \Gamma(m-1)} \int_0^1 (1-s)^{m-2} u(s) ds$$

$$s_2 = \frac{1}{\theta} \left[\int_0^1 \gamma_1(s) ds - \int_0^1 \gamma_2(s) ds \right] + \frac{1}{\Gamma(m)} \int_0^1 (1-s)^{m-1} u(s) ds$$

$$+ \frac{\vartheta}{\theta \Gamma(m-1)} \int_0^1 (1-s)^{m-2} u(s) ds$$

The unique solution of (2.1) is therefore follows as

$$y(t) = \int_0^t \left[\frac{\theta(t-s)^{m-1} + (\vartheta - \theta t)(1-s)^{m-1}}{\theta \Gamma(m)} + \frac{\vartheta(\vartheta - \theta t)(1-s)^{m-2}}{\theta^2 \Gamma(m-1)} \right] u(s) ds$$

$$+ \int_t^1 \left[\frac{(\vartheta - \theta t)(1-s)^{m-1}}{\theta \Gamma(m)} + \frac{\vartheta(\vartheta - \theta t)(1-s)^{m-2}}{\theta^2 \Gamma(m-1)} \right] u(s) ds$$

$$+ \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 \gamma_1(s) ds + (\vartheta + \theta t) \int_0^1 \gamma_2(s) ds \right]$$

$$= \int_0^1 G_m(t,s) u(s) ds + \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 \gamma_1(s) ds + (\vartheta + \theta t) \int_0^1 \gamma_2(s) ds \right]$$

where $G_m(t,s)$ is given by (2.3). This is the required proof.

Lemma 4. The space $\mathcal{H} = \{y(t) \mid y \in C(\mathbb{R})\}$ is a Banach space under the defined norm $\|y\|_{\mathcal{H}} = \max_{t \in \mathbb{R}} |y(t)|$. Similarly, the norm on product space is defined as $\|(y, \bar{y})\|_{\mathcal{H} \times \mathcal{H}} = \|y\|_{\mathcal{H}} + \|\bar{y}\|_{\mathcal{H}}$. Obviously $(\mathcal{H} \times \mathcal{H}, \|\cdot\|_{\mathcal{H} \times \mathcal{H}})$ is a Banach space. Moreover the cone $\mathcal{U} \subset \mathcal{H} \times \mathcal{H}$ is defined as

$$\mathcal{U} = (y, \bar{y}) \in \mathcal{H} \times \mathcal{H} \mid y(t) \geq 0, \bar{y}(t) \geq 0.$$

Here system (1.1) is transformed into a fixed point problem. Let $\mathcal{S}: \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}$ be the operator defined as

$$\mathcal{S}(x, y)(t) = \begin{pmatrix} \int_0^1 G_m(t, s) k(s, y(s), \chi^\gamma(s)) ds + \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s)) ds \right. \\ \left. + (\vartheta + \theta t) \int_0^1 b_2(x(s)) ds \right] \\ \int_0^1 G_n(t, s) l(s, x(s), \chi^x(s)) ds + \frac{1}{\bar{\theta}^2} \left[(\bar{\theta}(1-t) + \bar{\vartheta}) \int_0^1 \bar{b}_1(y(s)) ds \right. \\ \left. + (\bar{\vartheta} + \bar{\theta} t) \int_0^1 \bar{b}_2(y(s)) ds \right] \end{pmatrix} \quad (2.4)$$

It follows that the fixed point of the operator coincides with the solution of the proposed coupled system (1.1).

To prove the existence and uniqueness of solutions to system (1.1), we require the subsequent conditions

(B1) For $t \in \mathfrak{N}$, $\exists \mu_1, \mu_2, \mu_3 \in C(\mathfrak{N}, \mathbb{R}^+)$ such that

$$|k(t, y(t), \chi^\gamma(t))| \leq \mu_1(t) + \mu_2(t)|y(t)| + \mu_3(t)|\chi^\gamma(t)|, \\ \forall y(t) \in C(\mathfrak{N}, \mathbb{R})$$

with

$$\mu_1^* = \sup_{t \in \mathfrak{N}} \mu_1(t), \mu_2^* = \sup_{t \in \mathfrak{N}} \mu_2(t), \mu_3^* = \sup_{t \in \mathfrak{N}} \mu_3(t).$$

Similarly, For $t \in \mathfrak{N}$, $\exists v_1, v_2, v_3 \in C(\mathfrak{N}, \mathbb{R}^+)$ such that

$$|l(t, x(t), \chi^x(t))| \leq v_1(t) + v_2(t)|x(t)| + v_3(t)|\chi^x(t)|, \\ \forall x(t) \in C(\mathfrak{N}, \mathbb{R})$$

with

$$v_1^* = \sup_{t \in \mathfrak{N}} v_1(t), v_2^* = \sup_{t \in \mathfrak{N}} v_2(t), v_3^* = \sup_{t \in \mathfrak{N}} v_3(t).$$

(B2) For $t \in \mathfrak{N}$, $\exists L_{b_1}, L_{b_2}$ such that

$$|b_1(x(t))| \leq L_{b_1}|x(t)| \text{ and } |b_2(x(t))| \leq L_{b_2}|x(t)|, \\ \forall x(t) \in C(\mathfrak{N}, \mathbb{R}).$$

Similarly, For $t \in \mathfrak{N}$, $\exists L_{\bar{b}_1}, L_{\bar{b}_2}$ such that

$$|\bar{b}_1(y(t))| \leq L_{\bar{b}_1}|y(t)| \text{ and } |\bar{b}_2(y(t))| \leq L_{\bar{b}_2}|y(t)|, \\ \forall y(t) \in C(\mathfrak{N}, \mathbb{R}).$$

(B3) For $t \in \mathfrak{N}$, $\exists L_{\hat{\omega}_1}, L_{\hat{\omega}_2}$ such that

$$\int_0^1 |\hat{\gamma}_1(t, s)| ds \leq L_{\hat{\omega}_1} \text{ and } \int_0^1 |\hat{\gamma}_2(t, s)| ds \leq L_{\hat{\omega}_2}.$$

(B4)

$$\mathcal{M}_1 = \int_0^1 |G_m(t, s)| \mu_1(s) ds < \infty, \mathcal{N}_1 = \int_0^1 |G_m(t, s)| \mu_2(s) ds + L_{\hat{\omega}_1} \int_0^1 |G_m(t, s)| \mu_3(s) ds \\ + \frac{(\theta + \vartheta)(L_{b_1} + L_{b_2})}{\theta^2} < \frac{1}{2} \\ \mathcal{M}_2 = \int_0^1 |G_n(t, s)| \nu_1(s) ds < \infty, \mathcal{N}_2 = \int_0^1 |G_n(t, s)| \nu_2(s) ds + L_{\hat{\omega}_2} \int_0^1 |G_n(t, s)| \nu_3(s) ds \\ + \frac{(\bar{\theta} + \bar{\vartheta})(L_{\bar{b}_1} + L_{\bar{b}_2})}{\bar{\theta}^2} < \frac{1}{2}$$

(B5) For all $\bar{y} \in C(\mathfrak{N}, \mathbb{R})$ and for each $t \in \mathfrak{N}$ there exist positive real constants $\mathfrak{R}_k, \mathfrak{R}_{\bar{k}}$, such that

$$|k(t, y(t), \chi^\gamma(t)) - k(t, \bar{y}(t), \bar{\chi}^\gamma(t))| \leq \mathfrak{R}_k|y - \bar{y}| + \mathfrak{R}_{\bar{k}}|\chi^\gamma - \bar{\chi}^\gamma|$$

Similarly, $\forall x, \bar{x} \in C(\mathfrak{N}, \mathbb{R})$ and for each $t \in \mathfrak{N}$ there exist positive constants $\mathfrak{R}_l, \mathfrak{R}_{\bar{l}}$, such that

$$|l(t, x(t), \chi^x(t)) - k(t, \bar{x}(t), \bar{\chi}^x(t))| \leq \mathfrak{R}_l|x - \bar{x}| + \mathfrak{R}_{\bar{l}}|\chi^x - \bar{\chi}^x|$$

(B6) For all $\bar{x} \in C(\mathfrak{N}, \mathbb{R})$ and for each $t \in \mathfrak{N}$ there exist positive real constants $\mathfrak{R}_{b_1}, \mathfrak{R}_{b_2}$, such that

$$|b_1(x) - b_1(\bar{x})| \leq \mathfrak{R}_{b_1}|x - \bar{x}| \text{ and } |b_2(x) - b_2(\bar{x})| \leq \mathfrak{R}_{b_2}|x - \bar{x}|$$

Similarly, For all $y, \bar{y} \in C(\mathfrak{N}, \mathbb{R})$ and for each $t \in \mathfrak{N}$ there exist positive real constants $\mathfrak{R}_{\bar{b}_1}, \mathfrak{R}_{\bar{b}_2}$, such that

$$|\bar{b}_1(y) - \bar{b}_1(\bar{y})| \leq \mathfrak{R}_{\bar{b}_1}|y - \bar{y}| \text{ and } |\bar{b}_2(y) - \bar{b}_2(\bar{y})| \leq \mathfrak{R}_{\bar{b}_2}|y - \bar{y}|$$

(B7) Let

$$(i). \sigma_1 = \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} \mathfrak{R}_l + \mathfrak{R}_{\theta\vartheta}, \text{ where } \mathfrak{R}_{\theta\vartheta} = \frac{(\theta + \vartheta)(\mathfrak{R}_{b_1} + \mathfrak{R}_{b_2})}{\theta^2}$$

and

$$\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} = \max_{t \in [0,1]} \left| \int_0^1 G_n(t, s) ds \right| \\ = \left| \int_0^t \left[\frac{\bar{\theta}(t-s)^{(n-1)} + (\bar{\vartheta} - \bar{\theta})(1-s)^{n-1}}{\bar{\theta}\Gamma(n)} + \frac{\bar{\vartheta}(\bar{\vartheta} - \bar{\theta}t)(1-s)^{n-2}}{\bar{\theta}^2\Gamma(n-1)} \right] ds \right| \\ + \left| \int_t^1 \left[\frac{(\bar{\vartheta} - \bar{\theta}t)(1-s)^{(n-1)}}{\bar{\theta}\Gamma(n)} + \frac{\bar{\theta}(\bar{\vartheta} - \bar{\theta}t)(1-s)^{n-2}}{\bar{\theta}^2\Gamma(n)} \right] ds \right| \\ = \frac{1}{\Gamma(n+1)} + \frac{2(\bar{\vartheta} + \bar{\theta})}{\Gamma(n+1)} + \frac{2(\bar{\vartheta}^2 + \bar{\theta}\bar{\vartheta})}{\bar{\theta}^2\Gamma(n)}$$

$$(ii). \sigma_2 = \mathcal{K}_{(m\theta\vartheta)} \mathfrak{R}_k + \mathfrak{R}_{\bar{\theta}\bar{\vartheta}}, \text{ where } \mathfrak{R}_{\bar{\theta}\bar{\vartheta}} = \frac{(\bar{\theta} + \bar{\vartheta})(\mathfrak{R}_{\bar{b}_1} + \mathfrak{R}_{\bar{b}_2})}{\bar{\theta}^2}$$

and

$$\mathcal{K}_{(m\theta\vartheta)} = \max_{t \in [0,1]} \left| \int_0^1 G_m(t, s) ds \right| \\ = \frac{1}{\Gamma(m+1)} + \frac{2(\theta + \vartheta)}{\Gamma(m+1)} + \frac{2(\vartheta^2 + \theta\vartheta)}{\theta^2\Gamma(m)}$$

EXISTENCE AND UNIQUENESS ANALYSIS

Theorem 3.1. If all of the prerequisites (B1) - (B7) are satisfied and $\sigma = \max(\sigma_1, \sigma_2) < 1$, so the suggested fractional order paired framework (1.1) has a unique solution.

Proof: For a positive number

$$\rho = \max\left(\frac{2\mathcal{M}_1}{1-2\mathcal{N}_1}, \frac{2\mathcal{M}_2}{1-2\mathcal{N}_2}\right),$$

we define a set

$$\mathcal{V} = \{(x, y) \in \mathcal{H} \times \mathcal{H} : \|(x, y)\|_{\mathcal{H} \times \mathcal{H}} \leq \rho\}.$$

First, to establish that $\mathcal{S}$ maps $\mathcal{V}$ into itself, we have

$$\begin{aligned} |\mathcal{S}_m(n, m)(t)| &\leq \int_0^1 |G_m(t, s)| |k(s, y(s), \chi^y(s))| ds \\ &\quad + \frac{1}{\alpha^2} \left[|(\theta(1-t) + \vartheta)| \int_0^1 |b_1(x(s))| ds + |(\theta + \vartheta t)| \int_0^1 |b_2(x(s))| ds \right] \\ &\leq \int_0^1 |G_m(t, s)| \mu_1(s) ds + \int_0^1 |G_m(t, s)| [\mu_2(s) |y(s)|] ds + \int_0^1 |G_m(t, s)| [\mu_3(s) |\chi^y(s)|] ds \\ &\quad + \frac{1}{\theta^2} \left[(\theta + \vartheta) \int_0^1 |b_1(x(s))| ds + (\theta + \vartheta) \int_0^1 |b_2(x(s))| ds \right] \\ &\leq \int_0^1 |G_m(t, s)| \mu_1(s) ds + \rho \left[\int_0^1 |G_m(t, s)| \mu_2(s) ds + L_{\omega_1} \int_0^1 |G_m(t, s)| \mu_3(s) ds \right]. \end{aligned}$$

Upon applying the maximization operation to both sides of the aforementioned inequality over the set $\mathcal{T}$, we derive the following expression:

$$\|\mathcal{S}_m(n, m)\|_{\mathcal{H}} \leq \frac{\rho}{2}.$$

Similarly,

$$\|\mathcal{S}_n(m, n)\|_{\mathcal{H}} \leq \frac{\rho}{2}.$$

Hence, we can conclude that,

$$\|\mathcal{S}(m, n)\|_{\mathcal{H} \times \mathcal{H}} \leq \rho. \quad (3.2)$$

The aforementioned inequality serves as confirmation that the mapping $\mathcal{S}$ effectively maps the set $\mathcal{V}$ into itself. Subsequently, to demonstrate that $\mathcal{S}$ qualifies as the contraction operator under the condition where $t \in \mathfrak{N}$, we proceed with the following analysis.

$$\begin{aligned} |\mathcal{S}_m(n, m)(t) - \mathcal{S}_m(\bar{y}, \bar{x})(t)| &\leq \int_0^1 |G_m(t, s)| |k(s, y(s), \chi^y(s)) - k(s, \bar{y}(s), \bar{\chi}^y(s))| ds \\ &\quad + \frac{1}{\theta^2} \left[|(\theta(1-t) + \vartheta)| \int_0^1 |b_1(x(s)) - b_1(\bar{x}(s))| ds \right. \\ &\quad \leq \mathcal{K}_{(m\theta\vartheta)} \mathfrak{R}_k |y(t) - \bar{y}(t)| \\ &\quad \left. + \frac{(\theta + \vartheta)(\mathfrak{R}_{b_1} + \mathfrak{R}_{b_2})}{\theta^2} |x(t) - \bar{x}(t)| \right]. \end{aligned}$$

When we operate the maximum on both sides of the above-mentioned inequality over $\mathfrak{N}$, we obtain

$$\|\mathcal{S}_m(y, x) - \mathcal{S}_m(\bar{y}, \bar{x})\|_{\mathcal{H}} \leq \mathcal{K}_{(m\theta\vartheta)} \mathfrak{R}_k \|y - \bar{y}\|_{\mathcal{H}} + \frac{(\theta + \vartheta)(\mathfrak{R}_{b_1} + \mathfrak{R}_{b_2})}{\theta^2} \|x - \bar{x}\|_{\mathcal{H}}. \quad (3.4)$$

Similarly, the following can be obtained

$$\|\mathcal{S}_n(x, y) - \mathcal{S}_n(\bar{x}, \bar{y})\|_{\mathcal{H}} \leq \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} \mathfrak{R}_l \|x - \bar{x}\|_{\mathcal{H}} + \frac{(\bar{\theta} + \bar{\vartheta})(\mathfrak{R}_{\bar{b}_1} + \mathfrak{R}_{\bar{b}_2})}{\bar{\theta}^2} \|y - \bar{y}\|_{\mathcal{H}}. \quad (3.5)$$

From the above two equations, we obtain

$$\|\mathcal{S}(x, y) - \mathcal{S}(\bar{x}, \bar{y})\|_{\mathcal{H} \times \mathcal{H}} \leq \sigma \|(x, y) - (\bar{x}, \bar{y})\|_{\mathcal{H} \times \mathcal{H}}. \quad (3.6)$$

Henceforth, it is evident that the operator $\mathcal{S}$ exhibits strict contraction properties. Employing Banach's fixed-point method, specifically designed for establishing a unique fixed point, becomes instrumental in determining the sole solution to the proposed coupled system (1.1). Consequently, the fulfillment of our primary objective is achieved.

ULAM'S STABILITY ANALYSIS

In this section, we endeavor to investigate the Ulam-type stability characteristics inherent in the envisaged coupled system denoted as (1.1).

For some real number $\varepsilon = \max(\varepsilon_m, \varepsilon_n) > 0$. We assume the inequality given as follows

$$\begin{cases} |{}^c D^m x(t) - k(t, y(t), \chi^y(t))| \leq \varepsilon_m, & t \in \mathfrak{N} \\ |{}^c D^n y(t) - l(t, x(t), \chi^x(t))| \leq \varepsilon_n, & t \in \mathfrak{N}. \end{cases} \quad (4.1)$$

The following definitions are inspired by Rus [19].

Definition 3: The Ulam-Hyers stability of the coupled system denoted by (1.1) is established under the condition that there exists a positive constant $\mathcal{K}_{(mn\theta\vartheta\bar{\theta}\bar{\vartheta})}$ such that, for any solution $(x, y) \in \mathcal{H} \times \mathcal{H}$ of the inequality (4.1), there exists a unique solution $(\zeta, \eta) \in \mathcal{H} \times \mathcal{H}$. This unique solution satisfies the inequality:

$$|(x, y)(t) - (\zeta, \eta)(t)| \leq \mathcal{K}_{(mn\theta\vartheta\bar{\theta}\bar{\vartheta})} \varepsilon, \quad t \in \mathfrak{N}$$

Definition 4: The generalized Ulam's-Hyers stability of the coupled system denoted by (1.1) is posited under the condition that there exists a function $\varpi(\varepsilon) \in C(\mathbb{R}^+, \mathbb{R}^+)$ with $\varpi(0) = 0$. This function is required to satisfy the condition that, for any solution $(x, y) \in \mathcal{H} \times \mathcal{H}$ of the inequality (4.1), there exists a unique solution $(\zeta, \eta) \in \mathcal{H} \times \mathcal{H}$ for the system. Additionally, the unique solution must adhere to the inequality:

$$|(x, y)(t) - (\zeta, \eta)(t)| \leq \varpi(\varepsilon), \quad t \in \mathfrak{N}$$

Remark 1. Consider the pair $(x, y) \in \mathcal{H} \times \mathcal{H}$ to be a solution of inequality (4.1) if there exist functions ϕ and ψ belonging to the space $C(\mathfrak{N}, \mathbb{R})$, contingent upon the variables x and y respectively, such that

$$(R1). |\phi(t)| \leq \varepsilon_m \text{ and } |\psi(t)| \leq \varepsilon_n, \quad t \in \mathfrak{N};$$

$$(R2). \text{ and}$$

$$\begin{cases} {}^c D^m x(t) = k(t, y(t), \chi^y(t)) + \phi(t), & t \in \mathfrak{N} \\ {}^c D^n y(t) = l(t, x(t), \chi^x(t)) + \psi(t), & t \in \mathfrak{N}. \end{cases} \quad (4.2)$$

Lemma 5. Let $(x, y) \in \mathcal{H} \times \mathcal{H}$ be deemed as a solution to inequality (4.1). Subsequently, the ensuing inequality is guaranteed to be satisfied:

$$\begin{cases} \left| x(t) - \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds \right. \right. \\ \quad \left. \left. + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds \right] \right| \leq \mathcal{K}_{m\theta\vartheta} \varepsilon_m, t \in \mathbb{N} \\ \left| y(t) - \int_0^1 G_n(t, s)l(s, x(s), \chi^x(s))ds - \frac{1}{\bar{\theta}^2} \left[(\bar{\theta}(1-t) + \bar{\vartheta}) \int_0^1 \bar{b}_1(x(s))ds \right. \right. \\ \quad \left. \left. + (\bar{\vartheta} + \bar{\theta} t) \int_0^1 \bar{b}_2(x(s))ds \right] \right| \leq \mathcal{K}_{n\bar{\theta}\bar{\vartheta}} \varepsilon_n, t \in \mathbb{N}. \end{cases}$$

Proof. By Remark 1 (R2), we have

$$\begin{cases} {}^c D^m x(t) = k(t, y(t), \chi^y(t)) + \phi(t), \quad t \in \mathbb{N} \\ {}^c D^n y(t) = l(t, x(t), \chi^x(t)) + \psi(t), \quad t \in \mathbb{N} \\ \theta x(0) + \vartheta x'(0) = \int_0^1 b_1(x(s))ds, \theta x(1) + \vartheta x'(1) = \int_0^1 b_2(x(s))ds \\ \bar{\theta} y(0) + \bar{\vartheta} y'(0) = \int_0^1 \bar{b}_1(y(s))ds, \bar{\theta} y(1) + \bar{\vartheta} y'(1) = \int_0^1 \bar{b}_2(y(s))ds. \end{cases} \quad (4.3)$$

By applying Lemma 3, the solution of (4.3) is given as follows:

$$\begin{cases} x(t) = \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds + \int_0^1 G_m(t, s)\phi(s)ds + \frac{1}{\theta^2} [(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds \\ \quad + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds], t \in \mathbb{N} \\ y(t) = \int_0^1 G_n(t, s)l(s, x(s), \chi^x(s))ds + \int_0^1 G_n(t, s)\psi(s)ds + \frac{1}{\bar{\theta}^2} [(\bar{\theta}(1-t) + \bar{\vartheta}) \int_0^1 \bar{b}_1(x(s))ds \\ \quad + (\bar{\vartheta} + \bar{\theta} t) \int_0^1 \bar{b}_2(x(s))ds], t \in \mathbb{N}. \end{cases} \quad (4.4)$$

Considering the first equation of system (4.4), we have

$$\begin{aligned} & \left| x(t) - \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds \right. \right. \\ & \quad \left. \left. + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds \right] \right| \leq \left| \int_0^1 G_\theta(t, s)\phi(s)ds \right| \\ & \leq \int_0^1 |G_\theta(t, s)\phi(s)|ds. \end{aligned}$$

By Remark 1 (R1) and using the condition of (B6), we get

$$\left| x(t) - \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds \right. \right. \\ \left. \left. + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds \right] \right| \leq \mathcal{K}_{(m\theta\vartheta)} \varepsilon_m. \quad (4.5)$$

By applying the same process for the second equation of system (4.4), we get

$$\left| y(t) - \int_0^1 G_n(t, s)l(s, x(s), \chi^x(s))ds - \frac{1}{\bar{\theta}^2} \left[(\bar{\theta}(1-t) + \bar{\vartheta}) \int_0^1 \bar{b}_1(x(s))ds \right. \right. \\ \left. \left. + (\bar{\vartheta} + \bar{\theta} t) \int_0^1 \bar{b}_2(x(s))ds \right] \right| \leq \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} \varepsilon_n. \quad (4.6)$$

Theorem 2. Assuming that conditions (B5) to (B7) are satisfied, the fractional order coupled system denoted by equation (1.1) can be deemed Ulam's-Hyers stable. This

assertion extends to the broader context of being generalized Ulam 's-Hyers stable.

$$(1 - \mathfrak{R}_{\theta\vartheta})(1 - \mathfrak{R}_{\bar{\theta}\bar{\vartheta}}) - \mathfrak{R}_{m\theta\vartheta}\mathfrak{R}_{n\bar{\theta}\bar{\vartheta}} \neq 0$$

Proof: Let $(x, y) \in \mathcal{H} \times \mathcal{H}$ be the solution of the system (12) and $(\zeta, \eta) \in \mathcal{H} \times \mathcal{H}$ be the unique solution to the following considered system:

$$\begin{cases} {}^c D^m \zeta(t) = k(t, \eta(t), \chi^\eta(t)), t \in \mathbb{N} \\ {}^c D^n \eta(t) = l(t, \zeta(t), \chi^\zeta(t)), t \in \mathbb{N} \\ \theta \zeta(0) + \vartheta \zeta'(0) = \int_0^1 b_1(\zeta(s))ds, \theta \zeta(1) + \vartheta \zeta'(1) = \int_0^1 b_2(\zeta(s))ds \\ \bar{\theta} \eta(0) + \bar{\vartheta} \eta'(0) = \int_0^1 \bar{b}_1(\eta(s))ds, \bar{\theta} \eta(1) + \bar{\vartheta} \eta'(1) = \int_0^1 \bar{b}_2(\eta(s))ds. \end{cases}$$

Using Lemma 3, the solution of the above system is

$$\begin{cases} \zeta(t) = \int_0^1 G_m(t, s)k(s, \eta(s), \chi^\eta(s))ds + \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(\zeta(s))ds \right. \\ \quad \left. + (\vartheta + \theta t) \int_0^1 b_2(\zeta(s))ds \right] \\ \eta(t) = \int_0^1 G_n(t, s)l(s, \zeta(s), \chi^\zeta(s))ds + \frac{1}{\bar{\theta}^2} \left[(\bar{\theta}(1-t) + \bar{\vartheta}) \int_0^1 \bar{b}_1(\eta(s))ds \right. \\ \quad \left. + (\bar{\vartheta} + \bar{\theta} t) \int_0^1 \bar{b}_2(\eta(s))ds \right], t \in \mathbb{N}. \end{cases}$$

We have

$$\begin{aligned} |x(t) - \zeta(t)| &= \left| x(t) - \int_0^1 G_m(t, s)k(s, \eta(s), \chi^\eta(s))ds - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(\zeta(s))ds \right. \right. \\ & \quad \left. \left. + (\vartheta + \theta t) \int_0^1 b_2(\zeta(s))ds \right] \right| \\ &\leq \left| x(t) - \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds \right. \right. \\ & \quad \left. \left. + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds \right] \right| \\ & \quad + \left| \int_0^1 G_m(t, s)k(s, y(s), \chi^y(s))ds - \int_0^1 G_m(t, s)k(s, \eta(s), \chi^\eta(s))ds \right| \\ & \quad + \left| \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(x(s))ds + (\vartheta + \theta t) \int_0^1 b_2(x(s))ds \right] \right. \\ & \quad \left. - \frac{1}{\theta^2} \left[(\theta(1-t) + \vartheta) \int_0^1 b_1(\zeta(s))ds + (\vartheta + \theta t) \int_0^1 b_2(\zeta(s))ds \right] \right| \\ &\leq \mathcal{K}_{m\theta\vartheta} \varepsilon_m + \mathfrak{R}_{m\theta\vartheta} |y(t) - \eta(t)| + \mathfrak{R}_{\theta\vartheta} |x(t) - \zeta(t)|, \end{aligned}$$

where $\mathfrak{R}_{m\theta\vartheta} = \mathcal{K}_{m\theta\vartheta} \mathfrak{R}_k$.

Hence, we get

$$(1 - \mathcal{K}_{\theta\vartheta}) \|x - \zeta\|_{\mathcal{H}} \leq \mathcal{K}_{(m\theta\vartheta)} \varepsilon_m + \mathfrak{R}_{m\theta\vartheta} \|y - \eta\|_{\mathcal{H}}. \quad (4.9)$$

Similarly, we have

$$(1 - \mathfrak{R}_{\bar{\theta}\bar{\vartheta}}) \|y - \eta\|_{\mathcal{H}} \leq \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} \varepsilon_n + \mathfrak{R}_{n\bar{\theta}\bar{\vartheta}} \|x - \zeta\|_{\mathcal{H}}, \quad (4.10)$$

where $\mathfrak{R}_{n\bar{\theta}\bar{\vartheta}} = \mathcal{K}_{n\bar{\theta}\bar{\vartheta}} \mathfrak{R}_l$.

From (4.9) and (4.10), it can be written as

$$\begin{cases} (1 - \mathfrak{R}_{\theta\vartheta}) \|x - \zeta\|_{\mathcal{H}} - \mathfrak{R}_{m\theta\vartheta} \|y - \eta\|_{\mathcal{H}} \leq \mathcal{K}_{(m\theta\vartheta)} \varepsilon_m \\ (1 - \mathfrak{R}_{\bar{\theta}\bar{\vartheta}}) \|y - \eta\|_{\mathcal{H}} - \mathfrak{R}_{n\bar{\theta}\bar{\vartheta}} \|x - \zeta\|_{\mathcal{H}} \leq \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})} \varepsilon_n. \end{cases} \quad (4.11)$$

The matrix representation of the above equation is as follows

$$\begin{pmatrix} (1 - \Re_{\theta\vartheta}) & -\Re_{m\theta\vartheta} \\ -\Re_{n\bar{\theta}\bar{\vartheta}} & (1 - \Re_{\bar{\theta}\bar{\vartheta}}) \end{pmatrix} \begin{pmatrix} \|x - \zeta\|_{\mathcal{H}} \\ \|y - \eta\|_{\mathcal{H}} \end{pmatrix} \leq \begin{pmatrix} \mathcal{K}_{(m\theta\vartheta)}\varepsilon_m \\ \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n \end{pmatrix}.$$

Simplifying the above inequality yields

$$\begin{pmatrix} \|x - \zeta\|_{\mathcal{H}} \\ \|y - \eta\|_{\mathcal{H}} \end{pmatrix} \leq \begin{pmatrix} \frac{(1 - \Re_{\bar{\theta}\bar{\vartheta}})}{\Delta} & \frac{\Re_{m\theta\vartheta}}{\Delta} \\ \frac{\Re_{n\bar{\theta}\bar{\vartheta}}}{\Delta} & \frac{(1 - \Re_{\theta\vartheta})}{\Delta} \end{pmatrix} \begin{pmatrix} \mathcal{K}_{(m\theta\vartheta)}\varepsilon_m \\ \mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n \end{pmatrix},$$

where $\Delta = (1 - \Re_{\theta\vartheta})(1 - \Re_{\bar{\theta}\bar{\vartheta}}) - \Re_{m\theta\vartheta}\Re_{n\bar{\theta}\bar{\vartheta}} \neq 0$. Further simplification gives

$$\|x - \zeta\|_{\mathcal{H}} \leq \frac{(1 - \Re_{\bar{\theta}\bar{\vartheta}})\mathcal{K}_{(m\theta\vartheta)}\varepsilon_m}{\Delta} + \frac{\Re_{m\theta\vartheta}\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n}{\Delta} \quad (4.12)$$

$$\|y - \eta\|_{\mathcal{H}} \leq \frac{\Re_{n\bar{\theta}\bar{\vartheta}}\mathcal{K}_{(m\theta\vartheta)}\varepsilon_m}{\Delta} + \frac{(1 - \Re_{\theta\vartheta})\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n}{\Delta}, \quad (4.13)$$

From equations (4.12) and (4.13), we have

$$\begin{aligned} \|x - \zeta\|_{\mathcal{H}} + \|y - \eta\|_{\mathcal{H}} &\leq \frac{(1 - \Re_{\bar{\theta}\bar{\vartheta}})\mathcal{K}_{(m\theta\vartheta)}\varepsilon_m}{\Delta} + \frac{\Re_{m\theta\vartheta}\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n}{\Delta} \\ &\quad + \frac{\Re_{n\bar{\theta}\bar{\vartheta}}\mathcal{K}_{(m\theta\vartheta)}\varepsilon_m}{\Delta} + \frac{(1 - \Re_{\theta\vartheta})\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}\varepsilon_n}{\Delta} \end{aligned}$$

Therefore, we have

$$\|(x, y) - (\zeta, \eta)\|_{\mathcal{H} \times \mathcal{H}} \leq \mathcal{K}_{(mn\theta\vartheta\bar{\theta}\bar{\vartheta})}\varepsilon, \quad (4.14)$$

where $\varepsilon = \max(\varepsilon_m, \varepsilon_n)$ and

$$\begin{aligned} \mathcal{K}_{(mn\theta\vartheta\bar{\theta}\bar{\vartheta})} &= \frac{(1 - \Re_{\bar{\theta}\bar{\vartheta}})\mathcal{K}_{(m\theta\vartheta)}}{\Delta} + \frac{\Re_{m\theta\vartheta}\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}}{\Delta} \\ &\quad + \frac{\Re_{n\bar{\theta}\bar{\vartheta}}\mathcal{K}_{(m\theta\vartheta)}}{\Delta} + \frac{(1 - \Re_{\theta\vartheta})\mathcal{K}_{(n\bar{\theta}\bar{\vartheta})}}{\Delta} \end{aligned}$$

Therefore, based on the preceding inequality, it can be deduced that the coupled system denoted as (1.1) exhibits Ulam 's-Hyers stability. Additionally, the aforementioned inequality can be expressed in the form:

$$\|(x, y) - (\zeta, \eta)\|_{\mathcal{H} \times \mathcal{H}} \leq \varpi(\varepsilon), \text{ where } \varpi(0) = 0.$$

Through this inequality, it becomes evident that the proposed coupled system (1.1) adheres to the principles of generalized Ulam 's-Hyers stability.

APPLICATION

Example. Here, we validate our findings related to existence, uniqueness, and stability by presenting a concrete example. Let the following fractional order connected coupled system as follows:

$$\begin{cases} {}^c D^{\frac{1}{2}} x(t) = \frac{1}{(t+)^2} \frac{|y(t)|}{1 + |y(t)|} + \int_0^t \frac{e^{-(s-t)}}{49} y(s) ds, m \in (0,1], t \in \mathbb{R} \\ {}^c D^{\frac{1}{2}} y(t) = \frac{1}{100} [t \cos x(t) + x(t) \sin t] + \int_0^t \frac{7}{s^2 + 7^2} x(s) ds, n \in (0,1], t \in \mathbb{R} \\ x(0) + x'(0) = \int_0^1 \frac{|x(s)|}{11 + |x(s)|} ds, x(1) + x'(1) = \int_0^1 \frac{|x(s)|}{13 + |x(s)|} ds \\ y(0) + y'(0) = \int_0^1 \frac{1}{22} [\cos y(s) + \sin y(s)] ds \\ y(1) + y'(1) = \int_0^1 \frac{1}{26} [\cos y(s) + \sin y(s)] ds. \end{cases}$$

By comparing coupled system (5.1) with system (1.1), the following values are derived:

$$\begin{aligned} m = n = \frac{1}{2}, \theta = \vartheta = \bar{\theta} = \bar{\vartheta} = 1, \Re_{b_1} = \Re_{\bar{b}_1} = \frac{1}{11} \\ \text{and } \Re_{b_2} = \Re_{\bar{b}_2} = \frac{1}{13}. \end{aligned}$$

Also,

$$k(t, y(t), \chi^y(t)) = \frac{1}{(t+7)^2} \frac{|y(t)|}{1 + |y(t)|} + \int_0^t \frac{e^{-(s-t)}}{49} y(s) ds$$

and

$$l(t, x(t), \chi^x(t)) = \frac{1}{100} [t \cos x(t) + x(t) \sin t] + \int_0^t \frac{7}{s^2 + 7^2} x(s) ds,$$

whereas $|k(t, y, \chi^y) - k(t, \bar{y}, \bar{\chi}^y)| \leq \left(\frac{1}{49}\right) \|y - \bar{y}\| + \|\chi^y - \bar{\chi}^y\|$ and $|l(t, x, \chi^x) - l(t, \bar{x}, \bar{\chi}^x)| \leq \frac{1}{50} \|x - \bar{x}\| + \|\chi^x - \bar{\chi}^x\|$.

Therefore, (B4) is satisfied with $\Re_k, \bar{\Re}_k = \frac{1}{49}$ and $\Re_l, \bar{\Re}_l = \frac{1}{50}$. Further, we have

$$\begin{aligned} \sigma_1 &= \left(\frac{1}{\Gamma(n+1)} + \frac{2(\bar{\theta} + \bar{\vartheta})}{\Gamma(n+1)} + \frac{2(\bar{\theta}^2 + \bar{\theta}\bar{\vartheta})}{\bar{\theta}^2 \Gamma(n)} \right) \Re_l + \frac{(\theta + \vartheta)(\Re_{b_1} + \Re_{b_2})}{\theta^2} \\ &= \left(\frac{1}{\Gamma(n+1)} + \frac{4}{\Gamma(n+1)} + \frac{4}{\Gamma(n)} \right) \frac{1}{50} + 2 \left(\frac{1}{11} + \frac{1}{13} \right) \\ &= \left(\frac{5}{\Gamma(n+1)} + \frac{4}{\Gamma(n)} \right) \frac{1}{50} + \left(\frac{48}{143} \right) = \left(\frac{5}{\Gamma(\frac{1}{2}+1)} + \frac{4}{\Gamma(\frac{1}{2})} \right) \frac{1}{50} + \left(\frac{48}{143} \right) \\ &= \left(\frac{10}{\sqrt{\pi}} + \frac{4}{\sqrt{\pi}} \right) \frac{1}{50} + \left(\frac{48}{143} \right) = \left(\frac{14}{50\sqrt{\pi}} \right) + \left(\frac{48}{143} \right) < 1. \end{aligned}$$

Similarly,

$$\begin{aligned} \sigma_2 &= \left(\frac{1}{\Gamma(m+1)} + \frac{2(\theta + \vartheta)}{\Gamma(m+1)} + \frac{2(\theta^2 + \theta\vartheta)}{\theta^2 \Gamma(m)} \right) \Re_k + \frac{(\bar{\theta} + \bar{\vartheta})(\Re_{\bar{b}_1} + \Re_{\bar{b}_2})}{\bar{\theta}^2} \\ &= \left(\frac{1}{\Gamma(m+1)} + \frac{4}{\Gamma(m+1)} + \frac{4}{\Gamma(m)} \right) \frac{1}{49} + 2 \left(\frac{1}{11} + \frac{1}{13} \right) \\ &= \left(\frac{5}{\Gamma(m+1)} + \frac{4}{\Gamma(m)} \right) \frac{1}{49} + \left(\frac{48}{143} \right) = \left(\frac{5}{\Gamma(\frac{1}{2}+1)} + \frac{4}{\Gamma(\frac{1}{2})} \right) \frac{1}{49} + \left(\frac{48}{143} \right) \\ &= \left(\frac{10}{\sqrt{\pi}} + \frac{4}{\sqrt{\pi}} \right) \frac{1}{49} + \left(\frac{48}{143} \right) = \left(\frac{14}{49\sqrt{\pi}} \right) + \left(\frac{48}{143} \right) < 1. \end{aligned}$$

Consequently, it can be asserted that the coupled system denoted as (5.1) possesses a singular solution. Furthermore, the condition stipulated in Theorem 2, expressed as $(1 - \Re_{\theta\vartheta})(1 - \Re_{\bar{\theta}\bar{\vartheta}}) - \Re_{m\theta\vartheta}\Re_{n\bar{\theta}\bar{\vartheta}} \neq 0$, is also satisfied.

Consequently, the coupled system (5.1) demonstrates both Ulam-Hyers stability and generalized Ulam-Hyers stability.

CONCLUSION

We investigate Ulam-Hyers and generalized Ulam-Hyers stability for fractional integro-differential coupled systems with integral boundary conditions, an area where existing research has focused mainly on simpler, isolated equations rather than the coupled systems that appear more frequently in practice. Our approach differs by analyzing how these coupled systems behave when subjected to small perturbations, using fixed-point theorems and carefully constructed function spaces to establish both solution existence and uniqueness.

What makes this work particularly valuable is its relevance to real applications in engineering, finance, and physics, where model stability directly impacts whether computational results can be trusted for decision-making. Our quantitative analysis shows that when initial conditions vary slightly, the resulting solution changes remain bounded and predictable, exactly the kind of stability guarantee practitioners need. Beyond these immediate results, the mathematical framework we've developed opens doors to studying more complex systems, including those with neutral delays, impulsive effects, or various inclusion properties that arise when modeling systems with both finite and infinite time delays. Rather than simply filling theoretical gaps, this work provides concrete tools that researchers can apply when analyzing stability in increasingly sophisticated differential systems.

ACKNOWLEDGMENTS

The authors thank the reviewers for their valuable comments and suggestions that improved this manuscript.

All authors have read and approved the final manuscript.

AUTHORSHIP CONTRIBUTIONS

Brijendra Kumar Chaurasiya: Investigation, Methodology, Validation, Writing-original draft, Writing-review and editing. Avadhesh Kumar: Conceptualization, Investigation, Methodology, Supervision, Validation, Writing-original draft, 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] Chalishajar D, Kumar A. Existence, uniqueness and Ulam's stability of solutions for a coupled system of fractional differential equations with integral boundary conditions. *Mathematics* 2018;6:96. [\[CrossRef\]](#)
- [2] Kumar A, Kumar A, Vats RK, Kumar P. Approximate controllability of neutral delay integrodifferential inclusion of order $\alpha \in (1,2)$ with non-instantaneous impulses. *Evol Equ Control Theory* 2022;11:1635–1654. [\[CrossRef\]](#)
- [3] Yang Q, Chen D. Fractional calculus in image processing: a review. *Fract Calc Appl Anal* 2016;19:1222–1249. [\[CrossRef\]](#)
- [4] Magin RL. *Fractional calculus in bioengineering*. Redding (CA): Begell House; 2006.
- [5] Kilbas AA, Srivastava HM, Trujillo JJ. *Theory and applications of fractional differential equations*. Vol. 204. New York: Elsevier Science; 2006.
- [6] Miller KS, Ross B. *An introduction to fractional calculus and fractional differential equations*. New York: Wiley Interscience; 1993.
- [7] Hilfer R. *Applications of fractional calculus in physics*. Singapore: World Scientific; 2000. [\[CrossRef\]](#)
- [8] Khan A, Shah K, Li Y, Khan TS. Ulam type stability for a coupled system of boundary value problems of nonlinear fractional differential equations. *J Funct Spaces* 2017;2017:1–8. [\[CrossRef\]](#)
- [9] Ahmad B, Nieto JJ. Existence results for nonlinear boundary value problems of fractional integro-differential equations with integral boundary conditions. *Bound Value Probl* 2009;2009:1–11. [\[CrossRef\]](#)
- [10] Jafari H, Mehdinejadiani B, Baleanu D. Fractional calculus for modeling unconfined groundwater: applications in engineering, life and social sciences. *Mathematics* 2019;7:119–138. [\[CrossRef\]](#)
- [11] Oldham KB. Fractional differential equations in electrochemistry. *Adv Eng Softw* 2010;41:9–12. [\[CrossRef\]](#)
- [12] Jannelli A. Numerical solutions of fractional differential equations arising in engineering sciences. *Mathematics* 2020;8:215. [\[CrossRef\]](#)
- [13] Ionescu C, Lopes A, Copot D, Machado JT, Bates JH. The role of fractional calculus in modeling biological phenomena: a review. *Commun Nonlinear Sci Numer Simul* 2017;51:141–159. [\[CrossRef\]](#)

- [14] Ulam SM. A collection of the mathematical problems. New York: Interscience; 1960.
- [15] Hyers DH. On the stability of the linear functional equation. *Proc Natl Acad Sci U S A* 1941;27:222–224. [\[CrossRef\]](#)
- [16] Jung SM. Hyers-Ulam stability of linear differential equations of the first order, II. *Appl Math Lett* 2006;19:854–858. [\[CrossRef\]](#)
- [17] Wang J, Feckan M. A general class of impulsive evolution equations. *Topol Methods Nonlinear Anal* 2015;46:915–933. [\[CrossRef\]](#)
- [18] Muslim M, Kumar A, Feckan M. Existence, uniqueness and stability of solutions to second-order non-linear differential equations with non-instantaneous impulses. *J King Saud Univ Sci* 2016;28:1–9.
- [19] Rus IA. Ulam stabilities of ordinary differential equations in a Banach space. *Carpath J Math* 2010;26:103–107.
- [20] Ahmad B, Sivasundaram S. Existence and uniqueness results for nonlinear boundary value problems of fractional differential equations with separated boundary conditions. *Commun Appl Anal* 2009;13:121. [\[CrossRef\]](#)
- [21] Verma P, Kumar M. Analytical solution with existence and uniqueness conditions of non-linear initial value multi-order fractional differential equations using Caputo derivative. *Eng Comput* 2022;38:661–678. [\[CrossRef\]](#)
- [22] Van Bockstal K, Zaky MA, Hendy AS. On the existence and uniqueness of solutions to a nonlinear variable order time-fractional reaction-diffusion equation with delay. *Commun Nonlinear Sci Numer Simul* 2022;115:106755. [\[CrossRef\]](#)
- [23] Shafqat R, Niazi AUK, Jeelani MB, Alharthi NH. Existence and uniqueness of mild solution where $\alpha \in (1, 2)$ for fuzzy fractional evolution equations with uncertainty. *Fractal Fract* 2022;6:65. [\[CrossRef\]](#)
- [24] Sweis H, Shawagfeh N, Abu Arqub O. Fractional crossover delay differential equations of Mittag-Leffler kernel: existence, uniqueness, and numerical solutions using the Galerkin algorithm based on shifted Legendre polynomials. *Results Phys* 2022;41:105891. [\[CrossRef\]](#)
- [25] Younis M, Singh D. On the existence of the solution of Hammerstein integral equations and fractional differential equations. *J Appl Math Comput* 2022;68:1087–1105. [\[CrossRef\]](#)
- [26] Telli B, Soud MS, Alzabut J, Khan H. Existence and uniqueness theorems for a variable order fractional differential equation with delay. *Axioms* 2023;12:339. [\[CrossRef\]](#)
- [27] Mahdy AMS. Stability, existence, and uniqueness for solving fractional glioblastoma multiforme using a Caputo-Fabrizio derivative. *Math Methods Appl Sci* 2023;46:1–15.
- [28] Mehmood N, Abbas A, Akgül A, Abdeljawad T, Alqudah MA. Existence and stability results for coupled system of fractional differential equations involving AB-Caputo derivative. *Fractals* 2023;31:2340023. [\[CrossRef\]](#)
- [29] Zou J, Luo D, Li M. The existence and averaging principle for stochastic fractional differential equations with impulses. *Math Methods Appl Sci* 2023;46:6857–6874. [\[CrossRef\]](#)
- [30] Khan H, Alzabut J, Baleanu D, Alobaidi G, Rehman MU. Existence of solutions and a numerical scheme for a generalized hybrid class of n-coupled modified ABC-fractional differential equations with an application. *Math Methods Appl Sci* 2023;46:1–15.
- [31] Toprakseven S, Bengi R, Zeren Y. Lyapunov-type inequality for an anti-periodic fractional boundary value problem of the Riesz-Caputo derivative. *Sigma J Eng Nat Sci* 2023;41:807–811. [\[CrossRef\]](#)
- [32] Adigüzel H. A note on the asymptotic behavior of fractional differential equations. *Sigma J Eng Nat Sci* 2020;38:1061–1067.
- [33] Koşunalp HY, Gülsu M. Operational matrix for multi-order fractional differential equations with Hermite polynomials. *Sigma J Eng Nat Sci* 2024;42:1050–1057. [\[CrossRef\]](#)
- [34] Adigüzel H. A note on asymptotic behavior of fractional differential equations. *Sigma J Eng Nat Sci* 2020;38:1061–1067.
- [35] Jan R, Yüzbaşı Ş, Attaullah A, Jawad M, Jan A. Fractional derivative analysis of asthma with the effect of environmental factors. *Sigma J Eng Nat Sci* 2024;42:177–188. [\[CrossRef\]](#)
- [36] Sanatee AG, Rathour L, Mishra VN, Dewangan V. Some fixed point theorems in regular modular metric spaces and application to Caratheodory's type anti-periodic boundary value problem. *J Anal* 2023;31:619–632. [\[CrossRef\]](#)
- [37] Shahi P, Rathour L, Mishra VN. Expansive fixed point theorems for tri-simulation functions. *J Eng Exact Sci* 2022;8:14303–01e. [\[CrossRef\]](#)
- [38] Pathak VK, Mishra LN, Mishra VN, Baleanu D. On the solvability of mixed-type fractional-order non-linear functional integral equations in the Banach space $C(I)$. *Fractal Fract* 2022;6:744. [\[CrossRef\]](#)
- [39] Metwali MM, Mishra VN. On the measure of non-compactness in $L_p(\mathbb{R}^+)$ and applications to a product of s-integral equations. *Turk J Math* 2023;47:372–386. [\[CrossRef\]](#)
- [40] Bouzalmat I. Stochastic modeling of COVID-19 dynamics with emphasis on asymptomatic and isolated populations. *Proc Int Conf Math Model Mod Appl* 2024;1:1–12.
- [41] Paul SK, Mishra LN, Mishra VN, Baleanu D. An effective method for solving nonlinear integral equations involving the Riemann-Liouville fractional operator. *AIMS Math* 2023;8:17448–17469. [\[CrossRef\]](#)

- [42] Paul SK, Mishra LN, Mishra VN, Baleanu D. Analysis of mixed-type nonlinear Volterra-Fredholm integral equations involving the Erdélyi-Kober fractional operator. *J King Saud Univ Sci* 2023;35:102949. [\[CrossRef\]](#)
- [43] Bhat IA, Mishra LN, Mishra VN, Tunc C, Tunc O. Precision and efficiency of an interpolation approach to weakly singular integral equations. *Int J Numer Methods Heat Fluid Flow* 2024;34:1479–1499. [\[CrossRef\]](#)
- [44] Kumar K, Rathour L, Sharma MK, Mishra VN. Fixed point approximation for Suzuki generalized nonexpansive mapping using $B(\delta, \mu)$ condition. *Appl Math* 2024;13:215–227. [\[CrossRef\]](#)
- [45] Liu X, Zhou M, Ansari AH, Chakrabarti K, Abbas M, Rathour L. Coupled fixed point theorems with rational type contractive condition via c -class functions and inverse ck -class functions. *Symmetry* 2022;14:1663. [\[CrossRef\]](#)
- [46] Iqbal J, Mishra VN, Mir WA, Dar AH, Ishtyak M, Rathour L. Generalized resolvent operator involving $G(\cdot, \cdot)$ -co-monotone mapping for solving generalized variational inclusion problem. *Georgian Math J* 2022;29:533–542. [\[CrossRef\]](#)
- [47] Maheswari U, Ravichandran M, Anbarasan A, Rathour L, Mishra VN. Some results on coupled fixed point on complex partial b -metric space. *Ganita* 2021;71:17–27.
- [48] Deng J, Liu X, Sun Y, Rathour L. On almost type α -FZ-weak contraction in metric spaces. *Open J Appl Sci* 2022;12:528–540. [\[CrossRef\]](#)
- [49] Deng J, Liu XL, Sun Y, Rathour L. Some best proximity point results of several α - ψ interpolative proximal contractions. *Nonlinear Funct Anal Appl* 2022;27:533–551.
- [50] Mishra AK, Rathour Padmawati L, Mishra VN. Some fixed point theorem using generalized cone metric spaces with Banach algebra. *High Tech Lett* 2023;29:153–162.
- [51] Alam M, Zada A, Abdeljawad T. Stability analysis of an implicit fractional integro-differential equation via integral boundary conditions. *Alex Eng J* 2024;87:501–514. [\[CrossRef\]](#)



Research Article

In-sito co-pyrolysis of tea waste and polypropylene using microwave-assisted pyrolysis: Products analysis, energy consumption, and synergy effect

Husam Talib HAMZAH^{1,*}, Sridevi VELURU¹, Suhair Abdulhadi MAHDI²,
Thamer Adnan ABDULLAH³, M. TUKARAMBAI¹, Venkata Rao POIBA¹

¹Department of Chemical Engineering, College of Engineering, Andhra University, Visakhapatnam, 530003, India

²Department of Chemistry, College of Science, Mustansiriyah University, Baghdad, 10052, Iraq

³Applied Science Department, University of Technology, Baghdad, 10066, Iraq

ARTICLE INFO

Article history

Received: 05 May 2024

Revised: 02 July 2024

Accepted: 10 October 2024

Keywords:

Microwave Co-Pyrolysis;
Pyrolysis Index; Synergy Effect;
Polypropylene; Tea Waste

ABSTRACT

This study focuses on the microwave-assisted co-pyrolysis method, which recycles polypropylene (PP) and tea residues (TW). When utilising a KOH catalyst for co-pyrolyzing (PP: TW), the gaseous product yield dropped to 33.6%, but the oil and char yields rose to 34.8% and 31.6%, respectively. The maximum gas yield achieved by microwave co-pyrolysis without a KOH catalyst was 41.67 weight percent. In MACP of (PP: TW), 459 kJ of energy was consumed. Compared to co-pyrolysis without a catalyst, which had a synergistic impact on oil yield (-17.0) and char yield (-5.02), co-pyrolysis with a KOH catalyst has a synergistic effect that improves oil yield. While the maximal pyrolysis index (PI) without a catalyst is 88.12, it drops to 86.58 with the KOH catalyst. It shows how the KOH catalyst contributed to the pyrolysis process during the co-pyrolysis of polypropylene and tea waste. Uncatalyzed co-pyrolysis has more internal pore area. Finally, PP: TW char (123-196 m²/g) is created. The gasses created during co-pyrolysis would have affected the char "pore area." Aromatic hydrocarbon selectivity was 93.6 %. CO-pyrolysis oil was 27% cycloalkanes and alkenes. In co-pyrolysis, cyclic aliphatic molecules and benzene derivatives generated dispersion of carbon. Most compounds were C8 (31.5%), C15-C24 (22.4%), and C9 (14.4%).

Cite this article as: Hamzah HT, Velurus, Mahdi SA, Abdullah TA, Tukarambai M, Poiba VR. In-sito co-pyrolysis of tea waste and polypropylene using microwave-assisted pyrolysis: Products analysis, energy consumption, and synergy effect. Sigma J Eng Nat Sci 2025;43(5):1653–1662.

INTRODUCTION

There has been a correlation between rising global consumption and the usage of energy resources

dependent on fossil fuel utilization [1]. Because of this expansion, there is a greater demand for the discovery of alternative energy supplies that are devoid of any adverse effects on the planet, such as emissions of greenhouse

*Corresponding author.

*E-mail address: alwatanaliraqi@gmail.com

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



gases (GHG). One of the most dependable and environmentally friendly energy sources is carbon-based waste, such as biomass [2]. Wastes from industries such as manufacturing, agriculture, food production, and wastewater are all examples of diverse types of biomasses. There has been extensive industrial use of this technology for years, and it has also been the subject of laboratory-scale study into the potential for converting carbonaceous wastes towards clean energy. Some techniques include pyrolysis, gasification, thermal cracking, and anaerobic digestion [3]. Polypropylene (PP), a ubiquitous material used in packaging, automotive parts, and other products, accounts for over 30% of the world's plastic manufacturing. Based on current estimations, the global production of polypropylene (PP) exceeds 85 million tons per year, of which less than 10% is recycled. The remainder enters ecosystems as pollution. Tea waste is among the many organic wastes produced in large numbers. Tea trash is a significant by-product of the approximately 6 million tons of tea produced worldwide each year, the majority of which is either discarded or ends up polluting the environment. Environmental issues include greenhouse gas emissions, hazardous chemical leaching, and soil pollution are caused by both polypropylene products and tea waste [4,5]. Specifically, co-pyrolysis of PP and tea waste has been demonstrated to improve bio-oil production; studies have reported as much as a 25% increase in oil yield when compared to the pyrolysis of tea waste alone; the tea waste also improves plastic decomposition, leading to higher-quality bio-oil and as much as a 30% reduction in solid residue; microwave heating has also been shown to reduce energy consumption; reports have indicated that it can reduce energy use by 20-50% when compared to traditional heating methods [6]. The absence of oxygen in pyrolysis eliminates the need for a reaction, making this process more advantageous for industrial applications. Additionally, the temperature of the process, which ranges from 350 to 650 °C, can be lower than the temperature at which gasification occurs, which is greater than 900 °C. The most significant distinctions between gasification and pyrolysis can be found in these two fundamental ideas [7]. In various industrial applications, pyrolysis is an essential economic reaction without adverse environmental effects. The product's constituent parts can be classified into one of three groups, depending on the temperature at which they are implemented: Bio-oil, charcoal, and syngas [8]. Reported that the wastes of polystyrene (PS) and tea powder (WTP) were converted into far more valuable products through catalytic co-pyrolysis using a KOH catalyst. The production of the product and the heating rates (17–75 °C/min) were correlated. WTP was in charge of generating char, whereas PS was in charge of producing oil. According to (Erkus & Calhan 2022) [9], activated carbon (AC) with a large surface area is produced using microwave-assisted pyrolysis (MAP), a two-stage chemical activation process. For

this aim, a special precursor material made of Tangerine peel (TP) and spent tea waste (STW) is combined. TP and STW were combined to create AC through the first step of the process, which involved impregnating samples with alkali metal hydroxide. ZnCl_2 , H_3PO_4 , and KOH were then used as catalysts after this stage. Every experiment was run for twenty to thirty minutes, using a microwave power of between 600 and 700 watts.

Fu et al. [10] reported that to evaluate how the HZSM-5 and HY catalysts affected the co-pyrolysis performance and product distribution of polyethylene (CP73) and coffee grounds, a thermogravimetric analyzer equipped with Fourier transform infrared spectroscopy and pyrolysis-gas-chromatography-mass spectrometry was used. The incorporation of HZSM-5 and HY made the disintegration of CP73 easier and resulted in a decrease in the amount of leftover material it contained. Compared to CP73, HY-catalyzed CP73 (CP73/HY) has displayed a reduced residual content, with a reduction of 9.09%, eliminating the influence of the mixing ratio [10]. Ben Abdallah et al. [11] reported that a two-stage thermochemical method was studied for discarded green tea waste (GTW), with the first step being pyrolysis and the second being thermal or catalytic cracking. Two bench-scale reactors, one with a fixed bed and the other with a tubular design, were linked in parallel to execute this procedure. When applied to a fixed temperature at which pyrolysis occurs of GTW (550 °C), high cracking temperatures (700 and 800 °C) improved the gas product's yield and composition. Because of this, it might find numerous applications in the biofuel and chemical production processes, and it could even be partially recycled to improve efficiency further [11]. Waste is transformed into valuable resources through this process, which includes the production of chemicals and materials that absorb pollutants. To accurately anticipate product yields, it is essential to take into consideration the characteristics of the feedstock as well as the circumstances of the pyrolysis process. The pyro-product yields from the co-pyrolysis of biomass and polymers assisted by microwaves were predicted using a support vector machine (SVM). Following the training and validation of SVM models, they were applied to new data [12]. On the other hand, co-pyrolysis is a process that uses various waste materials as feedstock. In contrast to the abundant polymeric waste stream represented by tire debris, there is a correlation between the high consumption rate of coffee generally and the expansion of coffee waste material. Coffee and tire wastes can be combined to create a bio-oil manufacturing method that has the potential to be very successful. Specific conditions must be fulfilled to achieve the best possible outcomes [13]. MW-assisted pyrolysis is an efficient solid waste management approach. Hamzah et al. [6] reported that, converted Paraffin wax (PW) and rice straw (RS) into char, gas, and oil using MW-assisted catalytic co-pyrolysis. The catalyst and susceptor were KOH and graphite. RS and PW (0–10

g combination) were the feedstock. RS: PW co-pyrolysis yields varied in char content (9.8%–22.6% by wt.), oil production (34.1%–76.9%), and gas creation (13.2%–47.5%). Researchers also examined how different RS: PW ratios affected product yields, average heating rate, and feedstock conversion [6].

This work aims to study the effect of co-pyrolysis of tea waste and polypropylene using microwave-assisted co-pyrolysis on product yield, energy consumption, synergy effect, and pyrolysis index. Examining the effects of KOH catalyst on MW-assisted co-pyrolysis of tea waste and polypropylene, including the pyrolysis index, yields, composition, average heating rate, high heating value, energy consumption, and pyrolysis reaction time, is the main objective of this research.

MATERIALS AND METHODS

Materials

During this research, co-pyrolysis was carried out using two distinct feedstocks. These feedstocks were biomasses (tea waste) and heavy hydrocarbons (polypropylene) of Indian origin. The result was the production of high-quality oil, char, and syngas. Because microwave co-pyrolysis requires the identification of types of heavy hydrocarbons and biomass that can be utilized, the significance of this discovery becomes readily evident. When considering solid waste management, the selected feedstocks are of the utmost importance [5]. After the pyrolysis and co-pyrolysis of tea waste and polypropylene, the graphite employed as a susceptor in our studies, we examined the production of bio-oil, char, and gases and the quality of these products. 10 g each of tea waste and polypropylene were used in the laboratory experiments, 2 g KOH was used as catalyst, graphite was used as susceptor and the microwave power setting was 450 W. Our research aims to provide insight into the potential of microwave-assisted pyrolysis and co-pyrolysis for recovering energy and resources from recyclable biomass and heavy hydrocarbon polymers. The temperature was recorded every 30 seconds for the study of temperature and heating rate profile. The microwave energy has been calculated to be necessary for pyrolysis. To find out how much microwave power is needed, one uses the following equation 1[4]:

$$\text{Microwave energy} = \text{MW Power (450 W)} \times 60 \times \text{Total pyrolysis time (min)} \quad (1)$$

To find out how much of an impact synergistic effect, or the degree to which various biomass mixtures interacted during co-pyrolysis, we used the following equation 2:

$$\Delta X = X_{EXP} - X_{CAL} \quad (2)$$

X_{CAL} was calculated by using the Equation 3:

$$X_{CAL} = A1B1 + A2B2 \quad (3)$$

A1 and A2 represent the TW and PP weight percentages, respectively, whereas B1 and B2 denote the mixture's TW and PP mass fractions. If the ΔX value is more than 0, it indicates that the synergistic impact is negative. When the ΔX value is less than zero, the synergistic impact is positive [4]. It has been determined that the following equation Can be used to compute the pyrolysis index:

$$\text{Pyrolysis index (PI)} = F_m * F_{HHV} * F_X / P_{EC} * P_t \quad (4)$$

Where P_t , P_{EC} , F_m , F_X , F_{HHV} are pyrolysis time (s), pyrolysis energy consumption (J/s), the mass, of feedstock (g), conversion of feedstock (%), and the higher heating value of feedstock (J/g), respectively [4].

Analysis of Proximate and Elements for the Characterization of TW and PP

Elements and proxy Analysis are performed on dry samples of feedstocks, namely tea waste powder (TWP). The chemical composition of TWP is as follows: 79.3% carbon, 1.7% hydrogen, 3.4% nitrogen, 0.2% sulfur, and 15.2% oxygen, as stated in Table 1. The TWP has significant volatile matter (71.0%) and fixed carbon (15.57%). Meanwhile, TWP has little ash content and can work as a susceptor. The PP comprises 90.8 weight percent carbon, 7.1% hydrogen, 0.1 weight percent nitrogen, and 0.0 weight percent sulfur, while 2 weight percent oxygen makes up the compound. Its thermal value is high because of the low concentration of oxygen and ash in it. The PP's volatile matter and fixed carbon percentages are 93.7% and 6.3%, respectively. However, PP does not include a significant amount of ash. Ash, which is present in, permits this change in available products by acting as a susceptor. Table 1 shows the results of the elements and proximity analyses of TW and PP, respectively [4].

Table 1. Proximate and elemental analysis of (TW, PP)

Feedstock	Proximate Analysis (wt%)			Elemental Analysis (wt%)				
	VM	FC	Ash	C	H	N	S	O
Tea Waste Powder (TW)	19.8	71	9.2	79.3	1.7	3.4	0.2	15.3
Polypropylene (PP)	93.7	6.3	0.0	90.8	7.1	0.1	0.0	2.0

GC-MS Analysis

The organic composition of the bio-oil byproduct of MW-pyrolysis was examined employing gas chromatography/mass spectrometry (Agilent 7890, 5975C). With an internal diameter of 0.25 millimeters and a film thickness of 0.25 micrometers, the HP-5MS column, which is 30 meters long, was designed. In 69 steps, it separated the gaseous products of pyrolysis. To serve as the carrier gas, 0.8 millimeters per minute of ultra-high-grade helium, 99.9995%, was utilized. Peak mass spectra were compared to the NIST library, which allowed for the identification of the chemical components of bio-oil components. Other categories have been developed for organic compounds with at least 85 percent match factors. To achieve the mass balance, substances that were either unidentifiable or poorly matched were added to the range of already present compounds. Standardization was performed on the fraction of peak area for each chemical concerning bio-oil production. The peak area percentage is sometimes displayed as the relative selectivity [14,15].

Calorific Value

Feedstocks and products have calorific values, which is the energy produced when burned under standard conditions. The higher and lower heating values (HHV and LHV, respectively) indicate calorific values, and this reflects the energy of the feedstock or product. When water's latent heat of vaporization is considered, HHV represents the most significant heat extracted from fuel after burning. LHV does not consider the latent heat of vaporization. Due to the cooling process that occurs before testing, LHV is consistently lower than HHV. This method calculated hydrocarbon polymers, oil, biomass, and other components' HHVs [6].

Table 2. High heating value of feedstocks

Feedstock	HHV (MJ/Kg)
Tea waste (TW)	13.4
Polypropylene (PP)	44.7

RESULTS AND DISCUSSION

MW-Assisted Pyrolysis and Co-Pyrolysis of Polypropylene and Tea Waste

Here is the conventional method of testing using a bomb calorimeter to determine the heating value. Biomass (TW), heavy hydrocarbon polymers (PP), and oil yield were tested, and HHVs were identified using this method [6,14]. The feedstock utilized in our studies has a high heating value, as shown in Table 2, and HHVs of product oil yield in Table 3.

Table 3 indicates the experimental result of MW-assisted pyrolysis and co-pyrolysis of TW and PP at MW power, which was preferred at 450 W. The studied oil, gas, and char yields were reported. Benzene derivatives, like carboxylic acids linked to the benzene ring, are produced in MACP by combining the lignin and polyphenol structure of tea waste powder with polypropylene's hydrogen and carbonate content. The GCMS analysis of co-pyrolysis Oil of (PP:TW) reported the Benzene, toluene, phenol, furfural, furan, naphthalene, indene, and xylene are aromatic hydrocarbons that can be synthesized by manipulating biomass and polypropylene species with MACP. The recovered oil contained various aromatic compounds, including oxygenated molecules and mono-aromatics [4,14].

By co-pyrolyzing (PP: TW) with a KOH catalyst, the gaseous product yield dropped to 33.6%, but the oil and char yields rose to 34.8% and 31.6%, respectively. Using microwave co-pyrolysis without a catalyst resulted in a maximum gas yield of 41.67% by weight, see Figure 1. This matches with the study by (Hamzah et al., 2024a) where oil yield was increased using the KOH catalyst, when they are co-pyrolyzed rice straw and paraffin wax (plastic-like) [6].

Temperature and Heating Rate Profiles of MW-Assisted Co-Pyrolysis of (TW: PP)

The temperature profiles shown in Figure 2 correspond to the co-pyrolysis of (biomass: heavy hydrocarbons), tea waste, and polypropylene, as shown in Table 3. The temperature profiles were obtained by co-pyrolyzing the feedstock at a microwave power of 450 W. After 17 minutes of operation using graphite as the susceptor; the reactant temperature reached 701-718 C°. Experiment C4 (TW: PP) with a KOH catalyst had the highest oil yield (34.8 wt%)

Table 3. Experimental result of MW-assisted pyrolysis and co-pyrolysis of TW and PP

Exp. NO.	Plastic PP (g)	Tea Waste (g)	KOH catalyst (g)	AHR (°C/min)	Oil yield (wt.%)	Gas yield (wt.%)	Char yield (wt.%)	Conv. (wt.%)	Heating value (MJ/kg)	Pyrolysis time (min)	Pyrolysis Temp. (°C)	MW Energy (kJ)
C1	10	10	-	4.8	35	14.6	50.4	49.6	13.7	86.5	450	3114
C2	10	10	-	48.6	55.3	37.7	7	93	45.7	20	550	540
C3	10	10	-	40.4	27.95	41.67	30.38	69.62	36.7	17	718	459
C4	10	10	2	39.4	34.8	33.6	31.6	68.4	37.2	17	701	459

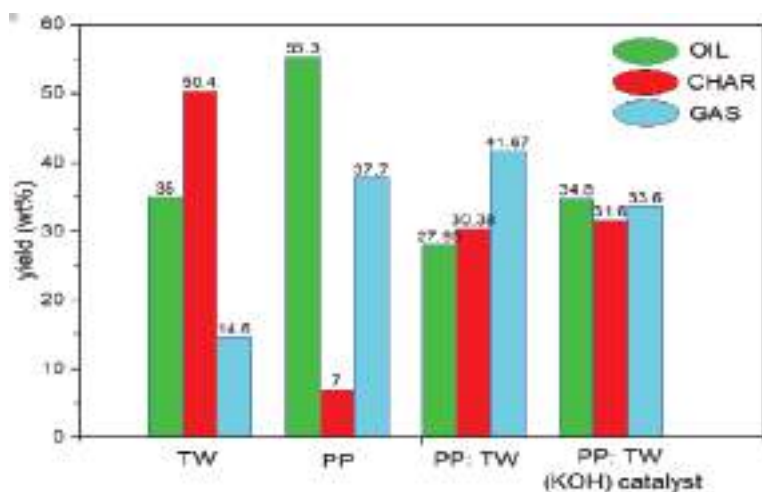


Figure 1. Oil, char, and gases yields of MW-assisted pyrolysis and co-pyrolysis of PP and TW.

and heating value (37.2 MJ/Kg), even though the average heating rates ranged from 39.4 to 40.4 °C min⁻¹. In contrast, the gas yield is at its highest (41.67 wt%), and char yield is at its lowest (31.6 wt%) during co-pyrolysis with no catalyst, these results are matches with the study by (Neha & Remya, 2022) [16]. Experiments involving microwave co-pyrolysis and their respective heating rate profiles are detailed in Table 3. The graphite susceptor shows great promise in Figure 2's heating rate profiles because of its excellent thermal stability, high average heating rate, and high heating rates (40.4 °C/min) attained in the first co-pyrolysis phase. The fact that non-condensable gases can be produced in large quantities by rapidly heating samples is well-known [17,18].

Microwave Energy Consumed and Higher Heating Values of Co-Pyrolyzed Biomass and Heavy Hydrocarbons

We used a bomb calorimeter to measure the higher heating values of MW Co-pyrolysis (tea waste and

polypropylene) oil, as shown in Table 3. All of the combustion tests were carried out within a specific minute range. All samples' HHVs are provided in Table 3. Biomass, heavy hydrocarbons, other feedstocks, and the KOH catalyst impacted the HHV of MW-pyrolysis oil [4]. Compared to separate pyrolysis, which requires a lot of energy and time, microwave energy consumption is the best (459 KJ) when used to co-pyrolyze tea waste and polypropylene. Additionally, the reaction time is lowered.

Synergy Effect on Product Yields of MW Co-Pyrolysis

A synergistic effect occurs when the effects of two or more substances are amplified when brought together. Results are more impressive when accumulated than when considered separately. Potentially, this is the end product of chemical processes involving carbon and hydrogen-containing compounds [6]. The term "synergistic effect" describes the positive or negative impact of an interaction between particles of different materials. Select the suitable

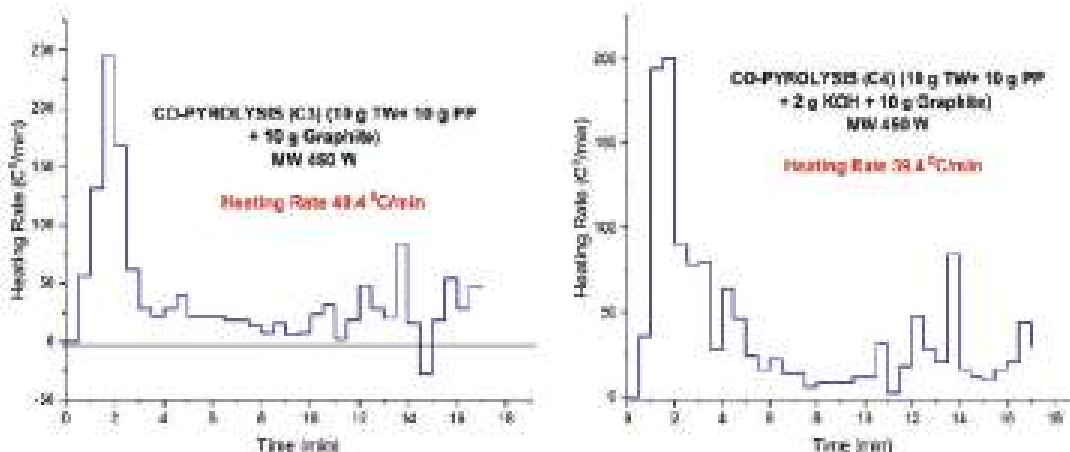


Figure 2. Temperature and Heating Rate Profiles of MW co-pyrolysis of PP: TW.

feedstock and optimize the process to get the most out of your pyrolytic products. Finding out how two materials synergize when pyrolyzed together is essential. Experimental values were derived using microwave-assisted co-pyrolysis of biomass and heavy hydrocarbons. One may obtain theoretical values for biomass (TW) and heavy hydrocarbons (PP) using separate pyrolysis procedures. To find out if the co-pyrolysis of biomass mixtures has a synergistic impact, one can compare the experimental and theoretical values of the parameters. By combining the effects of biomass and heavy hydrocarbons on the two feedstocks, their combined influence on pyrolysis was determined. Yields from pyrolysis, average heating rate, and conversion are defined below. The expected and measured product yields are displayed in Figure 3. Results for TW and PP from microwave

co-pyrolysis, as a result of synergy, are shown in Table 4. When compared to co-pyrolysis without a catalyst, which had a synergistic effect on oil yield (-17.2) and char yield (-5.02), co-pyrolysis with a KOH catalyst had a synergistic impact that improved oil yield (-17.0). Because the KOH catalyst improves the oil yield, and reduce the char (1.3); These results are matches with the study by [19].

Pyrolysis Index

Many variables affect the outcome of PI, including conversion, feedstock mass, outstanding heating value, pyrolysis time, and pyrolysis energy consumption. The value can be between zero and infinity, with more significant numbers indicating a higher pyrolysis efficiency. Table 5 displays the effect of MW-assisted pyrolysis individually and

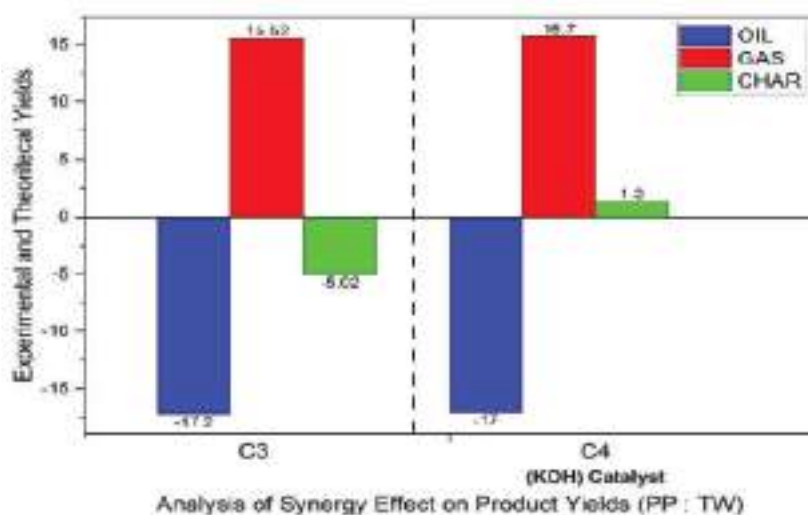


Figure 3. Effect of synergy on product yields of polypropylene and tea waste.

Table 4. Effect of synergy on product yields of microwave co-pyrolysis of (PP: TW)

Exp. NO.	Polypropylene PP (g)	Tea waste TW (g)	(KOH) catalyst (g)	Synergy effect on oil yield	Synergy effect on gas yield	Synergy effect on char yield
C4	10	10	2	-17.0	15.7	1.3
C3	10	10	-	-17.2	15.52	-5.02

Table 5. Pyrolysis index of MW-assisted pyrolysis and co-pyrolysis of TW and PP

Exp. NO.	Polypropylene (g)	Tea Waste (g)	KOH catalyst (g)	Pyrolysis Time (min)	Pyrolysis index
C1	-	10	-	86.5	4.4
C2	10	-	-	20	76.98
C3	10	10	-	17	88.12
C4	10	10	2	17	86.58

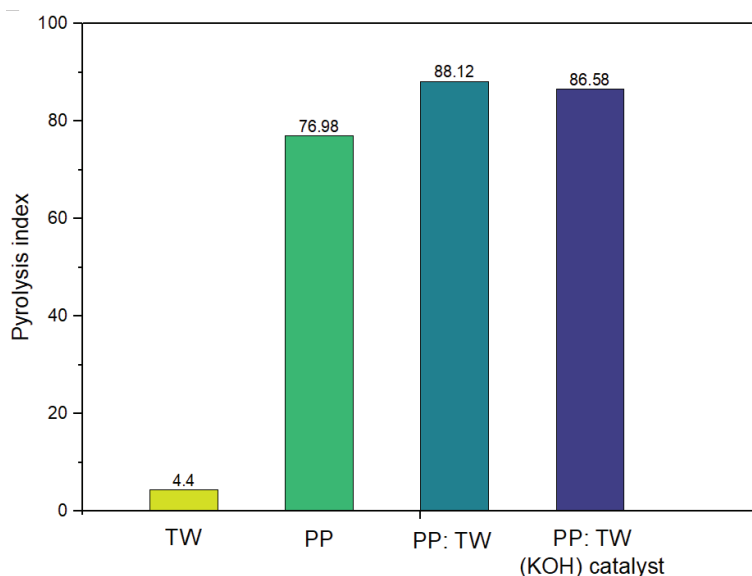


Figure 4. Pyrolysis index of MW-assisted pyrolysis and co-pyrolysis of TW and PP.

co-pyrolysis of (PP, TW) with and without KOH catalyst on the pyrolysis index (PI).

Figure 4 shows the maximum pyrolysis index of co-pyrolysis of TW: PP without using a catalyst was (88.12); it indicates that There is an inverse relationship with the reaction time, as the higher the PI, the lower reaction time, compared to if the pyrolysis was individual. The effect of the KOH catalyst was to reduce gas production, which contributed to reducing the conversion rate, and this affected the reduction of PI by a small percentage. A high PI of co-pyrolysis of (PP: TW) represents the efficiency of the process in terms of calorific value, reaction time, conversion rate, and energy consumed [20,21].

Analysis of Products

The blends of fifty-weight percent polypropylene and fifty-weight percent tea waste were pyrolyzed by microwave to explore the synergism that occurs during the co-pyrolysis process and its impact on the composition of the result. Ten grams of polypropylene and ten grams of tea waste powder were thoroughly combined in this particular instance. Consequently, a MW susceptor of 10 grams of graphite powder was added. After collecting the obtained liquid products, they were dissolved and then diluted with dichloromethane. GC–MS was used to analyze the phase of the fluid sample that had been extracted with dichloromethane. In microwave-assisted pyrolysis of polypropylene added to tea waste in a mass ratio of 10:10 grams, the surface area of the char produced after co-pyrolysis varied depending on whether or not a KOH catalyst was used. Compared to MW pyrolysis, which uses a KOH catalyst, the co-pyrolysis that does not involve the catalyst has a higher internal pore area. The char from PP: TW (123-196 m² /g) results from this. The char pore area

would have been dramatically affected by the developed gases during the co-pyrolysis process. Following the completion of the GCMS analysis for the oil that was produced by microwave co-pyrolysis of (PP: TW), it was discovered that aromatic hydrocarbons were the most prevalent group in both of the oils that were created by co-pyrolysis. (93.6 %) was the selectivity of aromatic hydrocarbons [22,23]. Cycloalkanes and alkenes were found in the oil produced through co-pyrolysis (27%). The synergy that resulted from the co-pyrolysis process led to the synthesis of cyclic aliphatic compounds and benzene derivatives. Co-pyrolysis exhibits a high selectivity for the following compounds: 1,2,3-trimethyl benzene (17.7%), O-xylene (7.5%), ethylene benzene (6.2%), ethylbenzene (5.8%), and naphthalene (4.7%). A comprehensive examination of the compounds discovered this. Co-pyrolysis resulted in the production of carbon numbers that fell between C6 and C24, according to the examination of the distribution of carbon numbers. Compounds of the chemical formula C8 (31.5%), C15-C24 (22.4%), and C9 (14.4%) were discovered to be the most prevalent in this range, these results are matches with the study by (Neha & Remya, 2022) [16]. Approximately 88% of the 20 carbon elements were found in the oil produced through co-pyrolysis. The elemental hydrogen level is estimated to be somewhere between 8.5 and 10.5%. Because of the significant amount of oxygen transferred from the waste tea into a gaseous fraction, the elemental oxygen content was low, ranging from 0.9% to 2.9%. As a result, the high carbon and hydrogen content of the oil produced by co-pyrolysis increased its energy density. The result matches with the study by [23,24] when they co-pyrolyzed rice straw and paraffin wax (plastic-like) [6].

CONCLUSION

Polypropylene and waste tea have emerged as potentially valuable feedstocks for producing oil, gasses, and char. The reaction had a temperature range of 700 to 718 °C. Approximately 17 minutes was the reaction time. In terms of oil production, it was discovered that the most effective method was the co-pyrolysis of TW: PP with a KOH catalyst. Using a KOH catalyst, the oil and char yield increased to 34.8 wt% and 31.6 wt%, respectively. It was discovered that the maximum pyrolysis index of co-pyrolysis of TW: PP without using a catalyst was 88.12 after 17 minutes of reaction time, and the maximum gas output was 41.68 wt%. The synergy of co-pyrolysis TW: PP improved the oil yield (-17.0) with the KOH catalyst and (-17.2) without the catalyst. Also, the synergy effect improved char yield (-5.02) without a catalyst. More of the internal pore area is present in the co-pyrolysis process that does not involve the catalyst. Consequently, the char from PP: TW (123-196 m²/g) is produced. During the co-pyrolysis process, the gases produced would have had a significant impact on the pore area of the char. Aromatic hydrocarbon selectivity was 93.6%. Cycloalkanes and alkenes made about 27% of co-pyrolysis oil. The synergy of co-pyrolysis produced cyclic aliphatic molecules and benzene derivatives. carbon number dispersion. In this range, C8 (31.5%), C15-C24 (22.4%), and C9 (14.4%) compounds were most common. This work needs further future study to fill in the gaps, for example the economic value of converting tea waste and polypropylene into a source of fuel or useful chemicals on a large scale.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests.

DATA AVAILABILITY

The data that has been used is confidential.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Husam Talib Hamzah], [Sridevi Veluru], [Suhair Abdulhadi Mahdi], [M. Tukarambai] and [Venkata Rao]. The first draft of the manuscript was written by [Husam Talib Hamzah] and [Thamer Adnan Abdullah] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

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

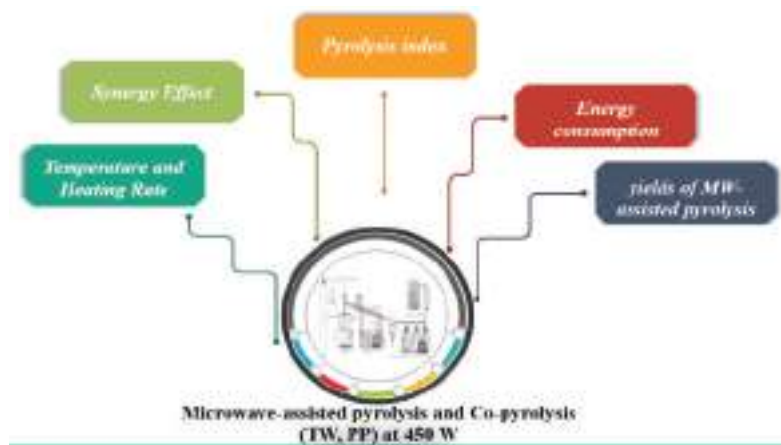
Artificial intelligence was not used in the preparation of the article.

REFERENCES

- [1] Hassan MA, Salem H, Bailek N, Kisi O. Random forest ensemble-based predictions of on-road vehicular emissions and fuel consumption in developing urban areas. *Sustainability* 2023;15:1503. [\[CrossRef\]](#)
- [2] Menzel CM. Effect of global warming on the yields of strawberry in Queensland: a mini-review. *Horticulturae* 2023;9:142. [\[CrossRef\]](#)
- [3] Lee J, Kim S, You S, Park YK. Bioenergy generation from thermochemical conversion of lignocellulosic biomass-based integrated renewable energy systems. *Renew Sustain Energy Rev* 2023;178:113240. [\[CrossRef\]](#)
- [4] Talib Hamzah H, Sridevi V, Seereddi M, Suriapparao DV, Ramesh P, Sankar Rao C, et al. The role of solvent soaking and pretreatment temperature in microwave-assisted pyrolysis of waste tea powder. *Bioresour Technol* 2022;363:127913. [\[CrossRef\]](#)
- [5] Hamzah HT, Sridevi V, Surya DV, Palla S, Yadav A, Rao PV. Conventional and microwave-assisted acid pretreatment of tea waste powder: analysis of functional groups using FTIR. *Environ Sci Pollut Res* 2023;30:1–12. [\[CrossRef\]](#)
- [6] Hamzah HT, Sridevi V, Surya DV, Ramesh P, Rao CS, Palla S, et al. Synergistic effects and product yields in microwave-assisted in-situ co-pyrolysis of rice straw and paraffin wax. *Process Saf Environ Prot* 2024;182:45–55. [\[CrossRef\]](#)
- [7] Osman AI, Qasim U, Jamil F, Al-Muhtaseb AH, Jrai AA, Al-Riyami M, et al. Bioethanol and biodiesel: bibliometric mapping, policies and future needs. *Renew Sustain Energy Rev* 2021;152:111677. [\[CrossRef\]](#)
- [8] Kan T, Strezov V, Evans TJ. Lignocellulosic biomass pyrolysis: a review of product properties and effects of pyrolysis parameters. *Renew Sustain Energy Rev* 2016;57:1126–1140. [\[CrossRef\]](#)
- [9] Erkus H, Calhan R. Co-pyrolysis of spent tea waste and tangerine peels with different catalysts by microwave-assisted pyrolysis. *Bioenergy Res* 2022;16:1099–1107. [\[CrossRef\]](#)
- [10] Fu J, Lin S, Cai B, Liang J, Chen Z, Evrendilek F, et al. Catalytic co-pyrolysis of coffee grounds and polyethylene: a comparison of HZSM-5 and HY. *Fuel* 2024;362:130815. [\[CrossRef\]](#)

- [11] Ben Abdallah A, Ben Hassen Trabelsi A, Veses A, García T, López JM, Navarro MV, et al. Enhancing the production of syngas from spent green tea waste through dual-stage pyrolysis and catalytic cracking. *Catalysts* 2023;13:1334. [\[CrossRef\]](#)
- [12] Potnuri R, Rao CS, Surya DV, Kumar A, Basak T. Utilizing support vector regression modeling to predict pyro product yields from microwave-assisted catalytic co-pyrolysis of biomass and waste plastics. *Energy Convers Manag* 2023;292:117387. [\[CrossRef\]](#)
- [13] Hilal L, Žabčić M, Rocha-Meneses L, Ghenai C, Inayat A. Production of bio-oil by co-pyrolysis of the coffee and tire wastes. *Proc World Congr New Technol* 2023;9:1–6. [\[CrossRef\]](#)
- [14] Talib Hamzah H, Suhair Abdulhadi M, Sridevi V, Thamer Adnan A. Microwave-assisted pyrolysis and co-pyrolysis: oil, char, and gases – a technological review. *El-Cezeri J Sci Eng* 2024;11:186–198. [\[CrossRef\]](#)
- [15] Liu Y, Liu Y, Chen R, Wang H, Gao H, Wang Y, et al. Microwave pyrolysis of sewage sludge for bio-oil production: effects of organic components and mechanisms. *Fuel Process Technol* 2024;253:108023. [\[CrossRef\]](#)
- [16] Neha S, Remya N. Co-production of biooil and biochar from microwave co-pyrolysis of food-waste and plastic using recycled biochar as microwave susceptor. *Sustain Energy Technol Assess* 2022;54:102892. [\[CrossRef\]](#)
- [17] Terapalli A, Kamireddi D, Sridevi V, Tukarambai M, Suriapparao DV, Rao CS, et al. Microwave-assisted in-situ catalytic pyrolysis of polystyrene: analysis of product formation and energy consumption using machine learning approach. *Process Saf Environ Prot* 2022;166:57–67. [\[CrossRef\]](#)
- [18] Hamzah HT, Veluru S, Hasan HM, Abed HY, Mahdi HS, Hamdullah NA, et al. Microwave assisted pyrolysis of biomass feedstock fundamentals and the effect of process parameters – a review. *J Phys Conf Ser* 2023;2550:012003. [\[CrossRef\]](#)
- [19] Burra K, Gupta A. Kinetics of synergistic effects in co-pyrolysis of biomass with plastic wastes. *Appl Energy* 2018;220:408–418. [\[CrossRef\]](#)
- [20] Suriapparao DV, Sridevi V, Ramesh P, Sankar Rao C, Tukarambai M, Kamireddi D, et al. Synthesis of sustainable chemicals from waste tea powder and polystyrene via microwave-assisted in-situ catalytic co-pyrolysis. *Bioresour Technol* 2022;362:127813. [\[CrossRef\]](#)
- [21] Raza M, Abu-Jdayil B, Al-Marzouqi AH, Inayat A. Kinetic and thermodynamic analyses of date palm surface fibers pyrolysis using Coats-Redfern method. *Renew Energy* 2022;183:67–77. [\[CrossRef\]](#)
- [22] Hamzah HT, Veluru S, Hasan HJ, Abdullah TA, Gudapati G. Innovative biosensors from agro-waste: laser and microwave approaches for current and future applications in environmental health. In: Neelancherry R, Gao B, Wisniewski A, editors. *Agricultural waste to value-added products*. Singapore: Springer; 2024. p. 1–12. [\[CrossRef\]](#)
- [23] Hamzah HT, Alawi NM, Hasan HM, Abed HY, Veluru S, Hamdullah NA, et al. Effect of acids and alkaline pretreatment of used tea waste powder under microwave-assisted pretreatment. *Key Eng Mater* 2023;960:149–160. [\[CrossRef\]](#)
- [24] Chen Z, Zhao M, Liu Y, Lv Y, Khan SN, Tariq G, et al. Comprehensive pyrolysis characteristic and kinetic analysis of industrial-sludge and rice straw: blending and pellet. *J Clean Prod* 2024;440:140887. [\[CrossRef\]](#)

Graphical abstract





Research Article

An integrative approach to agricultural challenges: Predictive modeling, crop alternatives and automation

Swapnil S. JADHAV^{1,*}, Lalit N. PATIL², Vikas S. PANWAR³, Sachin P. JADHAV⁴,
Digvijay B. KANASE⁵, Sayali S. JADHAV⁶, Neeraj D. PATEL², Khushi V. SALI²,
Srushti M. JOSHI², Ruchi N. NAGWANSHI²

¹Department of Mechanical Engineering, Adarsh Institute of Technology Research Centre, Maharashtra, 415311, India

²Department of Automation and Robotics, Dr. D. Y. Patil Institute of Technology, Pune, 411033, India

³Department of Mechatronics, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, India

⁴Department of Electrical Engineering, Rajarambapu Institute of Technology, Sakharale, Maharashtra, 415414, India

⁵Department of Electrical Engineering, Dr. D. Y. Patil Institute of Technology, Pune, 411033, India

⁶Department of Mechanical Engineering, Annasaheb Dange College of Engineering and Technology, Maharashtra, 416301, India

ARTICLE INFO

Article history

Received: 01 April 2024

Revised: 18 June 2024

Accepted: 13 October 2024

Keywords:

Automation; Data Science;
Internet of Things (IoT);
Machine Learning; Predictive
Modeling; Smart Agriculture

ABSTRACT

The present research will explore the possibility of an integrated method of enhancing agricultural practices in India that utilize machine learning, data science, and Internet of things (IoT) applications. This explorative research is a method of addressing the gap between traditional and modern farming technologies by integrating IoT and data science technologies and supplying the farmers with real-time monitoring and actionable insight and direction through data-driven decision-making. The study is separated into three sections that include predictive modeling, evaluation of business value, and Internet of things (IoT) implementation. This is the main finding of the study that helps develop a predictive model with up to 99% accuracy and make individual recommendations on crop choice and fertilizing in accordance with the soil qualities and environmental situation. Also, the Arduino-based system to an NPK sensor enables the real time checking of soil nutrients to optimize fertility of the soil and agricultural management. The results show that the possibilities of better harvest, fewer losses, and maximized returns can be achieved by using these technologies. The up-to-date interface created in the context of this study allows agriculture producers to make valid decisions and balances between tradition and innovation in the context of sustainable and resilient farming in India.

Cite this article as: Jadhav SS, Patil LN, Panwar VS, Jadhav SP, Kanase DB, Jadhav SS, Patel ND, Sali KV, Joshi SM, Nagwanshi RN. An integrative approach to agricultural challenges: Predictive modeling, crop alternatives and automation. Sigma J Eng Nat Sci 2025;43(5):1663–1682.

*Corresponding author.

E-mail address: swapniljadhav.9001@gmail.com

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



INTRODUCTION

Healthy agriculture is the backbone of the Indian economy; it is the occupation of about half the population in the country, meaning it is quite important. Its healthy states are mostly dependent on the state of the Indian soil, directly impacting the yield and quality of the crops. The better the soil, the better the agriculture. Reserve phytophthora control Mycotoxin solutions [1].

Nevertheless, even though it is crucial, the Indian farmers have numerous problems which impede them on the way to the enhancement of crop quality and sustainable agriculture. These stresses are posed through the varied weather, land degradation, fluctuating market prices and lack of access to information and resources [2].

Agricultural practices can be rooted in pre-historic concepts shared over thousands of years and may not keep abreast with current field or transformation of agriculture. Confidence on anecdotal data and hearsay also exposes farmers to incompetence and stumbling block in making sound judgments and productive in the farming industry. [3] The conventional practices that dominate agriculture today tend to yield negative results implying that innovativeness is necessary. In order to increase agricultural output, new applications and integration with the recent technologies have become the subject of attention. Unforeseeable rain patterns as a result of climate change is a motivating factor in the promotion of precision agriculture [4, 5].

There is need to take another step further in exploring this joint effort so as to appreciate the urgency of the challenges affecting the agriculture. Considering it is projected that the global population will reach over 9 billion people by 2050, the premises facing the world food supply are explicit. Traditional processes are in an unprecedented position facing a situation of limited resources and climate change. In this context, it is the incorporation of prediction models, variable farming as well as technology that brings forth some hope as it signifies a radical change in the sector of agriculture [6].

In order to guarantee the use of permaculture, the following technologies ought to be applied to include blockchain, the internet of things (IoT), big data, and machine learning (ML) [7, 8]. Amongst them, it will help to teach about agriculture which is the best approach to the problems present and in the future. If this big data on agriculture could be created and utilized in making decision in the agricultural sector, then most of the world food problems would be solved. As an example, when farmers have possibilities to develop information or maps about certain environmental conditions around their farms, they can apply such technologies as smart farming, farming, and vertical farming. Big data in the field of agriculture has been found to enhance yields, decrease expenses and enhance sustainability [9, 10]. In the focus of this paper lies collaborative approaches to allow the combination of learning

technologies, data science and the Internet of Things (IoT) to implement solutions to challenges of the Indian agriculture. It is better understood through the main aim of transforming traditional agriculture through empowerment of the farmers with insight and predictive analytics at three levels. In the first step, it is all about predictive modeling, which employs advanced mathematical algorithms to propose crops and fertilizers in consideration of other intricate conditions like the soil and climate conditions [11].

Data mining came to play full role in crop decision-making during the course of growth; it guided in choosing appropriate crops to use in certain applications as well as technical hitches. Complex neural networks are among the intelligent methods that have been appealing in developing agriculture as a whole. This paper will outline the history of artificial intelligence and show how it can make negative data meaningful and forecast the crop size based on the past market data records [12, 13].

What is more important is the fact that such a collaboration gives not only practical solutions but also proposes the shift towards the more efficient and effective agricultural sector where the role of technology is crucial in future agriculture development of India [14, 15].

The following part of this article describes the design, process, and findings of the study, and crop forecasting solutions, as a subcategory of soil fertility and crop friendly search and monitoring.

The third and the last step involves integrating Internet of Things (IoT) through the development of novel Internet of Things (IoT) devices that can be used to extract the vital data in a short time period. The device comprises NPK sensor, an-arduino Uno R4, and an OLED display that allows the process of transmitting data to the web to be analysed. The web site employs the capabilities of serial communication and Wi-Fi to execute on the predictive models developed in the initial step to deliver direct and on-time information to farmers in the fields [16, 17].

Along with this, a convenient web site was developed on Django framework to make it reachable and usable to the farmers as well. The site is a syndicating model that combines all three levels of the research, where farmers can benefit in an intuitive manner access to forecast models, market price forecasters, and topical information about the activities in real-time Internet of Things (IoT) inspection [18].

It is hoped that with the technology of machine learning, data science, Internet of Things (IoT) web application development, the knowledge and toolset that is used by Indian farmers to deal with the nuances of modern agriculture will be provided. The goal of the program is to enhance the agricultural performance stability, and the improvement of the health of the rural population by boosting leadership through personalized recommendation, predictive information, and real-time monitoring available on the web page [19].

Machine learning To automate agricultural processes using machine learning To find a universal solution to key

agrarian problems by developing a predictive model that will help optimise the nutritional structure and identify crops based on soil type [20].

The study seeks to maximize yield using the predictive modeling, suggest alternative crop choices, automate the farming process with the incorporation of the Internet of Things (IoT) and offer data-driven decision making to crop producers to improve sustainability and the profitability of the farm [21]. These goals form the core foundations of this study project, which regards a comprehensive exercise that would aim to maximize agricultural activities, foster sustainability and establish modern technological paradigms in the farming sector [22].

Smart Agriculture Overview

Agriculture has undergone numerous revolutions that have helped the industry to become lucrative and profitable. Domestication of the ancient species of plants (10,000 BC) had given birth to the first organization and civilization in the world. Agricultural mechanization (1900-1930) brought about machines and mechanization farming methods that helped in making work more efficient and made farmers more prolific [23]. The Green Revolution (approximately 1960s) provided farmers with the new crops and agricultural chemicals. Internet of Things (IoT) technology made it possible to develop plants with pre-selected features, including higher yields and resistance to insects, drought, and herbicides, in the late and early centuries (1990 to 2005). Digital transformation has been able to make people sustain and prosper in the long run [24]. Initial stages of digital transformation are concerned with automation technology involving some computerized tasks [25].

The historical context of the way agriculture has been developed over time along with technological achievements offers a manifestation of time of the farmers. Incorporation of ancient skills and technologies, e.g., machine learning, is crucial to sustainable and successful farming [26, 27].

History and development of farming; It depicts the critical moments when one should consider making adjustments due to the changes in technology, global warming, customer preference and international guidelines. The historical background points to the issues currently met by agriculture, in terms of initial agriculture to technical changes in the Industrial Revolution and the industrial age of the 20th century [28, 29]. Commercial technology pioneered precision agriculture, which is based on the analysis of data, the use of sensors and satellites and informs the attempts of farmers to better use resources and limit negative impact on the environment. The justification circles around the significant role of the AI in the agricultural sector because it is sensitive to the size of datasets, algorithm parameters, and terrain type and purpose prediction. Deep neural networks are a valuable component in artificial intelligence and are credited with the help they provide fertility prediction, crop strategy and land management particularly organic fields. Conclusively, the move towards

advanced agriculture is not only significant to be considered as a practical move but a major step towards ensuring the safety and efficiency towards the full-fledged cyclic agricultural activity [30, 31].

MATERIALS AND METHODS

Predictive Modeling

Model development

High performance predictive model was developed via the use of ensemble learning concepts like bagging to ensure recommendations on the type of products to apply alongside fertiliser applications based on soil properties and environmental conditions affecting the region.

The datasets used to develop both predictive models and market price predictions in dented this research constituted of series of data collected on different government agricultural websites. Particularly, information was gathered through Directorate of marketing & inspection (DMI) website (agmarknet.gov.in), India council of agricultural research (icar.org.in) and the ministry of agriculture and farmers welfare (agricoop.nic.in). Data was retrieved in areas like Kolhapur, Solapur, Satara, Sangli and Pune over the past 10 years between 2012 and 2022. The data consist of a set of about 37458 samples comprising of about 1 GB of soil characteristics and crop yield data, content of different variables concerning the tables, with region name, soil color, nitrogen content, phosphorus content, potassium content, pH value, rainfall amount, temperature, crop type, and crop yield being some of them. Also, about 500 MB of market price data were collected, 17467 samples relating to the historical prices of the major crops over the past 10 years (2010-2020). This data comprises such variables as the name of a region, a market, a product, a variety, a classification, a floor price, a ceiling price, a mode price, and a date.

In order to carry out an analysis, the following pre-processing methods have been used. To ensure the attributability of data incompleteness, missing values were imputed, and to ensure trait measures were scaled on a consistent basis, standardization was done. Nominal variables, e.g. field location, crop type and fertilizer type, were coded into numbers. This data was then separated into train (70%), and test (30%) in order to adequately assess model performance. Several algorithms were tested, such as, logistic regression, decision trees, random forests, gradient boosting, AdaBoost, SVM, K-nearest neighbors, naive Bayes, and multilayer perceptron with bagging model returning the highest accuracy, reaching 99 percent.

Preliminary data

This significant phase is the preparation of the data set to be analyzed with attention. The imputation of missing values, standardization of trait scales, coding of categorical variables are the methods used. The categorical behavior,

that is the names of regions, browns, crops, and fertilizers are quantified into numbers referring to relationships through the machine learning algorithms [32, 33].

Coding of categorical variables

Category variables like field location, crop type and type of fertilizer used were coded into numeric values that would be used within the predictive models. This has to be done through the creation of a dictionary whereby certain categorical based items are assigned analogue indices. The result encoding mode can be used to good effect to merge the distributed information into further machine learning models.

Specification of input and output variables

The model has a couple of input variables whose requirements include the name of the region, color of the soil, nitrogen, phosphorus, potassium, a crop, pH, rain, and temperature. The forecast targets crop yield as variable of interest.

Model selection

The SVM, RF, NN, Naive Bayes, decision tree, bagging and boost machine are used to enhance the accuracy and power of forecasting [34].

Model evaluation

Known material libraries of classified models in the well-known Sci Learning are dedicated to analysis. The data is categorically split into training (70 percent of the information) and the testing (the remaining 30 percent) data sets. This categorization approach enables one to obtain general information on modeling data that he/she has never encountered before.

Training data

Data on agricultural history of such regions as Kolhapur, Solapur, Satara, Sangli, and Pune are used to predict the education, and support various soils, weather conditions and crop preferences.

Input parameters

Entries on domain name, brown, nitrogen, phosphorus, potassium, pH, rainfall, and temperature, and other available options based on the products are accepted as input.

Model selection

The performance of different models can be compared in choosing before taking into consideration the evaluation. Models run available include logistic regression, decision trees, random forests, gradient boosting, AdaBoost, bagging, SVM, K-nearest neighbors, naive Bayes and multilayer perceptron. Subsequently, the Bagging model proved to be the most optimal one with the accuracy rate of up to 99%.

Model output

Once the parameters have been Keyed in, the model makes farmers specifically recommendable crop and fertilizer choices.

Recommendations

Where there is no crop presently growing on the land then the model would indicate that the crop is optimal and would involve regimes of fertilizing the soil in order to yield more. Where there are crops already established, application of suitable fertilizers with the aim of supplementing nutrients that exist and enhancing crop health and yields is advised. Other farms and fertilisers are also recommended to ensure farmers have choices and variety hence cutting on risks taken [35].

Implementation and impact

This is built into an easy to use interface that enables the farmers to use through the web or mobile platform.

User interface

Farmers are capable of making crop selection and fertilizing decisions that are suitable to their types of soil and environment. The aim of developing the forecasting model was to increase agricultural production levels, resource effectiveness and sustainable agriculture among farmers.

Strengths

The modeling phase is personalized and based upon the insights gleaned out of data and therefore the farmer is able to beat the status quo and get optimal out of his/her efforts which ultimately helps in the economic growth of the villages.

Market Price Prediction

Model development

In this step, the value of major crops will be estimated in the target by Long Short Term Memory (LSTM) network to estimate time series analysis. Apply pre-processing analysis such as anomaly detection, optimization to clean the dataset before training it to maintain quality and consistency in it.

Completely reliable information:

This information was received by dint of agricultural websites run by various governments.

Pre-processing Technology

Advanced detection

Employ statistical methods or the use of machine learning algorithms to determine and solve issues with products. Replicate the information This will assist in minimizing the influence of reported data on the performance of the model and raise the reliability of the forecast in general.

Normalization

Transform the data into normalized scale to measure characteristics to a new value (normally zero to one). Normalizing all the input parameters into the training model plays the part of equality in contributing to the training process and favorable convergence during optimization.

Model selection

LSTM neural networks have been selected because they are able to take capture a long span of the data series thus more suited to the task.

Training data

Sangli and Pune have a history of trade in crops such as Jowar, Gram, Ginger, fruit, maize and rice of the past decade.

Input parameters

This Model is Region Name, Market Name, Product, Variety, Class, Min Price(Rupee), Max Price (Rupee), Modal Price(Rupee) and Date Price.

Model output

The LSTM model is developed such that it would provide forecasts (quarterly) of the minimum, maximum, and crop prices of the target region.

Predictive insights

Because farmers know the market in great detail, they can anticipate fluctuations in the market and change farming practices, farming and sales strategy accordingly. In price forecasting, quarterly forecasting increases the crop profit-making of the farmer through optimum crop selection and harvest and market accessibility and eliminates costly financial risk of price fluctuations.

Implementation and impact

The development of the intuitive web applications is achieved based on Django framework Existing storage bags and LSTM models are built with the framework of the Django framework. The web has the capacity to predict design which is facilitated through this integration.

Web Platform Integration

Facilitate the provision of farmers with estimates of the market price on the Django web site within the easy to use interface to offer validity and validity. By becoming a part of a centralized system, farmers will be able to incorporate market intelligence in their decision-making process maximizing profits making it smarter, and more profitable, farming. The packaging mode employed in the system is using a machine learning mechanism that contends with various information such as soil quality, climate and crop patterns as examples. The analysis of this model can offer the choice of the best crop to the farmer of a certain agricultural context through the availability of satisfactory and timely market price predictions that enable informed choice of crops and that also enable farmers to make better choices about business opportunities and their involvement thus resulting in greater profits and livelihoods. This adoption of market price forecasting shows that the company is interested in ensuring that its farming community is empowered by giving them the resources they require to make and maintain a good business in agriculture in India.

Internet of Things (IoT) Integration

Device design and components

Nowadays, the Internet of Things (IoT) devices is designed and manufactured in such a way that any information could be gathered within the ground in a short period of time and then be transferred to a location where it could be analyzed. The device consists of numerous parts each playing a certain role in gathering and transferring of information.

NPK sensor

NPK sensor is also fitted into this device to enable it to know the nitrogen, phosphorus, and potassium that is found in the soil. The NPK sensor deployed in this system is specially made to be very sensitive to the amount of nitrogen, phosphorus and potassium in soil. It measures 0- 2000 mg/kg per nutrient with precision of 1 mg/kg - accuracy of 5 mg/kg. The sensor has a temperature range between 0-50 o C and conveys the information through a DC value with regards to the concentration of the nutrients in the form of an analog voltage signal. It is fed by a 5V DC and connected easily with Arduino analog input. NPK data is transferred to a database through a serial interface with data properly formatted and compiled. The machine learning data has features intact around the levels of NPK, soil moisture and temperature had 1000 samples taken in various soil types. This data is subject to preprocessing to address noise and missing values in order to use it as a proxy in training machine learning models. Findings are illustrated in graphs and tabular forms which are clear and would present evidence on how different NPK measurements influence soil health and plant growth with the help of legends as well as annotations. All technical specifications of the system are recorded, Arduino installation, sensor calibration, etc.; reference to literatures guarantees reliability and precision of the methodology used.

RS485 TTL converter

This TTL converter is included to ensure that there is a link between the NPK sensor and the microcontroller (Arduino Uno R4). This converter enables transmission of sensor data in the RS485 format so that it can be used with Arduino boards.

Arduino uno R4

Arduino Uno R4 micro-controller would serve as a central stand-in processor to IoT devices. It obtains information coming in NPK sensors through TTL transformers and begins transporting it on the central platform. Data collection, operation and communication occurs with the aid of programmed Arduino boards.

OLED screen

The OLED screen module is incorporated into the device in order to have instantaneous display on location information. The screen saves farmers the stress of

accessing the soil through a third party and making crucial decisions about a location as well as being able make immediate interventions when need be.

Data Transmission and Analysis

After the dynamics of locations is comprehensively gathered by Internet of Things (IoT) devices, it is transmitted to the central office where it will be analyzed in more detail. Redirect the processes of data transfer via the usage of the serial communication and Wi-Fi capabilities of the Arduino Uno R4 microcontroller.

Serial communication

The Arduino board plays a role of transferring the data to the central processing unit and this is done through connection. This communication procedure is required in case the data collected by the Internet of Things (IoT) devices is securely and efficiently sent to the platform.

Wi-Fi connection

Along with the communicational connection, there is also the Wi-Fi connection to Arduino Uno R4 through the use of the ESP32 module. This enables data to be transferred wirelessly into the central platform giving flexibility and convenience of data transfer.

Centralized platform

Information, which is sent down the ground is received and processed by the central system, which was developed using the Django framework. The platform contains a web access interface in which the data analysis and visualization tools could be employed. The framework of Django permits smooth integration of the information on different sources and enables farmers to receive immediate overview.

Implementation and impact

Application of Internet of Things (IoT) technology in agriculture has a number of advantages such as the possibility to gain an insight into the specifics of soil in real-time, prevent excessive nutrient content, and make decisions based on the analysis of the data.

Real-time monitoring

Internet of Things (IoT) devices track information about soil in real-time to allow farmers to evaluate the condition of the soil and change the way they fertilise it in time.

Proactive management

With the IoT devices, there is instant response to availability of food giving the farmers the chance to address food shortage and fertilizer consumption enhancing growth and production standards.

Data-driven decision making

This is because of the central platform as it assists in the management of data analysis and insights, thus farmers make real time decisions based on the information delivered through food. With the aid of the internet of things (IoT) technology, farmers can integrate real-time farming in their process to make those farming activities much more efficient and sustainable.

Circuit Diagram and Block Diagram

Figure 1 shows circuit diagram of the predictive model system performed during the research. This diagram gives a more in-depth visual breakdown of the parts and connections necessary in the implementing the predictive model. Its major elements are sensors that determine the content of soil nutrients, an Arduino microcontroller to process data and data communication interfaces to transfer information. The circuit also combines several aspects to make sure that

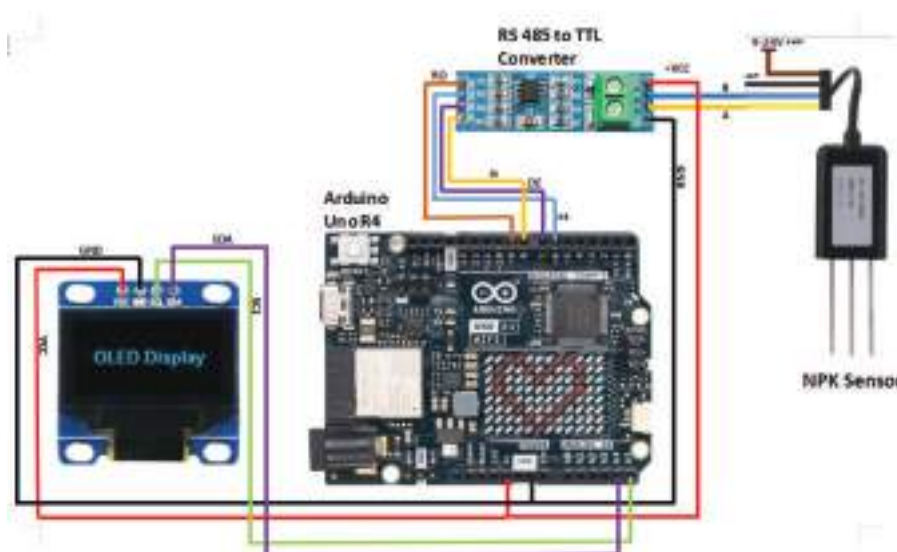


Figure 1. Circuit diagram of predictive model.

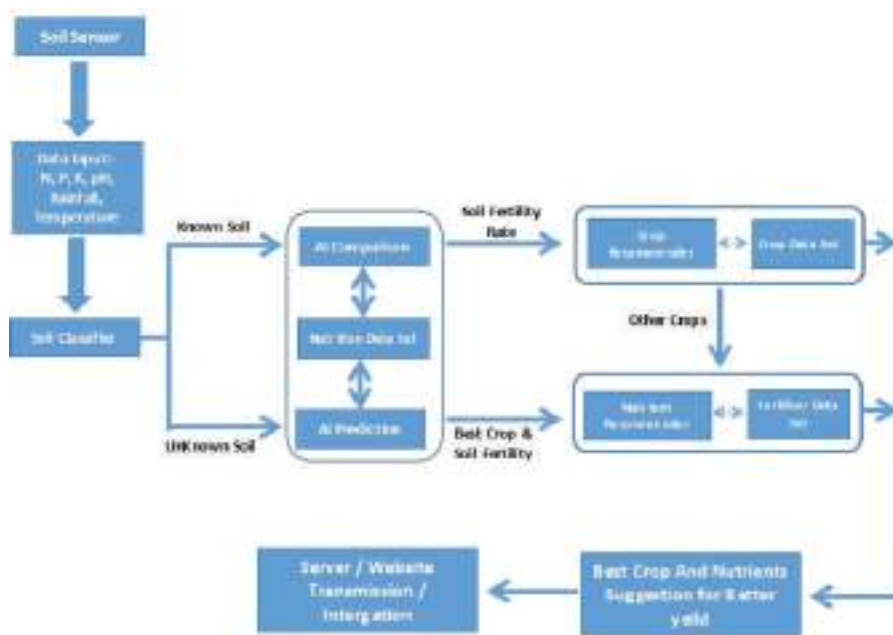


Figure 2. Block diagram of predictive model.

actual dataS in the field is precisely presented, processed and changed to centralization platform. Such an arrangement is the core of the predictive model that allows to make accurate crop and fertilizer recommendations characterizing soil conditions.

Figure 2 shows the block diagram of the proposed predictive model with its data flow direction and the connecting of all the modules. The diagram indicates the mechanism behind this step which includes data acquisition using the sensors to the data processing on the Arduino microcontroller to the analysis of data using the machine learning algorithms. The block diagram also indicates how

the hardware items and the software systems interact, and the chain of events that produces some insights that make predictions. This decoupled design—in that it is modular in nature and scaleable and flexible to incorporate new sensors and analytical capabilities to improve upon the models accuracy and functionality--will enable us to add more data points and increased accuracy and usefulness to the model.

Figure 3 shows the predictive system model that has been developed. This prototype is able to consolidate all the hardware components such as sensors, microcontrollers and communication modules in a consistent and useful framework. The prototype is a kind of model of the theoretical model which can be tested and validated in practice by the parameters of the system. It demonstrates implementation of the circuit and block diagrams in real life application confirming that the model can be practically used on the field to gather and analyse information in order to give crop and fertilizer recommendations.

RESULTS AND DISCUSSION

Predictive Modeling Results

With the predictive modeling process, encouraging results were achieved that offered recommendations on crop choices and fertilizer use by offering a 99% accurate forecast on environmental and soil conditions. Overall, the classification model scored accuracy of 99% on the testing dataset, which shows that it works effectively to make proper recommendations on the type of crops and fertilizers that farmers should grow.



Figure 3. Block diagram of predictive model.



Figure 4. Sample testing done on prototype.

The model is easy to read and displays appropriate crops and fertilizers according to such input as the name, terrain, rainfall, and temperature. The farmers mentioned that they were highly satisfied with the presented recommendations and were in agreement that it was environmental friendly and went along with local farming.

The illustration Figure 4 is the sample testing performed with the prototype model. This image illustrates the functioning of the prototype showing its capabilities to collect the soil data and process information with the help of the integrated system. The testing stage is significant to prove testing the accuracy and reliability of predictive model. It entails a comparison of the carbon footprints in the prototype to the actual field and crop performance, which further make the system more useful and able to give farmers with an actionable and reliable information. The prototype testing provides

the demonstration of the practical applicability and the efficacy in the agricultural working environment.

Impact on agriculture

The farmers who put into practice the recommendations of the model experience growing yields, less fertilizing and they become more efficient. The predictive model assists in adopting permaculture practices and to ensure economic improvement of farmers by proposing recommendations pertaining to particular farming areas.

Table 1 discusses the difference between accuracy, precision, and recall of the use of various machine learning models, including logistic regression, decision tree, random forest, and gradient boosting. Gradient Boosting and Random Forest have the best accuracy and hence can make the best recommendations on crops and fertilizers to be used to increase agricultural productivity and sustainability.

Table 1. Model compatibility result

	Logistic regression	Decision tree	Random forest	Gradient boosting
Accuracy precision recall	0.615953	0.992614	0.999261	0.998523
	0.410916	0.97885	0.997596	0.996422
	0.449497	0.992472	0.998922	0.996422

Table 2. Model compatibility results

	AdaBoost	Bagging	SVM	K-Nearest Neighbor
Accuracy precision recall	0.463811	0.996307	0.593058	0.937962
	0.264857	0.996814	0.351363	0.897532
	0.280842	0.99317	0.299619	0.888587

Table 3. Model compatibility results

	Naive bayes	Multilayer perceptron	XGBoost	CatBoost
Accuracy precision recall	0.655096	0.815362	1	0.994092
	0.549116	0.68206	1	0.993303
	0.632563	0.663501	1	0.986283

Table 2 compares the additional machine learning models by its accuracy, precision and recall: AdaBoost, Bagging, SVM and K-Nearest Neighbor. Bagging has the best performance reports implying that it is powerful in delivering the right crop and fertilizer prescription to support agriculture and sustainability in practice.

Table 3 also compares accuracy, precision, and recall of such models as the Naive Bayes, Multilayer Perceptron, XGBoost, and CatBoost. The excellent results in accuracy and recall here demonstrated by XGBoost and CatBoost demonstrate that techniques are efficient in predicting crop yields and making agricultural decisions to improve productivity.

The regression line in Figure 5 is the difference between the actual and the predictive result generated by machine learning model. Here the line gives the figure of how much the model is accurate in predicting the actual observation and the difference between the best line and the actual observation is the deviation. The regression line can help provide information on the bias and variability of the prediction model along with the distribution of the data points around the regression line may provide information regarding the accuracy of the consistent nature of the model. Regressions analysis can enable those who are conducting research to assess the overall effectiveness of

machine learning models and define where the accuracy can be increased in the prediction.

Market Price Prediction Results

The price forecast level that is market price forecast would enable the farmers to make decisions and decisions on crops, and marketing decisions based on an understanding of the differences in prices of major crops in the target region. The LSTM model reported reasonable outcomes in predicting the price of the quarter with an accuracy of 70% -75 % of all crops.

Price forecast accuracy

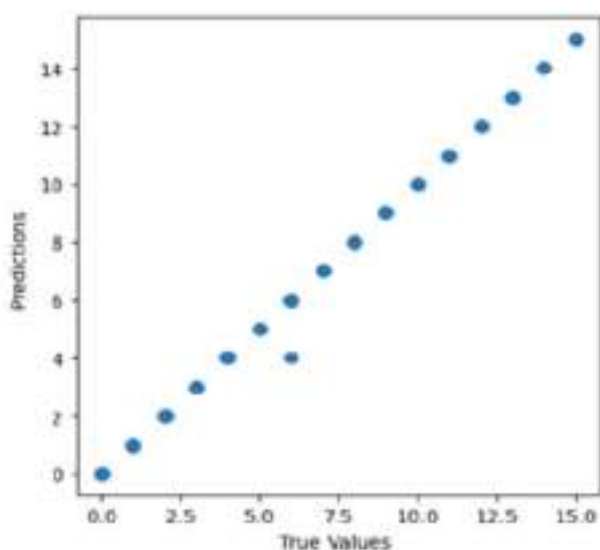
Perfectly contrasting the seasonal variation and the long-term market trends on the prices, the LSTM model offers a rationale to the farmers with regard to the minimum-upper-end and price range. It is recommended that farmers should utilize these forecasts to make better crop choice, timing of sales, and market engagement and thus augment their profitability and shield them the financial risk.

Value for farmers

Access to the right market prices helps farmers to negotiate better prices, long-term planning of production cycles and diversity of crops to meet the needs of the economy. By linking agriculture with the vibrant market, farmers answered this by increasing the level of the farm price prediction and improve the convenience and the competitiveness of farmers in business.

Figure 6 indicates the learning curve of the long short-term memory (LSTM) framework on the training process of predicting ginger prices. The curve shows the performance of the model in each successive epoch indicating the reduction in both training and validation loss. Convergent, well-behaved learning curves show that the model is learning well and there is good generalization highlighting the fact that the validation loss follows the training loss.

Figure 7 depicts the LSTM learning curve of gram: this is the model training, validation performance measured over different epochs. It demonstrates the incidence of the model loss under the impact of the training in the data, to minimize the variance of the predicted and actual values. The curve aids in determining whether the model is over-fitting, under-fitting or doing their best the curve can also aid in determining the model that is applicable in solving a certain task.

**Figure 5.** Actual vs. predicted values.

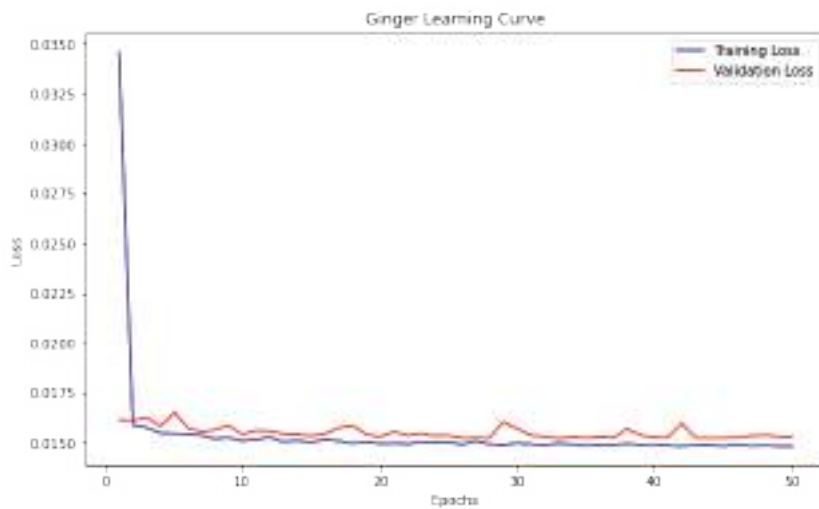


Figure 6. LSTM learning curve for ginger.

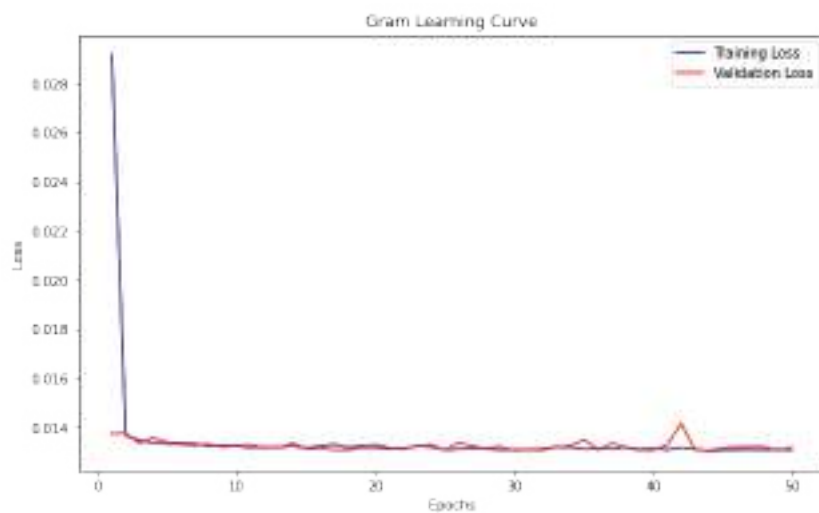


Figure 7. LSTM learning curve gram.

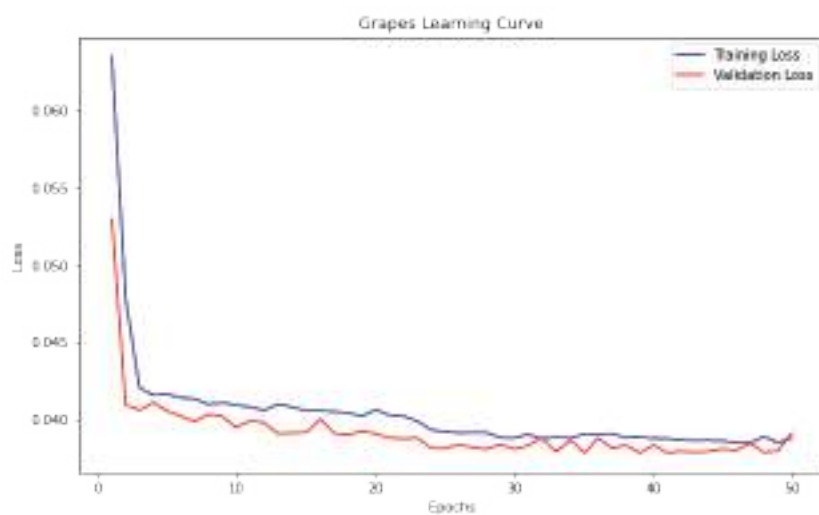


Figure 8. LSTM learning curve for grapes.

Figure 8 shows the loss values exhibit an LSTM learning curve during the training and validation of grapes. The consistent down turn in the values of losses indicates the growing precision by the model in forecasting the grape prices. The said curve gives an insight into the efficiency of the learning during the model and is used to refine the model to attain enhanced predictive capabilities.

The learning curve on maize depicts the improvement in the error at successive stages of training of the LSTM model in Figure 9. This figure plays an important role in assessing how far the given price trends of maize are reflected by the model. The closeness of the training and validation curve signifies that this model will be well suited to making predictions on new data and the effectiveness of this model in price forecasting.

Figure 10 illustrates the LSTM learning curve on Jowar to show the progression of loss by the model throughout training. The learning graph will be instrumental in determining the effectiveness of the model as well as the appropriate number of epochs in order to avoid overfitting. The tradeoff in training and validation loss is also satisfactory showing that a stable predictive model was compiled.

Figure 11 LSTM learning curve of wheat demonstrates the training and validation losses across the epochs that illustrate how the model learns. This value is critical to determine that the model is forming the regularities in the wheat price data. An aligned curve that shows that the values of loss fall is employed to indicate an effective training process with the possibility of the correct prediction of prices.

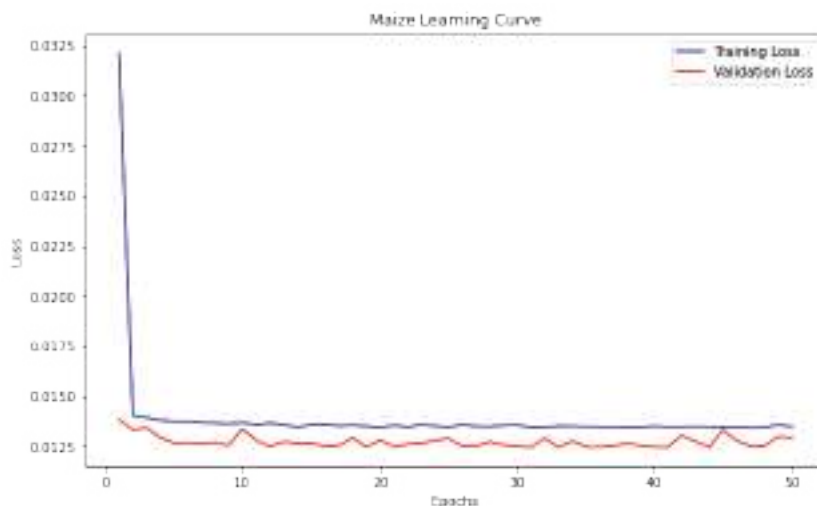


Figure 9. LSTM learning curve for maize.

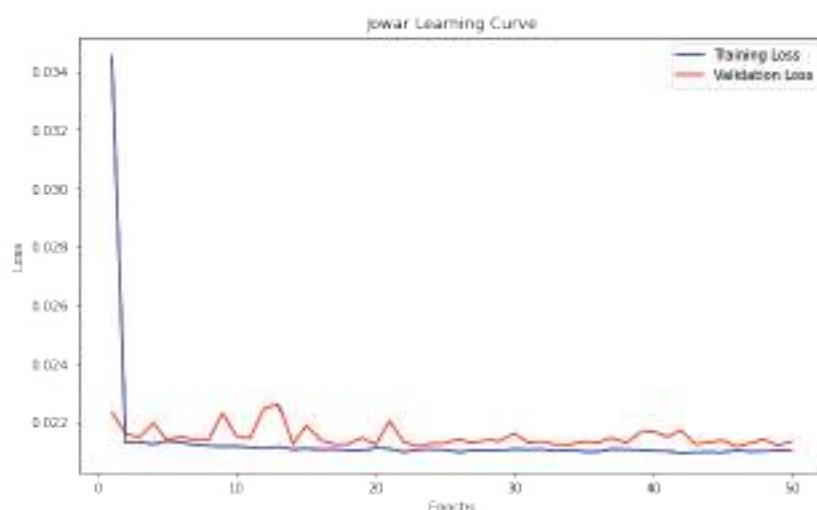


Figure 10. LSTM learning curve for Jowar.

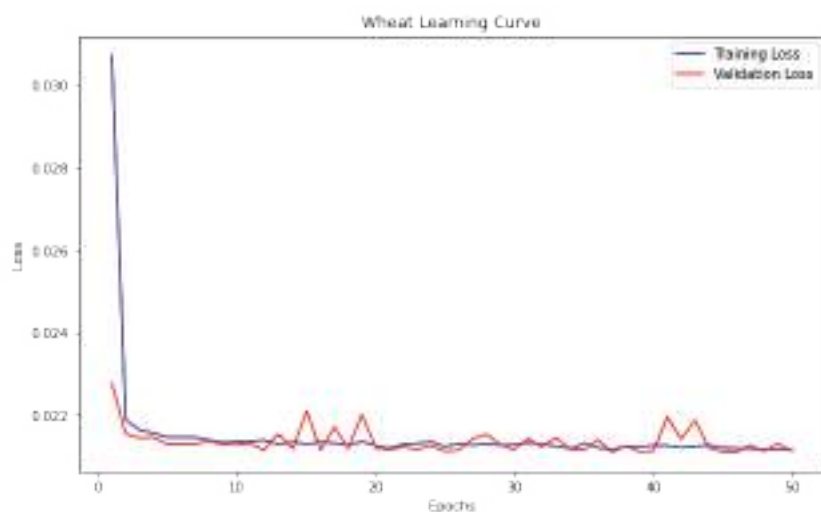


Figure 11. LSTM learning curve for wheat.



Figure 12. Average price change for maize.

Through Figure 12, the price change of maize can be averaged out over a specific time. It gives the visual depiction of price trends giving a clue on the changes to the prices and patterns that are important in understanding the market trend. Study of these changes can aid in decision making as far as maize cultivation and marketing concerns are concerned.

Figure 13 presented here indicates the average change in price of grapes which highlights the changes in the grape prices with time. Seasonal trends are important to market volatility, which can also help the growers and the traders in maximising their marketing and production measures. It is in knowing these price dynamics that profitability becomes the main area of focus.

Figure 14 represents the average change in prices of wheat and gives information about the temporal changes in

prices of wheat. Looking at these changes, stakeholders are in a better position to determine the factors that pressure wheat prices and adapt their planning and operations in one way or another. This is an essential analysis to participate well in the market or risk management.

Figure 15 provides a view of average price change in ginger over an applicable time. Such information is vital because ginger producers and marketers need to monitor the market forces, future price expectation, and adopt strategic management prompted by the willingness to maximize profit and market competitiveness.

To demonstrate the changes in gram prices at each point in time, Figure 16 shows average price change in gram. Interpretation of such changes enables farmers and traders to predict the market, and devise their production Kuch IE



Figure 13. Average price change for grapes.



Figure 14. Average price change for wheat.



Figure 15. Average price change for ginger.

bates as well as design their sales strategies to obtain better economic returns.

Figure 17 illustrates average Jowar sales price change representing the change in the price over a specific period of time. The trends in the sector are critical to be analyzed by stakeholders in order to regulate supply chains efficiently and take necessary decisions that will fit market conditions and consumer needs.

The research involves the actual versus predicted least and the maximum prices of a series of crops over a quarter period using long short-term memory (LSTM) models. These crops have been selected as Jowar, maize, gram and grapes, wheat and ginger.

Tables 4-7 demonstrate how the LSTM models performed in each of the crop forecasting minimum and maximum prices. This quarterly breakdown identifies the overall

prediction performance of this model overtime which can reveal the seasonality patterns or any market patterns.

Mal-predictions can be disastrous to the parties involved in the agricultural sector and thus accurate and precise predictions help in rational decision-making processes that can be used to subsidize on production, pricing, and the intervention process. In comparing the real prices and LSTM-predicted prices, it can determine the effectiveness of the models to pick the dynamism and fluctuation of prices in agricultural markets.

With this insight, hopes to contribute to gain traction in predictive analytics in the agricultural field, ensuring stronger supply chains and empowering farmers and policymakers to make informed decisions to drive sustainability in agricultural growth.

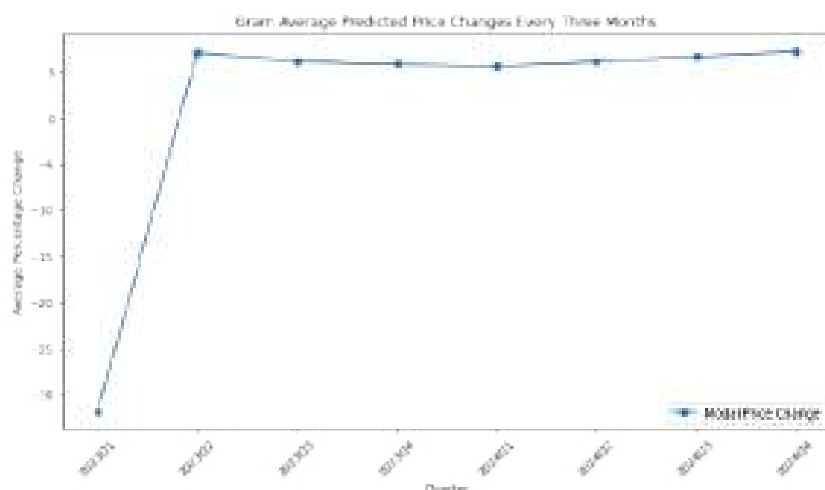


Figure 16. Average price change for gram.



Figure 17. Average price change for Jowar.

Table 4. Actual vs. predicted price from January to March

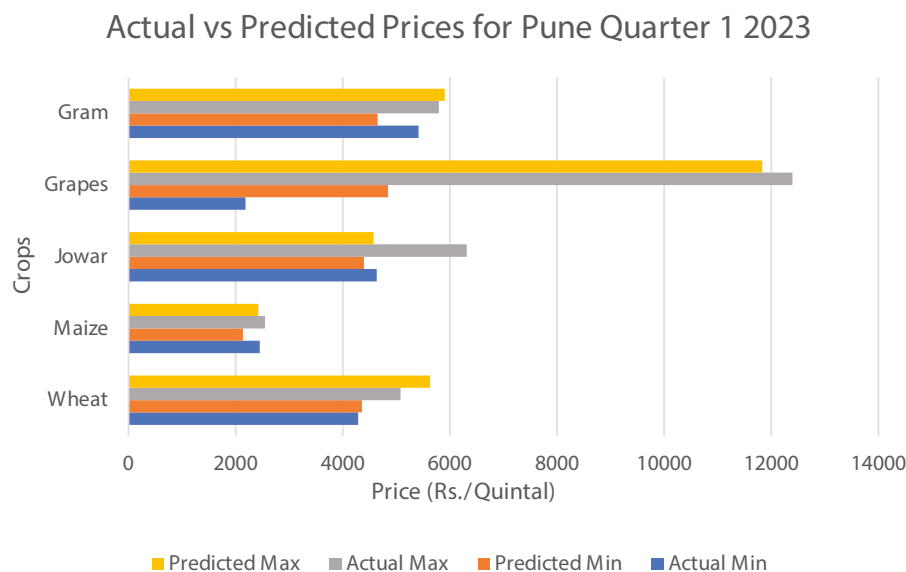
Quarter 1	Actual Min	Predicted Min	Actual Max	Predicted Max
Wheat	4289.55	4357.412	5080.59	5631.452
Maize	2450	2138.38	2550	2424.48
Jowar	4634.32	4395.06	6313.42	4576.92
Grapes	2186.56	4844.94	12395.52	11828.98
Gram	5414.92	4649.72	5792.53	5900.63

Table 4 compares real vs. forecasted prices of Wheat, Maize, Jowar, Grapes and Gram between January and March. The LSTM models demonstrate a good predictive power especially in Wheat and Gram, which enables the farmers to plan better crop sales markets and their engagements therein.

As Figure 18 shows, the predictive model is accurate since it shows a relatively close parallel between the actual and the predicted prices during the period of January to the month of March. The correspondence between the actual and the predicted values signifies the level at which

the model applies in predicting prices. It is these kinds of comparisons that are important in testing the model and enhancing the predictability of the model.

Table 5 shows the correlations between the actual and the estimated prices of the same crops in April, May and June. Compared to the LSTMs, the LSTM counterparts are more accurate in foreseeing the market conditions regarding Wheat and Maize and can serve the farmer as a predictive concept that would lead to appropriate responses in the area of strategic decision-making and financial benefit.

**Figure 18.** Actual vs. predicted price from January to March.**Table 5.** Actual vs. predicted price from April to June

Quarter 2	Actual min	Predicted min	Actual max	Predicted max
Wheat	3936.36	4457.13	4854.54	5666.6
Maize	2450	2138.38	2550	2424.48
Jowar	4333.65	4402.3	5372.72	4581.35
Grapes	3192.98	5332.57	11763.15	12682.87
Gram	5473.68	5662.35	5832.89	5735.85

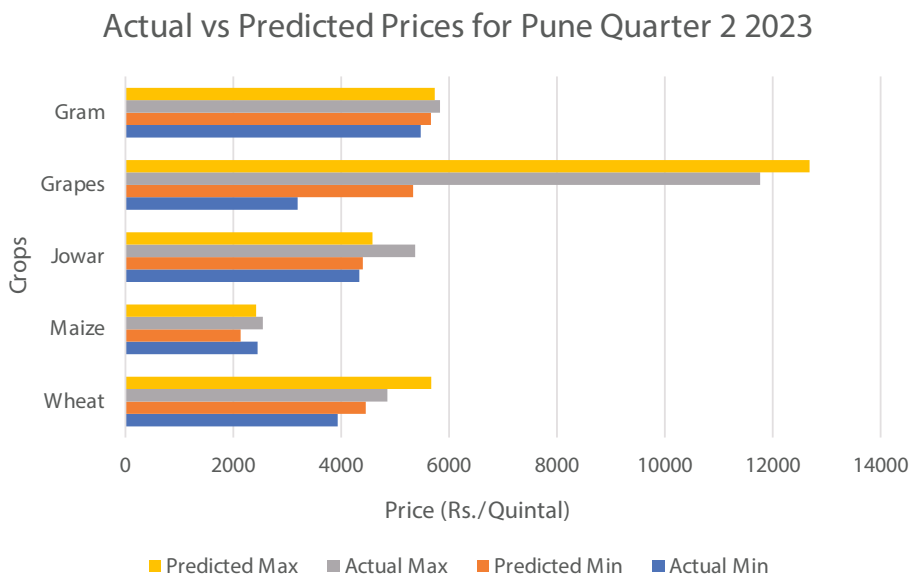


Figure 19. Actual vs. predicted price from April to June.

In Figure 19, the model forecast prices have been compared with actual prices between April and June indicating the way the model has performed over this period. The low difference between the actual and predicted prices shows the effectiveness of the model in capturing the trend in prices and eases the expectation of it being a powerful model making predictions relying on the past.

Table 6 presents the actual and the forecasted prices of crops between July and September. The LSTM models prove to be stable, especially in the case of Gram and Wheat, so that farmers can change their plans on production and sales operations depending on market tendencies in order to gain profits the most.

Various graphical representation is shown in figure 20 comparing actual and predicted price between July to September. The veracity of the model can be observed in the extent of similarity between the two groups of prices. Such observations play a crucial role in determining the performance of the model and making the relevant changes to boost or optimize its forecasting capacity.

Table 7 plots the comparison between the actual and forecasted minimum and maximum prices of Wheat, Maize,

Jowar, Grapes and Gram between October to December. The LSTM models have contrasting accuracy ranging well above actual prices, Wheat and Maize and below, Grapes, where farmers can make informed market decision making.

It can be seen in Figure 21 that the comparison of actual and predicted prices is at an equal level. The figure assists in the assessment of the model in prediction performance during this period. Such precision in predictions as expressed through the comparison underlies the usefulness of the model in price forecasting in real-time so as to help in making improved decisions by agricultural stakeholders.

This paper augments the content of the review of actual and forecasted lows and highs of prices of major crops with visual analysis through the inclusion of bar graphs. Such graphs, based on Tables 4 to 7, provide a brief and user-friendly picture of the comparative performance of the LSTM models over time spanning quarterly periods. By comparing the actual prices with the LSTM forecast prices graphically, stakeholders are able to identify the accuracy and efficiency of the forecasting models over time easier.

Table 6. Actual vs. predicted price from July to September

Quarter 3	Actual min	Predicted min	Actual max	Predicted max
Wheat	4257.89	4533.57	5256.57	5751.45
Maize	2400	2194.61	2500	2502.63
Jowar	5150.66	4404	6016	4582.2
Grapes	4032.25	6929.26	10919.35	17622.05
Gram	5773.68	6613.71	6189.47	6832.53

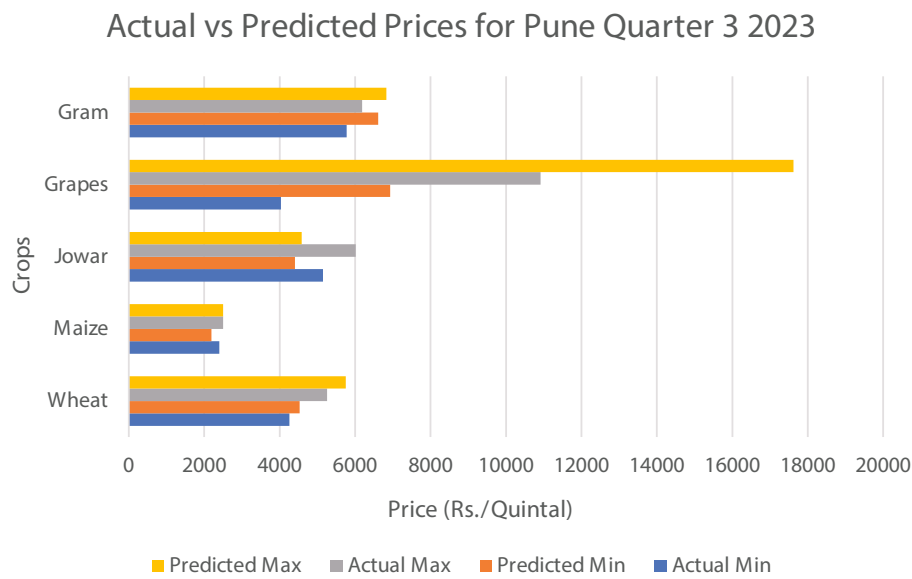


Figure 20. Actual vs. predicted price from July to September.

Table 7. Actual vs. predicted price from October to December

Quarter 4	Actual min	Predicted min	Actual max	Predicted max
Wheat	4384	4686.79	5464	5932.46
Maize	2500	2250.16	2650	2580.82
Jowar	5574.5	4404.56	6633.33	4582.51
Grapes	5000	8414.85	12750	22062.43
Gram	6381.94	7457.901	7393.05	7612.94

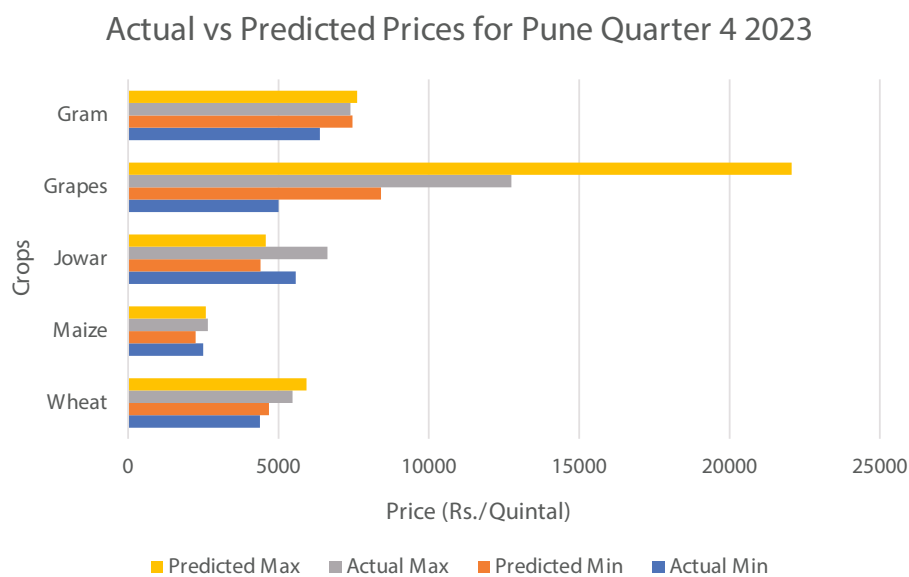


Figure 21. Actual vs. predicted price from October to December.

Bar graphs promote the readability of results and allow moving and comparing prices of various crops and quarters in a few seconds. Also, these visualizations play the role of an efficient communication tool, and complex analytical findings can be presented in a comprehensible and manageable way. Since bar graphs will be incorporated into this study, it will contribute to the better understanding of readers and, consequently, the better effectiveness of the conducted analysis, enabling the readers to draw practical conclusions related to agricultural decision-making and policy-making.

Overall Impact and Implications

Overall, the findings noted in the model forecast, the market value and the level of integration of the Internet of Things (IoT) reflect the dynamics of technological changes in the process of transformation. The program gives farmers custom recommendations, predictive data, and real-time monitoring and helps farmers overcome traditional obstacles, own and access permaculture benefits.

The adoption of technological products and services has enhanced the health status of the Indian farmers, improved their lives and also empowered the remote communities in the city. The program helps in preventing the weakening of agriculture due to the competition of this sector by enhancing precision agriculture and market-oriented agricultural approaches.

This dependence on both historical and real-time data presents a possible accuracy shortfall as a result of data quality problems, thus, impacting on model performance. The LSTM models, as much as they show great accuracy, still need to be refined more to give enough consideration to market dynamics and the conditions of the environment. Moreover, the technological framework that is needed an Internet of Things (IoT) implementation might not be easily available in the countryside, thus, precluding their large-scale usage. The way of scaling up to new crops and regions is problematic. It is recommended that future research looks into adding more detailed data sets, like satellite images and detailed weather predictions, to increase level of predictive accuracy. It is essential to further develop complex analytical tools, as well as use AI and blockchain technology to ensure greater data transparency and security, along with designing user-friendly interface design that would suit farmers. To ensure long term effects of long-term impacts of these intervention, longitudinal studies are also required as sustainable agricultural practices. The reviewed areas are the ones that should be worked on in order to promote the success of agricultural innovations and their inclusivity.

CONCLUSION

The work is responding to the high-priority problem of yield estimation in agriculture that preoccupies farmers and agriculture organizations conducts research, as well as uses

state-of-art predictive modeling, and Internet of Things (IoT) technologies. The study is one of the milestones in the history of Indian agriculture breaking new ground to revolutionize the native culture and the introduction of the latest technologies. The study is achieved by blending machine learning, Internet of Things (IoT), and data science to take the most critical challenges of the Indian farmers and deliver actionable insights and new tools to them so that they can optimize their farming activity. The resultant predictive models are very accurate; they give a prediction of 99 percent accuracy when it comes to prescribing crops and fertilizers in accordance with particular soil and climate conditions. Not only does this increase agricultural yields but also increases sustainable farming because this will reduce dependency on the use of fertilizers.

By incorporating the Long Short-Term Memory (LSTM) networks into market price prediction (accuracy of 70-75 percent), the farmers will be provided with an opportunity to gain strategic option in relation to the choice of crops, plan harvests, and making market decisions. This predictive ability promotes profitability and reduces the risks of losses in case of financial losses, and farmers make clear decisions in a rapidly changing market environment.

Additionally, Internet of Things (IoT) monitoring equipment to determine the soil conditions in real-time is implemented, which gives valuable data about nutrients. This allows taking proactive nutrient control and the centralized platform that is developed with the Django framework supports the same. Its interface is easy and convenient to use, so farmers will have no difficulties using this technology.

It is a holistic manner of harnessing resources and not only maximizing use of resources but also creates economic stability in the rural economy. The use of new technologies in every farming aspect motivates the economy and its absorption potential because the farmers tend to survive in an adaptive world of agriculture.

To sum it all up, this study establishes a strong background of sustainable and efficient future of Indian agriculture. Data-driven intelligence and cutting-edge tools empower farmers to be prepared to handle and face the challenges of modern agriculture.

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] Sk Al Zaminur Rahman, SM Mohidul Islam, Kaushik Chandra Mitra. Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series. In: 21st International Conference of Computer and Information Technology (ICCIT); 2018 Dec 21–23; Dhaka, Bangladesh.
- [2] Priya P, Muthaiah U, Balamurugan M. Predicting yield of the crop using machine learning Algorithm. *Int J Eng Sci Res Technol* 2018;7:2277–2284.
- [3] Ruß G, Kruse R, Schneider M, Wagner P. Estimation of neural network parameters for wheat yield prediction. In: Bramer M, editor. *Artificial Intelligence in Theory and Practice II*. IFIP International Federation for Information Processing. vol. 276. Berlin: Springer; 2008. p. 109–118. [\[CrossRef\]](#)
- [4] Bhat SA, Huang NF. Big Data and AI Revolution in Precision Agriculture: Survey and Challenges. *IEEE Access* 2021;9:110209–110222. [\[CrossRef\]](#)
- [5] Maya-Gopal PS, Chintala BR, Others. Big Data challenges and opportunities in agriculture. *Int J Agric Environ Inf Syst* 2020;11:48–66. [\[CrossRef\]](#)
- [6] Torky M, Hassanein AE. Integrating blockchain and the internet of things in precision agriculture: Analysis, opportunities, and challenges. *Comput Electron Agric* 2020;178:105476. [\[CrossRef\]](#)
- [7] Rao TVN, Manasa S. Artificial Neural Networks for Soil Quality and Crop Yield Prediction using Machine Learning. *Int J Future Revol Comput Sci Commun Eng* 2019;5:1
- [8] Pouyanfar S, Yang Y, Chen S-C, Shyu M-L, Iyengar SS. Survey on deep learning: algorithms, techniques and applications. *ACM Comput Surv* 2019;51:1–36. [\[CrossRef\]](#)
- [9] Yadav J, Chopra S, Vijayalakshmi M. Soil analysis and crop fertility prediction using machine learning. *Int J Innov Res Adv Eng* 2018;8:41–49. [\[CrossRef\]](#)
- [10] Trontelj J, Chambers O. Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. *Sensors* 2021;21:42208. [\[CrossRef\]](#)
- [11] Benos L, Tagarakis AC, Dolias G, Berruto R, Kateris D, Bochtis D. Machine learning in agriculture: a comprehensive updated review. *Sensors (Basel)* 2021;21:13758. [\[CrossRef\]](#)
- [12] Trendov NM, Varas S, Zeng M. Digital technologies in agriculture and rural areas: status report. Rome: Food and Agriculture Organization of the United Nations; 2019.
- [13] Khelifa B, Amel D, Amel B, Mohamed C, Tarek B. Smart irrigation using Internet of Things. In: 4th International Conference on Future Generation Communication Technology (FGCT); 2015; Luton, UK. p. 91–96. [\[CrossRef\]](#)
- [14] Sales N, Remedios O, Arsenio A. Wireless sensor and actuator system for smart irrigation on the cloud. In: World Forum on Internet of Things (WF-IoT); 2015; Milan, Italy. p. 693–698. [\[CrossRef\]](#)
- [15] Veenadhari S, Misra B, Singh C. Machine learning approach for forecasting crop yield based on climatic parameters. In: International Conference on Computer Communication and Informatics; 2014; Coimbatore, India. p. 1–5. [\[CrossRef\]](#)
- [16] Wu J, Olesnikova A, Song CH, Lee WD. The Development and Application of Decision Tree for Agriculture Data. In: Proceedings of IITSI; 2009. p. 16–20. [\[CrossRef\]](#)
- [17] Kiranmai C, Murali Krishna IV, Venugopal Reddy A. Data mining of geospatial database for agriculture related application. In: Map India; 2006; New Delhi, India. p. 83–96.
- [18] A Multiple Linear Regressions Model for Crop Prediction with Adam Optimizer and Neural Network. *Int J Adv Comput Sci Appl* 2020;11:253–257. [\[CrossRef\]](#)
- [19] Agrawal VK, Patil LN, Jadhav SS, Chavan K, Nimbalkar U, Gadhave S, Javanjal V. Statistical and mathematical modeling of temperature dynamics in automotive brake components: A systematic review. *Sigma J Eng Nat Sci* 2025;43:1417–1437. [\[CrossRef\]](#)
- [20] Khaki S, Wang L. Crop yield prediction using deep neural networks. *Front Plant Sci* 2019;10:621. [\[CrossRef\]](#)
- [21] Suruliandi A, Mariammal G, Raja SP. Crop prediction based on soil and environmental characteristics using feature selection techniques. *Math Comput Model Dyn Syst* 2021;27:117–140. [\[CrossRef\]](#)
- [22] Zhao Z, Chow TL, Rees HW, Yang Q, Xing Z, Meng FR. Predict soil texture distributions using an artificial neural network model. *Comput Electron Agric* 2009;65:36–48. [\[CrossRef\]](#)
- [23] Gadhave S, Kosode S, Javanjal V, Patil LN, Agrawal VK, Jadhav SS, Toke LK. Enhancing agricultural efficiency in India with IoT based smart boats. *Sigma J Eng Nat Sci* 2025;43:570–581. [\[CrossRef\]](#)
- [24] Vibhute AD, Kale KV. Mapping several soil types using hyperspectral datasets and advanced machine learning methods. *Results Optics* 2023;12:100503. [\[CrossRef\]](#)
- [25] Bejo S, Mustaffha S, Ishak W, Wan Ismail WI. Application of Artificial Neural Network in Predicting Crop Yield: A Review. *J Food Sci Eng* 2014;4:1–9.

- [26] Jadhav SS, Agrawal VK, Patil YM, Patil LN, Khan M, Pansare RM, et al. Bone fracture detection using image processing techniques. *Sigma J Eng Nat Sci* 2025;43:748–759. [\[CrossRef\]](#)
- [27] Patil LN, Jadhav SS, Waghulde KB, Gaikwad P, Duchal S, Patil S, et al. Precision mapping and navigation: A robotic restaurant management system using SLAM and ROS. *Sigma J Eng Nat Sci* 2025;43:1233–1247. [\[CrossRef\]](#)
- [28] Javanjal VK, Gadhave S, Patil LN, Mahajan KA, Jadhav SS. Efficiency improvement of semi-evaporative cooling systems through environmental analysis. *Sigma J Eng Nat Sci* 2025;43:1113–1123. [\[CrossRef\]](#)
- [29] Nugraha AT, Prayitno G, Hasyim AW, Roziqin F. Social Capital, Collective Action, and the Development of Agritourism for Sustainable Agriculture in Rural Indonesia. *Evergreen* 2021;8:1–12. [\[CrossRef\]](#)
- [30] Patil LN, Patil AA, Waghulde KB, Patil SA, Patil YM, Gadhave SL, Javanjal VK, Jadhav SS. Finite Element Analysis for Improved All-Terrain Vehicle Component Design. *Evergreen* 2023;10:1508–1521. [\[CrossRef\]](#)
- [31] Agrawal VK, Bogireddy S, Patil LN, Jadhav SS, Swami M, Shivaji S, Bhangade O, Jagtap A. Enhancing operational efficiency through robotic process automation in e-commerce. *Sigma J Eng Nat Sci* 2025;43:1248–1264. [\[CrossRef\]](#)
- [32] Onesimu JA, Kadam A, Sagayam KM, Elngar AA. Internet of things based intelligent accident avoidance system for adverse weather and road conditions. *J Reliable Intell Environ* 2021;7:299–313. [\[CrossRef\]](#)
- [33] Pramanik S, Sagayam KM, Jena OP. Machine Learning Frameworks in Cancer Detection. *E3S Web Conf* 2021;297:01073. [\[CrossRef\]](#)
- [34] Mhathesh TSR, Andrew J, Sagayam KM, Henesey L. A 3D Convolutional Neural Network for Bacterial Image Classification. In: Peter J, Fernandes S, Alavi A, editors. *Intelligence in Big Data Technologies—Beyond the Hype. Advances in Intelligent Systems and Computing*. vol. 1167. Singapore: Springer; 2021. [\[CrossRef\]](#)
- [35] Sundar GN, Narmadha D, Anton Jone A, Sagayam KM, Dang H, Pomplun M. Automated sleep stage classification in sleep apnoea using convolutional neural networks. *Informatics Med Unlocked* 2021;26:100724. [\[CrossRef\]](#)



Research Article

Unbiased weighted least squares estimation of Weibull modulus with small samples

Burak BİRGÖREN^{1,*}

¹Department of Industrial Engineering, TOBB University of Economics and Technology, 06510, Türkiye

ARTICLE INFO

Article history

Received: 18 March 2024

Revised: 12 June 2024

Accepted: 05 August 2024

Keywords:

Ceramics; Composite Materials; Fracture Strength; Reliability; Weibull Distribution; Weibull Modulus; Weighted Linear Least Squares; Unbiased Estimation

ABSTRACT

Weibull distribution has been used vastly for modeling the fracture strength of ceramic and composite materials. Estimating the Weibull modulus is an important problem, particularly in cases where the sample size is small due to high experimental costs. The Ordinary Least Squares (OLS) method is the most commonly used method by materials scientists. Numerous probability estimators have been proposed using the OLS method, and most of these studies focused on unbiased estimation. Weighted Least Squares (WLS) is a promising alternative to OLS, yet, there are only a few studies on this subject. This work followed a systematic analysis to develop a new probability index for unbiased estimation of Weibull modulus using WLS. For sample sizes less than 60, the performance of the new index is shown to be significantly superior to currently used OLS methods and is better or as good as the Maximum Likelihood Estimation method.

Cite this article as: Birgören B. Unbiased weighted least squares estimation of Weibull modulus with small samples. Sigma J Eng Nat Sci 2025;43(5):1683–1692.

INTRODUCTION

The Weibull distribution is one of the most widely used probability distributions in life testing and reliability studies. The distribution was developed by Weibull [1] and has found use in many different areas of application such as wind-speed data analysis [2–5], unemployment duration analysis, survival analysis as well as reliability analysis [6]. One major application area is the modelling statistical variation in the fracture strength of many materials such as advanced ceramics, metallic matrix composites and ceramic matrix composites [7–9], and strength of micro and nano structures [10]; it is also used to describe the

fracture toughness behavior of steels in ductile-brittle transition region [11], and fatigue behavior of metals [12,13].

The two-parameter Weibull distribution function is given by

$$F(s) = 1 - \exp \left[- \left(s / s_0 \right)^m \right] \quad (1)$$

where m is the shape parameter or Weibull modulus, s_0 is the scale parameter or characteristic strength of the distribution, and F is the fracture probability of the material at or below uniaxial tensile stress s . Weibull modulus, m , reflects the inherent variability in strength of a material: the higher

*Corresponding author.

*E-mail address: b.birgoren@etu.edu.tr

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



the m the lower the dispersion of fracture stress. It is generally independent of the size of the material, assuming the material composition and microstructure remain consistent. For example, if test specimens with different cross-sectional areas used in uniaxial compression and extension tests, their Weibull modulus remains the same. The scale parameter, on the other hand, is closely related to the mean fracture stress, and it changes with the cross-sectional area.

The Weibull parameters m and s_0 are estimated from a sample of strength measurements. High experimental costs limit the number of samples to be tested. Therefore, methods achieving high estimation precision with small sample sizes are highly desirable [14]. In practice, the Ordinary Least Squares (OLS) method and the Maximum Likelihood Estimation (MLE) method have been the most commonly used methods for estimating these parameters in general reliability applications [15]. In materials science, the OLS method remains the most popular due to its simplicity and ease of use [16], while the MLE method is considered a standard method in important references such as [17] due to its well-known distributional optimality properties in large samples [15]. Both of the methods produce biased estimators, therefore there are many studies for bias correction. One approach is using empirical correction factors, [18] is an example for MLE, and [19–21] are some recent studies for OLS; and the other approach is to satisfy unbiasedness through developing a probability index to be used in OLS. Davies [16,22] provides a detailed literature review and a thorough discussion of how such indices can be developed. Despite this focus on the two-parameter model, recent research has explored the potential of the three-parameter Weibull distribution in materials science, although its practical application remains limited [23–27].

One disadvantage of OLS is that it treats each measurement equally under the assumption that the variance of the error term is a constant. However, this assumption is violated for the OLS method used for Weibull parameter estimation. Several authors have proposed the Weighted Least Squares (WLS) method as a remedy, proposing different weight factors [28–33]. Their studies showed that the WLS methods show significantly superior performance as compared to the OLS methods.

On the other hand, the WLS methods also result in biased estimators, and there has been no study in the literature for unbiased estimation of Weibull parameters by WLS to the best of our knowledge. Therefore, the main purpose of this study is to develop one such method for the unbiased estimation of the Weibull modulus. As in the OLS case, empirical correction factors and developing a new probability index are two viable approaches. This study prefers the latter because it allows decreasing estimation variance both through weight factors and fine-tuning coefficients used in the probability index.

Finally, all the Monte Carlo simulations are coded and run in the R programming language which uses Mersenne-Twister random number generator as the default generator

whose cycle period is $2^{19937}-1$ [34]. The following sections are organized as follows: After a concise discussion of the OLS and WLS methods, a systematic analysis is introduced to develop a new probability index for unbiased estimation of Weibull modulus using Monte Carlo simulations as the basic tool. The probability index, proposed as a result of this analysis, is then compared with the MLE, the OLS and the WLS methods using different criteria. Finally, implementation on a practical example is discussed.

Ordinary Least Squares (OLS) Method

Eq. (1) becomes a straight line by a double logarithmic transformation:

$$\ln[-\ln(1-F(s))] = m \ln s - m \ln s_0 \quad (2)$$

In order to apply the OLS method, the measurements are ranked from the smallest to the largest. F -values are assigned according to the rank i of a measurement, s_i denoting the i th smallest: $s_1 \leq s_2 \leq \dots \leq s_n$. Then, in Eq. (2), $F(s)$ can be replaced by $F(s_i)$, and $\ln s$ can be replaced by $\ln s_i$. The most commonly used probability estimators of $F(s_i)$, which are simply denoted as F_i , are median rank,

$$F_i = (i - 0.3) / (n + 0.4), \quad (3a)$$

mean rank,

$$F_i = i / (n + 1), \quad (3b)$$

and hazen rank,

$$F_i = (i - 0.5) / n \quad (3c)$$

There are many other estimators proposed for this purpose [16]. Considering the familiar form of a regression equation, $Y = aX + b$, the left side of Eq. (2) corresponds to Y , $\ln s$ corresponds to X , m corresponds to a , and $-m \ln s_0$ corresponds to b . Using s_i and F_i pairs in Eq. (2), a and b are obtained by the OLS procedure. Then the Weibull parameter estimates are calculated as $\hat{m} = a$ and $\hat{\sigma}_0 = \exp(-b / \hat{m})$.

Weighted Least Squares (WLS) Method

Weighted least squares (WLS) extends ordinary least squares regression by incorporating weights into the sum of squared residuals. These weights, denoted by w_i for the i th observation, downplay the influence of data points s_i with higher variance or lower reliability. The regression model in Eq. (2) has non-constant error variance. As a result use of the weighted least squares method gives estimates with better statistical properties: Bergman [30], Faucher and Tyson [29] and Hung [31] proposed using the following weight factors in the WLS method, respectively:

$$w_i = [(1 - F_i) \ln(1 - F_i)]^2 \quad (4a)$$

$$w_i = 3.3F_i - 27.5 \left[1 - (1 - F_i)^{0.025} \right] \quad (4b)$$

where

$$z_i = \ln s_i \quad (7c)$$

$$w_i = \frac{[(1 - F_i) \ln(1 - F_i)]^2}{\sum [(1 - F_i) \ln(1 - F_i)]^2} \quad (4c)$$

$$y_i = \ln[-\ln(1 - F_i)] \quad (7d)$$

where hazen rank is used in Eq. (4a), and median rank is used in Eq. (4b) and (4c). Lu et al. [28] showed that $Z = -\ln(1 - F(s))$ in Eq. (2) follows a standard exponential distribution. Then Eq. (2) takes the form of

$$\ln(Z_i) = m \ln(s_i) - m \ln(s_0) \quad (5)$$

where Z_i is the i th order statistic of the standard exponential variable. Using the expectation and variance of this variable, Lu et al. [28] proposed the following approximate formula for the weights:

$$w_i = [E(Z_i)]^2 / \text{Var}(Z_i) \quad (6)$$

where $E(Z_i) = \sum_{j=1}^i \frac{1}{n-j+1}$ and $\text{Var}(Z_i) = \sum_{j=1}^i \frac{1}{(n-j+1)^2}$.

Note that this formula does not involve a probability index. Also, they went on to compare these four WLS methods: They pointed out that the denominator of Hung's weight factor in Eq. (4c) is constant, thus Hung's weight factor is the same as Bergman's in Eq. (4a). They concluded that Bergman's weight factor (as well as that of Hung) generates a larger Mean Squared Error (MSE) than the others. Faucher and Tyson's weight factor in Eq. (4b) and their method show similar levels of performance.

As a result, we decided to use Faucher and Tyson's weight factors (Eq. 4b) in this study due to its ease of use. These weight factors were derived independently from median ranks [29] and should approximately work for any probability rank value between 0 and 1.

WLS computations are performed by various statistics packages such as MINITAB[®] and SPSS[®]. Microsoft Excel[®] spreadsheets can also be used for this purpose using the following closed-form formulae:

$$\hat{m} = \frac{(\sum w_i)(\sum w_i z_i y_i) - (\sum w_i z_i)(\sum w_i y_i)}{(\sum w_i)(\sum w_i z_i^2) - (\sum w_i z_i)^2} \quad (7a)$$

and

$$\hat{s}_0 = \exp \left(- \frac{\sum w_i y_i - \hat{m} \sum w_i z_i}{\hat{m} \sum w_i} \right) \quad (7b)$$

and w_i are computed by Eq. (4b). As a side note, w_i are set to 1 for OLS computations.

In both OLS and WLS $\hat{m}/m$ is a pivotal statistic. In Monte Carlo simulations, therefore, choosing $m=1$ will suffice for all practical purposes. Similarly, the scale parameter will be fixed as $s_0 = 1$.

A New Probability Index Formula for Unbiased WLS

This section and following sections develop a new probability index producing an unbiased estimator of m , that is, $E(\hat{m}) = m$. Numerous probability indices have been proposed for F in addition to Eq. (3a)-(3c). Davies[16] has recently provided a comprehensive list of them. These probability indices are all developed for the ordinary least squares method and conform to the general form of

$$F_i = (i - a) / (n + b) \quad (8)$$

where $0 \leq a \leq 1$ and $b > -a$ (because $1 - F_i$ in Eq. (2) has to be between 0 and 1 for all $i = 1, \dots, n$ to avoid negative terms in the logarithms). For example, $a = 0.3$ and $b = 0.4$ for median rank in Eq. (3a). The coefficients a and b are usually constant [16]; however, some studies aiming at minimizing the coefficient of variation of $\hat{m}$ [35,36] or making $\hat{m}$ unbiased [16,37,38] proposed a and b values as functions of n . One problem with this approach is that the values change inconsistently with changing n which leads to the lack of expressions for F which are valid for all materials engineering relevant values of n [16]. Therefore, this study aims to compute a as functions of n , and b as a constant such that WLS produces $\hat{m}$ values that are unbiased and have minimum variation.

The basic simulation procedure employed in this study involves generating a sample of n values from a Weibull distribution with parameters $m=1$ and $s_0 = 1$; estimating $\hat{m}$ using WLS with certain a and b values in Eq. (8) (or any index function with two constants such as in Eq. (10) as will be discussed below) and repeating this R times to compute the sample mean and the standard deviation of $\hat{m}$.

In this procedure the sample mean of $\hat{m}$, $\hat{m}_{ave}$, can be considered as a stochastic function $g(a, b, n, R)$. For a given n , finding the values of a and b producing an unbiased estimator ($\hat{m}_{ave} = 1$) is equivalent to finding the root of $g(a, b, n, R) - 1$. The root-finding procedure employed to compute the a - b combination in Eq. (8) is (or in Eq. (10) as will be discussed shortly) as follows: While keeping b at a certain value b' , a is increased incrementally by δ , and at any step going from a' to $a' + \delta$, if $\hat{m}_{ave}$, estimated by the basic simulation procedure, goes from a value less than 1 to a value

greater than 1, or the vice versa, this indicates that there exists an unbiased estimator for $b = b'$ and the values of $a' + \delta/2$ and b' can be considered as the approximate values producing an unbiased estimator. The precision of the procedure depends on the choice of δ . This is repeated for all b values in a given search interval in increments of δ : $\dots b' - \delta, b', b' + \delta, \dots$

As an initial attempt to produce approximate a and b values in Eq. (8) a small group sample sizes $n = 6, 15, 30, 75, 100$ and 150 , are selected. In Eq. (8) b is usually between 0 and 1 [16], hence $0 \leq a \leq 1$ and $0 \leq b \leq 1$. The root-finding procedure was repeated for each sample size with $\delta = 0.01$ and the simulation run number is $R = 10^5$. However, in order to avoid numerical errors, the values of 0 and 1 for a were avoided throughout all the simulations in this study; 0 is always replaced by 0.001 and 1 by 0.999 ($a = 0.001, 0.01, 0.02, \dots, 0.99, 0.999$ and $b = 0, 0.01, 0.02, \dots, 0.99, 1$). This simulation, unfortunately, was not able to produce a constant b value producing an unbiased estimator for all n values. Further analysis with b values larger than 1 also showed that as the b value increased the amount of bias became larger.

Therefore, the probability index, used for OLS, cannot be used for WLS. This necessitates another index form suitable for WLS. Another such form used in probability plots, in general, has the following form [28,39]:

$$F_i = (i - a) / (n - 2a + 1) \quad (9)$$

When this formula is tried for the same group of sample sizes, the root finding procedure was able to find an unbiased estimator for each sample size with $\delta = 0.01$ and $R = 10^5$ (consider $b = 1$). Next, we decided to try a more general formula with two constants as in Eq. (7):

$$F_i = (i - a) / (n - 2a + b) \quad (10)$$

This allows choosing among a range of b values producing an unbiased estimator with the smallest variance. Following the discussion on bounds for Eq. (8), $0 \leq a \leq 1$ and $a < b$.

Search for a Value Of b with the Smallest Variance

The root-finding procedure is repeated for Eq. (10) with the same group of sample sizes, changing b from 1 to 3, a from 0 to 1 (strictly speaking from 0.001 to 0.999) with $\delta = 0.01$ and $R = 10^5$. This procedure produced an unbiased estimator except for $b > 2.6$. Among all b values between 1 and 2.6, the standard deviations, multiplied by $\sqrt{n}$ for normalization, are averaged over all sample sizes. The smallest average value is produced by $b = 1.57$. Figure 1 shows the normalized standard deviations for selected b values; note that the minimums are achieved at different values of a for each n , which are not presented here. It indicates that change in b may cause significant differences in the standard deviation for $n \leq 30$; also, the value $b = 1.5$, which is

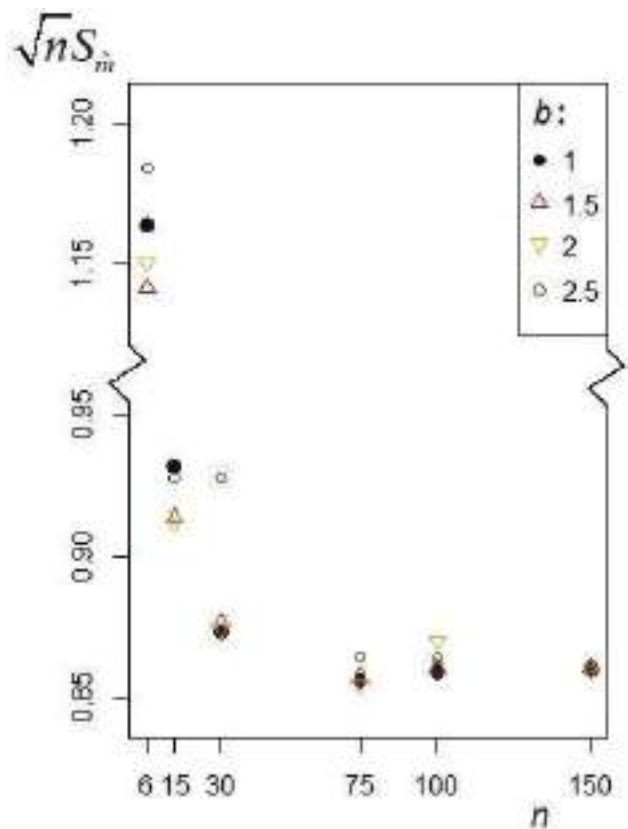


Figure 1. Samples size vs. normalized standard deviations of $\hat{m}$ for selected b values.

close to value of $b = 1.57$, consistently produces small standard deviation values for each sample size. This can be seen in Figure 1 for $n = 6$ and 15 . As a result of this analysis $b = 1.57$ was fixed as a constant in Eq. (10).

Computing Values of a as a Function of n When $b = 1.57$

The next task is to compute values of a as a function of n . Our initial trials with the newly computed index function

$$F_i = (i - a) / (n - 2a + 1.57) \quad (11)$$

revealed the stochastic function $\hat{m}_{ave} = g(a, b, n, R)$ is a considerably flat function with respect to a . When $\delta = 0.01$ and $R = 10^5$ were used, this caused the function value to fluctuate above and below 1 for neighboring values of a , making it difficult to specify a precise a value producing an unbiased estimator. Increasing δ and R simultaneously and running the simulations for all n are expected to result in total simulation times that will be prohibitively long.

A two-step solution is proposed for this problem: In the first step, $R = 10^5$ was increased to $R = 10^6$ while keeping $\delta = 0.01$ for a larger set of selected sample sizes $n = 6, 8, 10, 12, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140$ and 150 . Then, using the root-finding procedure, approximate values of a were computed. Using them, values

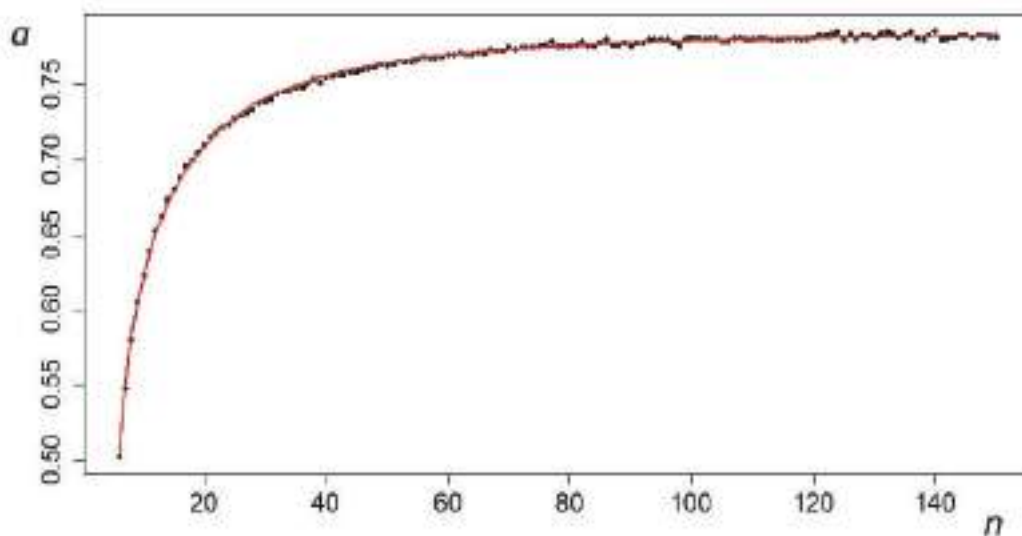


Figure 2. Sample size vs. a values, provided in the appendix, and the fitted curve in Eq. (12).

for all n were approximated: for example, if $a = 0.63$ for $n = 10$ and $a = 0.65$ for $n = 12$, then as $a \approx 0.64$ for $n = 11$.

In the second step, $\delta = 0.01$ was reduced to $\delta = 0.001$ along with $R = 10^6$. The root finding procedure was applied to an interval of length 0.024 centering the previously computed values of a for all $n = 6, \dots, 150$. The length 0.024 was determined by some trial-and-error; it was to be kept to a minimum because newly selected values of δ and R considerably increased the simulation time. This particular interval length managed to produce values of a for all n ; values of a are provided in the appendix.

In order to fit the values to the sample size, the following function is proposed to be used in Eq. (11):

$$a = -7.19247 + 7.97925 \exp(-(n^{-1.5})) + 4.52723 \exp(-n^{0.6}) \quad (12)$$

The fit is considerably good with an R^2 value of 0.9989, residual standard error of 0.0015 and a maximum percent error of 0.45. The values and fitted function are shown in Figure 2, which visually shows a very good fit.

Consequently, this study proposes using the WLS methods with the probability index in Eq. (11) together with coefficients in Eq. (12) using the weights in Eq. (4b).

COMPARISON WITH OTHER METHODS

It is tricky to select methods for comparison purposes, because there is no WLS method developed for unbiased estimation of the Weibull modulus m . We selected Davies' [16] probability index formula as the first method of comparison; it is designed for unbiased estimation of the mean of m (he proposed others for median and mode) to be used with OLS. While OLS can be seen as a special case of WLS, where all the weight are one, OLS and WLS are treated as

two separate methods in this study. Davies [16] showed that the formula results in unbiased estimates with low variation. As the second and third methods of comparison, the MLE method and Faucher and Tyson's WLS method [29] are selected. While their estimates are biased, they may still produce estimates close to the true value, which can be measured by the Mean Squared Error (MSE): $MSE = E(\hat{m} - m)^2$. MSE is also related with bias: $MSE = Var(\hat{m}) + (E(\hat{m}) - m)^2$, where the second part of the summation is the square of the bias of $\hat{m}$. Note that this part is zero for the proposed method and Davies' method because of their unbiasedness.

The basic simulation procedure described previously is used with the selected methods for comparison and selected sample sizes; that is, parameters are set as $m=1$ and $s_0 = 1$ for Weibull random number generation; the simulation run number is set as $R=10^6$. The MSE values as well as the sample means of $\hat{m}$ are shown in Table 1.

First, let's look at the precision of the unbiasedness: the first column of the proposed method in Table 1 should be ideally 1, but there are slight deviations which are due to the simulation error and the errors incurred during the root finding procedure. Davies' method is designed for $n \geq 10$, this explains the relatively high deviations for $n = 6$ and 8.

Before comparing the methods let's note that for two estimators, $\hat{m}_1$ and $\hat{m}_2$, the ratio of their MSEs are called relative efficiency, defined as $RE[\hat{m}_1, \hat{m}_2] = MSE(\hat{m}_1)/MSE(\hat{m}_2)$. A RE value of 0.80 implies that the necessary sample size for the method using $\hat{m}_1$ is 80% of that needed for $\hat{m}_2$ to achieve approximately equal MSE value [15]. This may translate as the approximate percent saving in testing costs when the method of $\hat{m}_1$ is preferred to that of $\hat{m}_2$.

When the MSE's of the two unbiased methods are compared, Table 1 shows that Davies' method has significantly larger values than the proposed method for all

sample sizes. It is to the degree that the levels achieved by the proposed method can be achieved with an additional five test specimens (e.g. $n = 15$ vs $n = 20$) for $n \geq 20$, and with an additional 10 specimen for $n \geq 35$; this result can be seen by comparing MSE values for different n values, and also agrees with the RE column (note that RE produces an approximate result). Note that, since these methods are unbiased, MSE of $\hat{m}$ is equivalent to the sample variance of $\hat{m}$, S_m^2 (strictly speaking, they are approximately equal due to small deviations from unbiasedness).

While the MLE method has been a popular alternative to OLS and WLS, it suffers from severe overestimation bias in small samples. For instance, in Table 1, the MLE estimate for $n = 10$ (1.1698) overestimates the unknown Weibull modulus by 16.98%. Table 2 decomposes MSE values in Table 1 into bias and variance components, revealing that the bias becomes negligible (below 2%) for sample sizes exceeding $n = 70$; note that $MSE = (Bias)^2 + Variance$ (here, for $n = 10$, $0.1517 = 0.1698^2 + 0.1229$). Table 2 highlights the substantial bias in MLE estimates for small samples significantly impacts the overall MSE .

The MSE values of the MLE method are at least 10% larger than those of the proposed method for $n \leq 30$. The difference is so severe that the levels achieved by the proposed method for $n \leq 12$ can be achieved by the MLE method with an addition of at least three test specimens. However, as the sample size increases, the MLE methods starts to show better performance as a result of its optimality properties in large samples. For $n \geq 60$, the MLE method starts to perform better, and the MSE for the proposed method becomes 14% larger than that of the MLE method at $n = 150$.

When compared with the proposed method, Faucher and Tyson's method has smaller MSE values for all sample sizes, and the bias is not severe as can be seen in Table 2. The proposed method's MSE is about 6% above that of Faucher and Tyson's method for $n \leq 15$, but as the sample size increases, this percentage goes below 2% after $n \geq 35$ and below 1% after $n \geq 60$. These results do not indicate that Faucher and Tyson's method is better than the proposed method, because it has significant bias. It underestimates $\hat{m}$ and approximately 5% underestimation should be expected for $n = 6$ -20. The percentage of underestimation

Table 1. Sample mean and MSE values of $\hat{m}$ for the compared methods

n	Proposed (P)		Davies (D)		RE(P,D)	MLE (M)		RE(P,M)	Fau&Tys (F)		
	Mean	MSE	Mean	MSE		Mean	MSE		Mean	MSE	RE(P,F)
6	1.0008	0.2130	1.0368	0.2420	0.8802	1.3341	0.4848	0.4394	0.9752	0.2031	1.0487
8	1.0000	0.1289	1.0111	0.1453	0.8871	1.2249	0.2399	0.5373	0.9511	0.1193	1.0805
10	0.9996	0.0924	1.0030	0.1067	0.8660	1.1698	0.1517	0.6091	0.9459	0.0859	1.0757
12	0.9988	0.0719	0.9997	0.0854	0.8419	1.1358	0.1081	0.6651	0.9454	0.0676	1.0636
15	0.9998	0.0545	0.9998	0.0670	0.8134	1.1055	0.0749	0.7276	0.9495	0.0519	1.0501
20	1.0000	0.0389	0.9998	0.0498	0.7811	1.0762	0.0482	0.8071	0.9558	0.0376	1.0346
25	1.0003	0.0303	1.0005	0.0399	0.7594	1.0598	0.0352	0.8608	0.9617	0.0295	1.0271
30	1.0001	0.0249	1.0004	0.0334	0.7455	1.0489	0.0276	0.9022	0.9661	0.0244	1.0205
35	1.0004	0.0212	1.0006	0.0288	0.7361	1.0418	0.0227	0.9339	0.9701	0.0208	1.0192
40	1.0002	0.0184	1.0002	0.0254	0.7244	1.0362	0.0191	0.9634	0.9730	0.0182	1.0110
45	1.0000	0.0163	0.9999	0.0226	0.7212	1.0318	0.0166	0.9819	0.9753	0.0161	1.0124
50	1.0001	0.0147	0.9998	0.0205	0.7171	1.0286	0.0146	1.0068	0.9776	0.0145	1.0138
55	1.0001	0.0133	0.9999	0.0187	0.7112	1.0261	0.0131	1.0153	0.9795	0.0132	1.0076
60	0.9999	0.0122	0.9996	0.0172	0.7093	1.0236	0.0118	1.0339	0.9808	0.0121	1.0083
65	0.9999	0.0112	0.9999	0.0159	0.7044	1.0217	0.0108	1.0370	0.9821	0.0112	1.0000
70	0.9999	0.0105	0.9999	0.0148	0.7095	1.0202	0.0099	1.0606	0.9833	0.0104	1.0096
75	1.0000	0.0098	1.0001	0.0139	0.7050	1.0189	0.0092	1.0652	0.9844	0.0097	1.0103
80	0.9998	0.0092	1.0000	0.0131	0.7023	1.0176	0.0085	1.0824	0.9852	0.0091	1.0110
85	1.0000	0.0086	1.0001	0.0123	0.6992	1.0166	0.0080	1.0750	0.9862	0.0086	1.0000
90	0.9999	0.0082	1.0001	0.0117	0.7009	1.0156	0.0075	1.0933	0.9868	0.0081	1.0123
95	1.0000	0.0077	1.0001	0.0111	0.6937	1.0148	0.0070	1.1000	0.9876	0.0077	1.0000
100	0.9999	0.0073	1.0001	0.0105	0.6952	1.0140	0.0067	1.0896	0.9881	0.0073	1.0000
125	1.0000	0.0059	0.9999	0.0085	0.6941	1.0113	0.0052	1.1346	0.9905	0.0059	1.0000
150	0.9999	0.0049	0.9996	0.0071	0.6901	1.0093	0.0043	1.1395	0.9920	0.0049	1.0000

becomes less severe as the sample size increases and goes below 2% for $n \geq 60$.

Practical Example

An application of the proposed method was carried out on the results of an experimental study [40]: 19 identical composite specimens were prepared from quasi-isotropic carbon-epoxy sheets with $(0^\circ)_3$ configuration, 0.89 mm thickness, and 295 gr/m² weight and the tension experiments were carried out using an Instron 8516+ universal testing machine. The measured fracture strength values are presented in Table 3.

To enhance the application study in terms of sample size, in addition to the original data set given in Table 3, two random subsets of sizes 6, 10, and 15 were drawn from the same data set. These subsets consist of {522, 532.7, 476.5, 521.6, 439, 507.3} for $n = 6$, {513.6, 552.519, 521.6, 439, 450.9, 463.5, 497.5, 476.5, 477} for $n = 10$, and {532.7, 502.5, 442, 519, 502.7, 477, 552, 522, 439, 513.6, 521.6, 450.9, 476.5, 507.3, 463.5} for $n = 15$.

To implement the proposed method, Eq. (7) is used where F_i is computed by Eq. (11) and Eq. (12). The following Table 4 gives the estimated Weibull modulus values. As n increases, the MSE decreases as can be seen from Table 1,

Table 2. Decomposition of MSE values of $\hat{m}$ for biased estimation methods

n	MLE (M)				Fau&Tys (F)			
	Mean	MSE	Bias	Variance	Mean	MSE	Bias	Variance
6	1.3341	0.4848	33.41%	0.3732	0.9752	0.2031	-2.48%	0.2025
8	1.2249	0.2399	22.49%	0.1893	0.9511	0.1193	-4.89%	0.1169
10	1.1698	0.1517	16.98%	0.1229	0.9459	0.0859	-5.41%	0.0830
12	1.1358	0.1081	13.58%	0.0897	0.9454	0.0676	-5.46%	0.0646
15	1.1055	0.0749	10.55%	0.0638	0.9495	0.0519	-5.05%	0.0493
20	1.0762	0.0482	7.62%	0.0424	0.9558	0.0376	-4.42%	0.0356
25	1.0598	0.0352	5.98%	0.0316	0.9617	0.0295	-3.83%	0.0280
30	1.0489	0.0276	4.89%	0.0252	0.9661	0.0244	-3.39%	0.0233
35	1.0418	0.0227	4.18%	0.0210	0.9701	0.0208	-2.99%	0.0199
40	1.0362	0.0191	3.62%	0.0178	0.9730	0.0182	-2.70%	0.0175
45	1.0318	0.0166	3.18%	0.0156	0.9753	0.0161	-2.47%	0.0155
50	1.0286	0.0146	2.86%	0.0138	0.9776	0.0145	-2.24%	0.0140
55	1.0261	0.0131	2.61%	0.0124	0.9795	0.0132	-2.05%	0.0128
60	1.0236	0.0118	2.36%	0.0112	0.9808	0.0121	-1.92%	0.0117
65	1.0217	0.0108	2.17%	0.0103	0.9821	0.0112	-1.79%	0.0109
70	1.0202	0.0099	2.02%	0.0095	0.9833	0.0104	-1.67%	0.0101

Table 3. Fracture strength of carbon-epoxy composite material specimens (megapascals)

Test No.	1	2	3	4	5	6	7	8	9	10
Fracture strength [MPa]	532.7	502.5	442	473	519	502.7	477	510	522	552
Test No.	11	12	13	14	15	16	17	18	19	
Fracture strength [MPa]	522	439	513.6	497.5	521.6	450.9	476.5	507.3	463.5	

Table 4. Weibull modulus estimates ($\hat{m}$) for the fracture strength data

n	Proposed (P)	Davies (D)	MLE (M)	Fau&Tys (F)
6	16.0764	14.5886	21.9916	15.3496
10	13.8474	15.2852	15.6205	13.2601
15	15.6385	16.2970	17.3370	14.9150
19	17.6756	18.1770	18.8625	16.9160

hence estimates with larger sample sizes have higher precision. Therefore, the estimates for $n = 19$ should be closer to the true, unknown Weibull modulus value.

The estimates for $n = 19$ are close to each other due to their small *MSE* values (check the values for $n = 20$ in Table 1). However, the smallest *MSE*'s belong to the proposed WLS method and Faucher and Tyson's WLS method, thus their estimates, 17.6756 and 16.916 can be argued to be closer to the true m value.

For smaller n values the estimates are farther away from each other, due to increased *MSE* and increased bias. Consider $n = 6$: *MSE* is a measure of expected distance of an estimate from the true value, and they are 5-10 times as large as those for $n = 20$. Also, due to the substantial overestimation problem in the MLE method, which can be seen in Table 2, MLE estimates are larger than the others in Table 4.

CONCLUSION

This research introduced a novel probability index that, for the first time, offers unbiased WLS estimators of the Weibull modulus with a satisfactory degree of precision. To this end, a systematic study was carried out to compute parameters of a probability index to be used with the WLS method for the unbiased estimation of the Weibull modulus. First Faucher and Tyson's weights were selected among several weighting schemes. Then two parameters of a generic probability index were computed by using a step-by-step approach. The result is a simple index formula presented in Eq. (10) wherein the single parameter a , produced from simulations, is fitted as a function of the sample size in Eq. (11). Use of two parameters in this study allowed variance reduction of the estimator as well as satisfying unbiasedness.

The proposed method requires only the use of the new probability index along with Faucher and Tyson's weights. It is the only WLS method with unbiased Weibull modulus estimator in the literature. When compared with a recent OLS method proposing a new probability index producing unbiased Weibull modulus estimators, it exhibited significantly smaller variance. *MSE* is used for comparison among methods producing biased estimators. Results showed that the proposed method performs better than the MLE method for sample sizes less than 60, and its performance is close to that of Faucher and Tyson's method, but with slightly larger *MSE* values. The advantage of the proposed method is most prominent in the range of sample sizes smaller than 125 where the percent bias in the aforementioned biased methods is greater than 1% and they should be used with caution.

In materials science, high experimental costs limit the number of samples to be tested; and in general, smaller the sample size are preferred. In general, sample sizes are less than 60 in such experiments, which makes the proposed method advantageous in terms of its smaller *MSE* and unbiasedness property. While this paper focuses on

materials science applications, it should be pointed out that the proposed method can be used in other application areas such as wind speed analysis and reliability analysis.

In conclusion, the implementation of the proposed WLS method is almost as simple as that of the OLS method and we hope that this helps materials scientists and engineers who generally prefer OLS and MLE in practice.

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] Weibull W. A statistical distribution function of wide applicability. *J Appl Mech* 1951;1-5.
- [2] Kaplan YA. Comparison of the performance of the methods used to find the weibull parameters at different heights. *Arab J Sci Eng* 2021;46:12145–53. [CrossRef]
- [3] Usta I, Arık I, Yenilmez I, Kantar YM. A new estimation approach based on moments for estimating Weibull parameters in wind power applications. *Energy Convers Manag* 2018;164:570–578. [CrossRef]
- [4] Kantar YM, Usta I, Yenilmez I, Arık I. A study on estimation of wind speed distribution by using the Modified Weibull distribution. *Bilişim Teknol Derg* 2016;9:63.
- [5] Kantar YM, Arık I, Usta I, Yenilmez I. Comparison of some estimation methods of the two parameter Weibull distribution for unusual wind speed data cases. *Bilişim Teknol Derg* 2016;9:81.
- [6] Lai C-D, Murthy D, Xie M. Weibull distributions and their applications. *Springer Handb.*, Springer; 2006, p. 63–78. [CrossRef]
- [7] Barbero E, Fernández-Sáez J, Navarro C. Statistical analysis of the mechanical properties of composite materials. *Compos Part B Eng* 2000;31:375–381. [CrossRef]

- [8] Lei W-S, Yu Z, Zhang P, Qian G. Standardized Weibull statistics of ceramic strength. *Ceram Int* 2021;47:4972–4993. [CrossRef]
- [9] Lei W-S, Zhang P, Yu Z, Qian G. Statistics of ceramic strength: Use ordinary Weibull distribution function or Weibull statistical fracture theory? *Ceram Int* 2020;46:20751–20768. [CrossRef]
- [10] Bernal RA. On the application of Weibull statistics for describing strength of micro and nanostructures. *Mech Mater* 2021;162:104057. [CrossRef]
- [11] Barbero E, Fernandez-Saez J, Navarro C. Statistical distribution of the estimator of Weibull modulus. *J Mater Sci Lett* 2001;20:847–849. [CrossRef]
- [12] Yang X, Xie L, Yang Y, Zhao B, Li Y. A comparative study for parameter estimation of the Weibull distribution in a small sample size: An application to spring fatigue failure data. *Qual Eng* 2022;1–13. [CrossRef]
- [13] Liu J, Zhang Z, Zhu L, Wang J, He Y. Research on fatigue life evaluation method of shaft parts based on small sample. *Int J Struct Integr* 2023;14:709–732. [CrossRef]
- [14] Yalçinkaya M, Birgören B. Shortest confidence intervals for weibull modulus for small samples in materials reliability analysis. *Gazi Univ J Sci* 2023;1. [CrossRef]
- [15] Genschel U, Meeker WQ. A comparison of maximum likelihood and median-rank regression for Weibull estimation. *Qual Eng* 2010;22:236–255. [CrossRef]
- [16] Davies IJ. Unbiased estimation of Weibull modulus using linear least squares analysis—A systematic approach. *J Eur Ceram Soc* 2017;37:369–380. [CrossRef]
- [17] Composite Materials Handbook. Vol. 1. Polymer matrix composites guidelines for characterization of structural materials, SAE International on behalf of CMH-17, a division of Wichita State University. SAE International; 2012.
- [18] Hirose H. Bias correction for the maximum likelihood estimates in the two-parameter Weibull distribution. *IEEE Trans Dielectr Electr Insul* 1999;6:66–68. [CrossRef]
- [19] Davies I. Empirical correction factor for the best estimate of Weibull modulus obtained using linear least squares analysis. *J Mater Sci Lett* 2001;20:997–999. [CrossRef]
- [20] Davies I. Best estimate of Weibull modulus obtained using linear least squares analysis: an improved empirical correction factor. *J Mater Sci* 2004;39:1441–1444. [CrossRef]
- [21] Jia X, Xi G, Nadarajah S. Correction Factor for Unbiased Estimation of Weibull Modulus by the Linear Least Squares Method. *Metall Mater Trans A* 2019;50:2991–3001. [CrossRef]
- [22] Davies IJ. Unbiased estimation of the Weibull scale parameter using linear least squares analysis. *J Eur Ceram Soc* 2017;37:2973–2981. [CrossRef]
- [23] Garrido A, Caro-Carretero R, Jimenez-Octavio JR, Carnicero A, Such M. A new approach to fitting the three-parameter Weibull distribution: An application to glass ceramics. *Commun Stat-Theory Methods* 2021;50:3403–3420. [CrossRef]
- [24] Deng B, Jiang D, Gong J. Is a three-parameter Weibull function really necessary for the characterization of the statistical variation of the strength of brittle ceramics? *J Eur Ceram Soc* 2018;38:2234–2242. [CrossRef]
- [25] Yang F, Ren H, Hu Z. Maximum likelihood estimation for three-parameter Weibull distribution using evolutionary strategy. *Math Probl Eng* 2019;2019:6281781. [CrossRef]
- [26] Nohut S. Three-parameter (3P) weibull distribution for characterization of strength of ceramics showing R-Curve behavior. *Ceram Int* 2021;47:2270–2279. [CrossRef]
- [27] Guo J, Kong X, Wu N, Xie L. Weibull parameter estimation and reliability analysis with small samples based on successive approximation method. *J Mech Sci Technol* 2023;37:5797–5811. [CrossRef]
- [28] Lu H, Chen C, Wu J. A note on weighted least-squares estimation of the shape parameter of the Weibull distribution. *Qual Reliab Eng Int* 2004;20:579–586. [CrossRef]
- [29] Faucher B, Tyson W. On the determination of Weibull parameters. *J Mater Sci Lett* 1988;7:1199–2203. [CrossRef]
- [30] Bergman B. Estimation of Weibull parameters using a weight function. *J Mater Sci Lett* 1986;5:611–614. [CrossRef]
- [31] Hung W. Weighted least-squares estimation of the shape parameter of the Weibull distribution. *Qual Reliab Eng Int* 2001;17:467–469. [CrossRef]
- [32] Mert Kantar Y. Estimating variances in weighted least-squares estimation of distributional parameters. *Math Comput Appl* 2016;21:7. [CrossRef]
- [33] Kantar YM. Generalized least squares and weighted least squares estimation methods for distributional parameters. *REVSTAT-Stat J* 2015;13:263–282.
- [34] Dutang C, Wuerzt D. A note on random number generation 2009. <https://cran.r-project.org/web/packages/randtoolbox/vignettes/fullpres.pdf>.
- [35] Song L, Wu D, Li Y. Optimal probability estimators for determining Weibull parameters. *J Mater Sci Lett* 2003;22:1651–1653. [CrossRef]
- [36] Wu D, Lu G, Jiang H, Li Y. Improved estimation of Weibull parameters with the linear regression method. *J Am Ceram Soc* 2004;87:1799–1802. [CrossRef]
- [37] Tiryakioğlu M. An unbiased probability estimator to determine Weibull modulus by the linear regression method. *J Mater Sci* 2006;41:5011–5013. [CrossRef]
- [38] Tiryakioğlu M, Hudak D. On estimating Weibull modulus by the linear regression method. *J Mater Sci* 2007;42:10173–10179. [CrossRef]
- [39] D'Agostino R. Goodness-of-fit-techniques. Routledge; 2017. [CrossRef]
- [40] Dirikolu MH, Aktas A, Birgoren B. Statistical analysis of fracture strength of composite materials using Weibull distribution. *Turk J Eng Environ Sci* 2002;26:45–48.

APPENDIX

Table A. values of a to be used in Eq. (11) for WLS

n	a	n	a	n	a	n	a	n	a
6	0.5025	35	0.7485	64	0.7695	93	0.7795	122	0.7825
7	0.5475	36	0.7485	65	0.7715	94	0.7795	123	0.7835
8	0.5805	37	0.7505	66	0.7705	95	0.7795	124	0.7845
9	0.6055	38	0.7535	67	0.7735	96	0.7795	125	0.7795
10	0.6235	39	0.7515	68	0.7725	97	0.7775	126	0.7835
11	0.6395	40	0.7545	69	0.7725	98	0.7755	127	0.7805
12	0.6525	41	0.7555	70	0.7745	99	0.7795	128	0.7825
13	0.6625	42	0.7565	71	0.7735	100	0.7815	129	0.7825
14	0.6735	43	0.7565	72	0.7745	101	0.7805	130	0.7815
15	0.6805	44	0.7585	73	0.7745	102	0.7795	131	0.7835
16	0.6885	45	0.7585	74	0.7745	103	0.7805	132	0.7845
17	0.6955	46	0.7595	75	0.7755	104	0.7805	133	0.7835
18	0.6995	47	0.7605	76	0.7765	105	0.7785	134	0.7815
19	0.7045	48	0.7615	77	0.7785	106	0.7785	135	0.7825
20	0.7095	49	0.7635	78	0.7755	107	0.7805	136	0.7845
21	0.7145	50	0.7625	79	0.7755	108	0.7815	137	0.7805
22	0.7175	51	0.7635	80	0.7755	109	0.7795	138	0.7795
23	0.7215	52	0.7655	81	0.7765	110	0.7795	139	0.7825
24	0.7235	53	0.7645	82	0.7785	111	0.7805	140	0.7855
25	0.7275	54	0.7645	83	0.7755	112	0.7805	141	0.7795
26	0.7295	55	0.7665	84	0.7755	113	0.7795	142	0.7805
27	0.7315	56	0.7685	85	0.7775	114	0.7795	143	0.7815
28	0.7335	57	0.7675	86	0.7795	115	0.7795	144	0.7825
29	0.7375	58	0.7675	87	0.7765	116	0.7805	145	0.7825
30	0.7385	59	0.7685	88	0.7765	117	0.7805	146	0.7815
31	0.7395	60	0.7695	89	0.7775	118	0.7795	147	0.7825
32	0.7435	61	0.7695	90	0.7755	119	0.7805	148	0.7825
33	0.7455	62	0.7715	91	0.7775	120	0.7815	149	0.7815
34	0.7465	63	0.7695	92	0.7775	121	0.7825	150	0.7815



Research Article

Designing of triple rings metasurface based perfect absorber for C-band and X-band applications

Jigar M. PATEL¹, Rahul D. MEHTA^{1*}

¹Department of Electronics & Communication Engineering, Government Engineering College, Gandhinagar, Gujarat Technological University, Gujarat, 382424, India

ARTICLE INFO

Article history

Received: 20 May 2024

Revised: 18 July 2024

Accepted: 22 October 2024

Keywords:

Absorption; C- band and X – band; Metasurface; Perfect Absorber; Triple Circular Ring

ABSTRACT

The present investigation aims to design, simulate and fabricate Triple Rings Metasurface-based Perfect Absorber for C – band and X-band applications. The proposed structure is designed and simulated in CST Simulator software and fabricated hardware of 6x6 array and 10x10 array with the unit cell size 9x9 mm. In simulated results, the maximum absorption for unit cells is achieved for normal EM waves of 96.62% at 7.97 GHz frequency. For the 6x6 Array, the maximum absorption is achieved at 99.04 % and 99.21% at 9.87 GHz and 11.58 GHz respectively with the substrate's thickness of 1.8 mm. For the 6x6 Array, the maximum absorption is achieved at 97.82% at 7.83 GHz respectively with a substrate thickness of 0.8 mm. For the 10x10 Array, the maximum absorption is achieved at 99.04 % and 99.21% at 9.87 GHz and 11.58 GHz respectively with the substrate's thickness of 1.8 mm. For the 10x10 Array, the maximum absorption is achieved at 92.49% at 10.84 GHz with a substrate thickness of 0.8 mm. The proposed structure is investigated by varying the substrate's thickness and the rings' variation. The experimental results show maximum absorption in the C – and X – bands. Due to these results, the proposed design will be used for C-band and X-band applications.

Cite this article as: Patel JM, Mehta RD. Designing of triple rings metasurface based perfect absorber for C-band and X-band applications. Sigma J Eng Nat Sci 2025;43(5):1693–1703.

INTRODUCTION

Recent years have witnessed a lot of work on the design and investigation of metamaterial absorbers for their interesting electromagnetic features and application prospects in C, X, Ku, L, and S-band frequencies. Metamaterial absorbers are absorbers whose structures are designed and can absorb

electromagnetic waves fully, therefore have been applied in EMI shielding, stealth technology as well and energy harvesting.

In the year 2023, Patel and Mehta proposed a comparative review on metasurface-based perfect absorbers, where the researchers mainly focused on the performance of the metasurfaces through simulations.[1] An X-shaped triple

*Corresponding author.

*E-mail address: rdmehta@hotmail.com

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



split ring resonator metamaterial absorber is designed and analysed for the quad-band insensitivity of the incidence and polarization angles by Jahan et al. [2], for C, X, and Ku bands. More progress was shown by Jahan et al. [3] with a quad-band metamaterial perfect absorber with a double X-shaped ring resonator for having high shielding effectiveness.

Also, in Ullah et al. [4], 2024, we have a four-band compact metamaterial absorber for electromagnetic energy harvesting application and a metamaterial absorber with triple band microwave proposed by Hossain et al. [5] which uses double E-shaped symmetric split ring resonators for EMI shielding with stealth application. Liu et al. [6] added to designing an ultra-thin absorber with fragmented magnetic structures for L, S, and part of C bands.

Shahid et al. [7] and Rajyalakshmi et al. (2023) designed MMAs for L, S and C bands with particular attention made on the angle dependency [8-10]. Some of the other important contributions are on polarization-independent, wide-angle, broadband metasurface absorber by Amer et al. [11], 2022 using Resistor Loaded Split Ring Resonator and on Dual-band metamaterial for 5G sub-6GHz by Hasan et al. [12] 2022.

Specifically, Hakim et al. [13] presented a detailed study of dual and quad-band metamaterial absorbers for K and Ku [14] bands sensing applications and Kaur & Singh [15] reported a compact ultrathin polarization- and Incident angle -Independent triple band Metamaterial absorber. Deng et al. [16] paid attention to the ultrathin and tri-band MMAs with WIAA for conformal usages in X and Ku bands.

Zhao et al. [17] investigated the features of a V-shaped metamaterial absorber and its possible uses and Singh and Gupta [18] designed a wrenched-square shaped polarization insensitive and wide angle stable ultra-thin metamaterial absorber for S, X and Ku band usages. Finally, Naqvi et al. [19] designed a new meander line metamaterial absorber for 5G at 24 GHz and 28 GHz.

Shakiba et al. [20] introduced a novel, wide operating frequency, multiband PMA that incorporated the switching functionality of phase change materials for biosensing. Tunability and high absorption efficiency of the structure over multiple bands is because of the employment of PCMs which make the absorber suitable for precise biosensing application. As has been pointed out in the given research, metamaterial absorbers have demonstrated promising development in medical and environmental applications.

S, C and X band metamaterial absorber is presented by Afsar et al. [21] which is designed based on one of the fundamental symmetries of the solar system, rotational symmetry. Moreover, due to the helix structure of the proposed design, the efficiency of the power absorption is optimized for wireless communication and the EMI shielding with near-perfect power absorption is confirmed by the simulation and the experiment. Therefore, the study shows the absorber's suitability for effective and consistent electromagnetic wave absorption in communication systems and stealth technologies.

Still, there are some open questions that one has to pose to further improve the concept and application of metamaterial absorbers. Different design solutions can provide multi-band absorbing properties; however, to design an object that gives the true wideband absorbing characteristics over a wider range of frequencies, additional studies are to be made to increase overall flexibility and usability. In a similar regard, recently angle-independent and polarization-insensitive absorbers have also been implemented, however, more enhancements are required to render functionality invariant to operation conditions. Forming very thin layer geometries that still exhibit high absorption efficiency, conformal, and flexible is still a problem, important for integrating technologies into clothing and intricate structures. This is in terms of new materials and techniques of manufacture that are also much more efficient and less expensive but which offer these performances consistently.

DESIGN AND SIMULATION

Design

The basic structure of the Meta surface based on a perfect absorber is shown in Figure 1. The transmitted EM wave (E_t) will not pass due to the absorbing nature of the metamaterial. Also, there is no reflection of the EM wave (E_r). When reflected power and transmitted power have been minimized simultaneously, absorptivity will be maximum.

The simulation performed using CST software incorporating electromagnetic simulation with frequency solver enables the calculation of absorbance ($A(\lambda)$) from S-parameter results obtained. Thus, for this particular design where a full metal layer is used, the analysis reduces to $A(\lambda) = 1 - |S_{11}|^2$ [1].

For metasurface based perfect absorber, the resonant condition should be matched which is interaction between the thickness of substrate and triple ring structure. The relation between these parameters is shown as below:

$$2\eta_{eff}t_s = m\lambda \quad (1)$$

Where η_{eff} is effective refractive index of the substrate, t_s is the thickness of substrate, λ is the wavelength and m are the order of resonance.

If a thick substrate is used, the optical path length for reflected waves will increase and might lead to a change in resonance wavelength. However, a thicker substrate can support several resonant modes — higher orders m , corresponding to multiple absorption peaks. The thickness of the substrate plays a critical role for realizing effective impedance matching between the metamaterial and free space.

In metasurface, absorption $A(\lambda)$ is also a function of the impedance matching between the metamaterial and free space. The reflection coefficient $R(\lambda)$ is related to the impedance mismatch:

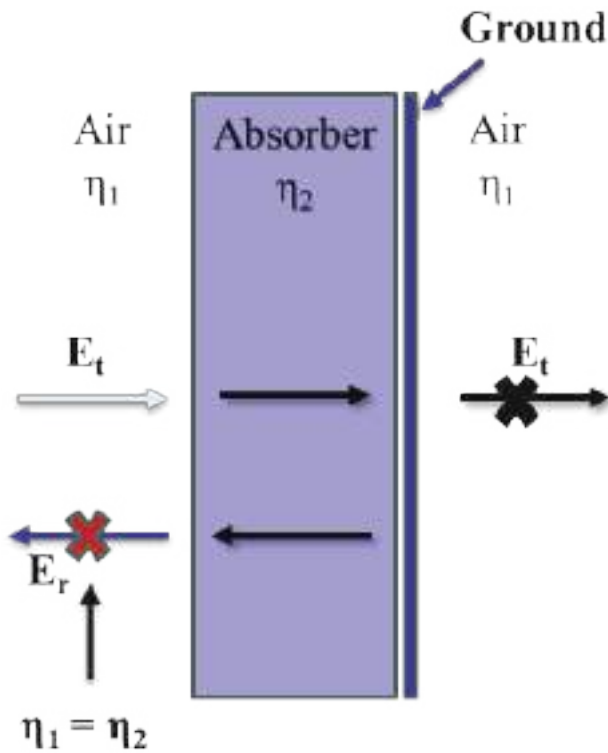


Figure 1. The structure of metasurface absorber.

$$R(\lambda) = \left| \frac{Z_{eff} - Z_0}{Z_{eff} + Z_0} \right|^2 \quad (2)$$

In other terms,

$$A(\lambda) = 1 - \left| \frac{Z_{eff} - Z_0}{Z_{eff} + Z_0} \right|^2 \quad (3)$$

For perfect absorption, Z_{eff} should be matched with Z_0 .

The basic block diagram of metasurface based perfect absorber is shown in Figure 2 which shows the basic design of the absorber. Based on this, the first structure is designed which is shown in Figure 3 which contains the three circular rings, made of copper material at top, FR4 dielectric material at middle layer and Copper material at bottom layer.

A triple circular ring metamaterial structure with a circular ring topology is depicted in Figure 3 to demonstrate the structure construction. The physical form of the substrate used for the structure is FR4 and has a dielectric constant of 4.6 and a loss tangent of 0.019. A diagram illustrating one possible metamaterial unit cell suggested for the design is shown below. It is made up of two metallic laminae separated by a dielectric lamina. The topmost metallic layer is constructed using three circular rings with different radii and widths for the triple resonance copper. These rings are spaced by two distinct gaps and both are individually tailored through full wave optimization. The ground plane is made of copper and is connected throughout the device from the lowest metallic level. The substrate dimensions of the unit cell are 9 mm X 9 mm X 0.8 mm in the x direction, y direction and thickness, respectively, where the radii of the circular rings denoted by are 2.33 mm, 2.78 mm, and 3.29 mm, respectively, which plays an important role in the electrical resonance. The widths of the rings are 0.28 mm for the first two and the third one a little bit smaller – 0.26. The first gap is 0.23 mm and the second one is 0.17 mm corresponding to the two rings. One at a time the rings are inserted – inner ring, middle ring and outer ring as illustrated in Figure 4, Figure 5 and Figure 6 respectively.

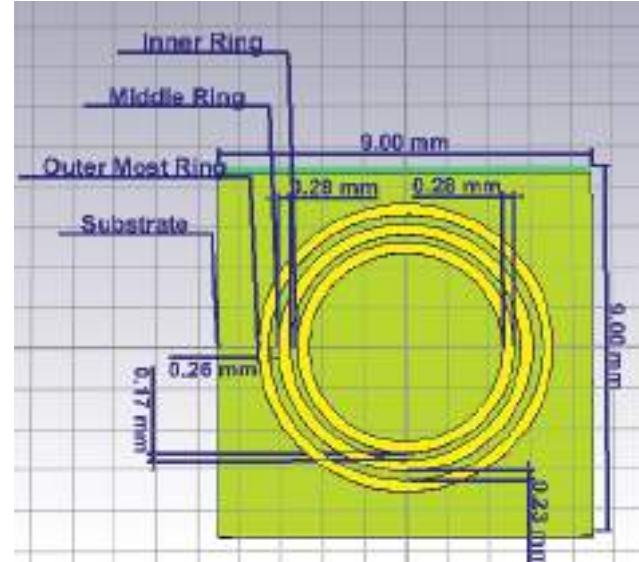


Figure 3. Design: 1 The geometry of the unit cell of the proposed metamaterial absorber.

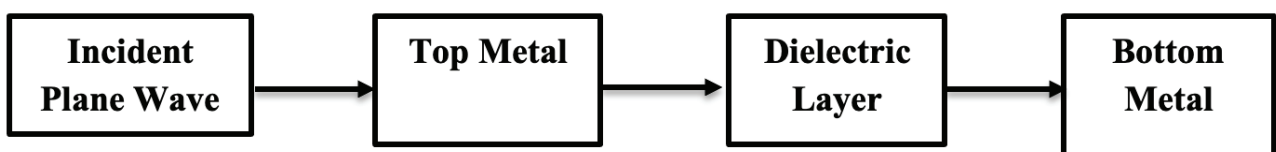


Figure 2. Basic block diagram of metasurface based perfect absorber.

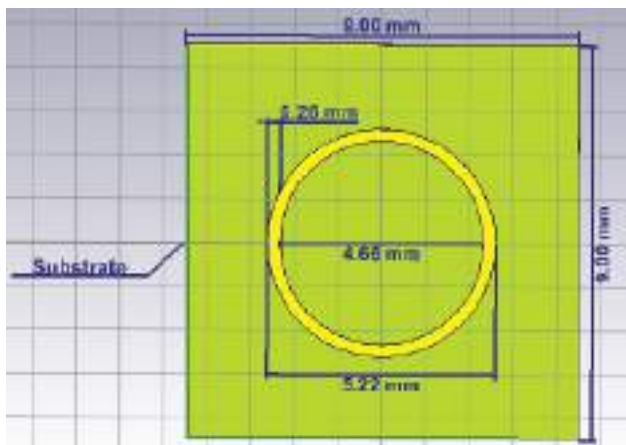


Figure 4. Design: 2 The geometry of the unit cell of the proposed metamaterial absorber with an inner ring.

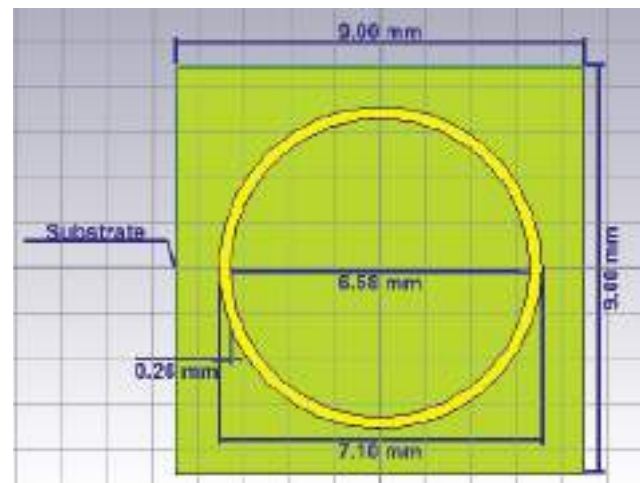


Figure 6. Design: 4 The geometry of the unit cell of the proposed metamaterial absorber with outer ring.

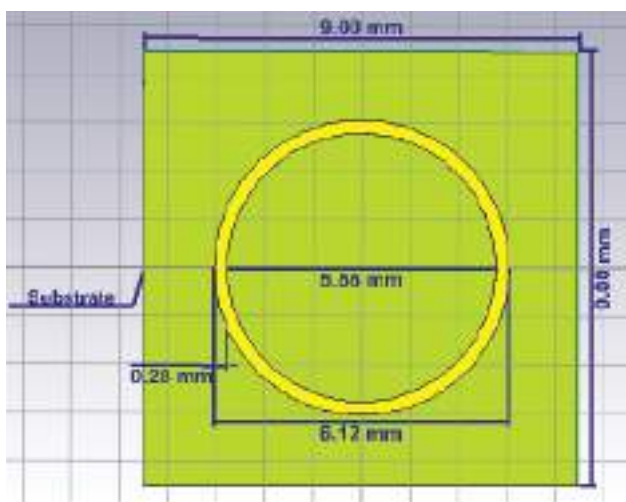


Figure 5. Design: 3 The geometry of the unit cell of the proposed metamaterial absorber with the middle ring.

Simulation

The simulation results for the triple circular ring metamaterial absorber with normal incident waves are presented in terms of high absorbance at three resonance frequencies indicated by $f = 7.865$ GHz, $f = 9.855$ GHz, and $f = 11.445$ GHz, and the corresponding absorbance are 96.62%, 78.668%, and 69.797%, respectively. The first, second, and third triple resonances are associated with the three smallest, medium, and biggest circular circles, respectively.

Figure 7 presents the absorbance values of the proposed metamaterial absorber for the TE-polarized incident wave obtained from simulations. The calculated resonant absorbance at resonant frequencies 7.865 GHz, 9.855 GHz, and 11.445 GHz are 96.62 %, 78.668 %, and 69.797 %. In so doing, the resonant frequencies remain almost constant, but there is a slight increase in the values; this occurs in the inner ring, in which case the resonances become 11.265 GHz with the corresponding absorbance levels as 96.32%

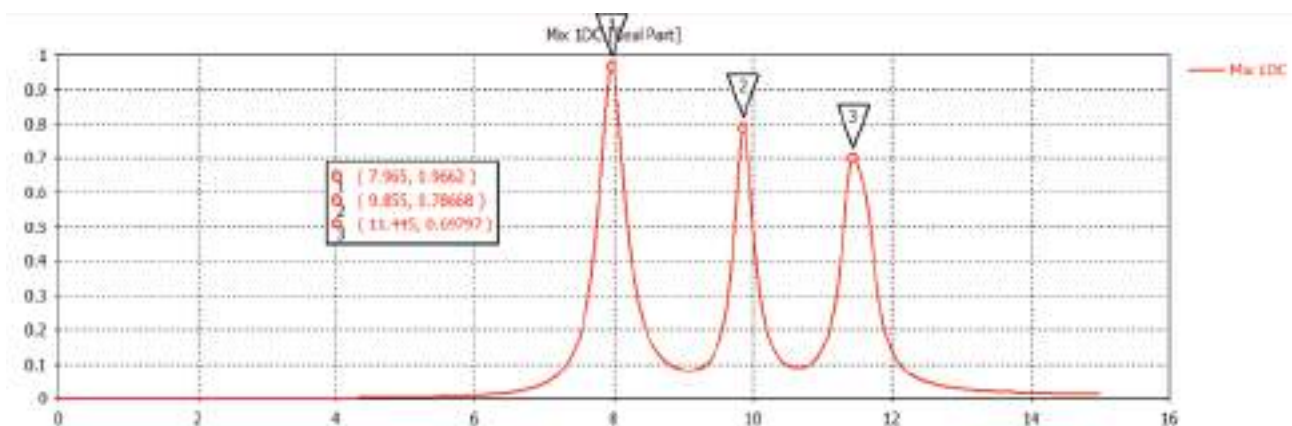


Figure 7. The frequency Vs absorption graph for design 1.

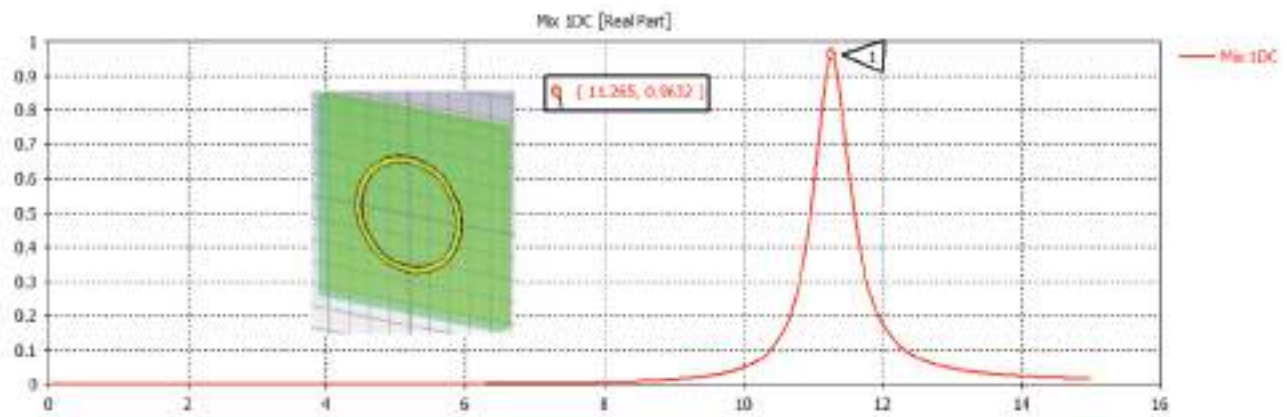


Figure 8. The frequency Vs absorption graph for design 2.

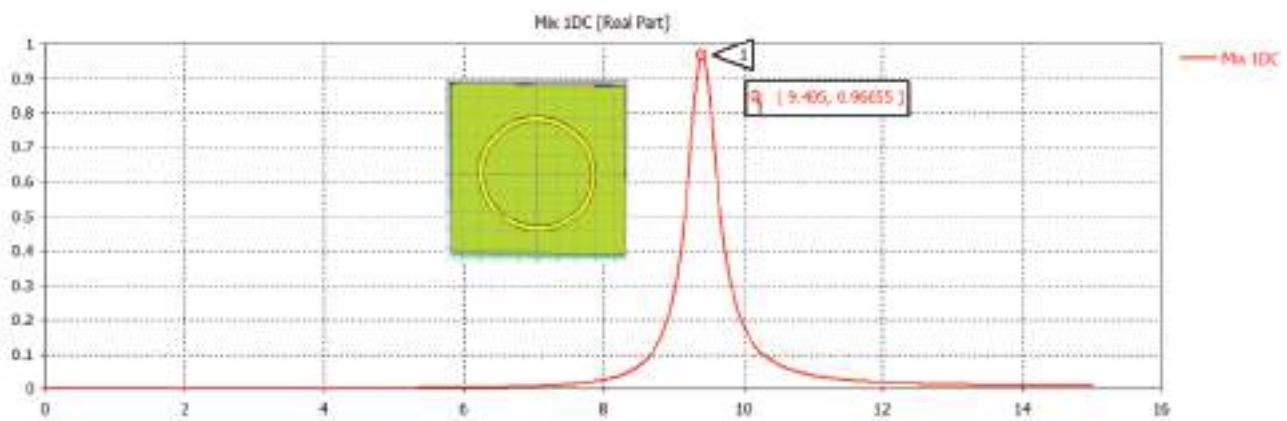


Figure 9. The frequency vs. absorption graph for design 3.

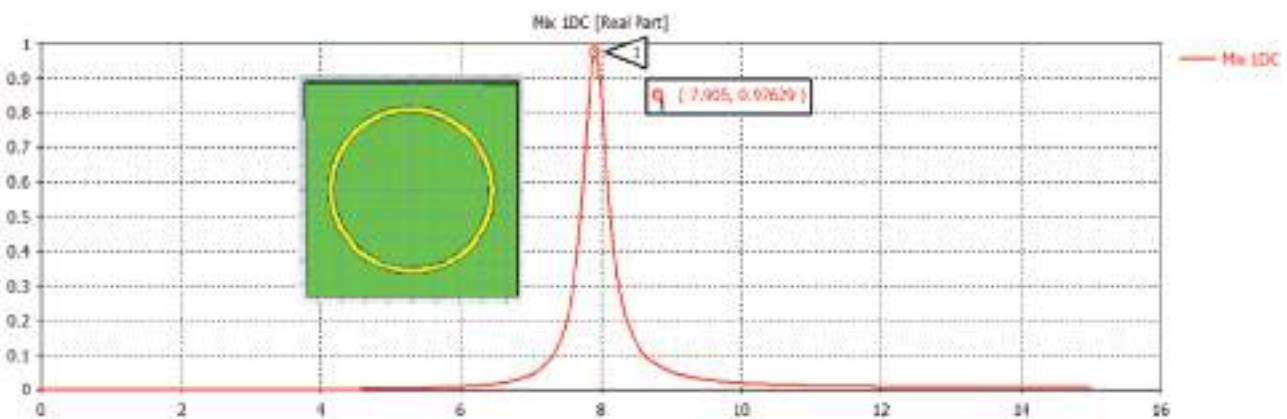


Figure 10. The frequency Vs absorption graph for design 4.

as shown in Figure 8. Similarly for middle ring subsequently the observed values of resonant frequencies are at 9.405 GHz along with the absorbance values are 96.655% as shown in Figure 9. At last, the measurement of resonant frequencies for outer ring is 7.905 GHz, and the absorbance

level corresponding to Figure 10 is 97.629% with the help of outer ring.

Table 1 shows the results of a comparative analysis that was carried out to assess the influence that the incorporation of circular ring structures in the metamaterial had on

Table 1. Comparative analysis of four deigns

No. of rings	Frequency in GHz	Absorption in %
Inner ring	11.265	96.32
Middle ring	9.405	96.66
Outer ring	7.905	97.63
All rings	7.965	96.62
	9.855	78.67
	11.445	69.80

Table 2. Comparative Analysis of deigns 2, 3 & 4

No. of rings	Frequency in GHz	Absorption in %
Inner ring	10.52	84.28
Middle ring	7.23	88.23
Outer ring	6.9	95.56

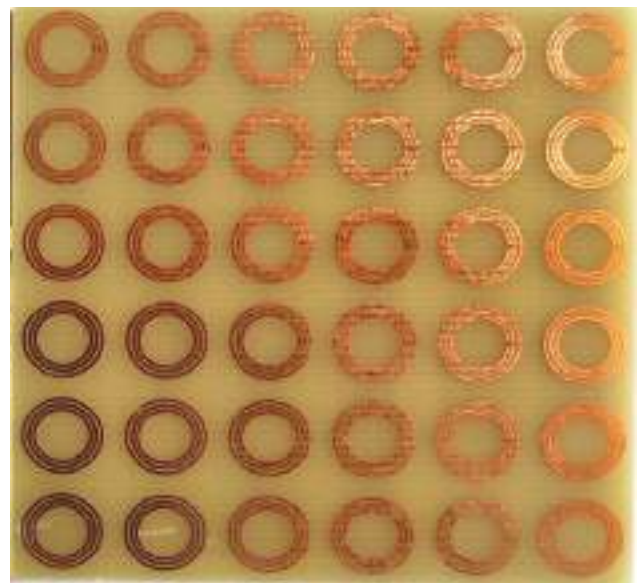
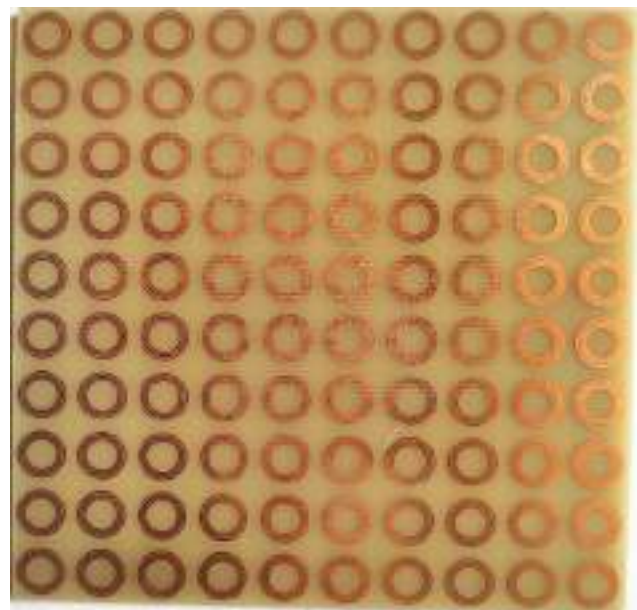
the overall design. When the rings in the structure were raised, the investigation noted the occurrence of resonance frequency; the distinct frequency at which maximum absorption takes place as 7.965 GHz. Moreover, this specific design showed a high degree of absorbance with 96.62% marked as a number.

A simulation study has been performed to investigate the impact of changing the diameters of the three rings in a metamaterial absorber structure on the resonance frequency and absorption properties of the proposed metamaterial absorber structure by providing the observation results in Table 2.

FABRICATION AND MEASUREMENT

The selection of the dimensions for the three circular rings of the metamaterial absorber is achieved through simulation. The fabricated triple circular ring metamaterial absorber is 54 mm × 54 mm corresponding to a 6 × 6 array of triple circular ring elements and 90 mm × 90 mm corresponding to a 10 × 10 array of triple circular ring elements using copper metallic traces printed on an FR-4 dielectric substrate using a printed circuit board (PCB) [Figure respectively (Fig. 11, Fig. 12). The structure is then put under measurement to confirm its EM absorbing features. Details of the measurement setup are shown in Figure 13 and include a microwave anechoic chamber. There are two horn antennas located on the semi-circle track which can be adjusted to specific angles. These antennas are also connected to a vector network analyzer (VNA) at port 1 (input) and port 2 (output). These electromagnetic waves are emitted by the horn antenna at port 1 and are radiated onto the sample sheet, while the horn antenna at port 2 receives the reflected waves. The VNA will read the reflection magnitude on the sample sheet. In addition, a pyramidal microwave absorber surrounds the sample

sheet to prevent unnecessary coupling of signals between the horn antennas. The transmitting and receiving horn antennas are connected with the VNA which measures reflection spectra from 1 GHz to 15 GHz. An aluminium board that is sectioned into two and acts as the ideal reference reflector is employed to calibrate the reflection measurement. For focusing only on reflection measurements, the metamaterial absorber sample is placed between four square aluminium plates of the same dimension – one square meter each. This structure helps to reduce transmission at all frequencies.

**Figure 11.** Fabricated 6 x 6 array.**Figure 12.** Fabricated 10 x 10 array.

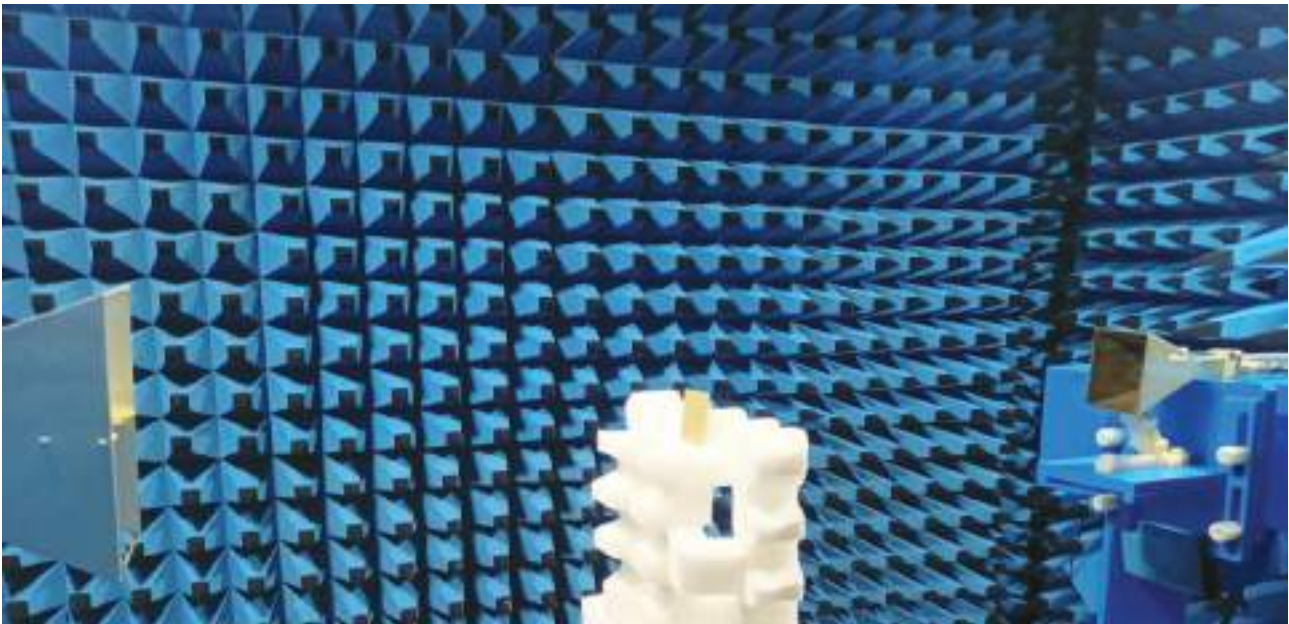


Figure 13. Experimental arrangement for validating the electromagnetic (EM) absorption properties of the metamaterial absorber.

Figure 14 and Figure 15 depict the measured EM absorbance of TE polarized wave under normal incidence for two arrays of 6×6 and 10×10 , respectively. The experimental data tallies with the simulation result, for most of the cases, with some deviations. It should be noted that the described object has several resonance frequencies: 7.832 GHz, 9.596 GHz, and 11 GHz. When the operating frequency was set to 248 GHz, the calculated absorbance was

97.802% of 77,909% and 80.351%, in that order. In the measurements, the above resonant frequencies are altered marginally to 7.8 GHz, 9.57 GHz, and 11.31 GHz having absorbance of 96.78%; 76.28%; and 76%, 65%, respectively. Figure 14 and Figure 15 represent the measured values for 6×6 array and 10×10 array metamaterial absorbers with substrate 0.8 mm thickness of polyoxymethylene. Figure 16 and Figure 17 give the measured values for metamaterial

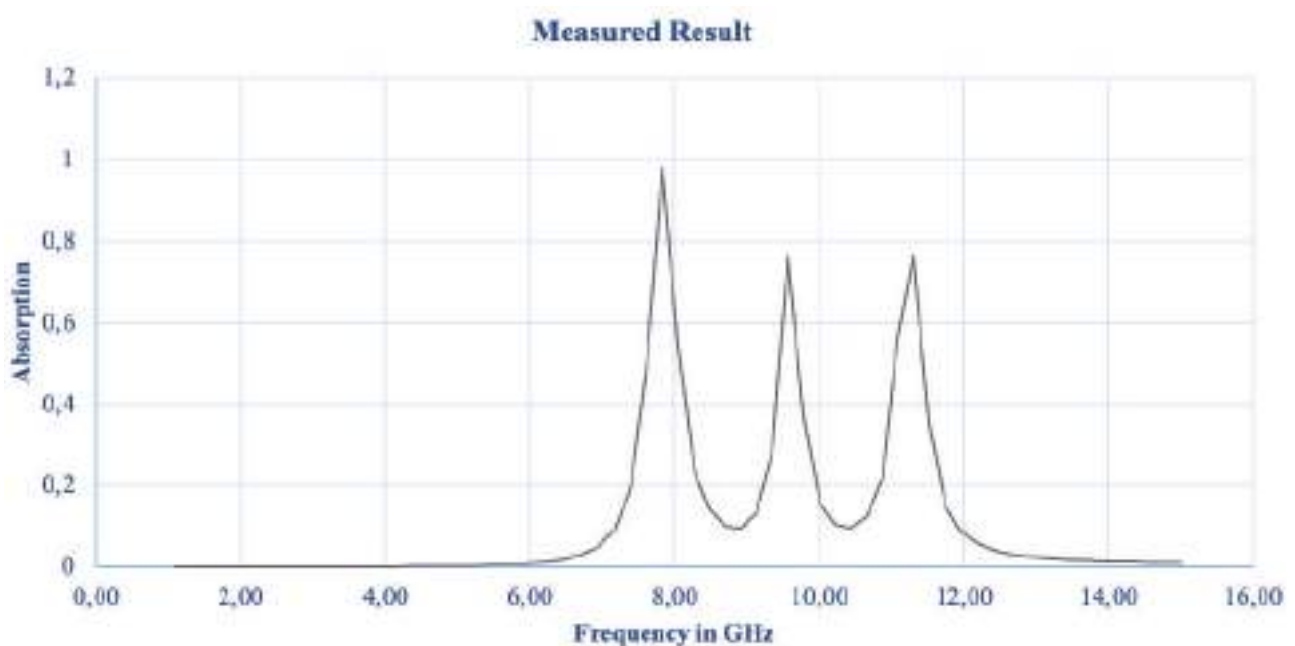


Figure 14. Measured Absorption (A) for 6x6 Array with thickness 0.8 mm.

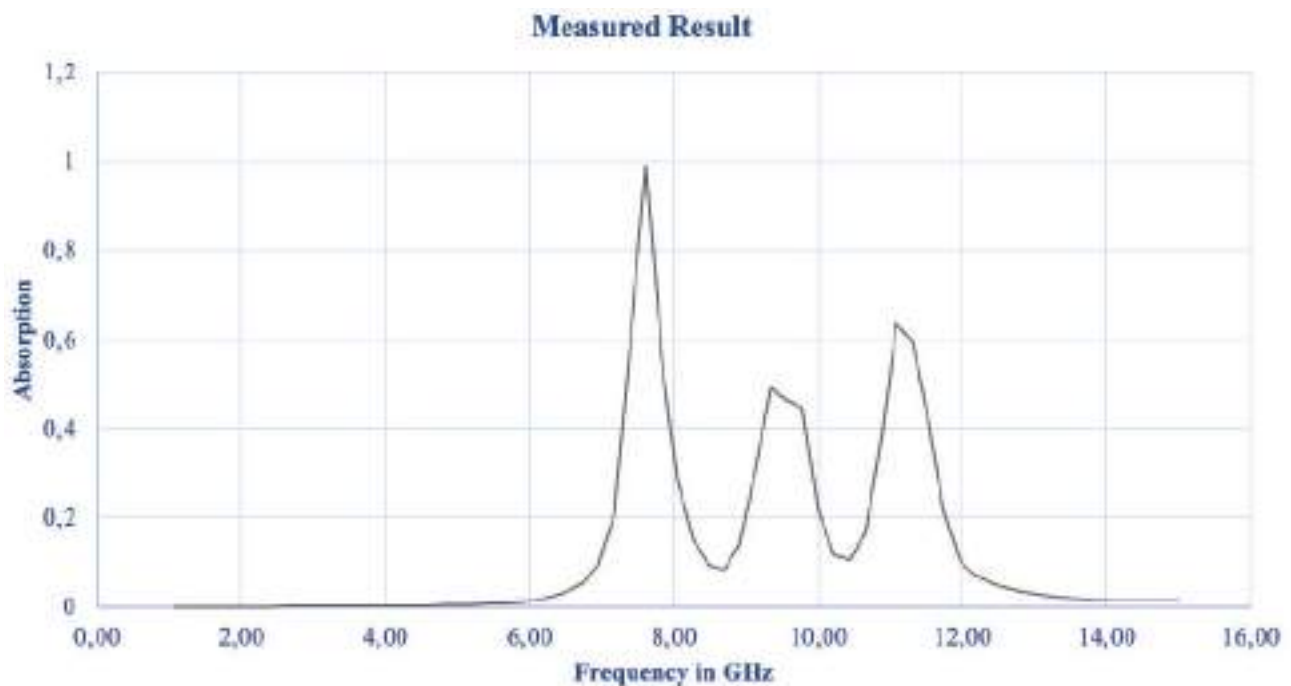


Figure 15. Measured Absorption (A) for 10x10 Array with thickness 0.8 mm.

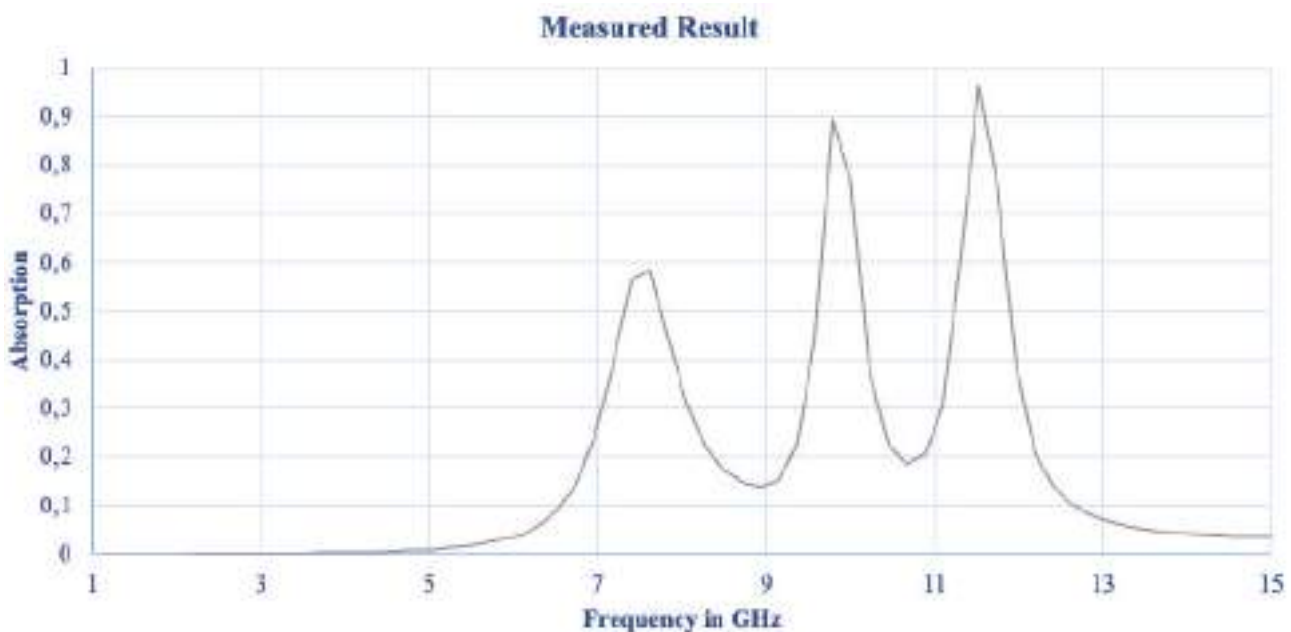


Figure 16. Measured Absorption (A) for 6x6 Array with thickness 1.6 mm.

absorbers with 6x6 array and 10x10 array structures with a 1.6 mm thickness of the substrate. The values of absorbance obtained from the simulation results and the experiment agreed with each other to a minor degree which might be due to mutual coupling of the horn antennas and scattering from the structure. Tolerances in fabrication and dispersion of substrate dielectric also affect the frequency shift.

Two metamaterial absorbers, one with a 6x6 array and the other with a 10x10 array are fabricated using substrates of 0.8mm and 1.6mm thickness, respectively. Experimental measurements are carried out, and the outcomes are contrasted in Table 3. The comparative analysis of stimulated and measured results is shown in Table 4.

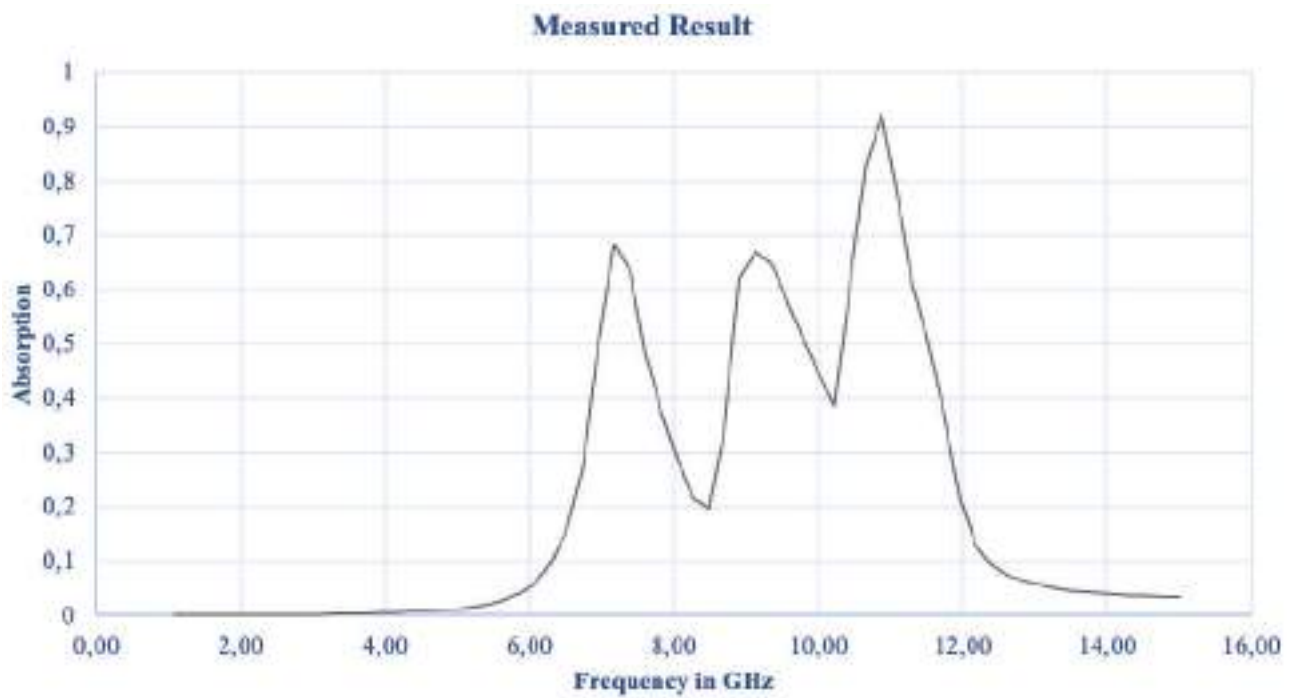


Figure 17. Measured Absorption (A) for 10x10 Array with thickness 1.6 mm.

Table 3. Comparative analysis with the thickness of the substrate

Thickness of Substrate	6 x 6 Array				10 x 10 Array			
	Simulated		Measured		Simulated		Measured	
	Freq. in GHz	Absorption in %	Freq. in GHz	Absorption in %	Freq. in GHz	Absorption in %	Freq. in GHz	Absorption in %
t = 1.6 mm	7.515	60.165	7.6	58.36	7.23	69.393	7.18	68.49
	9.87	99.04	9.78	89.09	9.05	68.89	9.14	67.03
	11.58	99.221	11.52	96.44	10.842	92.49	10.18	92.17
t = 0.8 mm	7.832	97.802	7.8	96.78	7.58	99.726	7.61	98.90
	9.596	77.909	9.57	76.28	9.4	49.61	9.35	49.23
	11.248	80.351	11.31	76.65	11.164	65.52	11.09	63.69

Table 4. Comparative Analysis of the proposed design

Ref.	Size (mm)	Frequency Range (GHz)	Design	Absorption (%)	Variation of Substrate Thickness (mm)	Array
[5]	10 x 10	5.376 10.32 12.25	a double E-shaped symmetric SRRs	Above 99%	No	No
[7]	35 x 32	1.21 3.64 5.30	a modified square-shape closed-loop resonator	94 % 90 % Above 99 %	No	No
Proposed design	9 x 9	7.515 9.87 11.58	Triple Circular Ring	60.17 Above 99 % Above 99 %	Yes t = 0.8 mm t = 1.6 mm	Yes 6 x 6 10 x 10

CONCLUSION

In conclusion, all steps connected with the design, fabrication, and evaluation of a triple circular ring metamaterial absorber have been completed. Moreover, a contribution towards understanding the relationship between the number of circular rings in the structure and the substrate thickness on absorbance has been made. The evaluation of the absorber is made in terms of absorbance and thickness of the substrate. Measured and simulated results also show fair comparison with small frequency shifts. Studies on the structural parameters of the triple circular ring metamaterial absorber are presented using analyses of the substrate thickness. Future research will focus on implementing structures that effectively work at high absorbance levels at multi-band frequencies without compromising at least 95% absorbance of the electromagnetic waves that hit them at right angles.

The strengths of this work are the high absorption of EM in three frequencies, the crystalline structure analysis at the atomic level, and the satisfactory match between the theoretical and the real model. Still, the drawbacks are found in the form of minor variations either in frequency or fabrication tolerance. Moreover, one can mention that subsequent research work should be devoted to improving the absorbance in other spectral regions, investigation of various materials and constructions, as well as, stability and effectiveness of PEF coatings in real conditions. Possible future studies like the development of absorbers for various angles of incidence and polarization, as well as the integration of absorber elements into actual goods. Further research on efficient manufacturing techniques is needed to encourage the use of the said absorbers. Discussions upon these kinds of challenges and novel trends can result in great advancements in the design of metamaterial absorbers where the present study facilitates the improvement of the given area of research.

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] Patel J, Mehta R. Simulation-based comparative review on a metasurface-based perfect absorber. *J Adv Eng Comput* 2023;7:204–209. [\[CrossRef\]](#)
- [2] Jahan MI, Faruque MRI, Hossain MB. An X-shaped triple split ring resonator-based metamaterial perfect absorber with quad-band incident and polarization angle insensitivity for C, X, and Ku band applications. *J Magn Magn Mater* 2023;170940. [\[CrossRef\]](#)
- [3] Jahan MI, Faruque MRI, Hossain MB, Khandaker MU, Elsayed F, Salman M, et al. Quad-band metamaterial perfect absorber with high shielding effectiveness using double X-shaped ring resonator. *Materials* 2023;16:4405. [\[CrossRef\]](#)
- [4] Ullah N, Islam MS, Hoque A, Yong WH, Soliman MS, Islam MT. A compact-sized four-band metamaterial-based perfect absorber for electromagnetic energy harvesting applications. *Opt Laser Technol* 2024;168:109836. [\[CrossRef\]](#)
- [5] Hossain MB, Faruque MRI, Islam MT, Singh M, Jusoh M. Triple band microwave metamaterial absorber based on double E-shaped symmetric split ring resonators for EMI shielding and stealth applications. *J Mater Res Technol* 2022;18:1653–1668. [\[CrossRef\]](#)
- [6] Liu Z, Zhang J, Xu Z, Yuan Z, Mei Z, Niu T. Realization of an ultra-thin absorber with fragmented magnetic structure at L-, S-, and partial C-bands. *J Appl Phys* 2023;134. [\[CrossRef\]](#)
- [7] Shahid M, Shanmuganantham T, Kumar SA. Metamaterial absorber for L-band, S-band, and C-band applications. Preprint 2023. [\[CrossRef\]](#)
- [8] Jahan MI, Faruque MRI, Hossain MB, Abdullah S. An ultra-thin, triple-band, incident angle-insensitive perfect metamaterial absorber. *Materials* 2023;16:1623. [\[CrossRef\]](#)
- [9] Ullah N, Islam MS, Hoque A, Yong WH, Soliman MS, Albadran S, et al. A compact complementary split ring resonator (CSRR) based perfect metamaterial absorber for energy harvesting applications. *Eng Sci Technol Int J* 2023;45:101473. [\[CrossRef\]](#)
- [10] Rajyalakshmi G, Kumar YR, Krishna DR, Rao KS. Angle independent metamaterial absorber for S and C band application. *Prog Electromagn Res C* 2023;130:241–254. [\[CrossRef\]](#)
- [11] Amer AAG, Sapuan SZ, Alzahrani A, Nasimuddin N, Salem AA, Ghoneim SSM. Design and analysis of polarization-independent, wide-angle, broadband metasurface absorber using resistor-loaded split-ring resonators. *Electronics* 2022;11:1986. [\[CrossRef\]](#)

- [12] Hasan M, Islam MT, Almalki SHA, Alharbi AG, Alsaif H, Islam M, et al. Polarization insensitive dual band metamaterial with absorptance for 5G sub-6 GHz applications. *Sci Rep* 2022;12:1–20. [\[CrossRef\]](#)
- [13] Hakim ML, Alam T, Almutairi AF, Mansor MF, Islam MT. Polarization insensitivity characterization of dual-band perfect metamaterial absorber for K band sensing applications. *Sci Rep* 2021;11:17829. [\[CrossRef\]](#)
- [14] Hakim ML, Alam T, Islam MT, Baharuddin MH, Alzamil A, Islam MS. Quad-band polarization-insensitive square split-ring resonator (SSRR) with an inner Jerusalem cross metamaterial absorber for Ku- and K-band sensing applications. *Sensors* 2022;22:4489. [\[CrossRef\]](#)
- [15] Kaur M, Singh HS. Design and analysis of a compact ultrathin polarization- and incident angle-independent triple band metamaterial absorber. *Microw Opt Technol Lett* 2020;62:1920–1929. [\[CrossRef\]](#)
- [16] Deng G, Lv K, Sun H, Yang J, Yin Z, Li Y, et al. An ultrathin, triple-band metamaterial absorber with wide-incident-angle stability for conformal applications at X and Ku frequency band. *Nanoscale Res Lett* 2020;15:1–10. [\[CrossRef\]](#)
- [17] Zhao G, Bi S, Cui Y. Study on the characteristics of a V-shaped metamaterial absorber and its application. *AIP Adv* 2019;9:075016. [\[CrossRef\]](#)
- [18] Singh RK, Gupta A. A wrenched-square shaped polarization-independent and wide-angle stable ultra-thin metamaterial absorber for S-band, X-band and Ku-band applications. *AEU Int J Electron Commun* 2021;132:153648. [\[CrossRef\]](#)
- [19] Naqvi SA, Baqir MA, Gourley G, Iftikhar A, Khan MS, Anagnostou DE. A novel meander line metamaterial absorber operating at 24 GHz and 28 GHz for the 5G applications. *Sensors* 2022;22:3764. [\[CrossRef\]](#)
- [20] Shakiba L, Salehi MR, Emami F. A multiband perfect metamaterial absorber based on phase change material for switching and biosensing applications. *Opt Commun* 2024;560:130498. [\[CrossRef\]](#)
- [21] Afsar MSU, Faruque MRI, Abdullah S, Hossain MJ. Rotational symmetric solar system shaped triple band perfect metamaterial absorber for S-, C-, and X-band application. *Sens Actuators A Phys* 2024;365:114839. [\[CrossRef\]](#)



Research Article

The effects of using lithium on biochemical factors of urinary beta-2 microglobulin, creatinine and urea serum in chronic bipolar patients

Seyed Reza MOMENI SHIDEH^{1,*}, Amireh Ebrahimi MOGHADAM², Zainab WETWET³,
Hayder Abbas Hussein AL MAHDI⁴

¹Department of toxicology, Faculty of Pharmacy, Islamic Azad University, Shahreza, 1477893855, Iran

²Department of Biology, Faculty of Science, Arak University, Arak 384817758, Iran

³College of Materials Engineering, University of Babylon, Babylon, 51002, Iraq

⁴College of Arts, University of Babylon, Babylon, 51002, Iraq

ARTICLE INFO

Article history

Received: 09 July 2022

Revised: 14 November 2022

Accepted: 04 April 2023

Keywords:

Chronic Bipolar Patients,
Creatinine; Lithium, Urea Serum;
Urinary Beta-2 Microglobulin

ABSTRACT

In this research, the effects of using lithium on biochemical factors of urinary beta-2 microglobulin, creatinine and urea serum in chronic bipolar patients hospitalized in Razi psychiatric center are investigated. For this purpose, random sampling method was used and among hospitalized patients in Tehran's Razi psychiatric center. Then 87 patients were selected and urinary beta-2 microglobulin tests by ELISA method, urea and blood serum by enzymatic method measured in 42 bipolar patients (including 18 females - 24 males) that using lithium and 45 bipolar patients (including 21 females - 24 males) that using drugs other than lithium. According to studies, lithium-treated bipolar patients have a substantially greater urine beta-2 microglobulin concentration than individuals who are not lithium-treated ($P_{\text{vale}} > 0.028$). The results demonstrated that the longer time of using lithium will cause higher amount of urinary beta-2 microglobulin secretion ($P_{\text{val}} > 0.027$). Also, the results showed that the value of creatinine and urea serum were not significantly different between two groups. Finally, it seems that urinary beta-2 microglobulin is a better indicator than urea and creatinine serum to evaluate the side effects of primary kidney damages caused by lithium, and its main role in diagnosis of the disease requires further researches in this field.

Cite this article as: Momeni Shideh SR, Moghadam AE, Wetwet Z, Al Mahdi HAH. The effects of using lithium on biochemical factors of urinary beta-2 microglobulin, creatinine and urea serum in chronic bipolar patients. Sigma J Eng Nat Sci 2025;43(5):1704–1711.

INTRODUCTION

One of the most important psychiatric drugs that causes side effects is lithium, and there is a concern that treatment with lithium can cause kidney problems and, in rare cases,

dialysis of patient [1]. Therefore, psychiatrists who use this drug to treat bipolar patients, in order to reduce toxicities and drug side effects, check the blood serum lithium levels of patients on a monthly basis. First, kidney dysfunction in bipolar patients treated with lithium reported in 1970 [2,3].

*Corresponding author.

*E-mail address: r.momenishideh1978@gmail.com

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



Lithium is known as the first-line treatments for the manic and depressive phases of bipolar disorder. It protects against depression and mania as well as being the only therapy known to decrease suicide risk in this patient population [3] with a composite measure of suicide plus deliberate self-harm lowered in patients receiving lithium [4,5]. Even with the availability of newer treatments, lithium is still considered the first-line treatment when prescribing a relapse-prevention drug in bipolar disorder. It has efficacy in the prevention of both manic and depressive episode relapse or recurrence [6-10], having the most robust evidence for long-term relapse prevention [11].

In one study, as we need to know about the effect of lithium on the kidney stated that, lithium is a valuable drug in recent decades for the treatment of bipolar patients, although therapeutic use of lithium due to its toxicity in various body systems is problematic. Kidney side effects include polyuria, nephrogenic diabetes, proteinuria, kidney tubular acidosis, and decreased kidney filtration. Pathologically, chronic toxicity of lithium in the kidney is diagnosed by interstitial nephritis with small glomerulus cysts. Although this toxicity is not common, but the risk increases with the dose of lithium usage and the duration of use [12-15]. In an experimental model, lithium has a protective and restorative role in proteinuria diseases. This duality or inconsistency may be due to the short time use or low dose of the drug. However, long term use and high doses of lithium can cause nephrotoxic agents. Short time and low dose of lithium usage may be a point for improving kidney function [9]. In Poznan, Poland, a group of 111 patients (25 male, 76 female) who received lithium for 5-39 (mean 18) years regarding clinical factors associated with the prophylactic efficacy of lithium was assumed. It is found a better effect in patients with later illness onset, no family history of mood disorders, and family members on prophylactic lithium, as well as in women with comorbid anxiety disorders and men who did not abuse alcohol [10]. Recently, British researchers performed a meta-analysis of 71 studies with a total of 12,000 patients regarding clinical factors associated with quality of prophylactic lithium response. They identified such positive features as a mania-depression-remission episode sequence, non-rapid-cycling, no psychotic episodes, a family history of bipolar disorder, short illness duration prior to lithium initiation, and later illness onset. Fewer episodes and hospitalizations prior to lithium treatment, as well as lithium treatment in other family members were also described as favorable factors [16,17].

In a 2020 study, Nunes *et al.* [23] assessed 1,266 lithium-treated patients from seven European and Canadian centers. In agreement with previous results, the authors demonstrated that an episodic clinical course and non-rapid-cycling were the most significant factors for a good lithium response [18-20].

Beta-2 microglobulin, is also a small molecule with a molecular weight of 11,800 daltons that is observable on the surface of human nucleated cells and in part of MHC

calss1 (Major Histocompatibility Complex). Increased beta-2 microglobulin occurs in kidney malignancies, immune system defects and autoimmune diseases. Beta-2 microglobulin dependence is evident in cardiac death, CKD and hemodialysis as well as thrombotic and vascular diseases. However, the association between MIA syndrome (Malnutrition, Inflammation, and Atherosclerosis) and beta-2 microglobulin has not been well studied [21-24]. Therefore, the diagnosis of drug therapies' side effects, especially the side effects of lithium in patients, and taking timely and effective measures to minimize these side effects in addition to reducing physical and behavioral problems, increases the patient's motivation to use the drug [25-27]. Finally, the objective of this research is evaluating the effects of using lithium on biochemical factors of urinary beta-2 microglobulin, creatinine and urea serum in chronic bipolar patients hospitalized in Razi psychiatric center by conducting several tests on specific persons.

MATERIAL AND METHODS

Materials

This research is a retrospective or comparative scientific research. The statistical population of this research includes all patients with bipolar disorder who are hospitalized in different wards of psychiatric center. The sample size of this study is 87 patients (male and female) with chronic bipolar disorder. Out of the sample number, 43 bipolar patients treated with lithium were randomly selected as experimental group. And 45 bipolar patients treated with drugs other than lithium were placed in the control group. The target group consisted of 24 males and 18 females and the control group included 24 males and 21 females. The method of testing conduction is as follows:

Beta-2 Microglobulin Test

Beta-2 microglobulin can be measured by quantitative luminescence and ELISA methods, we used ELISA method.

Urea Test

Urea can be measured enzymatically (indirectly) and chemically (directly). In this research, the chemical method has been used.

Creatinine Test

Creatinine is similar to urea can be measured in two methods: chemical and enzymatic methods. In this article, we used the chemical method (Pars test kits).

For statistical analysis, SPSS-20 computer software was used and for this purpose, to test the research hypotheses through t-test for independent groups was used.

RESULTS AND DISCUSSION

The linear regression model that used here is also known as covariance analysis models. In this model, the variables

of gender and type of used drug are considered as the factor and the variables of age and duration of drug usage are considered as continuous predictor variables or urgency. Tables (1) to (9) demonstrate the results of fitting the linear regression model with each of the beta-2 microglobulin, creatinine and urea variables as response variables.

Beta-2 Microglobulin

In order to evaluate the variables of gender, duration of use, patient age and type of used drug, the amount of beta-2 microglobulin measured and linear regression model was used. The variance analysis of the regression model to evaluate the relation between mentioned variables and the amount of beta-2 microglobulin in patients is shown in Table (1). The variance of a variable shows the amount of fluctuations or variability of a variable. In a homogeneous population where no factor affects the decrease or increase of beta-2 microglobulin, no variability will be observed in the beta-2 microglobulin variable. These factors can be gender, age of the patient, duration of drug usage and type of drug or even other factors that have not been considered by the researcher. Other factors in the modeling literature are called error or residual factors. Therefore, in the absence of any effective factor on beta-2 microglobulin (factors considered in the research and error factor), the variance of this variable will be zero.

Table (1) demonstrates the contribution of each variable in increasing the variance of the response variable in the sum of square column. If the contribution of one factor is large enough compared to the contribution of other factors (error component), we say that the factor has a significant

relation with the response variable, because by changing the response variable (increasing or decreasing it) the variance increases. According to the table above, the effect of drug type as well as the interaction effect of drug type and duration of usage is significant at the test level of 0.05. The effect of other factors is not significant at the test level of 0.05.

It should be noted that the tests carried out in Table (2) are based on hypotheses, the most important of which are the assumption of variance homogeneity of the response variable and the normality of the error component distribution in the regression model. If either of these two hypotheses is not observed, the results presented in the table above will not be valid. Between these two assumptions, deviation from the assumption that the error component is normal when the sample size is large enough has less effect on the validity of the reported results. The variance homogeneity hypothesis states that the variance of the error component is not related to any of the predictor variables in the regression model. This hypothesis can be tested by modeling the quadratic power of the previous regression model with the predictor variables shown in Table (2). If at least one of the predictor variables has a significant relation with the quadratic power of the regression model, the assumption of variance homogeneity will be rejected.

The result of variance homogeneity test is reported in Table (2). The null hypothesis in the hypothesis test performed in this table is the variance homogeneity of the error component. As can be seen, the assumption of variance homogeneity at the test level of 0.05 was not rejected.

Therefore, according to the assumption of variance homogeneity, the results reported in Table (2) can be cited

Table 1. The variance analysis of regression model for evaluating beta-2 microglobulin

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value	Power of the test
Type of drug	4/192	1	4/192	7/682	0/007	0/782
Gender	0/954	1	0/954	1/748	0/190	0/257
Interaction effect between drug type and gender	0/049	1	0/049	0/090	0/765	0/060
Age	0/036	1	0/036	0/067	0/797	0/057
Duration of use	0/001	1	0/001	0/001	0/994	0/050
Interaction effect between drug type and duration of use	2/766	1	2/766	5/069	0/027	0/604
Error	43/652	80	0/546			
Revised total	49/421	86				

Table 2. Test of variance homogeneity hypothesis in regression model

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value
Revised model	436/383	6	72/730	1/033	0/410
Error	5630/157	80	70/377		
Revised total	6066/540	86			

that based on the result, it was observed that the type of drug used as well as the interaction effect between the type of drug used and the duration of the use at the test level of 0.05 is significant. In order to determine the relation between these variables and the level of beta-2 microglobulin in the blood serum of patients, we will refer to the table for estimating the parameters of the regression model. Parameters estimation of this regression model is reported in Table (3). In this test table, the hypothesis related to each of the model parameters is reported in two methods, parametric and non-parametric. The non-parametric tests were performed using the Bootstrap re-sampling algorithm. Bootstrap algorithm is a non-parametric method based on data re-sampling to estimate the analysis distribution of estimators to provide confidence interval, bias calculation, standard error and hypothesis testing. In Table (3), the reported asymptotic result is that if the error component is not assumed to be normal, the results obtained are almost correct if the sample size is large.

According to the table above, the effect of drug type as well as the interaction effect of drug type and duration of usage at the test level of 0.05 was significant with both parametric and non-parametric test methods. According to the estimation of the parameters, it can be inferred that the

mean of beta-2 microglobulin in patients who have used lithium-containing drugs is on average 0.759 higher than patients who have used lithium-free drugs.

Given to the parameter estimation of the interaction effect between the duration of use and the type of drug, it can be inferred that with each year of increasing the period of drug use, the effect of lithium drug increases by an average of 0.037 mg / μ l.

Creatinine

In order to evaluate the effect of gender, duration of usage, patient age and type of used drug variables on the amount of creatinine measured in the serum of patients, a linear regression model was used. Table (4) demonstrates the result of variance analysis of this regression model.

Table (4) demonstrates the contribution of each variable in increasing the variance of the response variable in the sum of squares column. According to the table above, the probability value of gender effect and age of the patient is close to the test level of 0.05.

The assumption of variance homogeneity in this regression model has been studied by modeling the quadratic power of its residuals and the result is reported in Table (5). If at least one of the predictor variables has a significant

Table 3. Parameters estimation of regression model for evaluating beta-2 microglobulin

Parameter	Estimation	Parametric (asymptotic)			Non-parametric (Bootstrap)	
		Standard error	T	Probability value	Standard error	Probability value
Intercept	0/215	0/409	0/526	0/601	0/353	0/544
Type of drug	0/759	0/339	2/239	0/028	0/365	0/048
Gender (male)	0/268	0/228	1/178	0/242	0/177	0/139
Type of drug (lithium) $\times$ Gender (male)	0/103	0/343	0/300	0/765	0/346	0/760
Age	0/002	0/008	0/258	0/797	0/007	0/743
Duration of use	0/019	0/013	1/481	0/142	0/015	0/192
Type of drug (lithium) $\times$ Duration of use	0/037	0/017	2/252	0/027	0/018	0/044

Table 4. Variance analysis of regression model for evaluating the creatinine amount in patients

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value	Power of the test
Type of drug	0/043	1	0/043	0/514	0/475	0/109
Gender	0/312	1	0/312	3/685	0/058	0/475
Interaction effect between drug type and gender	0/000	1	0/000	0/000	0/985	0/050
Age	0/318	1	0/318	3/761	0/056	0/483
Duration of use	0/108	1	0/108	1/275	0/262	0/200
Interaction effect between drug type and duration of use	0/075	1	0/075	0/887	0/349	0/154
Error	6/763	80	0/085			
Revised total	8/150	86				

Table 5. Test of variance homogeneity hypothesis in regression model for evaluating the creatinine amount in patients

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value
Revised model	0/137	6	0/023	0/829	0/551
Error	2/206	80	0/028		
Revised total	2/343	86			

Table 6. Parameters estimation of regression model for evaluating the creatinine amount in patients

Parameter	Estimation	Parametric (asymptotic)			Non-parametric (Bootstrap)	
		Standard error	t	Probability value	Standard error	Probability value
Intercept	0/759	0/161	4/713	0/000	0/172	0/001
Type of drug (lithium)	0/071	0/134	0/530	0/597	0/121	0/558
Gender (male)	0/123	0/090	1/367	0/175	0/111	0/272
Type of drug (lithium) × Gender (female)	0/003	0/135	0/019	0/985	0/138	0/976
Age	0/006	0/003	1/939	0/056	0/002	0/044
Duration of use	0/008	0/005	1/523	0/132	0/005	0/115
Type of drug (lithium) × Duration of use	0/006	0/007	0/942	0/349	0/006	0/296

relationship with the quadratic power of the model residues shown in Table (5), the assumption of variance homogeneity will be rejected. Given to the result reported in this table, the hypothesis of variance homogeneity at the test level of 0.05 is not rejected.

In order to determine the relation between the variables of gender and age of the patient with the level of creatinine in the urine of patients, we will refer to the table of estimation parameters of the regression model. Parameters estimation of this regression model is reported in Table (6). In this test table, the hypothesis related to each of the model parameters is reported in two methods, parametric and non-parametric. The non-parametric tests were performed using a Bootstrap sampling algorithm.

According to the result of non-parametric Bootstrap test in the table above, only the effect of patient age is significant

at the test level of 0.05. Note that the results reported by the non-parametric method are more conservative than the parametric method. In this study, because the sample size is not large enough, the result reported by the non-parametric method is the decision criterion. According to the estimation of the parameters, it can be inferred that the creatinine level in the serum of patients increases with an average of 0.006 mg / dl with each year of increase in age.

Urea

In order to evaluate the effect of gender, duration of usage, patient age and type of used drug variables on the amount of urea measured in the serum of patients, linear regression model was used. Table (7) demonstrates the result of variance analysis of this regression model.

Table (7) demonstrates the contribution of each variable in increasing the variance of the response variable in the

Table 7. Variance analysis of regression model for evaluating the urea amount in patients

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value	Power of the test
Type of drug	45/382	1	45/382	0/582	0/448	0/117
Gender	32/246	1	32/246	0/413	0/522	0/097
Interaction effect between drug type and gender	0/504	1	0/504	0/006	0/936	0/051
Age	2/418	1	2/418	0/031	0/861	0/053
Duration of use	896/204	1	896/204	11/485	0/001	0/917
Interaction effect between drug type and duration of use	135/550	1	135/550	1/737	0/191	0/256
Error	6242/450	80	78/031			
Revised total	7997/103	86				

Table 8. Test of variance homogeneity hypothesis in regression model for evaluating the urea amount in patients

Source of variations	Sum of square	Degree of freedom	Mean of square	F	Probability value
Revised model	200757/210	6/000	33459/535	1/613	0/154
Error	1659291/998	80/000	20741/150		
Revised total	1860049/208	86/000			

Table 9. Parameters estimation of regression model for evaluating the urea amount in patients

Parameter	Estimation	Parametric (asymptotic)			Non-parametric (Bootstrap)	
		Standard error	t	Probability value	Standard error	Probability value
Intercept	24/911	4/892	5/092	0/000	4/367	0/001
Type of drug (lithium)	-2/494	4/056	-0/615	0/540	3/418	0/461
Gender (male)	-1/426	2/724	-0/523	0/602	2/578	0/575
Type of drug (lithium) × Gender (male)	0/330	4/105	0/080	0/936	4/137	0/933
Age	0/017	0/095	0/176	0/861	0/086	0/839
Duration of use	0/280	0/063	2/254	0/047	0/126	0/025
Type of drug (lithium) × Duration of use	0/261	0/198	1/318	0/191	0/228	0/263

sum of square column. According to the table above, the duration of use effect on urea level in patients' serum at the test level of 0.05 is significant.

The assumption of variance homogeneity in this regression model is evaluated by modeling the quadratic power of the residuals of this model and the result is reported in Table (8). If at least one of the predictor variables has a significant relationship with the quadratic power of the model residues shown in Table (7), the assumption of variance homogeneity will be rejected. According to the result reported in this table, the hypothesis of variance homogeneity at the test level of 0.05 is not rejected.

In order to determine the relation between the duration of use variable and the level of urea in the urine of patients, we will refer to the table of parameters estimation of the regression model. Parameters estimation of this regression model is reported in Table (9). In this test table, the hypothesis related to each of the model parameters is reported in two methods, parametric and non-parametric. The non-parametric tests were performed using the Bootstrap resampling algorithm.

According to the table above, the test performed by non-parametric Bootstrap method also confirms the result obtained in the analysis of variance table. In the table above, only the effect of duration of usage at the test level of 0.05 is significant. According to the parameter estimation column, it can be inferred that the level of urea in the serum of patients increases with an average of 0.28 mg / dl with increase of each year of use.

CONCLUSION

In the present research, the results demonstrated that patients using lithium had a higher mean of urinary beta-2 microglobulin than those who did not use it. Although this difference was not observable for urea and serum biomarkers, but it can also be inferred from this study that the mean of urinary beta-2 microglobulin in bipolar patients using lithium. On average, 0.258 µg / ml was higher in men and 0.283 µg / ml in women than in patients who did not use lithium. A similar study in 2013 reported that women and men who used lithium had an increase in urinary beta-2 microglobulin and a decrease in urinary specific gravity and GFR compared to those who did not use lithium [15]. In our study, it was evident that the longer the lithium is used, the higher the urinary beta-2 microglobulin secretion, and given the significance of interaction effect between the duration of use and the type of drug, it can be inferred that with each year increasing lithium use, the amount of beta-2 microglobulin to 0.037 µg / ml increases. In our study, urinary beta-2 microglobulin amount were lower in women with bipolar disorder than in men but were not significant. In the case of creatinine serum, as mentioned, no significant difference was observed between the two groups, but the effect of age on creatinine serum was greater, i.e. it can be inferred that with increasing age each year, the average of 0.006 mg/dl the amount of creatinine serum increases. In each group separately, creatinine serum was higher in men than women and P value was close to 0.05. In the case of urea, the difference between the two groups was not significant, but the duration of drug usage effect was significant and with the increase of each year of

usage, the amount of urea increases by an average of 0.28 mg/dl. In this study, we also concluded that lithium use increases the secretion of beta-2 microglobulin in urine and the longer it is used, the higher the secretion of beta-2 microglobulin, but this difference was not observed in urea and serum creatinine. These results suggest that urinary beta-2 microglobulin is a better indicator for assessing primary kidney damages caused by using lithium.

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] Chang CWL, Ho CSH. Lithium use in a patient with bipolar disorder and end-stage kidney disease on hemodialysis: a case report. *Front Psychiatry* 2020;11:6. [\[CrossRef\]](#)
- [2] McKnight RF, Adida M, Budge K, Stockton S, Goodwin GM, Geddes JR. Lithium toxicity profile: a systematic review and meta-analysis. *Lancet* 2012;379:721–728. [\[CrossRef\]](#)
- [3] Davis J, Desmond M, Berk M. Lithium and nephrotoxicity: a literature review of approaches to clinical management and risk stratification. *BMC Nephrol* 2018;19:305. [\[CrossRef\]](#)
- [4] Yatham LN, Kennedy SH, Parikh SV, Schaffer A, Bond DJ, Frey BN, et al. Canadian network for mood and anxiety treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord* 2018;20:97–170. [\[CrossRef\]](#)
- [5] Gomes-da-Costa S, Marx W, Corponi F, Anmella G, Murru A, Pons-Cabrera MT, et al. Lithium therapy and weight change in people with bipolar disorder: a systematic review and meta-analysis. *Neurosci Biobehav Rev* 2022;134:104266. [\[CrossRef\]](#)
- [6] Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. *Lancet* 2013;381:1672–1682. [\[CrossRef\]](#)
- [7] Decker BS, Goldfarb DS, Dargan PI, Friesen M, Gosselin S, Hoffman RS, et al. Extracorporeal treatment for lithium poisoning: systematic review and recommendations from the EXTRIP Workgroup. *Clin J Am Soc Nephrol* 2015;10:875–887. [\[CrossRef\]](#)
- [8] Nielsen RE, Kessing LV, Nolen WA, Licht RW. Lithium and renal impairment: a review on a still hot topic. *Pharmacopsychiatry* 2018;51:200–205. [\[CrossRef\]](#)
- [9] Ferencztajn-Rochowiak E, Chłopocka-Woźniak M, Rybakowski JK. Ultra-long-term lithium therapy: all-important matters and a case of successful 50-year lithium treatment. *Braz J Psychiatry* 2020;43:407–413. [\[CrossRef\]](#)
- [10] Dembinska-Krajewska D, Kliwicki S, Chłopocka-Woźniak M, Rybakowski J. The effectiveness of prophylactic use of lithium in bipolar disorder and schizotypal traits. *Pharmacother Psychiatr Neurol* 2012;28:153–158.
- [11] Hui TP, Kandola A, Shen L, Lewis G, Osborn DP, Geddes JR, et al. A systematic review and meta-analysis of clinical predictors of lithium response in bipolar disorder. *Acta Psychiatr Scand* 2019;140:94–115. [\[CrossRef\]](#)
- [12] Nunes A, Ardaub R, Berghöfer A, Bocchetta A, Chillotti C, Deiana V, et al. Prediction of lithium response using clinical data. *Acta Psychiatr Scand* 2020;141:131–141. [\[CrossRef\]](#)
- [13] Wu HC, Lee LC, Wang WJ. Associations among serum beta 2 microglobulin, malnutrition, inflammation, and advanced cardiovascular event in patients with chronic kidney disease. *J Clin Lab Anal* 2017;31:e22056. [\[CrossRef\]](#)
- [14] Alsady M, Baumgarten R, Deen PM, de Groot T. Lithium in the kidney: friend and foe. *J Am Soc Nephrol* 2016;27:1587–1595. [\[CrossRef\]](#)
- [15] Rybakowski JK, Abramowicz M, Chłopocka-Woźniak M, Czekalski S. Novel markers of kidney injury in bipolar patients on long-term lithium treatment. *Hum Psychopharmacol* 2013;28:615–618. [\[CrossRef\]](#)
- [16] Sato T, Akatsuka H, Yamaguchi Y, Miyashita K, Tanaka M, Tamaki T, et al. Establishment of β -2 microglobulin deficient human iPS cells using CRISPR/Cas9 system. *Integr Mol Med* 2015;2:373–377. [\[CrossRef\]](#)
- [17] Chen P, Lin WL, Liu XY, Li SJ, Chen RF, Hu ZH, et al. D30 alleviates β 2-microglobulin-facilitated neurotoxic microglial responses in isoflurane/surgery-induced cognitive dysfunction in aged mice. *Lab Invest* 2025;105:102190. [\[CrossRef\]](#)

- [18] Baldemir H, Akin A, Akin O. Two modified forms of the SAIR model with a fuzzyfied vaccination effectiveness parameter. *Sigma* 2025;43:887–898. [\[CrossRef\]](#)
- [19] Li L, Dong M, Wang XG. The implication and significance of beta 2 microglobulin: a conservative multifunctional regulator. *Chin Med J* 2016;129:448–455. [\[CrossRef\]](#)
- [20] Meyer JM. Demystifying lithium therapy – A primer for clinicians: lithium-related polyuria. *Psychiatr Times* 2025;42:1.
- [21] Wu HC, Lee LC, Wang WJ. Associations among serum beta 2 microglobulin, malnutrition, inflammation, and advanced cardiovascular event in patients with chronic kidney disease. *J Clin Lab Anal* 2017;31:e22056. [\[CrossRef\]](#)
- [22] Shi F, Sun L, Kaptoge S. Association of beta-2-microglobulin and cardiovascular events and mortality: a systematic review and meta-analysis. *Atherosclerosis* 2021;320:70–78. [\[CrossRef\]](#)
- [23] Abrams Z. Diagnosing and treating bipolar spectrum disorders. *Monit Psychol* 2022;53:36–41.
- [24] Malhi GS, Gessler D, Outhred T. The use of lithium for the treatment of bipolar disorder: recommendations from clinical practice guidelines. *J Affect Disord* 2017;217:266–280. [\[CrossRef\]](#)
- [25] Nguyen T, Seiler N, Brown E, O'Donoghue B. The effect of clinical practice guidelines on prescribing practice in mental health: a systematic review. *Psychiatry Res* 2020;284:112671. [\[CrossRef\]](#)
- [26] Khawagi WY, Steinke DT, Nguyen J, Pontefract S, Keers RN. Development of prescribing safety indicators related to mental health disorders and medications: modified e-Delphi study. *Br J Clin Pharmacol* 2021;87:189–209. [\[CrossRef\]](#)
- [27] Nikolova VL, Pattanaseri K, Hidalgo-Mazzei D, Taylor D, Young AH. Is lithium monitoring NICE? Lithium monitoring in a UK secondary care setting. *J Psychopharmacol* 2018;32:408–415. [\[CrossRef\]](#)



Research Article

Mathematical analysis of fractional order Covid 19 epidemic model

Pınar AÇIKGÖZ¹, Mehmet MERDAN^{1,*}

¹Department of Mathematical Engineering, Faculty of Engineering and Natural Science, Gümüşhane University, Gümüşhane, 29100, Türkiye

ARTICLE INFO

Article history

Received: 24 May 2024

Revised: 03 July 2024

Accepted: 04 October 2024

Keywords:

Conformable Derivative;
Conformable Differential
Transformation; Covid 19
Model; Variational Iteration
Method

ABSTRACT

This article aims to examine the dynamics of the fractional epidemic model of covid 19. The model in question incorporates the notion of a compatible derivative. Given their consistent and worldwide characteristics, fractional derivatives are presently employed to address numerous practical issues. In addition, we employ two numerical techniques, namely the conformal differential transform and the variational iteration approach, to provide an approximate solution for the given model. The research closes by providing an in-depth analysis and visually representing the numerical findings. Furthermore, it has been demonstrated that the solution obtained is convergent.

Cite this article as: Açıkgöz P, Merdan M. Mathematical analysis of fractional order Covid 19 epidemic model. Sigma J Eng Nat Sci 2025;43(5):1712–1724.

INTRODUCTION

Epidemiology, also known as the “science of epidemics,” is a crucial discipline within the subject of public health. Epidemiology, commonly referred to as “epidemia,” is a field of study that examines the causative elements behind the occurrence and transmission of diseases within a specific timeframe and setting. Mathematical epidemiology has become a valuable scientific discipline for assessing the occurrence and death rates of epidemics, as well as identifying their underlying causes.

Epidemic disease models are a crucial component of contemporary mathematical modeling investigations. Several renowned organizations, such as the World Health Organization, the Centre for the Mathematical Modeling of Infectious Disease in England, and the BC Center for

Disease Control in Canada, engage in comprehensive research on disease modeling.

Mathematical models are employed across various fields such as biology, chemistry, psychology, and economics to enhance our understanding of real-world phenomena. Financial models often use stochastic equations, while ordinary differential equation systems are essential for forecasting the progression of infectious diseases. It is important to highlight the application of fractional analysis in studying the history of the event. This is because the current trend in mathematical modeling involves utilizing different types of fractional derivatives and solution methods to accurately, approximately, or numerically solve the system under investigation.

Compartment models for epidemiology are models that look at the kinetics of transmission and have been explored

*Corresponding author.

*E-mail address: mehmetmerdan@gmail.com

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



since the early 1900s [1-3]. These models have been used to generate pretty accurate estimates for many [4-6] and are still in use today for the Covid-19 [7-8]. Compartment models establish discrete divisions, known as compartments, within a society that is isolated or self-contained. Each compartment is designated by a capital letter. For example, the variable S represents the total number of individuals who are at risk of infection. The variable E represents the total number of individuals who are infected but not yet contagious. The variable I represents the total number of individuals who are infected and capable of spreading the infection. The variable R represents the total number of individuals who have recovered from the infection. The concept derives its name from the compartments, which are named after the specific disease-related states of individuals. The SIR model, sometimes referred to as the classical compartmental model, can be employed to develop new models for a wide range of diseases and situations. The SEIR model incorporates an edible compartment E (exposed) to represent the stage when a patient has been exposed to the disease but has not yet developed active symptoms in the body. Furthermore, the integration of different situations allows for the creation of epidemic illness models that encompass a range of compartments, including SIS, SEIR, SEIRS, and MSEIR.

They have formulated the model with five compartments. The mentioned model is given by:

$$\begin{aligned}\frac{dS}{dt} &= \Lambda - \phi_c IS - \mu S + v_3 R, \\ \frac{dE}{dt} &= \phi_c IS - (\sigma + \mu)E, \\ \frac{dI}{dt} &= \sigma E - (\alpha + v_2 + \delta + \mu)I + v_1 R, \\ \frac{dQ}{dt} &= \alpha I - (\theta + \mu)Q \\ \frac{dR}{dt} &= \theta Q - (v_1 + v_3 + \mu)R + v_2 I\end{aligned}\quad (1)$$

In this section, we present a mathematical system for the population at the time of COVID-19 and divide it into five parts [9]. S is susceptible, E is exposed, I is infected, Q is quarantined, and R is recovered.

It is important to mention that scholars have previously and now utilized integer derivatives to simulate a diverse array of biological and physical phenomena. Fractional mathematics has gained significant attention in the scientific world in the past two decades due to its importance. Mathematicians have diligently endeavored to formulate and analyze a diverse array of biological phenomena by employing fractional derivatives. Utilizing fractional derivatives offers numerous benefits. The Caputo and Riemann-Liouville definitions of fractional derivatives have nonlocal characteristics, such as the ability to retain information from the past and predict future behavior. It is important to

remark that the aforesaid topic is being more and more utilized in engineering and research domains such as optics, signal processing, viscoelasticity, fluid mechanics, electrochemistry, biological population models, and electromagnetics. Scientific and technical phenomena have been represented using fractional order differential operators to improve understanding of these phenomena. Moreover, as mentioned in reference [10], noninteger order systems have been utilized to accurately examine damping effects. The authors utilized fractional derivatives [11-14] to study chaotic systems. They also conducted studies on epidemic models [15-20] using fractional order derivatives. These derivatives are valuable for making predictions about past events and implementing necessary interventions in epidemics. Various authors have employed the fractional derivative to investigate the delay [21-22]. In a similar manner, they compose written content. The authors in reference [23] employed fractional derivatives to analyze systems of stochastic differential equations. These operators, known as fractional-order operators, have demonstrated consistent behavior over the last few decades, enabling the solution of complex problems in various scientific fields. Findings derived from fractional order calculations are more precise compared to those obtained from whole numbers. By consulting references [24-28], one can acquire comprehensive information regarding fractional calculations.

Non-integer differential equation structures offer the benefit of increased degrees of freedom and the inclusion of the memory effect in the model. We employ fractional conformable order differential equations [29-30] in conjunction with a mathematical model for Covid-19. CD does not contain any memory terms and is occasionally seen as a fractional order local differential operator. A number of academics have lately employed this operator in their research; for further details, refer to [31-32]. The conformable fractional operator [31-32] addresses the limitations of prior fractional operators and provides the following benefits: it enables standard calculus operations such as Rolle's theorem, the product of two functions, the derivative of the quotient of two functions, and the mean value theorem.

Furthermore, mathematical models including fractional derivatives are built upon memory systems that are present in several biological systems. The references [33-38] encompass the latest research conducted in this particular area.

The conformable derivative (CD) is a fractional local differential operator that does not include a memory term. A recent project has been finished. Based on the above logic, employing fractional differential operators to examine diverse real-world issues is more accurate than utilizing regular derivatives. Due to their reduced flexibility and non-local nature compared to integer derivatives, fractional order derivatives offer a higher degree of freedom in approximating real data than classical derivatives. Given the presence of these characteristics, we employed the fractional derivative formulation of the Covid 19 model. In this

context, the current study extends the Covid-19 model (1) from integer to fractional order. We analyze the model in the context of CD. Model (2) encapsulates our cognitive framework.

Equation (1) can be rewritten as follows:

$$\begin{aligned} T_{\zeta_0}^q S(t) &= \Lambda - \phi_c I(t) S(t) - \mu S(t) + v_3 R(t), \\ T_{\zeta_0}^q E(t) &= \phi_c I(t) S(t) - (\sigma + \mu) E(t), \\ T_{\zeta_0}^q I(t) &= \sigma E(t) - (\alpha + v_2 + \delta + \mu) I(t) + v_1 R(t), \quad (2) \\ T_{\zeta_0}^q Q(t) &= \alpha I(t) - (\theta + \mu) Q(t), \\ T_{\zeta_0}^q R(t) &= \theta Q(t) - (v_1 + v_3 + \mu) R(t) + v_2 I(t), \end{aligned}$$

In addition, the initial conditions are described as

$$S(0)=S_0, E(0)=E_0, I(0)=I_0, Q(0)=Q_0, R(0)=R_0 \text{ and } \begin{cases} S(0)=S_0 \geq 0 \\ E(0)=E_0 \geq 0 \\ I(0)=I_0 \geq 0 \\ Q(0)=Q_0 \geq 0 \\ R(0)=R_0 \geq 0 \end{cases} \quad (3)$$

Fractional differential problems exhibit a greater level of flexibility, as previously mentioned. However, obtaining an accurate analytical solution for each problem is a highly laborious endeavor. Furthermore, calculating the analytical and closed solutions to the aforementioned challenges can frequently be a challenge. Therefore, there is a continuing demand for the most effective numerical approximation techniques. In order to analyze the area stated above for simulations, scientists have developed a range of computational tools. Due to our ability to understand the resolution of these difficulties and investigate their potential applications Several numerical approaches for fractional order difficulties have already been developed and are commonly referred to as [38-40]. The qualitative theory is a powerful tool for determining the existence of a solution to the problem with fractional order. Furthermore, the presence and singularity of the solution are vital stages in the procedure. In order to obtain the desired findings, we commonly utilize fixed point theory and functional analytic methods [41-44].

In addition, we employ a numerical method outlined in reference [45] to approximately solve the given model (4) in order to replicate our findings. We will employ Euler's methodology, which is the most straightforward technique in the literature, to estimate the solution. Furthermore, the method proposed in [46-47] has been used to prove the positivity and boundedness of the solution. Recently, researchers have devised efficient techniques for solving nonlinear fractional integral equations and nonlinear Volterra-Fredholm integral equations [48-50]. Studies on random differential have been undertaken in recent years [51-53].

The study is designed as follows: Chapter 2 contains basic definitions. Chapter 3 contains the theoretical

analysis of the Covid 19 model. Chapter 4 contains the solutions obtained from the conformable fractional differential transformation method (CFDTM) and variational iteration method (VIM) and the comparison of these solutions. The fifth section contains the conclusion.

PRELIMINARIES

Definition 1. The conformable derivative of function $\aleph: [\zeta_0, \infty) \rightarrow \mathbb{R}$ with order $p \in (0, 1]$ is given [31-32,39-40]

$$T_{\zeta_0}^p \aleph(\zeta) = \lim_{h \rightarrow \zeta_0} \frac{\aleph(\zeta + h\zeta^{1-p}) - \aleph(\zeta)}{h}, \text{ for every } \zeta > \zeta_0, \text{ on condition that if } T_{\zeta_0}^p \chi(\zeta) = \lim_{\zeta \rightarrow \zeta_0} D_p^{\zeta_0} \aleph(\zeta).$$

Definition 2. The conformable fractional integral of a function function $\aleph: [\zeta_0, \infty) \rightarrow \mathbb{R}$ with order $p \in (0, 1]$ is given by

$$I_{\zeta_0}^{\zeta} \aleph(\zeta) = \int_{\zeta_0}^{\zeta} (\Phi - \zeta_0)^p \aleph(\Phi) d\Phi, \text{ for every } \zeta > \zeta_0, \text{ on condition that the integral on the right exists.}$$

Lemma 1. Under the continuity of $\aleph: [\zeta_0, \infty) \rightarrow \mathbb{R}$, we have the given results.

$$T_{\zeta_0}^p [I_{\zeta_0}^{\zeta} \aleph(\zeta)] = \aleph(\zeta). \quad p \in (0, 1], \text{ for every } \zeta > \zeta_0.$$

Lemma 2. Besides, we have the following result.

$$I_{\zeta_0}^{\zeta} [T_{\zeta_0}^p \aleph(\zeta)] = \aleph(\zeta) - \aleph(\zeta_0), p \in (0, 1].$$

THEORETICAL ANALYSIS OF THE COVID-19 MODEL

Positivity and Boundedness

We follow the procedure described in [46-47] and we can conclude that the solution of the considered model is positive:

$$\begin{cases} T_{\zeta_0}^p S(\zeta)|_{S=0} = \Lambda + v_3 R > 0 & (4a) \\ T_{\zeta_0}^p E(\zeta)|_{E=0} = \phi_c I S > 0 & (4b) \\ T_{\zeta_0}^p I(\zeta)|_{I=0} = \sigma E + v_1 R > 0 & (4c) \\ T_{\zeta_0}^p Q(\zeta)|_{Q=0} = \alpha I > 0 & (4d) \\ T_{\zeta_0}^p R(\zeta)|_{R=0} = \theta Q + v_2 I > 0. & (4e) \end{cases}$$

Thus, the Covid-19 model is positive if the initial conditions are positive.

Theorem 3.1.1. The Covid-19 model is bounded and the existence holds in the feasible region given by

$$\Omega = \left\{ (S, E, I, Q, R) \in R_+^5 : S + E + I + Q + R \leq \frac{\Lambda}{\mu} \right\}. \quad (5)$$

Proof: Let N be total population at time say ζ . Thus, we have

$$N(\zeta) = S(\zeta) + E(\zeta) + I(\zeta) + Q(\zeta) + R(\zeta). \quad (6)$$

Taking CFOD $p > 0$ of (6), one has

$$\begin{aligned} T_{\zeta_0}^p[N(\zeta)] &= T_{\zeta_0}^p[S(\zeta)] + T_{\zeta_0}^p[E(\zeta)] + T_{\zeta_0}^p[I(\zeta)] \\ &\quad + T_{\zeta_0}^p[Q(\zeta)] + T_{\zeta_0}^p[R(\zeta)] \\ &= \Lambda - \mu N(\zeta) - \delta I(\zeta) \leq \Lambda - \mu N(\zeta). \end{aligned} \quad (7)$$

Since $0 < I(\zeta) \leq N(\zeta)$ the term $\delta I(\zeta)$ is neglected. As a result, the solution is limited, and we can express the appropriate region as shown in (5). Taking conformable Laplace transform of both sides, and using $\zeta \rightarrow \infty$, we have $N(\zeta) \leq \frac{\Lambda}{\mu}$.

Equilibrium Points

The disease epidemic and free equilibrium points have been computed in [9] as $E_0 = (\frac{\Lambda}{\mu}, 0, 0, 0, 0)$ and $E^* = (S^*, E^*, I^*, Q^*, R^*)$ where

$$\begin{aligned} S^* &= \frac{a_2(a_1a_3a_4 - a\theta v_1 - a_3v_1v_2)}{\phi_c\sigma a_3a_4} \\ E^* &= \frac{(\Lambda\phi_c\sigma a_3a_4 + a\mu\theta a_2v_1 - \mu a_1a_2a_3a_4 + \mu a_2a_3v_1v_2)(a\theta v_1 - a_1a_3a_4 + a_3v_1v_2)}{(a\sigma\theta v_3 + a\theta a_2v_1 + \sigma a_3v_2v_3 - a_1a_2a_3a_4 + a_2a_3v_1v_2)\phi_c\sigma a_3a_4} \\ I^* &= \frac{\mu a_1a_2a_3a_4 - \Lambda\phi_c\sigma a_3a_4 - a\mu\theta a_2v_1 - \mu a_2a_3v_1v_2}{\phi_c(a\sigma\theta v_3 + a\theta a_2v_1 + \sigma a_3v_2v_3 - a_1a_2a_3a_4 + a_2a_3v_1v_2)} \\ Q^* &= \frac{a(\mu a_1a_2a_3a_4 - \Lambda\phi_c\sigma a_3a_4 - a\mu\theta a_2v_1 - \mu a_2a_3v_1v_2)}{\phi_c(a\sigma\theta v_3 + a\theta a_2v_1 + \sigma a_3v_2v_3 - a_1a_2a_3a_4 + a_2a_3v_1v_2)a_3} \\ R^* &= \frac{(\mu a_1a_2a_3a_4 - \Lambda\phi_c\sigma a_3a_4 - a\mu\theta a_2v_1 - \mu a_2a_3v_1v_2)}{(a\sigma\theta v_3 + a\theta a_2v_1 + \sigma a_3v_2v_3 - a_1a_2a_3a_4 + a_2a_3v_1v_2)} \times \frac{(a\theta + a_3v_2)}{\phi_c a_3a_4} \end{aligned}$$

We get the equations. Here, $a_1 = (\alpha + v_2 + \delta + \mu)$, $a_2 = (\sigma + \mu)$, $a_3 = (\theta + \mu)$ and $a_4 = (v_1 + v_3 + \mu)$

The Basic Reproduction Number ($\mathfrak{R}_0$)

$\mathfrak{R}_0$ denotes the mean number of secondary cases that arise from a patient, who is infected by that patient, in a community where everyone is vulnerable to the infectious disease. The term “ $\mathfrak{R}_0$ ” denotes the fundamental reproduction number of the Covid-19 model. It is employed in [9] for the purpose of determining the next generation matrix

technique. The following equation provides the numerical value.

$$\mathfrak{R}_0 = \frac{\phi_c \Lambda \sigma}{\mu(\sigma + \mu)(\alpha + v_2 + \delta + \mu)}$$

Based on the value of $\mathfrak{R}_0$, there are three potential outcomes for the transmission or decline of a disease, which are as follows:

If the value of $\mathfrak{R}_0$ is less than or equal to 1, each existing infection results in less than one new infection. In this scenario, the sickness undergoes a gradual decrease and ultimately comes to an end. If the basic reproduction number $\mathfrak{R}_0$ is equal to 1, each existing function results in a new infection. The sickness remains extant, although it does not propagate. If the basic reproduction number $\mathfrak{R}_0$ is greater than 1, each existing infection leads to the occurrence of numerous new infections. The pathogen is transmitted to humans, leading to an outbreak of the disease. Figure 1 displays the three-dimensional profile of $\mathfrak{R}_0$. We utilize the quantitative values presented in Table 2.

Analyzed was the impact of parameter modifications on the Basic Reproduction Number, utilizing the forward normalized sensitivity index. Contour plotting is a valuable and efficient graphical method frequently employed to present results of the Basic Reproduction Number. These

Table 1. Description of variables and parameters of the model (1)

Variable	Description
$S(t)$	At time t, the number of susceptible humans was
$E(t)$	Number of people who were exposed at time t
$I(t)$	At time t, the number of infectious humans was
$Q(t)$	Humans in quarantine at time t
$R(t)$	At time t, the number of recovered humans was

Table 2. Parameter values used in the model [18-19] [obtained from authors solution]

Parameter	Definition	Value
Λ	Recovery rate of individuals	750
ϕ_c	Contact rate for Covid-19 transmission	0.0000124
σ	Exposure rate of individuals	0.000011618
μ	Rate of natural deaths	0.00324588
v_2	Relapse rate of individuals	0.001466848
δ	Death rate from Covid-19	0.00286
θ	Survival rate of individuals in quarantine	0.0766169
α	Detection rate of infectious individuals	0.010939586
v_1	Survival rate of infectious people	0.1109289
v_3	Rate of returning individuals to the susceptible class	0.0022927

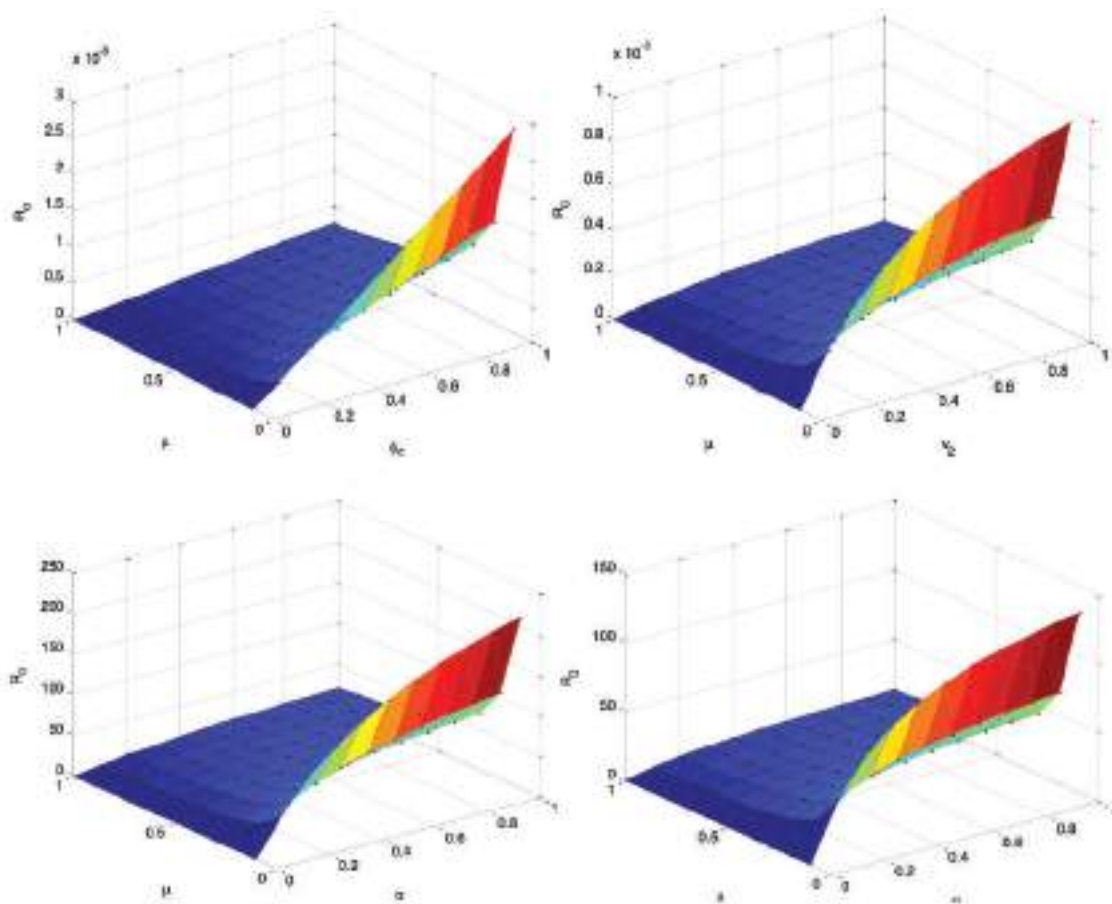


Figure 1. Three-dimensional representation of the $\mathcal{R}_0$ value in the Covid-19 model

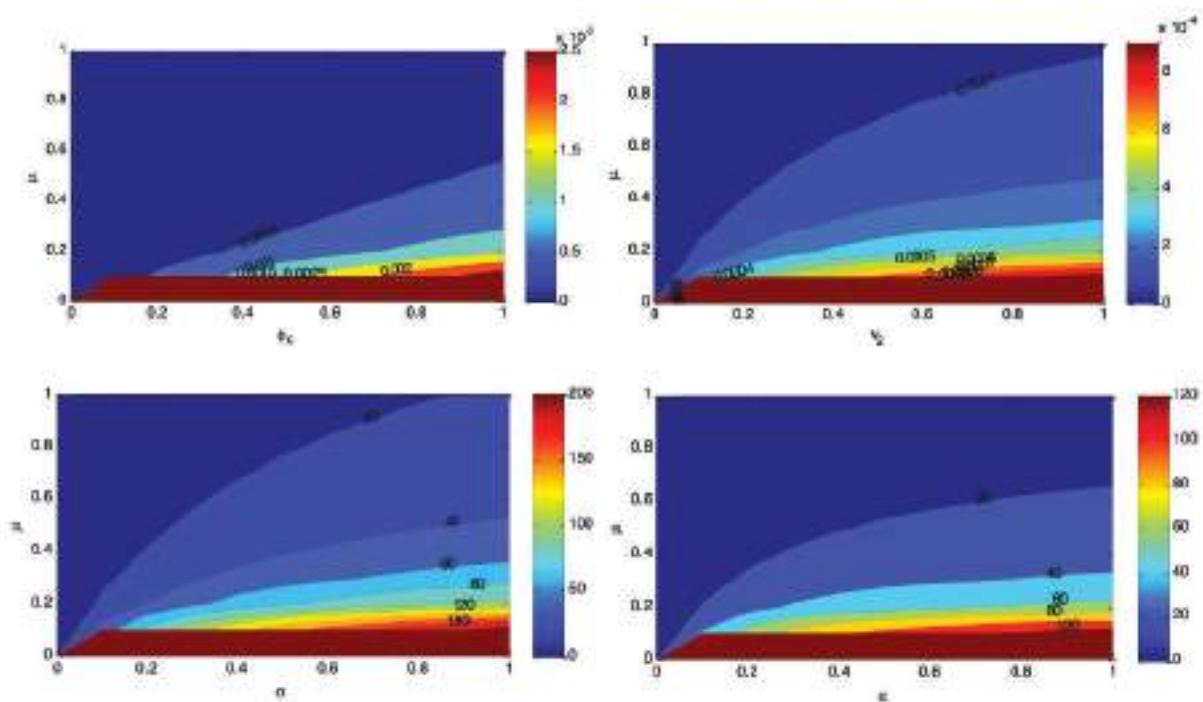


Figure 2. Contour plot of representation for $\mathcal{R}_0$ of the Covid-19 model.

plots are especially valuable for identifying and analyzing the spread of shocks and abrupt changes.

NUMERICAL SCHEME FOR THE COVID-19 MODEL

The Conformable Fractional Differential Transformation Method

Assume that $\mathcal{F}$ is an infinitely differentiable function in $[20, 22]$ for some q with $0 < q \leq 1$. The fractional power series expansion of $\mathcal{F}(w)$ around the point $w = 0$ is as follows:

$$\mathcal{F}(w) = \sum_{k=0}^{\infty} \frac{w^{qk}}{q^{k!}} \left[(T_q \mathcal{F})^{(k)} \right]_{w=0}, \quad 0 < w < \mathcal{R}^{1/q}, \mathcal{R} > 0. \quad (8)$$

where, $\left[(T_q \mathcal{F})^{(k)} \right]_{w=0}$ denotes the application of the fractional derivative for k times. Conformable fractional differential transform of $\mathcal{F}(w)$ is defined as

$$\mathcal{H}_q(k) = \frac{1}{q^{k!}} \left[(T_q \mathcal{F})^{(k)} \right]_{w=0} \quad (9)$$

The definition of the inverse conformable fractional differential transform of $\mathcal{H}(k)$ is as follows:

$$\mathcal{F}(w) = \sum_{k=0}^{\infty} w^{qk} \mathcal{H}_q(k) = \sum_{k=0}^{\infty} \frac{w^{qk}}{q^{k!}} \left[(T_q \mathcal{F})^{(k)} \right]_{w=0} \quad (10)$$

The conformable differential transform approach involves performing fundamental mathematical operations.

(i) If $\mathcal{F}(w) = (w - w_0)^p$, then $\mathcal{H}_q(k) = \delta(k - (p/\alpha))$, where

$$\delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \quad (11)$$

(ii) If $\mathcal{F}(w) = T_q u(w)$, then $\mathcal{H}_q(k) = q(k+1)U_q(k+1)$.

Application of conformable fractional differential transform Method for Covid-19 model

$$\begin{aligned} q(k+1)S_q(k+1) &= \Lambda S(k) - \phi_c \sum_{r=0}^k I_q(r)S_q(k-r) - \mu S_q(k) + v_3 R_q(k) \\ q(k+1)E_q(k+1) &= \phi_c \sum_{r=0}^k I_q(r)S_q(k-r) - (\sigma + \mu)E_q(k) \\ q(k+1)I_q(k+1) &= \sigma E_q(k) - (\alpha + v_2 + \delta + \mu)I_q(k) + v_1 R_q(k) \\ q(k+1)Q_q(k+1) &= \alpha I_q(k) - (\theta + \mu)Q_q(k) \\ q(k+1)R_q(k+1) &= \theta Q_q(k) - (v_1 + v_3 + \mu)R_q(k) + v_2 I_q(k) \end{aligned} \quad (12)$$

Hence, recurrence relation is obtained as

$$\begin{aligned} S_q(k+1) &= \frac{1}{q(k+1)} \left[\Lambda S(k) - \phi_c \sum_{r=0}^k I_q(r)S_q(k-r) - \mu S_q(k) + v_3 R_q(k) \right] \\ E_q(k+1) &= \frac{1}{q(k+1)} \left[\phi_c \sum_{r=0}^k I_q(r)S_q(k-r) - (\sigma + \mu)E_q(k) \right] \\ I_q(k+1) &= \frac{1}{q(k+1)} \left[\sigma E_q(k) - (\alpha + v_2 + \delta + \mu)I_q(k) + v_1 R_q(k) \right] \\ Q_q(k+1) &= \frac{1}{q(k+1)} \left[\alpha I_q(k) - (\theta + \mu)Q_q(k) \right] \\ R_q(k+1) &= \frac{1}{q(k+1)} \left[\theta Q_q(k) - (v_1 + v_3 + \mu)R_q(k) + v_2 I_q(k) \right] \end{aligned} \quad (13)$$

With initial conditions, $S(0) = 5000$, $E(0) = 2003$, $I(0) = 416$, $Q(0) = 404$, $R(0) = 115$ and parameter values are given in table 2.

Variational Iteration Method

To illustrate the fundamental concept of the variational iteration method, let us examine the given nonlinear differential equation in operator form:

$$H(u(t)) + N(u(t)) = h(t). \quad (14)$$

H represents the linear operator, N represents the nonlinear operator, and h represents any real function referred to as the inhomogeneous term. The correction function corresponding to equation (14) is presented as follows:

$$u_{n+1}(t) = u_n(t) + \int_0^t \lambda \{ H u_n(s) + N \tilde{u}_n(s) - h(s) \} ds, \quad (15)$$

Here, λ is the general Lagrange multiplier [54-57], which is best described by variational theory [44], and $N \tilde{u}_n$ is constrained variation, i.e. $\delta N \tilde{u}_n = 0$.

Theorem 1. Consider a Banach space B equipped with a norm $\|\cdot\|$. Let the series $\sum_{i=0}^{\infty} u_i$ be defined in B . Assume that the initial guess $y_0 = u_0$ is contained within the solution ball $\gamma(x)$. The series solution $\sum_{i=0}^{\infty} u_i$ converges if there is a value of r for which the inequality [54] $\|u_{n+1}\| \leq r \|u_n\|$ holds.

Theorem 2. Examine the second conformable differential equation. The variational iteration formula is expressed as follows:

$$\begin{aligned} S_{n+1}(t) &= S_n(t) - I_q \{ T_q S_n(t) - \Lambda + \phi_c I_n(t) S_n(t) + \mu S_n(t) - v_3 R_n(t) \}, \\ E_{n+1}(t) &= E_n(t) - I_q \{ T_q E_n(t) - \phi_c I_n(t) S_n(t) + (\sigma + \mu) E_n(t) \}, \\ I_{n+1}(t) &= I_n(t) - I_q \{ T_q I_n(t) - \sigma E_n(t) + (\alpha + v_2 + \delta + \mu) I_n(t) - v_1 R_n(t) \}, \\ Q_{n+1}(t) &= Q_n(t) - I_q \{ T_q Q_n(t) - \alpha I_n(t) + (\theta + \mu) Q_n(t) \}, \\ R_{n+1}(t) &= R_n(t) - I_q \{ T_q R_n(t) - \theta Q_n(t) + (v_1 + v_3 + \mu) R_n(t) - v_2 I_n(t) \}, \end{aligned} \quad (16)$$

where S_n , E_n , I_n , Q_n , R_n are n th approximation, T_q is the conformable derivative of order q , and I_q is the fractional integral of order $q \in (0, 1]$

Proof. Equation (2) may be rewritten equivalently as

$$\begin{aligned}
T_q S(t) - \Lambda + \phi_c I(t) S(t) + \mu S(t) - v_3 R(t) &= 0, \\
T_q E(t) - \phi_c I(t) S(t) + (\sigma + \mu) E(t) &= 0, \\
T_q I(t) - \sigma E(t) + (\alpha + v_2 + \delta + \mu) I(t) - v_1 R(t) &= 0, \\
T_q Q(t) - \alpha I(t) + (\theta + \mu) Q(t) &= 0, \\
T_q R(t) - \theta Q(t) + (v_1 + v_3 + \mu) R(t) - v_2 I(t) &= 0,
\end{aligned} \quad (17)$$

Multiplying equation (17) by a general Lagrange multiplier $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_3(t)$, $\lambda_4(t)$ and $\lambda_5(t)$ yields

$$\begin{aligned}
\lambda_1(t) \{T_q S(t) - \Lambda + \phi_c I(t) S(t) + \mu S(t) - v_3 R(t)\} &= 0, \\
\lambda_2(t) \{T_q E(t) - \phi_c I(t) S(t) + (\sigma + \mu) E(t)\} &= 0, \\
\lambda_3(t) \{T_q I(t) - \sigma E(t) + (\alpha + v_2 + \delta + \mu) I(t) - v_1 R(t)\} &= 0, \\
\lambda_4(t) \{T_q Q(t) - \alpha I(t) + (\theta + \mu) Q(t)\} &= 0, \\
\lambda_5(t) \{T_q R(t) - \theta Q(t) + (v_1 + v_3 + \mu) R(t) - v_2 I(t)\} &= 0,
\end{aligned} \quad (18)$$

Now, upon applying I_q to the both sides of equation (18) will give

$$\begin{aligned}
I_q [\lambda_1(t) \{T_q S(t) - \Lambda + \phi_c I(t) S(t) + \mu S(t) - v_3 R(t)\}] &= 0, \\
I_q [\lambda_2(t) \{T_q E(t) - \phi_c I(t) S(t) + (\sigma + \mu) E(t)\}] &= 0, \\
I_q [\lambda_3(t) \{T_q I(t) - \sigma E(t) + (\alpha + v_2 + \delta + \mu) I(t) - v_1 R(t)\}] &= 0, \\
I_q [\lambda_4(t) \{T_q Q(t) - \alpha I(t) + (\theta + \mu) Q(t)\}] &= 0, \\
I_q [\lambda_5(t) \{T_q R(t) - \theta Q(t) + (v_1 + v_3 + \mu) R(t) - v_2 I(t)\}] &= 0,
\end{aligned} \quad (19)$$

Then, the correction functional of equation (2) will be read as follows:

$$\begin{aligned}
S_{n+1}(t) &= S_n(t) + I_q [\lambda_1(t) \{T_q S_n(t) - \Lambda + \phi_c \bar{S}_n(t) \bar{I}_n(t) + \mu \bar{S}_n(t) - v_3 \bar{R}_n(t)\}], \\
E_{n+1}(t) &= E_n(t) + I_q [\lambda_2(t) \{T_q E_n(t) - \phi_c \bar{I}_n(t) \bar{S}_n(t) + (\sigma + \mu) \bar{E}_n(t)\}], \\
I_{n+1}(t) &= I_n(t) + I_q [\lambda_3(t) \{T_q I_n(t) - \sigma \bar{E}_n(t) + (\alpha + v_2 + \delta + \mu) \bar{I}_n(t) - v_1 \bar{R}_n(t)\}], \\
Q_{n+1}(t) &= Q_n(t) + I_q [\lambda_4(t) \{T_q Q_n(t) - \alpha \bar{I}_n(t) + (\theta + \mu) \bar{Q}_n(t)\}], \\
R_{n+1}(t) &= R_n(t) + I_q [\lambda_5(t) \{T_q R_n(t) - \theta \bar{Q}_n(t) + (v_1 + v_3 + \mu) \bar{R}_n(t) - v_2 \bar{I}_n(t)\}],
\end{aligned} \quad (20)$$

In this case, calculating the value of $\lambda_i(t)$, for $i = \{1, 2, 3, 4, 5\}$ from equation (20), which is a fractional integral functional, is difficult. Equation (20) can be written as:

$$\begin{aligned}
S_{n+1}(t) &= S_n(t) + \int_0^t \tau^{q-1} \left[\lambda_1(\tau) \left\{ \tau^{1-q} \frac{d}{d\tau} S_n(\tau) - \Lambda + \phi_c \bar{S}_n(\tau) \bar{I}_n(\tau) + \mu \bar{S}_n(\tau) - v_3 \bar{R}_n(\tau) \right\} \right] d\tau \\
E_{n+1}(t) &= E_n(t) + \int_0^t \tau^{q-1} \left[\lambda_2(\tau) \left\{ \tau^{1-q} \frac{d}{d\tau} E_n(\tau) - \phi_c \bar{I}_n(\tau) \bar{S}_n(\tau) + (\sigma + \mu) \bar{E}_n(\tau) \right\} \right] d\tau \\
I_{n+1}(t) &= I_n(t) + \int_0^t \tau^{q-1} \left[\lambda_3(\tau) \left\{ \tau^{1-q} \frac{d}{d\tau} I_n(\tau) - \sigma \bar{E}_n(\tau) + (\alpha + v_2 + \delta + \mu) \bar{I}_n(\tau) - v_1 \bar{R}_n(\tau) \right\} \right] d\tau \\
Q_{n+1}(t) &= Q_n(t) + \int_0^t \tau^{q-1} \left[\lambda_4(\tau) \left\{ \tau^{1-q} \frac{d}{d\tau} Q_n(\tau) - \alpha \bar{I}_n(\tau) + (\theta + \mu) \bar{Q}_n(\tau) \right\} \right] d\tau \\
R_{n+1}(t) &= R_n(t) + \int_0^t \tau^{q-1} \left[\lambda_5(\tau) \left\{ \tau^{1-q} \frac{d}{d\tau} R_n(\tau) - \theta \bar{Q}_n(\tau) + (v_1 + v_3 + \mu) \bar{R}_n(\tau) - v_2 \bar{I}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (21)$$

Where $\bar{S}_n$, $\bar{E}_n$, $\bar{I}_n$, $\bar{Q}_n$, and $\bar{R}_n$ are the restricted variations with $\delta \bar{S}_n = 0$, $\delta \bar{E}_n = 0$, $\delta \bar{I}_n = 0$, $\delta \bar{Q}_n = 0$ and $\delta \bar{R}_n = 0$. From equation (21), we obtain

$$\begin{aligned}
\delta S_{n+1}(t) &= \delta S_n(t) + \delta \int_0^t \left[\lambda_1(\tau) \left\{ \frac{d}{d\tau} S_n(\tau) - \Lambda \tau^{q-1} \right. \right. \\
&\quad \left. \left. + \phi_c \tau^{q-1} \bar{S}_n(\tau) \bar{I}_n(\tau) + \mu \tau^{q-1} \bar{S}_n(\tau) - v_3 \tau^{q-1} \bar{R}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (22)$$

$$\begin{aligned}
\delta E_{n+1}(t) &= \delta E_n(t) + \delta \int_0^t \left[\lambda_2(\tau) \left\{ \frac{d}{d\tau} E_n(\tau) - \phi_c \tau^{q-1} \bar{I}_n(\tau) \bar{S}_n(\tau) \right. \right. \\
&\quad \left. \left. + (\sigma + \mu) \tau^{q-1} \bar{E}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (23)$$

$$\begin{aligned}
\delta I_{n+1}(t) &= \delta I_n(t) + \delta \int_0^t \left[\lambda_3(\tau) \left\{ \frac{d}{d\tau} I_n(\tau) - \sigma \tau^{q-1} \bar{E}_n(\tau) \right. \right. \\
&\quad \left. \left. + (\alpha + v_2 + \delta + \mu) \tau^{q-1} \bar{I}_n(\tau) - v_1 \tau^{q-1} \bar{R}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (24)$$

$$\begin{aligned}
\delta Q_{n+1}(t) &= \delta Q_n(t) + \delta \int_0^t \left[\lambda_4(\tau) \left\{ \frac{d}{d\tau} Q_n(\tau) - \alpha \tau^{q-1} \bar{I}_n(\tau) \right. \right. \\
&\quad \left. \left. + (\theta + \mu) \tau^{q-1} \bar{Q}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (25)$$

$$\begin{aligned}
\delta R_{n+1}(t) &= \delta R_n(t) + \delta \int_0^t \left[\lambda_5(\tau) \left\{ \frac{d}{d\tau} R_n(\tau) - \theta \tau^{q-1} \bar{Q}_n(\tau) \right. \right. \\
&\quad \left. \left. + (v_1 + v_3 + \mu) \tau^{q-1} \bar{R}_n(\tau) - v_2 \tau^{q-1} \bar{I}_n(\tau) \right\} \right] d\tau
\end{aligned} \quad (26)$$

Using integration by parts, equation (22) becomes for equations (23), (24), (25) and (26) are similar to proof of equation (22)

$$\begin{aligned}
\delta S_{n+1}(t) &= \delta S_n(t) + \delta \left(\lambda_1(t) S_n(t) - \int_0^t \lambda_1'(\tau) S_n(\tau) d\tau \right) \\
&= \left(1 + \lambda_1(t) \delta S_n(t) - \delta \int_0^t \lambda_1'(\tau) S_n(\tau) d\tau \right) \\
\delta E_{n+1}(t) &= \left(1 + \lambda_2(t) \delta E_n(t) - \delta \int_0^t \lambda_2'(\tau) E_n(\tau) d\tau \right) \\
\delta I_{n+1}(t) &= \left(1 + \lambda_3(t) \delta I_n(t) - \delta \int_0^t \lambda_3'(\tau) I_n(\tau) d\tau \right) \\
\delta Q_{n+1}(t) &= \left(1 + \lambda_4(t) \delta Q_n(t) - \delta \int_0^t \lambda_4'(\tau) Q_n(\tau) d\tau \right) \\
\delta R_{n+1}(t) &= \left(1 + \lambda_5(t) \delta R_n(t) - \delta \int_0^t \lambda_5'(\tau) R_n(\tau) d\tau \right)
\end{aligned} \quad (27)$$

The lagrange multipliers $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_3(t)$, $\lambda_4(t)$ and $\lambda_5(t)$ can be obtained by $\lambda_i'(t) = 0$ for all $i = \{1, 2, 3, 4, 5\}$ with boundary condition: $1 + \lambda_i(t) = 0$ for all $i = \{1, 2, 3, 4, 5\}$. Solving the last initial value problem for λ_i for all $i = \{1, 2, 3, 4, 5\}$, the general Lagrange multiplier λ_i is found to be

$$\lambda_i = -1, \text{ for all } i = \{1, 2, 3, 4, 5\} \quad (28)$$

Thus, substituting λ_i into the corresponding smoothing function (20) yields the iteration formula:

$$\begin{aligned}
S_{n+1}(t) &= S_n(t) - I_q \{ T_q S_n(t) - \Lambda + \phi_c \tilde{S}_n(t) \tilde{I}_n(t) + \mu \tilde{S}_n(t) - v_3 \tilde{R}_n(t) \} \\
E_{n+1}(t) &= E_n(t) - I_q \{ T_q E_n(t) - \phi_c \tilde{I}_n(t) \tilde{S}_n(t) + (\sigma + \mu) \tilde{E}_n(t) \} \\
I_{n+1}(t) &= I_n(t) - I_q \{ T_q I_n(t) - \sigma \tilde{E}_n(t) + (\alpha + v_2 + \delta + \mu) \tilde{I}_n(t) - v_1 \tilde{R}_n(t) \} \quad (29) \\
Q_{n+1}(t) &= Q_n(t) - I_q \{ T_q Q_n(t) - \alpha \tilde{I}_n(t) + (\theta + \mu) \tilde{Q}_n(t) \} \\
R_{n+1}(t) &= R_n(t) - I_q \{ T_q R_n(t) - \theta \tilde{Q}_n(t) + (v_1 + v_3 + \mu) \tilde{R}_n(t) - v_2 \tilde{I}_n(t) \}
\end{aligned}$$

Application of the variational iteration method

Applying the variational iteration method using Theorem 1,

$$\begin{aligned}
S_{n+1}(t) &= S_n(t) - \int_0^t \tau^{q-1} \left\{ \tau^{1-q} \frac{d}{d\tau} S_n(\tau) - \Lambda + \phi_c I_n(\tau) S_n(\tau) + \mu S_n(\tau) - v_3 R_n(\tau) \right\} d\tau \\
E_{n+1}(t) &= E_n(t) - \int_0^t \tau^{q-1} \left\{ \tau^{1-q} \frac{d}{d\tau} E_n(\tau) - \phi_c I_n(\tau) S_n(\tau) + (\sigma + \mu) E_n(\tau) \right\} d\tau \\
I_{n+1}(t) &= I_n(t) - \int_0^t \tau^{q-1} \left\{ \tau^{1-q} \frac{d}{d\tau} I_n(\tau) - \sigma E_n(\tau) + (\alpha + v_2 + \delta + \mu) I_n(\tau) - v_1 R_n(\tau) \right\} d\tau \quad (30) \\
Q_{n+1}(t) &= Q_n(t) - \int_0^t \tau^{q-1} \left\{ \tau^{1-q} \frac{d}{d\tau} Q_n(\tau) - \alpha I_n(\tau) + (\theta + \mu) Q_n(\tau) \right\} d\tau \\
R_{n+1}(t) &= R_n(t) - \int_0^t \tau^{q-1} \left\{ \tau^{1-q} \frac{d}{d\tau} R_n(\tau) - \theta Q_n(\tau) + (v_1 + v_3 + \mu) R_n(\tau) - v_2 I_n(\tau) \right\} d\tau
\end{aligned}$$

With initial conditions, $S(0) = 5000$, $E(0) = 2003$, $I(0) = 416$, $Q(0) = 404$, $R(0) = 115$ and parameter values are given in table 2, we obtain

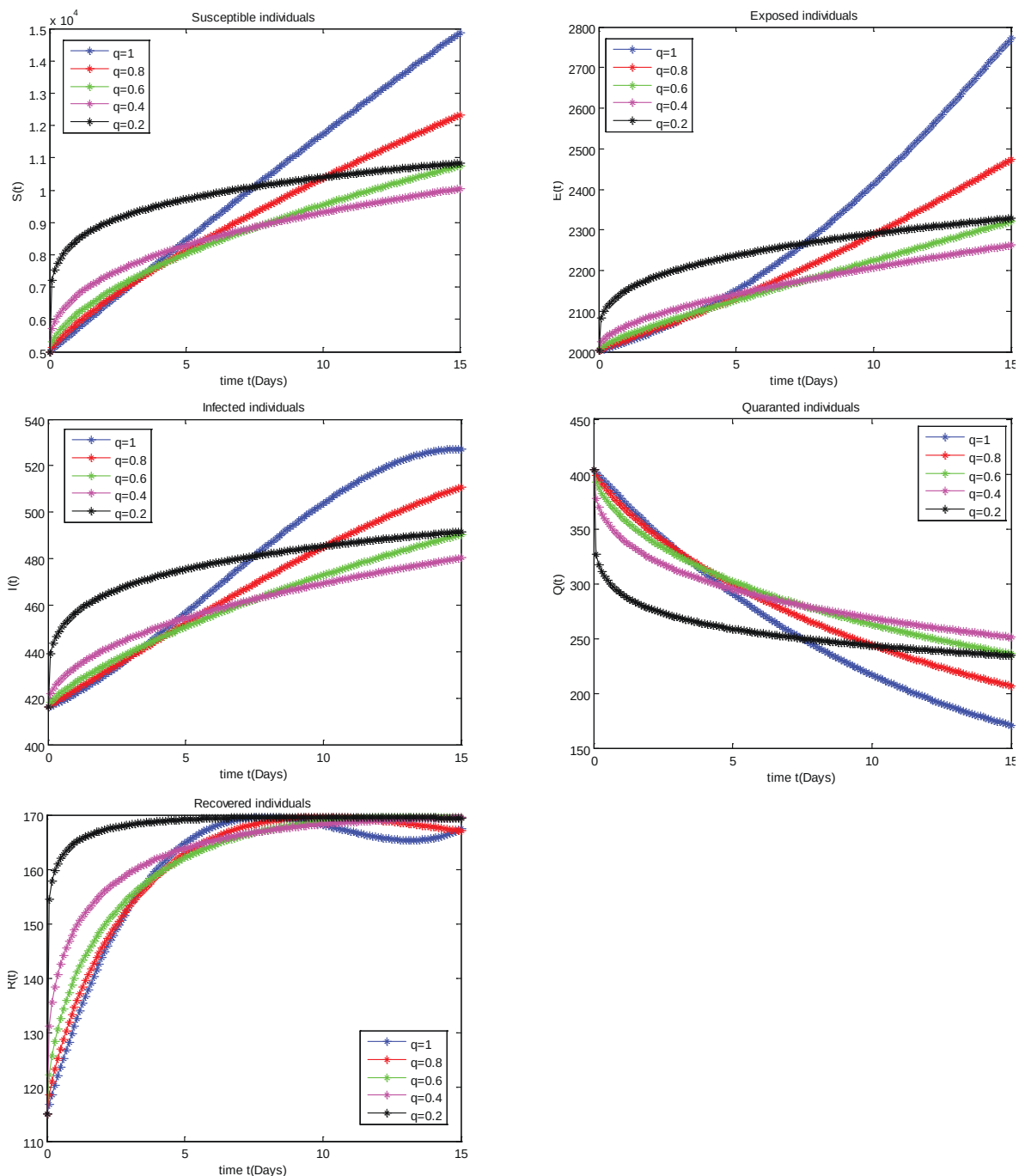


Figure 2. The CDTM solutions of for system (13) at different q values.

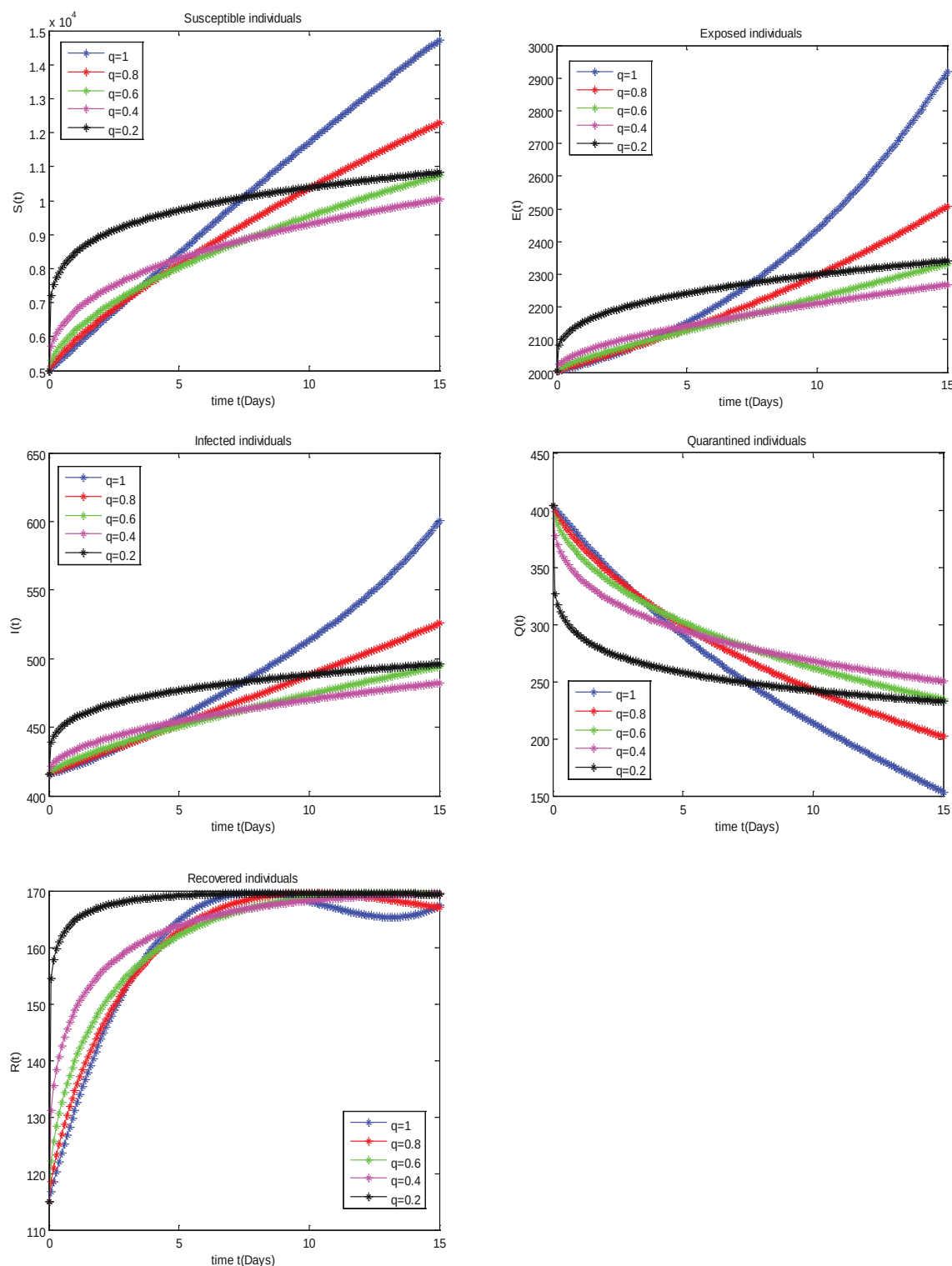


Figure 3. The VIM solutions of for system (30) at different q values.

Figure 2 illustrates that the number of susceptible, exposed, and infected individuals grows as the fractional order derivative drops during the first seven days. However, after this period, the number of individuals in these

categories decreases. The number of individuals who have recovered increases over time, whereas the number of individuals under quarantine decreases as the fractional order derivative decreases.

Figure 3 displays the VIM solutions for system (30) at various q values.

The result of equation (4)'s epidemic system is completely in line with the result of the conformable fractional differential transformation approach. Figure 2-3. Figure 3 demonstrates that as the fractional order derivative declines, there is an increase in the number of susceptible, exposed, and infected individuals during the initial 7 days, followed by a reduction after 7 days. As the value of the fractional order derivative declines, there is an observed increase in the number of persons who have recovered over time, while the number of individuals in quarantine decreases. It is evident that the solution acquired from the CFDTM in Figure 2 and the solutions produced from the VIM method in Figure 3 are compatible for various values of the fractional order q [58].

Below, the approximate analytical $S(t)$ solution obtained from the VIM method and the convergence analysis for this solution are given.

$$\begin{aligned} S(t) = & 5000. + 708.2422605t - 0.02762284236t^3 \\ & - 3.112751154t^2 - .4759874044e - 6t^6 \\ & + .1590043949e - 3t^5 - .6340772020e - 3t^4 \\ & - .1437275197e - 20t^{11} + .3080875470e - 14t^{10} \\ & + .8428734232e - 12t^9 - .2762956380e - 9t^8 \\ & - .7677240689e - 9t^7 \end{aligned}$$

If we examine the convergence by choosing $t = 1, q = 1$.

$$\begin{aligned} r_1: \frac{\|S_2\|}{\|S_1\|} &= \frac{\|-3.112751154t^2\|}{\|708.2422605t\|} = 0.0044 < 1 \\ r_2: \frac{\|S_3\|}{\|S_2\|} &= \frac{\|-0.02762284236t^3\|}{\|-3.112751154t^2\|} = 0.0089 < 1 \\ r_3: \frac{\|S_4\|}{\|S_3\|} &= \frac{\|-.6340772020e - 3t^4\|}{\|-0.02762284236t^3\|} = 0.0230 < 1 \\ r_4: \frac{\|S_5\|}{\|S_4\|} &= \frac{\|+.1590043949e - 3t^5\|}{\|-.6340772020e - 3t^4\|} = 0.2508 < 1 \end{aligned}$$

⋮

It converges because its values are less than 1 [54-55].

Table 3. Table for numerical solutions of the Covid-19 model for $q = 0.9$ with CFDTM

t	$S(t)$	$E(t)$	$I(t)$	$Q(t)$	$R(t)$
0.0	5000.	2003.	416.	404.	115.
0.1	5096.186981	2005.678857	416.7191853	400.2668463	117.4063418
0.2	5181.260454	2008.163002	417.4071349	397.0205933	119.4249899
0.3	5262.087376	2010.629044	418.1047249	393.9860365	121.2472211
0.4	5340.155114	2013.111987	418.8165406	391.1008331	122.9215621
0.5	5416.169854	2015.627333	419.5424076	388.3340727	124.4747571
0.6	5490.546773	2018.183646	420.2805261	385.6667078	125.9251287
0.7	5563.559537	2020.786264	421.0283768	383.0855310	127.2867593
0.8	5635.402119	2023.438816	421.7830748	380.5806647	128.5712595
0.9	5706.219127	2026.143916	422.5415349	378.1443197	129.7886583
1.0	5776.122446	2028.903568	423.3005544	375.7701077	130.9479015

Table 4. Table for numerical solutions of the Covid-19 model for $q = 0.9$ with VIM

t	$S(t)$	$E(t)$	$I(t)$	$Q(t)$	$R(t)$
0.0	5000.	2003.	416.	404.	115.
0.1	5099.008371	2005.733396	416.7290149	400.1455378	117.5004699
0.2	5184.657369	2008.162796	417.3897214	396.8428549	119.6005287
0.3	5265.844571	2010.521633	418.0410085	393.7391713	121.5381523
0.4	5344.237772	2012.851421	418.6923496	390.7670265	123.3608975
0.5	5420.598995	2015.170437	419.3474747	387.8951332	125.0915734
0.6	5495.365463	2017.488820	420.0081702	385.1052755	126.7438936
0.7	5568.820372	2019.812931	420.6753546	382.3854843	128.3271556
0.8	5641.161773	2022.147076	421.3494972	379.7272671	129.8481617
0.9	5712.535743	2024.494343	422.0308120	377.1242679	131.3121552
1.0	5783.054353	2026.857034	422.7193594	374.5715396	132.7233368

Comparison

In order to compare accuracy of the obtained results by CFDTM and VIM, Table 3,4 are presented.

Once again, it has been noted that the solution derived from the Conformable Fractional Differential Transformation Method is consistent with the solution acquired from the variational iteration method.

CONCLUSION

In this study, we employed two numerical approaches to provide approximate solutions for the Covid 19 epidemic model. These techniques rely on conformable derivatives, which have recently become prominent. By employing the q -derivative, we initially redefined the concerted differential transformation method and variational iteration approach. Through the application of the suggested techniques to the Covid 19 epidemic model, we have showcased their effectiveness and precision. Our variational iteration method successfully obtained the approximate solution matching to the concerted differential transformation method, which is in total agreement with it. Moreover, our methods can be readily applied to a diverse array of epidemic models, such as fractional order avian influenza, swine flu, SIR, SEIR/SEIRS, and SIS, due to their simplicity in mathematical application to many physical problems.

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.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

ETHICS

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

REFERENCES

- [1] Abbey H. An examination of the Reed Frost theory of epidemics. *Hum Biol* 1952;24:201.
- [2] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A Math Phys Sci* 1927;115:700–721. [\[CrossRef\]](#)
- [3] Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;42:599–653. [\[CrossRef\]](#)
- [4] Anderson RM. The population dynamics of infectious diseases: theory and applications. New York: Springer; 2013.
- [5] Brauer F, Castillo-Chavez C. Mathematical models in population biology and epidemiology. 2nd ed. New York: Springer; 2012. p. 508. [\[CrossRef\]](#)
- [6] Murray JD. Mathematical biology: I. An introduction. 3rd ed. New York: Springer; 2007. Vol. 17.
- [7] Oliveira G. Refined compartmental models, asymptomatic carriers and COVID-19. *arXiv Preprint*. 2020. doi:10.1101/2020.04.14.20065128 [\[CrossRef\]](#)
- [8] Abou-Ismaïl A. Compartmental models of the COVID-19 pandemic for physicians and physician scientists. *SN Compr Clin Med*.2020;1. [\[CrossRef\]](#)
- [9] Abioye AI, Peter OJ, Ogunseye HA, Oguntolu FA, Oshinubi K, Ibrahim AA, Khan I. Mathematical model of COVID-19 in Nigeria with optimal control. *Results Phys* 2021;28:104598. [\[CrossRef\]](#)
- [10] Ray SS, Atangana A, Noutchie SC, Kurulay M, Bildik N, Kilicman A. Fractional calculus and its applications in applied mathematics and other sciences. *Math Probl Eng* 2014;849395. [\[CrossRef\]](#)
- [11] Li C, Peng G. Chaos in Chen's system with a fractional order. *Chaos Solit Fractals* 2004;22:443–450. [\[CrossRef\]](#)
- [12] Owolabi KM, Gómez-Aguilar JF, Fernández-Anaya G, Lavín-Delgado JE, Hernández-Castillo E. Modelling of chaotic processes with Caputo fractional order derivative. *Entropy* 2020;22:1027. [\[CrossRef\]](#)
- [13] Petráš I. A note on the fractional-order Chua's system. *Chaos Solit Fractals*. 2008;38:140–147. [\[CrossRef\]](#)
- [14] Atangana A, Koca I. Chaos in a simple nonlinear system with Atangana–Baleanu derivatives with fractional order. *Chaos Solit Fractals* 2016;447–454. [\[CrossRef\]](#)
- [15] Gozalez-Parra G, Arenas AJ, Chen-Charpentier BM. A fractional order epidemic model for the simulation of outbreaks of influenza A(H1N1). *Math Methods Appl Sci* 2014;37:2218–2226. [\[CrossRef\]](#)
- [16] Paul S, Mahata A, Mukherjee S, Roy B. Dynamics of SIQR epidemic model with fractional order derivative. *Partial Differ Equ Appl Math* 2022;5:100216. [\[CrossRef\]](#)
- [17] Area I, Batarfi H, Losada J, Nieto JJ, Shammakh W, Torres A. On a fractional order Ebola epidemic model. *Adv Differ Equ* 2015:278. [\[CrossRef\]](#)
- [18] Din A, Abidin MZ. Analysis of fractional-order vaccinated Hepatitis-B epidemic model with Mittag-Leffler kernels. *Math Model Numer Simul Appl* 2022;2:59–72. [\[CrossRef\]](#)
- [19] Ding Y, Wang Z, Ye H. Optimal control of a fractional-order HIV-immune system with memory. *IEEE Trans Control Syst Technol* 2011;99:1–7.

- [20] El-Shahed M, Alsaedi A. The fractional SIAC model and influenza A. *Math Probl Eng*. 2011;4:480378. [\[CrossRef\]](#)
- [21] Birs I, Muresan C, Nascu I, Ionescu C. A survey of recent advances in fractional order control for time delay systems. *IEEE Access* 2009;7:30951–30965. [\[CrossRef\]](#)
- [22] Chen L, Chai Y, Wu R, Ma T, Zhai H. Dynamic analysis of a class of fractional-order neural networks with delay. *Neural Netw* 2013;111:190–194. [\[CrossRef\]](#)
- [23] Zou G, Wang B. On the study of stochastic fractional-order differential equation systems. *arXiv:1611.07618* 2016. doi: 10.48550/arXiv.1611.07618
- [24] Ross B. Fractional calculus and its applications; Proceedings of the International Conference Held at the University of New Haven, June. Berlin: Springer Verlag; 1974. [\[CrossRef\]](#)
- [25] Samko S, Kilbas A, Marichev O. Fractional integrals and derivatives: theory and applications. Amsterdam: Gordon & Breach Science Publishers; 1993.
- [26] Sher M, Shah K, Sarwar M, Alqudah MA. Mathematical analysis of fractional order alcoholism model. *Alex Eng J* 2023;78:281–291. [\[CrossRef\]](#)
- [27] Khalil R, Horani MA, Yousef A, Sababheh M. A new definition of fractional derivative. *J Comput Appl Math* 2014;264:65–70. [\[CrossRef\]](#)
- [28] Podlubny I. Fractional differential equations. Mathematics in science and engineering. New York: Academic Press; 1999.
- [29] Abdeljawad T. On conformable fractional calculus. *J Comput Appl Math*. 2015;279:57–66. [\[CrossRef\]](#)
- [30] Mostafa E, Rezazadeh H. The first integral method for Wu–Zhang system with conformable time-fractional derivative. *Calcolo* 2015;53:1–11. [\[CrossRef\]](#)
- [31] Chung WS. Fractional Newton mechanics with conformable fractional derivative. *J Comput Appl Math* 2015;290:150–158. [\[CrossRef\]](#)
- [32] Ünal E, Gökdoğan A. Solution of conformable fractional ordinary differential equations via differential transform method. *Optik* 2017;128:264–273. [\[CrossRef\]](#)
- [33] Shehzada K, Ullah A, Saifullah S, Akgül A. Fractional generalized perturbed KdV equation with a power law kernel: a computational study. *Res Control Optim* 2023;12:100298. [\[CrossRef\]](#)
- [34] Ahmad S, Saifullah S. Analysis of the seventh-order Caputo fractional KdV equation: applications to the Sawada–Kotera–Ito and Lax equations. *Commun Theor Phys* 2023;75:085002. [\[CrossRef\]](#)
- [35] Li B, Zhang T, Zhang C. Investigation of financial bubble mathematical model under fractal–fractional Caputo derivative. *Fractals* 2023;31:2350050. [\[CrossRef\]](#)
- [36] Li B, Eskandari Z. Dynamical analysis of a discrete-time SIR epidemic model. *J Frank Inst* 2023;360:7989–8007. [\[CrossRef\]](#)
- [37] He Q, Rahman M, Xie C. Information overflow between monetary policy transparency and inflation expectations using multivariate stochastic volatility models. *Appl Math Sci Eng* 2023;31:2253968. [\[CrossRef\]](#)
- [38] Zhu X, Xia P, He Q, Ni Z, Ni L. Coke price prediction approach based on dense GRU and opposition-based learning salp swarm algorithm. *Int J Bio Inspired Comput* 2023;21:106–121. [\[CrossRef\]](#)
- [39] Hanert E. A comparison of three Eulerian numerical methods for fractional-order transport models. *Environ Fluid Mech* 2010;10:7–20. [\[CrossRef\]](#)
- [40] 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. [\[CrossRef\]](#)
- [41] Khan A, Gomez-Aguilar JF, Khan TS, Khan H. Stability analysis and numerical solutions of fractional order HIV/AIDS model. *Chaos Solitons Fractals* 2019;122:119–128. [\[CrossRef\]](#)
- [42] Khan H, Alam K, Gulzar H, Etemad S, Rezapour S. A case study of fractal fractional tuberculosis model in China: existence and stability theories along with numerical simulations. *Math Comput Simul* 2022;198:455–473. [\[CrossRef\]](#)
- [43] Nawaz Y, Arif MS, Shatanawi W. A new numerical scheme for time fractional diffusive SEAIR model with non-linear incidence rate: an application to computational biology. *Fractal Fract* 2022;6:78. [\[CrossRef\]](#)
- [44] Sinan M, Leng J, Shah K, Abdeljawad T. Advances in numerical simulation with a clustering method based on K-means algorithm and Adams–Bashforth scheme for fractional order laser chaotic system. *Alex Eng J* 2023;75:165–179. [\[CrossRef\]](#)
- [45] Sinan M, Ali A, Shah K, Assiri TA, Nofal M. Stability analysis and optimal control of COVID-19 pandemic SEIQR fractional mathematical model with harmonic mean type incidence rate and treatment. *Results Phys* 2021;22:103873. [\[CrossRef\]](#)
- [46] Mohammadnezhad V, Eslami M, Rezazadeh H. Stability analysis of linear conformable fractional differential equations system with time delays. *Bol Soc Paran Mat* 2020;38:159–171. [\[CrossRef\]](#)
- [47] Lin W. Global existence theory and chaos control of fractional differential equations. *J Math Anal Appl* 2007;332:709–726. [\[CrossRef\]](#)
- [48] Paul SK, Mishra LN. Stability analysis through the Bielecki metric to nonlinear fractional integral equations of n-product operators. *AIMS Math* 2024;9:7770–7790. [\[CrossRef\]](#)
- [49] Paul SK, Mishra LN, Mishra VN, Baleanu D. Analysis of mixed type nonlinear Volterra–Fredholm integral equations involving the Erdélyi–Kober fractional operator. *J King Saud Univ Sci* 2023;35:102949. [\[CrossRef\]](#)

- [50] Paul SK, Mishra LN. Approximation of solutions through the Fibonacci wavelets and measure of noncompactness to nonlinear Volterra–Fredholm fractional integral equations. *Korean J Math* 2024;32:137–162.
- [51] Merdan M, Anaç H, Kesemen T. The new Sumudu transform iterative method for studying the random component time-fractional Klein–Gordon equation. *Sigma J Eng Nat Sci* 2019;10:343–354.
- [52] Anaç H, Merdan M, Bekiryazıcı Z. Global stability in a pathogen-specific CD8 immune response prediction model. *Sigma J Eng Nat Sci* 2019;10:325–342.
- [53] Şişman Ş, Merdan M. Global stability of susceptible diabetes complication (SDC) model in discrete time. *Sigma J Eng Nat Sci* 2021;39:290–312. [\[CrossRef\]](#)
- [54] Yadav S, Kumar D, Singh J, Baleanu D. Analysis and dynamics of fractional order COVID-19 model with memory effect. *Results Phys* 2021;24:104017. [\[CrossRef\]](#)
- [55] Inokuti M. General use of the Lagrange multiplier in nonlinear mathematical physics. In: Nemat-Nasser S, ed. *Variational method in the mechanics of solids*. Oxford: Pergamon Press; 1978. p. 156–162. [\[CrossRef\]](#)
- [56] He JH. Some asymptotic methods for strongly nonlinear equations. *Int J Mod Phys B* 2006;20:1141–1199. [\[CrossRef\]](#)
- [57] Bildik N. General convergence analysis for the perturbation iteration technique. *Turk J Math Comput Sci* 2017;6:1–9.
- [58] Odibat ZM. A study on the convergence of variational iteration method. *Math Comput Model* 2010;51:1181–1192. [\[CrossRef\]](#)



Research Article

Global solar radiation prediction using machine learning approaches

M. MOHANDES^{1,3}, Hilal H. NUHA², Satria Akbar MUGITAMA², S. REHMAN^{3,*},
A. AL-SHAILKHI¹

¹Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia

²Telcom University, Bandung, 40257, Indonesia

³King Fahd University of Petroleum & Minerals (KFUPM), Interdisciplinary Research Center for Sustainable Energy Systems (IRC-SES), Dhahran, 31261, Saudi Arabia

ARTICLE INFO

Article history

Received: 15 July 2024

Revised: 22 August 2024

Accepted: 11 September 2024

Keywords:

BiLSTM; ConvNet; GSR; GRU

ABSTRACT

Global Solar Radiation (GSR) stands as a crucial renewable energy source for electricity and heat generation without emitting greenhouse gases. Fuelled by escalating fossil fuel prices, the necessity to curb greenhouse gases (GHG) emissions, and the rapid advancement of solar technology; the role of GSR becomes pivotal in shaping the energy landscape. So, it becomes imperative to understand the variability and availability of GSR on various time scales in the temporal domain. This research conducts an in-depth comparative analysis of various machine learning models, including Long-Short Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (ConvNet), Multilayer Perceptron (MLP), Generalized Additive Model (GAM), Gaussian Process Regression (GPR), and Linear Regression (LR) for GSR prediction to recommend the best method/s for the purpose. Employing robust evaluation metrics such as: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Biased Error (MBE), and the coefficient of determination (R^2), the study examines the predictive capabilities of these models. The numerical experimental results show that BiLSTM emerges as the standout performer, having minimal deviation from actual values and slightly positive bias. Its remarkable R^2 value (99.26%) highlights its predictive capability.

Cite this article as: Mohandes M, Nuha HH, Mugitama SK, Rehman S, Al-Shailkhi A. Global solar radiation prediction using machine learning approaches. Sigma J Eng Nat Sci 2025;43(5):1725–1736.

INTRODUCTION

Global solar radiation is important for a number of reasons. First, it is a renewable energy source that can be used to generate electricity and heat without emitting greenhouse gases (GHG). Second, it is a distributed source that

is available in most parts of the world. Third, it is a cost-effective and is becoming increasingly affordable as solar technology continues to improve. In 2022, the global solar energy market was worth \$160 billion. It is expected to grow to \$280 billion by 2027 [1]. The growth of the solar energy

*Corresponding author.

*E-mail address: srehman@kfupm.edu.sa; mohandes@kfupm.edu.sa

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



market is being driven by a number of factors, including the increasing cost of fossil fuels, the need to reduce GHG emissions, and the growing availability of solar technology.

Predicting GSR ahead of time is important for several reasons. First, it can help to better understand the cycle of the Earth's energy balance and climate. Second, it can be used to optimize the design and operation of solar energy systems. Third, it can help to mitigate the effects of climate change. However, despite its periodicity, GSR modeling is challenging. As Fig. 1 shows, there are variations in hourly GSR averages over 24 hours on different days (Fig. 1a) and monthly GSR averages over 12 months in different years (Fig. 1b). The GSR data was collected at King Fahd University of Petroleum & Minerals (KFUPM) from 2012 to 2020. Research on GSR prediction is ongoing globally. Scientists are working to develop more accurate and reliable models that can be used to support the development and operation of solar energy systems [2–5]. Researchers are making efforts to develop models to estimate the effects of GSR intensities on climate change.

In recent years, numerous studies have been conducted to investigate solar radiation prediction models and their applications in various fields [6]. Sun et al. [7] studied the effect of solar radiation and other meteorological variables on the its predictability using autoregressive moving average combined with exogenous variable-generalized autoregressive conditional heteroscedasticity (ARMAX-GARCH) and multivariate GARCH (MGARCH) models. Authors reported good correlations between GSR and sunshine duration compared to ambient temperature. Alsina et al. [8] used Artificial Neural Networks (ANNs) to predict Monthly Average Daily GSR (MADGSR) over Italy and achieved better performance with MAPE ranging from 1.67% to 4.25%. The ANN best configuration included

seven relevant input parameters, highlighting the efficacy of ANN for solar radiation prediction.

The authors [9] experiments in remote areas for milk storage systems using solar thermal energy and adsorption cooling. The authors [10] research by creating a pollution monitoring and forecasting system for healthy breathing and living. Sharnil The authors [11] conducted research to build a framework based on sensor fusion and deep learning for virus outbreaks. The authors [12] build a framework for the MQTT multicast messaging protocol for IoT-related problems. The authors [13] conducted research on the effect of the filling and emptying process of PCM with paraffin and Al_2O_3 additives at three locations. The authors [14] carried out experiments to analyse and theorise on exhaust manifolds with uncoated ceramics and with Al_2O_3 , TiO_2 and ZrO_2 coatings.

The authors [15] conducted research to study the drying of coriander seeds in a photovoltaic thermal collector with a solar dryer. The authors [16] undertook research into the design and manufacture of a two-stage, energy-efficient, nickel-chromium-reinforced pyrolysis reactor for plastic waste applications. The authors [17] carried out research on the effects of emissions from DI-based sources that may cause adverse health effects. Kumar et al. [18] reported global challenges and emerging trends being faced by the scientific and engineering community in forecasting thermal performance degradation of WTG gearboxes. Thermal performance of WTG gearboxes degradation in relation to plant loads was carried out by experimental measurements of temperature flux measurements with condition monitoring systems by Kumar et al. [19] Dhabale et al. [20] conducted on the multi-objective optimisation of friction stir welding process parameters for copper particle reinforcement in aluminium alloys

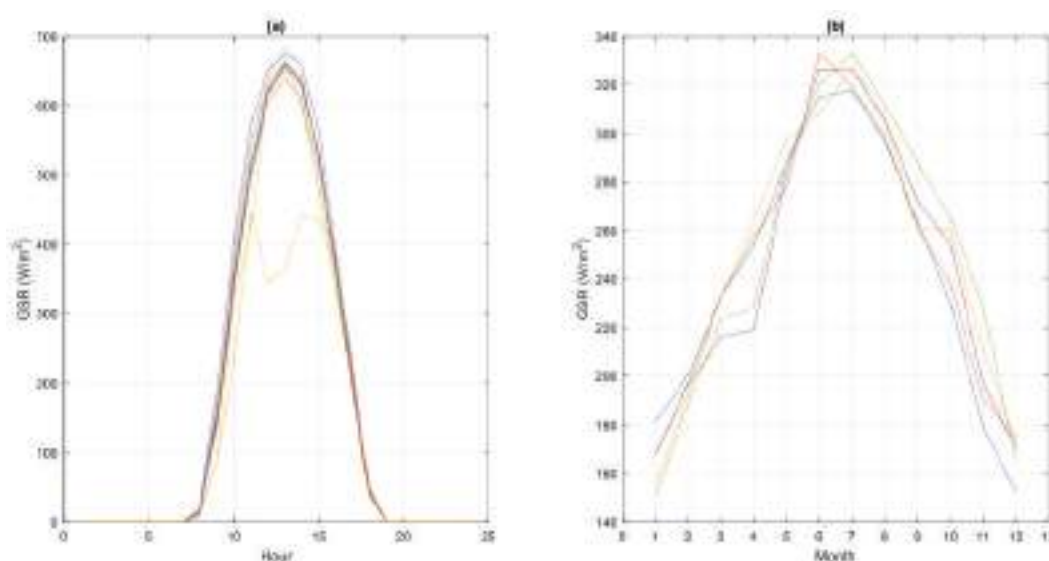


Figure 1. (a) Hourly and (b) Monthly GSR samples [6].

using Taguchi-based Grey Relational Analysis (GRA) and ANOVA. Raghavendrakumar et al. [21] investigated the analysis of central atmospheric temperature variability data from the SABER satellites.

A comprehensive perspective on machine-learning models for GSR was presented by reviewing 232 papers, exploring input parameters, feature selection methods, and model development techniques [22]. The authors reviewed machine-learning models for GSR forecasting and sources of input data. Additionally, they reviewed. Rabehi and Lalmi [23] compared the performance of different models (multi-layer perceptron (MLP), boosted decision tree, and hybrid models combining MLP with linear regression) for predicting solar radiation. The results showed that the MLP model performed the best with a normalized RMSE of 0.033 and an R^2 of 97.7%.

Ramedani et al. [24] developed four models (Support Vector Regression (SVR) with polynomial and radial basis function (RBF) kernels) to predict GSR. Results showed that SVR with RBF had better accuracy compared SVR with polynomial, ANFIS, and ANN models.

The present research compares the predictive performance of LSTM, BiLSTM, GRU, ConvNet, MLP, GAM, GPR, and LR for GSR prediction. Present study selected recurrent models (LSTM, BiLSTM, and GRU) based on their better capability of capturing long-term temporal dependencies, which is crucial for time dependent predictions. For benchmarking purpose, the study includes the GPR and GAM regression models. For model performance evaluation, RMSE, MAE, MBE, and R^2 error metrics is used.

METHODOLOGY

This section provides the description of methods used in this paper. The methods include LSTM, BiLSTM, GRU, ConvNet, MLP, GPR, GAM, and LR. The performance of

these methods is compared using the error measures, mentioned earlier.

LONG-SHORT-TERM MEMORY (LSTM) AND BIDIRECTIONAL LSTM (BiLSTM)

The Long-Short-Term Memory (LSTM) neural networks [25] technique is employed in the study for GSR forecasting. This model is an advanced form of RNN and incorporates a recurrent network delay which enables the association of the network's output for a given t -th sample with prior inputs, as illustrated in Figure 2. The LSTM architecture addresses the issue of diminishing gradients through the implementation of specialized logic gates, specifically the input (i), forget (f), and output (o) gates [26]. Moreover, the model is equipped with weights that link the gates, which are the input weights U , consisting of U^f , U^i , U^o and U^c , the recurrent weights W , comprising of W^f , W^i , W^o and W^c , and the bias terms b , including b_f , b_i , b_o and b_c [26]. The input at the t -th sample, denoted as X_t along with the hidden state from the preceding sample h_{t-1} , are processed by the LSTM cell to produce the current hidden state h_t using the steps outlined below:

- (1) The forget gate (f_t) determines the information to be discarded from the unit state, which is executed according to the subsequent equation:

$$f_t = \sigma(U^f X_t + W^f h_{t-1} + b_f) \quad (1)$$

where $\sigma(z)$ is the sigmoid function given by:

$$\sigma(z) = \frac{1}{1+e^{-z}}. \quad (2)$$

- (2) The input gate (i_t) identifies the information that should be retained in the unit and selects the values using the following equation:

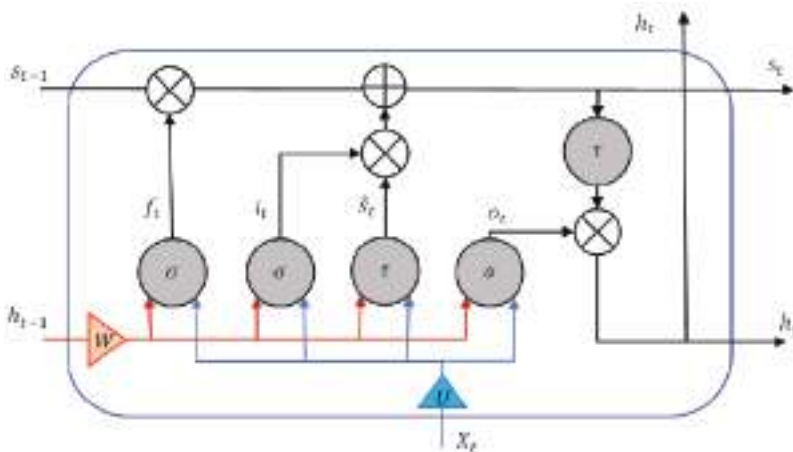


Figure 2. LSTM unit structure.

$$i_n = \sigma(U^i X_t + W^i h_{t-1} + b_i) \quad (3)$$

(3) The unit state is updated with the candidate value $\hat{s}_t$ based on the following equation 4:

$$\hat{s}_t = \tau(U^c X_t + W^s h_t + b_s) \quad (4)$$

where $\tau(z)$ is a tangent-hyperbolic function given by:

$$\tau(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (5)$$

The candidate unit state is further used to determine the unit state values as follows:

$$s_t = s_{t-1} \otimes f_t \oplus i_t \otimes \hat{s}_t \quad (6)$$

where are the operators $\oplus$ and $\otimes$ represent element-wise summation and product, respectively.

(4) Next, the output o_t and the hidden state h_t is determined using the following calculation:

$$o_n = \sigma(U^o X_t + W^o h_{t-1} + b_o). \quad (7)$$

while the hidden state of the unit is given by:

$$h_t = o_t \otimes \tau(s_t). \quad (8)$$

The Bidirectional LSTM (BiLSTM) (Fig. 3) is an extension of the LSTM unit where the unit has two-way access. BiLSTM utilizes two LSTM models instead of just one, training in opposite directions on the input sequence. This means that while the first LSTM is designed to understand the sequence in a forward manner, the second LSTM is tasked with learning the sequence in reverse order. In the present approach, the LSTM is trained using LM method

[27] for fast convergence. The LM algorithm is deemed particularly appropriate for this application, which necessitates a moderate quantity of units and layers. The update of weights (Δu) within the LM method is characterized by the following:

$$\Delta u = -[J^T J + \lambda I]^{-1} J^T (y - \hat{y}) \quad (9)$$

where J represents the Jacobian matrix that maps the partial derivatives of the error function relative to the weight vector of the model [28]. The error function is evaluated by the discrepancy between the actual values (y) and the predicted outputs ($\hat{y}$). The scalar (λ) controls the magnitude of the step size, which is augmented if the updates lead to a reduction in the error function and is reduced otherwise. The methodology employs a maximum number of iterations set at $T = 200$.

GATED RECURRENT UNIT (GRU)

The GRU is a type of RNN architecture that is designed to capture long-term dependencies in sequential data while mitigating the vanishing gradient problem [29]. Mathematically, the GRU (Figure. 4) can be described as follows:

(1) Update Gate (z_t): The update gate determines how much of the previous hidden state (h_{t-1}) should be retained and combined with the new candidate hidden state ($\tilde{h}_t$) at the current time step (t). It takes input features x_t and h_{t-1} and produces a value between 0 and 1 to represent the proportion of information to keep.

$$z_t = \sigma(W_z \times x_t + U_z \times h_{t-1} + b_z) \quad (10)$$

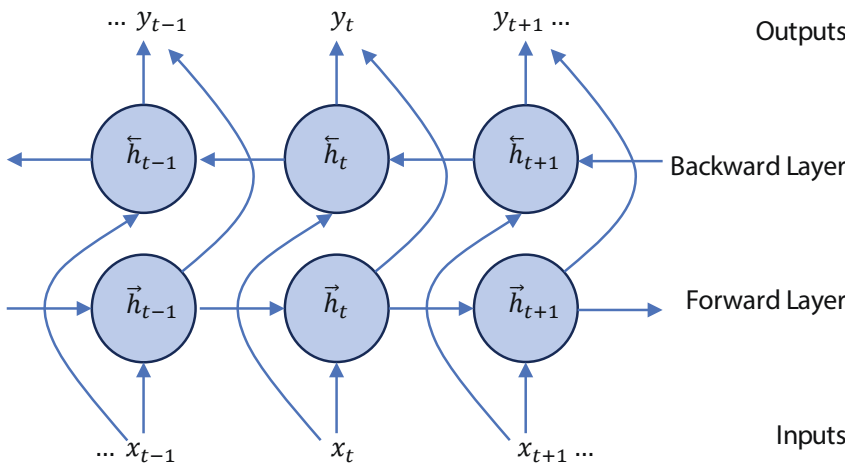


Figure 3. Bidirectional Unit.

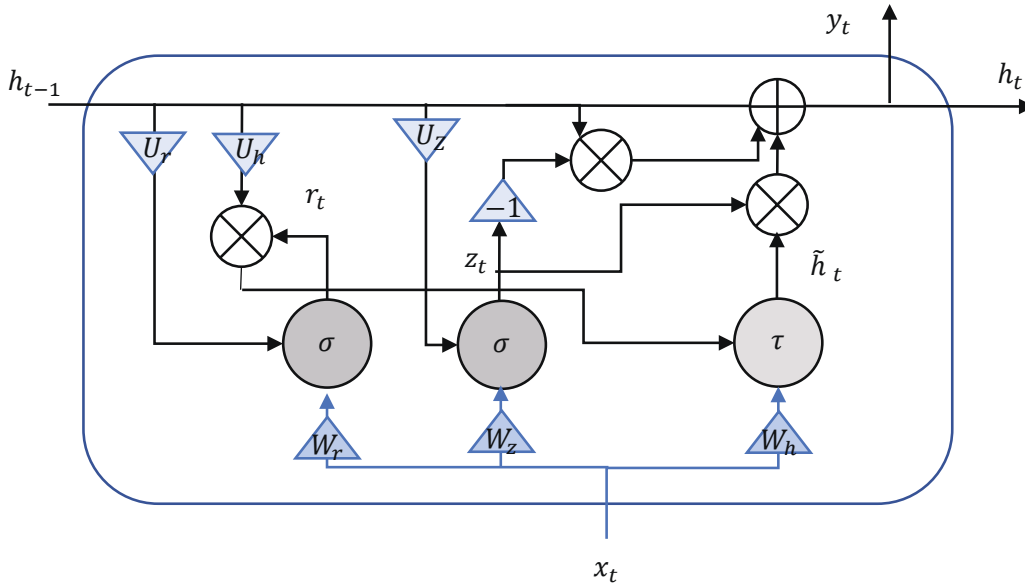


Figure 4. GRU architecture.

- (2) Reset Gate (r_t) determines how much of the previous hidden state (h_{t-1}) should be ignored while calculating the new state ($\tilde{h}_t$). It produces values between 0 and 1.

$$r_t = \sigma(W_r \times x_t + U_r \times h_{t-1} + b_r) \quad (11)$$

- (3) Candidate Hidden State ($\tilde{h}_t$) is the new hidden state at time step t , which combining the current input features (x_t) and previous hidden state (h_{t-1}) based r_t .

$$\tilde{h}_t = \tau(W_h \times x_t + U_h \times (r_t \times h_{t-1}) + b_h) \quad (12)$$

- (4) Final Hidden State (h_t) at time step t is a combination of h_{t-1} and $\tilde{h}_t$ weighted by z_t . It determines how much information from the previous hidden the candidate hidden state should be incorporated.

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (13)$$

To use the GRU to predict GSR at time step x_{t+1} , given the previous GSR values $x_1, x_2, \dots, x_t$, the following steps are taken:

- Initialize state h_0 to a vector of zeros.
- For each time step t from 1 to T , compute the update gate (z_t), reset gate (r_t), candidate hidden state $\tilde{h}_t$, and the final hidden state (h_t) using the formulas mentioned above.
- At time step $t + 1$, use the final hidden state h_t as the context or representation of the sequence $x_1, x_2, \dots, x_t$. Then, compute the predicted GSR value x_{t+1} using this context h_t as input to a fully connected layer or regression layer.

$$\tilde{h}_t = \tau(W_h \times x_t + U_h \times (r_t \times h_{t-1}) + b_h) \quad (14)$$

BPTT is used to train the model along with weights and biases to minimize the error. These steps are repeated over multiple epochs till the model achieves satisfactory performance.

CONVOLUTIONAL NEURAL NETWORKS (ConvNet)

The ConvNet [30], originally designed for image classification, has shown exceptional results surpassing other leading methods in that domain. Given the 2D ConvNet's impressive outcomes in image-centric tasks, a 1D variant is created to analyse 1D signals and sequences [30]. This is achieved by isolating significant features at both high and low levels within the input data sequences. The ConvNet structure consists of a sequence input layer, a convolutional layer, a fully connected layer, and a regression layer. In this context, the width of the time series, denoting the feature count, is represented by K , and its length by N [31]. The initial phase involves a convolution operation that progresses along a set direction from the start to the end of the time series, utilizing convolutional filters that match the series width but may vary in length. The primary function of the convolutional layer is to extract features from the time series, followed by the subsequent pooling stage where vectors from the convolutional layer are aggregated into a new vector in what is known as the pooling layer.

This model was enhanced by incorporating an adaptive moment estimation (ADAM) optimizer [32]. The update of each weight w_i is obtained as follows:

$$w_i = w_{i-1} - \alpha \frac{\sqrt{1-\beta_1^i}}{1-\beta_1^i} m_i \oslash \sqrt{v_i} \oplus \epsilon. \quad (15)$$

where α and β are the learning and decay rates. The operation of element-wise vector division is indicated by the symbol $\oslash$, and the calculation of the first moment m is executed in the subsequent manner.

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \nabla_w e(w_{i-1}). \quad (16)$$

where $\nabla_w e(w)$ symbolizes the vector differential operation of the error function e with respect to the weight w . The computation of the second moment vector v is carried out in the following way.

$$v_i = \beta_2 v_{i-1} + (1 - \beta_2) \nabla_w e(w_{i-1}) \otimes \nabla_w e(w_{i-1}). \quad (17)$$

The learning rate α can assume any positive real number value, whereas the decay rates β_1 and β_2 are real numbers that are greater than 0 but less than 1.

MULTI-LAYER PERCEPTRON (MLP)

The MLP [33] is utilized as a standard for comparing the performance of LSTM. MLP possesses a notably simpler architecture compared to LSTM. Unlike LSTM, MLP employs a feed-forward structure that processes data in a sequential manner from the input layer to the output layer. The output of the k -th hidden unit, denoted as $h_k(t)$, for the t -th input vector, is mathematically expressed as follows.

$$h_k(t) = \tau(V_k x(t) + b_k) \quad (18)$$

where V_k represents the input weight matrix. In this paper, the number of hidden units used is $K = 20$. With these values for the hidden units established, the output of the MLP is determined by the following equation.

$$y(t) = U h(t) + b \quad (19)$$

This study employs gradient descent with adaptive learning rate (lr) backpropagation for training the MLP [27]. This algorithm is applied to compute the derivatives of the error function, denoted as dE , relative to the weight and bias variables w . The performance of this approach is expressed as follows.

$$E = (y - \hat{y})^2 \quad (20)$$

Each weight is updated based on the following gradient descent equation.

$$w_i = w_{i-1} + lr \frac{\partial E}{\partial w} \quad (21)$$

During each iteration, if the output diverges from the target, the learning rate is augmented by a factor of $\Delta lr = 1.05$. Conversely, if the error escalates beyond the threshold of $\Delta E_{max} = 1.04$, the adjustment that caused the increase in error is reversed, and the learning rate is reduced by a factor of $\Delta lr = 0.7$. The training process concludes once the maximum iteration count, set at $T = 1200$, is attained.

GAUSSIAN PROCESS REGRESSION (GPR)

Gaussian Process Regression (GPR) [34] is a non-parametric Bayesian regression method that builds the relationship between the target variable and the independent variables as a Gaussian process. GPR is useful with small datasets. Mathematically, GPR can be expressed as follows:

$$y_{t+1} \sim N(m_{t+1}, K_{t+1}) \quad (22)$$

where m_{t+1} is the mean function at time step $t + 1$ given by the historical GSR values $x_1, x_2, \dots, x_t$. K_{t+1} is the covariance matrix at time step $t + 1$. To use GPR for predicting GSR at time step $t + 1$ (x_{t+1}) given the previous GSR values $x_1, x_2, \dots, x_t$, you need to estimate the mean function m_{t+1} and the covariance matrix K_{t+1} . This involves finding the hyperparameters of the mean and covariance functions that best fits the training data. Once the mean function m_{t+1} and the covariance matrix K_{t+1} are estimated, one can make predictions for the GSR value at time step $t + 1$ using the conditional distribution of y_{t+1} given the observed historical GSR values $x_1, x_2, \dots, x_t$. The predicted GSR value at time step $t + 1$ is the mean of this conditional distribution, which is given by m_{t+1} .

GENERALIZED ADDITIVE MODEL (GAM)

Generalized Additive Model (GAM) [35] is a type of generalized linear model that extends linear regression to accommodate non-linear relationships between predictors and the response variable. Mathematically, a GAM can be represented as follows. For a regression task, let's consider a single response variable y and p predictor variables ($X_1, X_2, \dots, X_p$). The general form of a GAM can be written as.

$$y = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) + \epsilon \quad (23)$$

Where y is the response variable, β_0 is the intercept term, ($f_1(X_1), f_2(X_2), \dots, f_p(X_p)$) are smooth functions of the predictor variables. These functions are represented by splines, polynomials, or others. ϵ is the error term and follows a normal distribution for regression tasks. The GAM fitting process involves estimating the parameters β_0 and the smooth functions $f_i(X_i)$ from the training data. This is estimated by minimizing a loss function that penalizes the complexity of the model to prevent overfitting. The penalization term discourages overly complex functions, encouraging smoothness in the fitted functions.

LINEAR REGRESSION (LR)

Linear regression [36] is a simple and widely used statistical method for modelling the relationship between a dependent variable and one or more independent variables. In the context of predicting global solar radiation (GSR) at time step $t + 1$ (x_{t+1}) given previous GSR values $x_1, x_2, \dots, x_p$, linear regression can be used to establish a linear relationship between the historical GSR values and the target GSR value at time $t + 1$.

$$y = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) + \epsilon \quad (24)$$

Here, $\beta_0, \beta_1, \dots, \beta_t$ are the coefficients of the model that represent the intercept and the slopes associated with each of the historical GSR values. ϵ represents the error term, which captures the discrepancy between the predicted GSR value and the actual GSR value at time $t+1$. The resulting y_{t+1} represents the predicted GSR value at time $t+1$. The coefficient vector β is calculated using the following formula.

$$\beta = (X^T X)^{-1} X^T y \quad (25)$$

where X^T is the transpose of the design matrix X . The inverse of the matrix $(X^T X)^{-1}$ is generated from the multiplication of X^T and X . The matrix $X^T y$ is resulting from the multiplication of X^T and the target vector y .

DATA AND ERROR MEASURES

The measured meteorological data was obtained from the station located at $26^\circ 18' 26''$ N latitude and $50^\circ 08' 29''$ E longitude at King Fahd University of Petroleum & Minerals,

Dhahran, Saudi Arabia. The meteorological parameters (global solar radiation intensity-GSR, ambient temperature, relative humidity, wind speed, wind direction, and surface pressure) are scanned every five seconds and saved as averages over one hour. The data, used in the present study, spans over a length of nine years from 01/01/2012 to 12/31/2020. The experiment in this paper relies on a single dataset from one location, which may limit the generalizability of our findings. However, the dataset covers a substantial period of 8 years, offering robust temporal coverage that effectively captures various patterns and trends in solar radiation.

A sample of data set of GSR values (W/m^2), between 01/01/2015 and 31/12/2017 is shown in Figure 5. It is evident from the sample data set that the GSR intensity values vary from a minimum at the start of the year (winter time) and after reaching a maximum during summer time touches back the minima at the end of the year. This process is repeated from year to year but with varying magnitudes. Similarly, the daily variation of GSR intensity values, shown in Figure 6, start increasing from the early hours of the day and then keep on increasing with the passage of time and after reaching a maximum at around noon time tend to touch the minima back again at sun set. This process, as indicated in Figure 6, keeps on repeating day after day but with different values of GSR intensities.

Since the dataset is a time series, so is not shuffled when loaded. The data is divided into training, validation, and test portions of 60%, 20%, and 20%, respectively. The 24 hours lag GSR-values are considered input to give a desired input structure into a model, and the model predicts the next hour. Error measures, such as RMSE, MAE, and R^2 , are commonly used to assess the accuracy and performance of predictive models, including those used for GSR prediction.

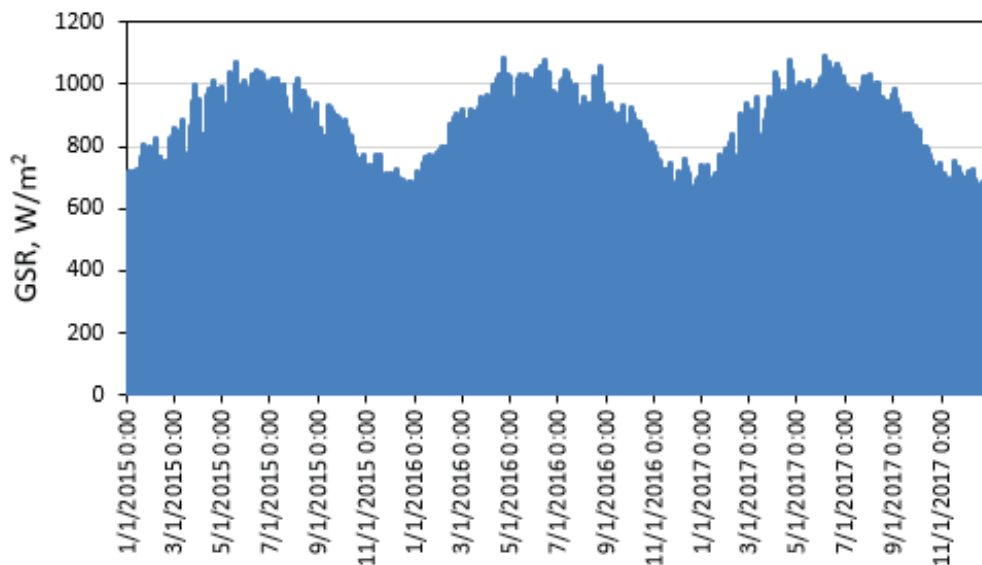


Figure 5. GSR intensity values over a period of three years (01/01/2015 to 31/12/2017).

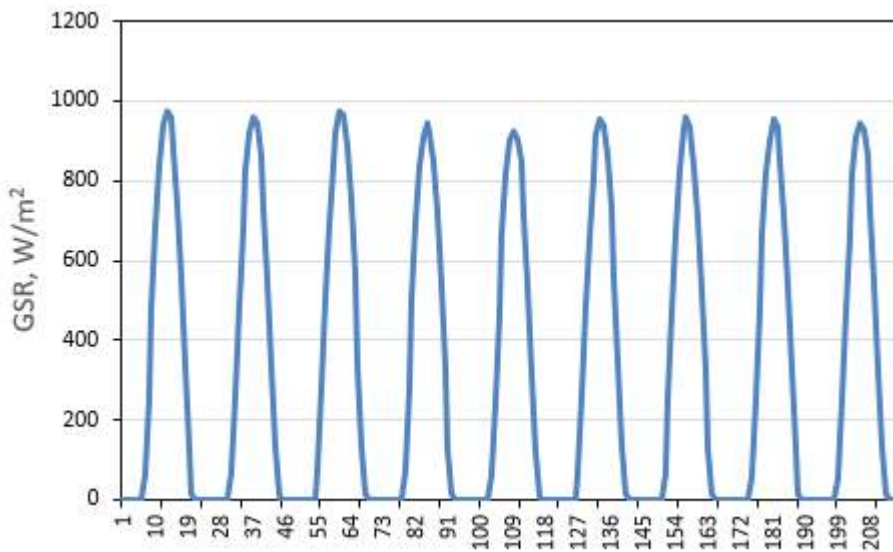


Figure 6. GSR intensity values over a period of nine days (22/06/2016 to 30/06/2016).

RMSE [37] is calculated by taking the square root of the average of the squared differences between the predicted GSR values (y_p) and the actual GSR values (y).

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (y_p(n) - y(n))^2}{N}} \quad (26)$$

MAE [38] is another error metric used to measure the average absolute difference between the predicted GSR values and the actual GSR values.

$$MAE = \frac{\sum_{n=1}^N |y_p(n) - y(n)|}{N} \quad (27)$$

MBE is used to measure the average difference between the predicted GSR values and the actual GSR values. It provides the model's tendency to under and overestimation.

$$MBE = \frac{\sum_{n=1}^N (y_p(n) - y(n))}{N} \quad (28)$$

R^2 [39–41] measures the proportion of variance in the GSR values that can be explained by the predictive model. The R^2 values range between 0 and 1.

$$R^2 = 1 - \frac{\sum_{n=1}^N (y_p(n) - y_t(n))^2}{\sum_{n=1}^N (y_t(n) - y_m)^2} \quad (29)$$

RESULTS AND DISCUSSION

The GSR values predicted using different proposed models are compared with the measured values and the resulting error parameters are summarized in Table 1. The error values obtained for different methods show that BiLSTM performs the best in terms of RMSE, indicating that the predicted values have the smallest overall deviation from the measured

Table 1. Models Accuracy

No	Methods	RMSE	MAE	MBE	R^2 (%)
1	BiLSTM [31]	30.84	17.21	11.65	99.26
2	LSTM [25]	45.87	23.97	14.36	98.17
3	GRU [29]	45.039	21.99	-9.16	98.25
4	ConvNet [30]	57.75	36.86	21.68	97.23
5	MLP [27]	40.60	15.76	1.77	98.40
6	GPR [34]	43.77	14.69	-0.15	98.14
7	GAM [35]	41.09	17.98	2.01	98.36
8	LR [36]	44.83	19.75	3.33	98.05

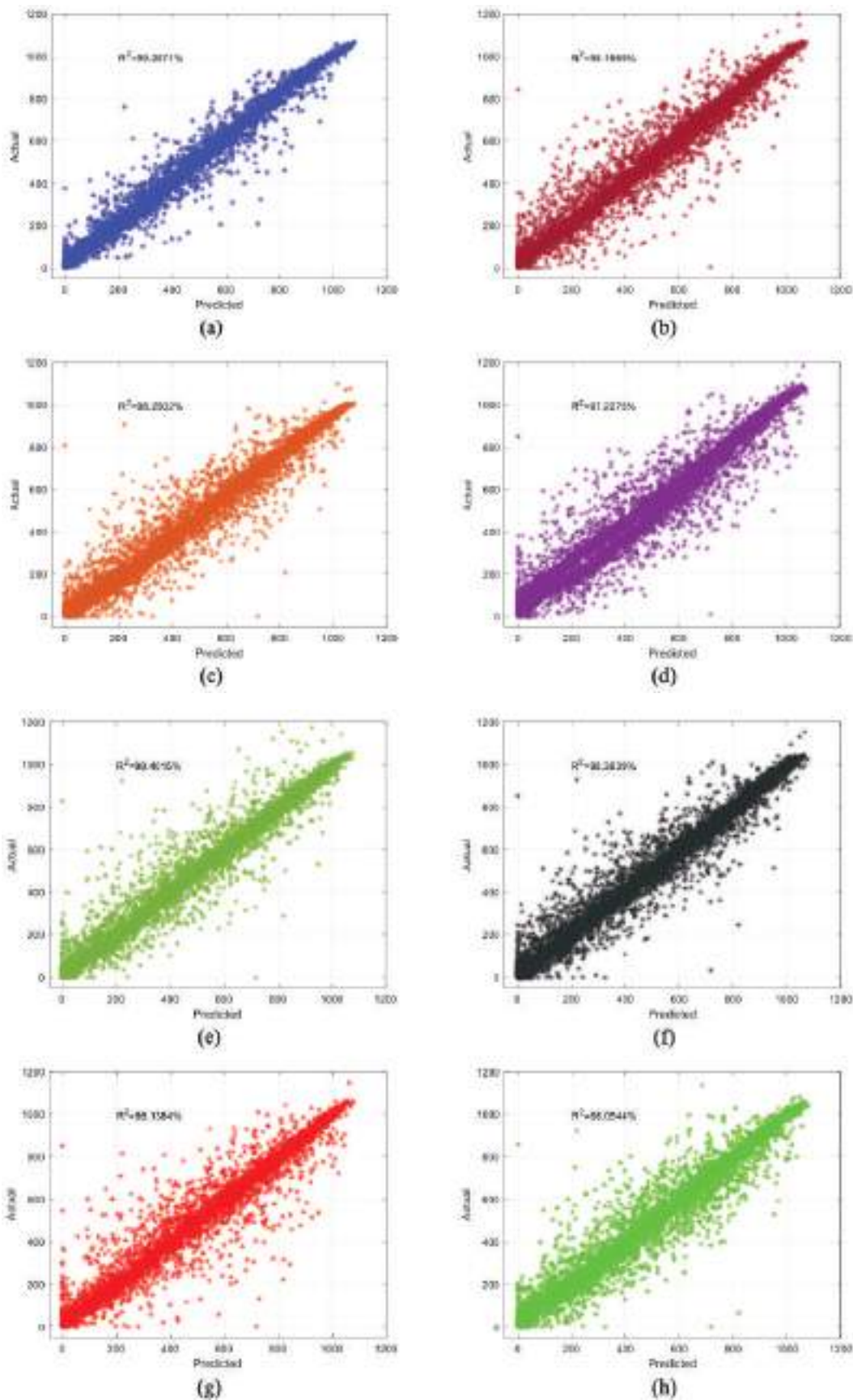


Figure 7. Scatter Plot of predicted and actual data using (a) BiLSTM, (b) LSTM, (c) GRU, (d) ConvNet, (e) MLP, (f) GPR, (g) GAM, and (h) LR.

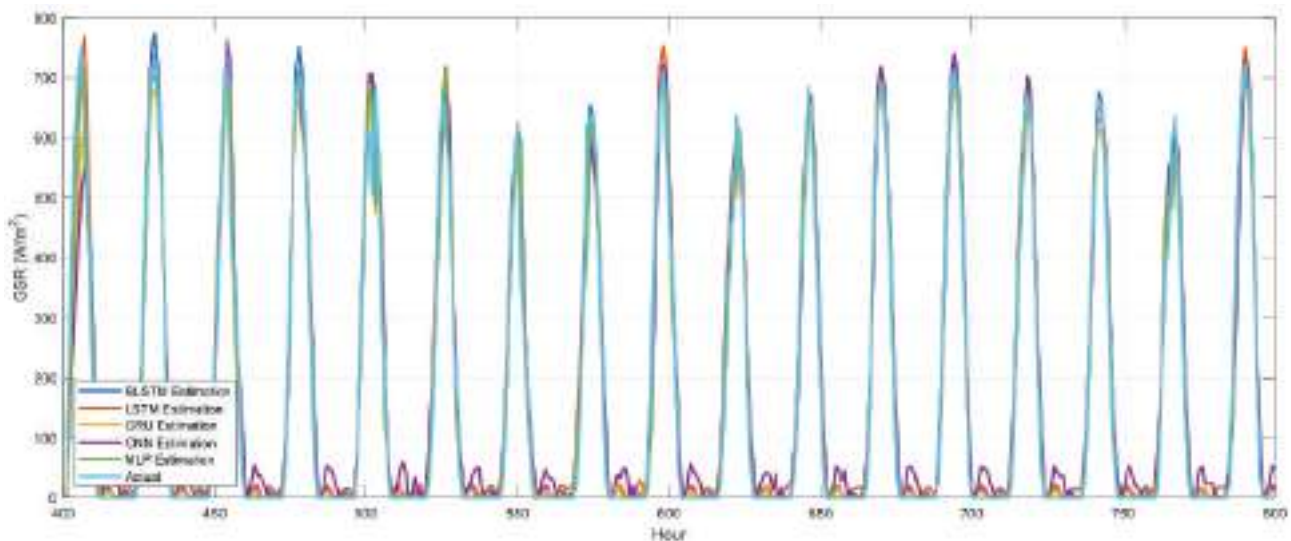


Figure 8. Comparison of predicted and actual data using BiLSTM, LSTM, GRU, ConvNet, and MLP.

values. Additionally, it has the highest R^2 , suggesting that BiLSTM method resulted in the closest agreement between the predicted and measures GSR values. LSTM is outperformed by BiLSTM, with larger values of RMSE and MAE relative to other models. However, R^2 of 98.17% is slightly lower than BiLSTM. GRU model performed poorly with high RMSE and negative MBE values. However, a R^2 value of 98.25% comparable to LSTM, represents a good fit. ConvNet method, slightly underperformed than the BiLSTM, LSTM, and GRU models in terms of RMSE and MAE. The associated R^2 value of 97.23% is observed to be the lowest, showing a weaker correlation, Table 1.

Traditional machine learning approach, like MLP, also demonstrated a relatively good performance with the lowest magnitude of MAE. The coefficient of determination value is found to be moderate with R^2 of 98.40%, suggesting a good fit to the data. GPR performed reasonably well with low MAE and RMSE values. The lower MBE and high R^2 value of 98.14% show a minor bias and a good fit to the data, respectively. GAM showed a good performance with low values of RMSE and MAE. The MBE and R^2 values of 2.01 and 98.36% show a positive bias and a good fit. The model LR performed slightly below expectations with higher RMSE and MAE values compared to some other techniques.

The plots in Figure 7, show that BiLSTM (Figure 7a) is the best- model, closely followed by MLP (Figure 7e) and GAM (Figure 7f). Overall, these models show better performance to predict GSR values with better accuracy. On the other hand, LSTM and GRU also performed reasonably well, but with varying degrees of biases and accuracies. Figure 8 shows the comparison of the predicted and actual GSR values obtained using BiLSTM, LSTM, GRU, ConvNet, and MLP methods.

CONCLUSIONS

The present study carried out state of the art analysis of predicting the Global Solar Radiation using different modern machine learning methods. Accurate GSR prediction ahead of time is crucial for better and commercial implementation of solar energy applications. These applications may include power production, solar cooling and heating, water desalination, green hydrogen production and so on. The study explored deep learning architectures (BiLSTM, LSTM, and GRU) and traditional machine learning methods (ConvNet, MLP, GPR, GAM, and LR). Among these, BiLSTM is found to be the best performing model, exhibiting the minimum RMSE and maximum R^2 values. It demonstrated the ability of retaining long-term dependencies of solar radiation patterns for accurate prediction of GSR ahead of time. MLP and GAM models followed in terms of robust performance and capturing temporal patterns of GSR. The recurrent models (BiLSTM, LSTM, and GRU) showed effectiveness in predicting the time-dependent patterns of GSR. Other deep network model, like ConvNet, also provided comparable performance. The identification of BiLSTM as the most accurate model, the present study provides a reliable method for predicting GSR with high accuracy. This, for sure, leads to more efficient resources allocation and utilization, energy management, and improved decision-making in solar energy applications.

ACKNOWLEDGEMENTS

This research is supported by King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia and PPM Telkom University, Indonesia.

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] Snapshot 2021 - IEA-PVPS. Available from: <https://iea-pvps.org/snapshot-reports/snapshot-2021/> Accessed on Dec 1, 2023.
- [2] Mutlu B, Özyörük B. A research on mathematical model approaches in biomass supply chain. *Sigma J Eng Nat Sci* 2024;42:945–955. [\[CrossRef\]](#)
- [3] İbrahim D, Sari M, Gülen S, Balacescu A. An investigation on the estimation of the impact factors of pandemic deaths with artificial neural network and multiple regression algorithms: Covid-19 case. *Sigma J Eng Nat Sci* 2024;42:667–678. [\[CrossRef\]](#)
- [4] Koçak E, Özsoy VS, Örkücü HH. Modeling of wind speed using differential evolution: Istanbul case. *Sigma J Eng Nat Sci* 2024;42:642–652. [\[CrossRef\]](#)
- [5] Kartal F, Kaptan A. Artificial neural network and multiple regression analysis for predicting abrasive water jet cutting of Al 7068 aerospace alloy. *Sigma J Eng Nat Sci* 2024;42:516–528. [\[CrossRef\]](#)
- [6] Olawoyin-Yussuf D, Mohandes M, Liu B, Rehman S. Solar radiation forecasting using attention-based temporal convolutional network. *Mater Res Proceed* 2024;43:88–95. [\[CrossRef\]](#)
- [7] Sun H, Zhao N, Zeng X, Yan D. Study of solar radiation prediction and modeling of relationships between solar radiation and meteorological variables. *Energy Convers Manag* 2015;105:880–890. [\[CrossRef\]](#)
- [8] Alsina E, Bortolini M, Gamberi M, Regattieri A. Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Convers Manag* 2016;120:320–329. [\[CrossRef\]](#)
- [9] Sur A, Sah RP, Pandya S. Milk storage system for remote areas using solar thermal energy and adsorption cooling. *Mater Today Proc* 2020;28:1764–1770. [\[CrossRef\]](#)
- [10] Pandya S, Sur A, Solke N. Pollution weather prediction system: smart outdoor pollution monitoring and prediction for healthy breathing and living. *Sensors* 2020;20:5448. [\[CrossRef\]](#)
- [11] Pandya S, Sur A, Solke N. COVIDSAVIO: a novel sensor-fusion and deep learning based framework for virus outbreaks. *Front Public Health* 2021;9:797808. [\[CrossRef\]](#)
- [12] Pamdya S, Mistry M, Kotecha K, Sur A, Parikh P, Shah K, et al. A novel multicast secure MQTT messaging protocol framework for IoT-related issues. In: *Proc 2nd Int Conf Comput Commun Cyber-Security*. 2021. p. 339–359. [\[CrossRef\]](#)
- [13] Muthuraja S. Effect of charging and discharging process of PCM with paraffin and Al₂O₃ additive subjected to three point temperature locations. *J Ecol Eng* 2022;23:32–42. [\[CrossRef\]](#)
- [14] Kumar KS, Kalos PS, Akhtar MN, Shaik S, Sundara V, Fayaz H, et al. Experimental and theoretical analysis of exhaust manifold by uncoated and coated ceramics (Al₂O₃, TiO₂ and ZrO₂). *Case Stud Therm Eng* 2023;50:103465. [\[CrossRef\]](#)
- [15] Baskaran S, Kumar KS. Drying investigation of coriander seeds in a photovoltaic thermal collector with solar dryer. *Therm Sci* 2023;27:659–668. [\[CrossRef\]](#)
- [16] 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;45:2288148. [\[CrossRef\]](#)
- [17] Naushad RP, SRS K, Edayadulla. The effect of emissions from DI-based sources causing hazardous health effects. In: *Recent advances in mechanical engineering*. 2024. p. 1–13. [\[CrossRef\]](#)
- [18] Kumar KS, Surakasi R, Mohan A, Tharanisrisakthi BT, Muniamuthu S, et al. Emerging trends and global challenges to predict drop in thermal performance of WTG gearbox. *J Therm Eng* 2024;10:657–669. [\[CrossRef\]](#)
- [19] Kumar KS, Muniamuthu S, DR P, Mohan KR. Measurement of temperature flow analysis by condition monitoring system for WTG gear box to evaluate the thermal performance associated with plant load factor. *J Therm Eng* 2022;9:970–978. [\[CrossRef\]](#)
- [20] Dhabale RB, Kurkute VK, Unde SS, Jadhav P, Dhabale VR, Sur A. Multi objective optimizations of friction stir welding process parameters for reinforcement of copper particles in aluminum alloy using Taguchi based grey relational analysis (GRA) and ANOVA. *J Phys Conf Ser* 2023;2604:012001. [\[CrossRef\]](#)

- [21] Raghavendrakumar E, Kamalakar V, Kumar KS. An investigation in temperature data analysis of middle atmospheric variation from SABER satellite. *Nat Environ Pollut Technol* 2024;23:443–450. [\[CrossRef\]](#)
- [22] Zhou Y, Liu Y, Wang D, Liu X, Wang Y. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. *Energy Convers Manag* 2021;235:113960. [\[CrossRef\]](#)
- [23] Rabehi MGA, Lalmi D. Hybrid models for global solar radiation prediction: a case study. *Int J Ambient Energy* 2020;41:31–40. [\[CrossRef\]](#)
- [24] Ramedani Z, Omid M, Keyhani A, Shamshirband S, Khoshnevisan B. Potential of radial basis function based support vector regression for global solar radiation prediction. *Renew Sustain Energy Rev* 2014;39:1005–1011. [\[CrossRef\]](#)
- [25] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–1780. [\[CrossRef\]](#)
- [26] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* 2019;323:203–213. [\[CrossRef\]](#)
- [27] Islam MS, Mohandes M, Rehman S. Vertical extrapolation of wind speed using artificial neural network hybrid system. *Neural Comput Appl* 2017;28:2351–2361. [\[CrossRef\]](#)
- [28] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 1963;11:431–441. [\[CrossRef\]](#)
- [29] Chung J, Gulcehre C, Cho K, Bengio Y. Gated feed-back recurrent neural networks. In: *Proc 32nd Int Conf Mach Learn*. 2015. p. 2067–2075.
- [30] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86:2278–2324. [\[CrossRef\]](#)
- [31] Lawal A, Alhems L. Wind speed prediction using hybrid 1D CNN and BLSTM network. *IEEE Access* 2021;PP:1. [\[CrossRef\]](#)
- [32] Al-Shaikhi A, Nuha H, Lawal A, Mohandes M. Vertical wind profile estimation using hybrid convolutional neural networks and bidirectional long short-term memory. *Arab J Sci Eng* 2023;13:1213. [\[CrossRef\]](#)
- [33] MathWorks. *traind* - gradient descent backpropagation. Available at: <https://www.mathworks.com/help/deeplearning/ref/traind.html> Accessed on Oct 18, 2024.
- [34] Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *J Math Psychol* 2018;85:1–16. [\[CrossRef\]](#)
- [35] Nuha H, Mohandes M, Rehman S, Al-Shaikhi A. Vertical wind speed extrapolation using regularized extreme learning machine. *FME Trans* 2022;50:412–421. [\[CrossRef\]](#)
- [36] Huang DMSJQ, Wang D. A simple method for partitioning total solar radiation into diffuse/direct components in the United States. *Int J Green Energy* 2018;15:497–506. [\[CrossRef\]](#)
- [37] Ahamed S, Tanino K, Guo H. Cloud cover-based models for estimation of global solar radiation: a review and case study. *Int J Green Energy* 2021;19. [\[CrossRef\]](#)
- [38] Rehman S, Nuha HH, Al-Shaikhi A, Akbar S, Mohandes M. Improving performance of recurrent neural networks using simulated annealing for vertical wind speed estimation. *Energy Eng* 2023;120:775–789. [\[CrossRef\]](#)
- [39] Mohandes M, Nuha H, Islam MS, Schulze FH. Accuracy of wind speed predictability with heights using recurrent neural networks. *FME Trans* 2021;49:908–918. [\[CrossRef\]](#)
- [40] Mohandes M, Rehman S, Nuha H, Schulze FH. Predictability of wind speed with heights using recurrent neural networks. In: *Proc 4th Int Symp Adv Electr Commun Technol* 2021. p. 1–4. [\[CrossRef\]](#)
- [41] Nuha H, Pahlevi R, Mohandes M, Rehman S, Al-Shaikhi A, Tella H. Vertical wind speed estimation using generalized additive model (GAM) for regression. In: *Proc CICN*. 2022. p. 691–695. [\[CrossRef\]](#)



Research Article

Dynamic water monitoring and management system using labview and machine learning

Rupali DESHMUKH¹, Rhea GOMEZ¹, Pramod GARHWAL¹, Disha PATIL¹,
Nilaj DESHMUKH^{2,*}

¹Department of Information Technology, Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai 400703, India

²Department of Mechanical Engineering, Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai 400703, India

ARTICLE INFO

Article history

Received: 18 March 2024

Revised: 13 May 2024

Accepted: 29 July 2024

Keywords:

LabVIEW; Machine learning;

Water management; Water

Supply; Water Level

ABSTRACT

Water plays a pivotal role in everyone's life. Water is necessary for all life to exist and grow. However, people frequently take its availability for granted, despite the fact that water is essential to existence. There are several problems regarding water that is concerning and needs to be addressed. In a residential area, the primary issues faced include water shortage, wastage of water, supply of poor quality water, ineffective water monitoring system, etc. All of this leads to poor water management and thus the residents suffer. Thus it's critical to develop an effective water monitoring and management system for proper water management. In this paper, the problem is addressed and solution is provided through dynamic water management system using LabVIEW and machine learning. the system offers predictive capabilities that allow users to estimate the total duration till the current volume of water is completely exhausted (available hours of water supply) and total duration till the tank refills completely (tank refill duration) based on the current water level in a tank. The main aim is to build a system that not only minimizes wastage of water but also enhances the overall effectiveness of water management. The data set was generated using LabVIEW software which simulated a water management system similar to the real world environment. Linear Regression machine learning algorithm was selected and applied to the data set for prediction of the required variables and the result obtained has 94.56% and 99.24% accuracy for Available Hours of Water Supply and Tank Refill Duration respectively. This system not only minimizes wastage of water but also enhances the overall effectiveness of water management.

Cite this article as: Deshmukh R, Gomez R, Garhwal P, Patil D, Deshmukh N. Dynamic water monitoring and management system using labview and machine learning. Sigma J Eng Nat Sci 2025;43(5):1737–1750.

INTRODUCTION

Water is supplied to the housing societies by the Municipality every day. This water is directly stored in the

underground tanks for storage purposes. At proper intervals, water is pumped to the overhead tanks of respective buildings to provide water supply to individual homes. Since the level of water in the tanks is not monitored, there

*Corresponding author.

*E-mail address: nilaj.deshmukh@fcr.it.ac.in

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



is a tendency for overflow or underflow of water, resulting in wastage and other water management problems. Also, often times, due to a water shortage, there are situations where less or no water is supplied by the Municipality on certain days. Thus, it becomes crucial to have knowledge of how many hours of water will be available based on the current water level or amount. This piece of information also remains important in terms of pumping water to overhead tanks so as to not exhaust the current volume of water carelessly. Therefore, there is a requirement for constant water level monitoring and management and predicting the availability of water supply in hours based on a certain water level. In this paper, the ultimate focus is on dynamically monitor and manage the water level in tanks and predicting the available hours of water supply along with tank refill time with the help of data taken from LabVIEW and implementing it with Machine learning. LabVIEW is an environment that simplifies hardware integration for engineering applications so that you can simulate your hardware to acquire data from NI and third-party hardware. The data obtained from this software can be treated as coincidental to the data obtained from an original hardware system. This software helps greatly in monitoring the current water level of individual tanks, thus preventing overflow or underflow of resources. Also, the data regarding water levels obtained from the simulation contributes largely to the prediction of available hours of water supply and tank refill time with the help of machine learning.

Different Water Management Systems, people are using for effective use of water. Durga et al. [1] developed by the system by identifying water flow in the main pipe line using a water flow sensor, informing the Arduino. If power is available, the system checks the water level in the tank using an ultrasonic sensor. If power is not available, the user is informed via SMS. If the water level is not up to MAX, the system triggers a motor to fill the tank. The incoming water line and power supply are monitored and alerted by the system, if the supply is stopped. Data collected in the cloud helps analyze water consumption and take necessary actions for future use. Johari et al. [2] developed the system for Students' hostels and observed, power failures can cause water levels to decrease, especially during weekends and public holidays. To prevent unexpected water shortages, a monitoring system is developed to alert the technician or person in charge. The system consists of water level detector circuitry integrated with a GSM module, sending SMS for further action. Rao et al. [3] design and developed automatic water level controllers to control the motor, ensuring a constant water reserve in the storage tank. The developed system automatically fill the overhead tank, monitor water levels, and start the motor when the water level drops below a certain level. They also shut off when sump water runs out before filling the upper tank or if the pump is dry. The system is versatile and can control multiple pumps or tanks. The carbon water level sensor provides contact water level measurement. The automatic system operates the water pump without user interaction,

eliminating the need for manual pumping. The system can be implemented at home or college using less expensive components. Automatic controls on water levels can minimize energy use, reduce water and energy loss, and prevent motor drying out, ensuring its longevity.

The systems developed by Ahmed et al. [4], and Herath et al. [5], were based on machine learning which includes various algorithms that enables machine to make accurate predictions for particular tasks. There are two main types of Machine Learning: Supervised and Unsupervised learning. Supervised learning includes two categories: Classification and Regression. The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y). Unsupervised learning focuses on grouping the objects based on the similarities, differences etc. Popular Machine Learning algorithms are Linear Regression, Convolutional Neural Networks(CNN), SVM etc. In the beginning, Prophet algorithm was selected to train the machine with the obtained data set. But the main drawback was that it required large datasets for training because of which ARIMA model was then selected. The same problem continued with ARIMA. Therefore, Linear Regression was finalized as it works well with the available datasets and produces the required result. The main aim of linear regression is to minimize the gap between predicted values and observed values.

Currently, there are several water controller systems [6], [7], [8], already existing in the market but most of them have some limitations in practice making the system inefficient. The main limitations of the existing systems are

- i. Existing systems lack the capability to calculate and forecast the Available Hours of Water Supply accurately.
- ii. They also lack the capability to calculate and forecast the Tank Refill Duration.
- iii. The capability to predict future water availability accurately is absent in these systems.
- iv. All the systems have complex interfaces that limits the accessibility for border use bases.
- v. Most of the system have manual monitoring and intervention, leading to delays and potential issues.
- vi. All the systems work on the fixed flow rates, leading to inefficient resource utilization.
- vii. Existing systems lack adaptability and struggle to cope with changing demands or unexpected events.

Hence, an attempt to overcome these limitations and increase the efficiency of the system was made by implementing a dynamic water monitoring and management system using LabVIEW and machine learning. To address these limitations, the system was designed and developed such that, given the level of water in the water tank, the hours of water availability can be predicted. The system is also developed to predict the duration for complete refilling of the tank. A simple user interface is created such that users can gain access to the water management system in an effortless fashion. Variable flow rates are supported in the system, thus taking into consideration the various scenarios and conditions

of water management and supply. With the help of Machine Learning, adapting to the changing trends and unexpected demands or events of the system can be fulfilled.

MATERIALS AND METHODS

Data Generation Using Labview

The main aim was to develop a system which can predict the future water availability accurately. For this purpose, the system makes use of :

- LabVIEW - to create the required data set.
- Machine Learning Algorithms - to make proper predictions based on the obtained data set from LabVIEW.

In order for the machine to be able to estimate the Available Hours of Water Supply and Tank Refill Duration prediction based on the water level in a tank, a dataset is required. However, a pre-defined dataset related to the system was not available on the internet. For this purpose, LabVIEW software was used to simulate a similar water management system and thus generate the required dataset for different conditions. Once the dataset was ready, Machine Learning algorithm was used to train the machine and predict the Available Hours of Water Supply and Tank Refill Duration based on the water level in a tank.

LabVIEW stands for Laboratory Virtual Instrument Engineering Workbench. It is a visual programming language and development environment utilized for creating applications in measurement, automation, and control systems [9]. It enables users to design programs graphically by connecting virtual instruments, facilitating tasks like data acquisition, analysis, and instrument control across various engineering and scientific domains. LabVIEW consists of two primary panels: Front panel and Back Panel. Front

panels are used as the user interface on which controls and indicators are constructed. Back Panels are used to develop graphical code by representing program logic with nodes and wires. Icons in the front panel are connected to the nodes in the back panel using wires. If the changes are made in front panel correspondingly it gets updated in the back panel.

Labview Front Panel

Fig 1 shows LabVIEW front panel structure. It includes:- [Text Wrapping Break]Two tanks - underground tank and an overhead tank. Water is supplied to the underground tank from Municipality. The connection is structured such that the outlet of the underground tank is connected to the inlet of the overhead tank. Finally, the outlet of the overhead tank is connected to the households.

There are flow-in and flow-out controllers which are used to control the flow-in and flow-out rates of the water. There are total 4 controllers (2 for underground tank, 2 for overhead tanks) and 4 indicators to show the current readings of flow rates. There are current water level indicators which show the current water level in the tanks. Cut-off activation indicators are colour indicators used for automating water supply cut-offs based on critical threshold values. For underground tank, if the water level goes below 1000 Litres, the lower indicator turns red and if the water goes above the 9000 Litres, then upper indicator will turn red otherwise it remains green. Similarly for the overhead tank, the minimum critical level is 100 Litres and the maximum critical level is 900 Litres. If the lower indicator turns red, it will automatically stop the water supply from the underground tank to overhead tank and if the upper indicator turns red it will halt the water supply. Further, there are two forecast indicators which show the Available Hours of Water Supply and Tank Refill Duration based on the water level in a tank.

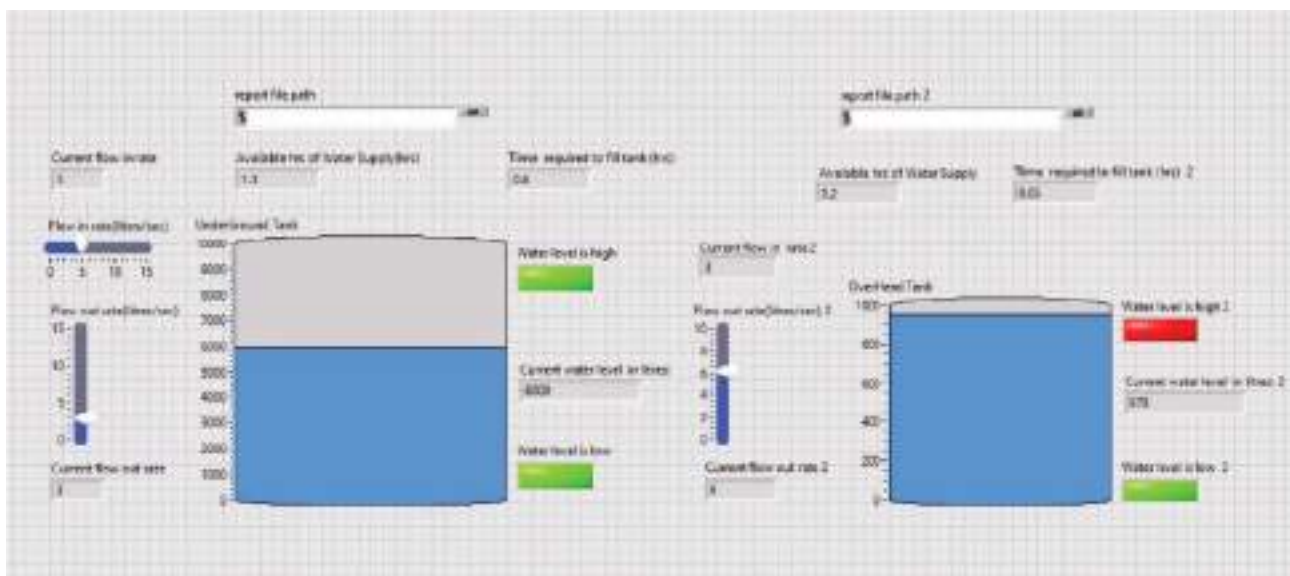


Figure 1. Front panel structure of LabVIEW.

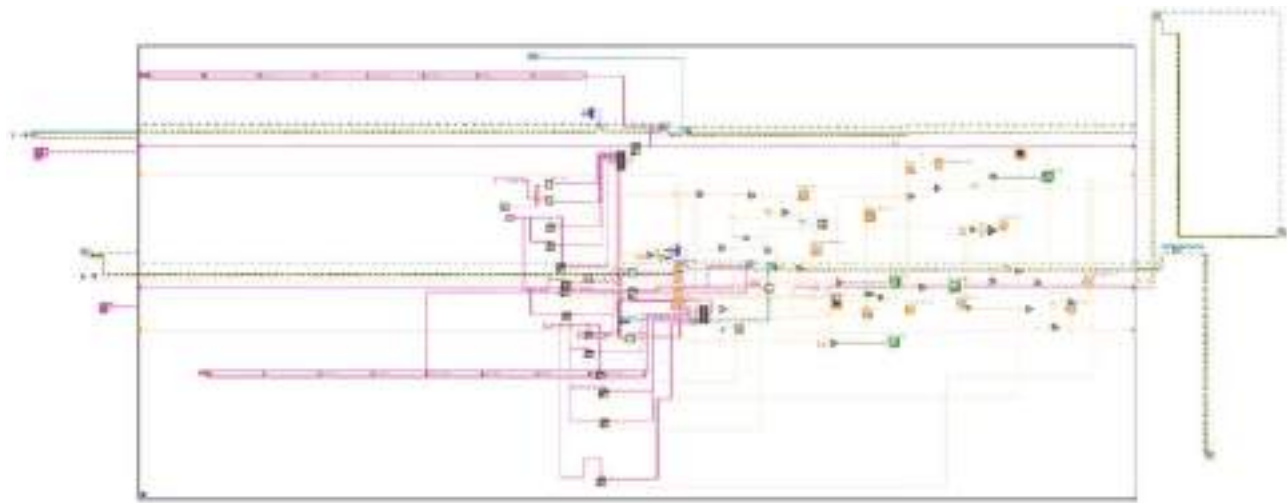


Figure 2. Architectural block diagram of LabVIEW.

Labview Block Diagram

Fig. 2 shows back panel structure. It includes, entire logical representation of the icons, controllers etc. placed in the front panel in the form of Block Diagram. The output nodes are connected to the Front Panel indicators and wires link the Front Panel controls to the appropriate nodes on the Block Diagram. When any value updates on the front panel, it is sent through the wire to the node in the back panel.

BACK PANEL BLOCK DIAGRAM TOOLS

Tools used for developing dynamic water controller system. Figure 3 shows the various components/tools used in LabVIEW to generate the front panel of the system.

Figure 3 (a) shows the While loop component. The code within the while loop is repeated until a specified condition is met. A While Loop always runs at least once. For continuous execution, the entire code is placed in a while loop. Figure 3 (b) shows Flow rate controllers for varying the rate of flow of water into and out of the tanks. Flow

rate controllers are horizontal and vertical pointer sliders. Sliders are used in the software to change numerical values. The tank component is shown in Figure 3 (c). To simulate the water tanks and display the numeric level, the tank function is used for developing program. The flow rate controls and current water level monitor are attached to the underground and overhead tanks in order to monitor and control the water levels in the tank.

Figure 4 (a) shows a numeric component. It functions as a numeric operator and is used to execute basic operations on numeric values, numerical functions are employed. Figure 4 (b) shows a Feedback Node. It acts as a memory register and stores data from one VI execution or loop iteration to the next. Figure 4(c) shows an In range component. The main function of this component is to determines whether x is inside the range defined by the upper and lower limits, and optionally coerces the value to be within the range. The underground tank's range is fixed to be at 0-10,000 and the overhead tank's range at 0-1000. Figure 4(d) shows a Wait component. It waits the specified number of milliseconds and returns the value of the millisecond timer.

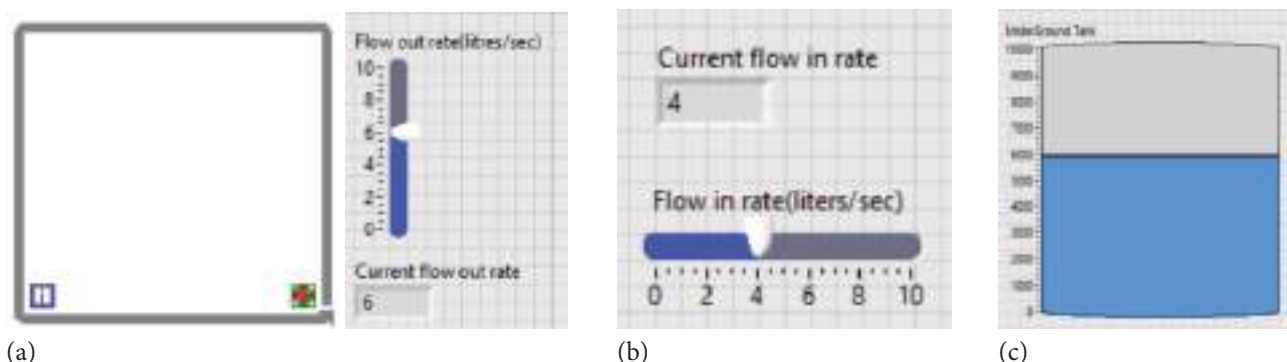


Figure 3. (a) While loop (b) Flow rate controllers (c) Tanks.

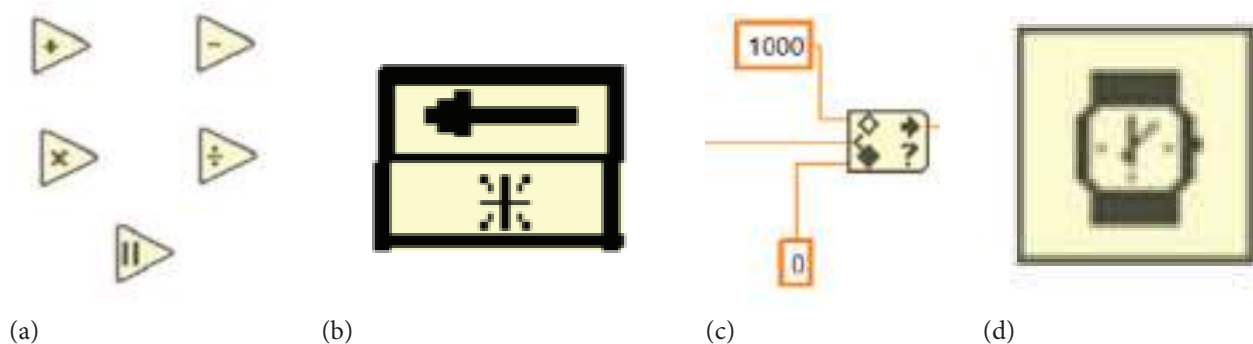


Figure 4. - (a) Numeric (b) Feedback node (c) In range (d) Wait.



Figure 5. Numeric comparator.

Figure 5 shows a numeric comparator component. This component also functions like an operator that is used to compare the numeric values in the given program. This component helps in setting conditions or threshold values for the upper and lower bound of cut-off activation.

Figure 6(a) shows a Build Array component. Build Array creates an array from one or more elements or arrays. Build array is used to integrate multiple parameters such as flow in, flow out, time, date, current water

level, cutoff activation, and so on into a single array and export the results to an Excel sheet. Figure 6(b) shows an Insert into Array component. The insert into array function inserts one or more elements or subarrays into an array. Figure 6(c) shows a Get Date/Time String component. This function enables to obtain the current time stamp for the dataset. Accurate date and time stamp will be of great help while training the machine with the acquired data values.

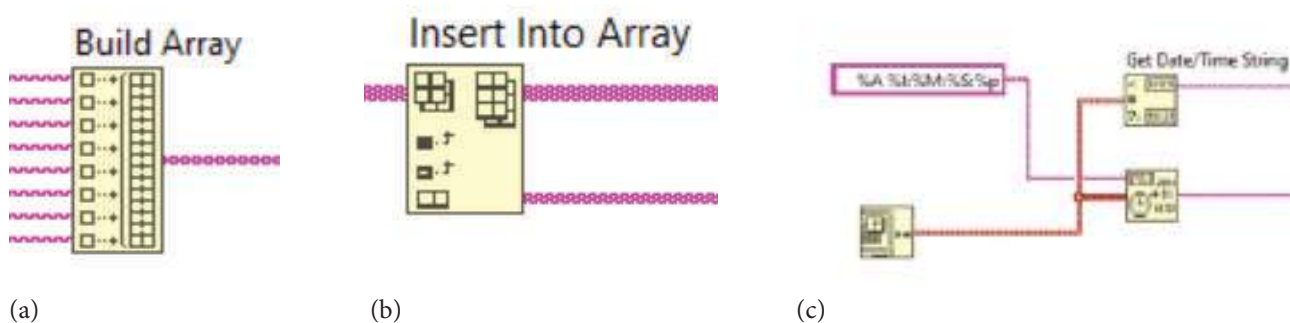


Figure 6. (a) Build array (b) Insert into array (c) Get date/time string.

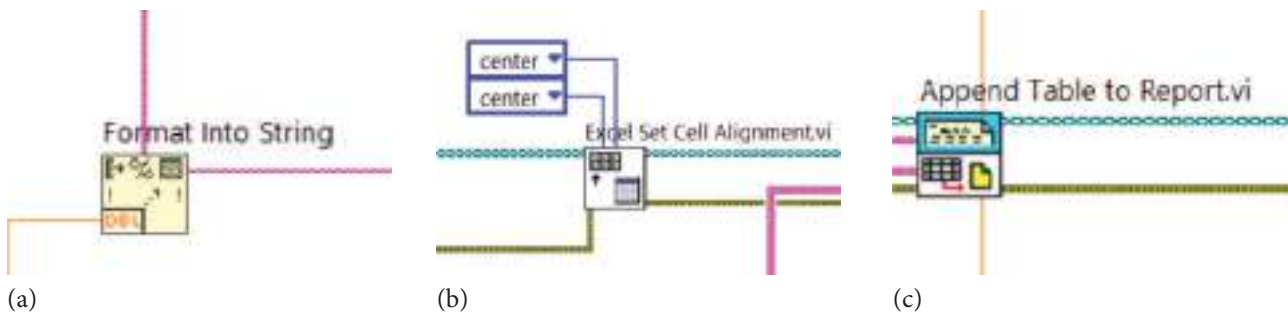


Figure 7. (a) Format to String (b) Excel Set Cell Alignment (c) Append Table to Report.

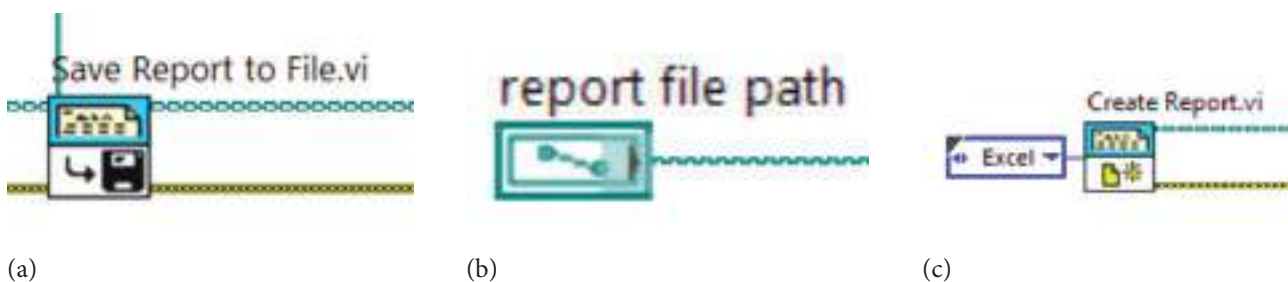


Figure 8. (a) Save Report to File (b) Report File Path (c) Create Report.

Figure 7 (a) shows Format to String component of LabVIEW. The Format to String describes how the function should convert the input parameters into the resulting string. Figure 7 (b) shows an Excel Set Cell Alignment component. Excel Set Cell Format VI takes in a reference to the report and sets format for the cells, based on the input given to the number format terminal. Figure 7 (c) shows an Append Table to Report component. This function enables to add rows and columns in the report.

Figure 8 (a) shows Save Report to File component. This is used to save report to excel sheet and save it in local drive. Figure 8 (b) shows a Report File Path component. It is used to give the path to the file to store it in local storage. Figure 8 (c) shows a Create Report component of LabVIEW. Create report function is used to create a excel report.

MACHINE LEARNING AND REGRESSION ANALYSIS

Developing an efficient water monitoring and management system requires the use of advanced predictive models to ensure accurate predictions of water supply availability and recharge times. In the initial stages of this research, examined various machine learning models to determine the most appropriate approach to analyzing data obtained from LabVIEW simulations [10].

Initial Model: Prophet

The Prophet model was initially selected due to its effectiveness in time series forecasting. Facebook created

Prophet, a powerful forecasting tool that can handle daily observations of time series data with potential seasonal effects, holiday patterns, and other temporal structures. The system is designed to handle huge amounts of data to predict precise projections using LabVIEW data for training to train the system. Prophet, has particular way of functioning and difficult to customize as per the requirements of this project. Due to this use of Prophet is inappropriate for the changing needs which is a main focus of this study.

Transition to Arima

To accomplished the requirement of changing demands the ARIMA (AutoRegressive Integrated Moving Average) model is used. ARIMA is useful and effective in time series forecasting and it has capacity to handle non-stationary data by differencing to make it stationary. This model demonstrated high accuracy in predicting values and interpolating missing data points. But, ARIMA model requires a large dataset to minimize errors and improve prediction accuracy [11-15].

Adoption of Linear Regression

Due to the limitations of Prophet and ARIMA, decided to use linear regression. This is simple and powerful machine learning model to fits the available dataset properly. In this models the linear relationship between a dependent variable (target) and one or more independent variables (features) is established. The linear regression is more simple and effective than Prophet and ARIMA models.

Mathematical Representation of Linear Regression:

The fundamental equation of Linear Regression [16, 17] is represented as:

$$Y = mX + C \quad (1)$$

Where,

Y is the dependent variable and it can be available Hours of Water Supply or Tank Refill Duration.

X is the independent variable and it can be current water level, flow in rate, flow out rate.

m is the slope of the line and represent the relationship between the dependent and independent variables.

C is the y-intercept and represent the value of y when x is zero.

Cost Function

The performance of the Linear Regression model is evaluated using the Cost Function, specifically Mean Squared Error (MSE). The average squared difference between the actual and predicted values is given by Cost Function, which help to quantify the model accuracy [18-20].

Mean Squared Error is given by:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - (mX_i + c))^2 \quad (2)$$

Where,

N represent number of data points.

Y_i represent actual value of the dependent variable for the ith data point.

X_i represent the value of the independent variable for the ith data point.

m and c represent slope and intercept of the model respectively.

Linear Regression is used to find the values of m and c which helps to minimize the Mean Square Error. It helps to reduce the gap between predicted and observed values. Deep Learning can be used in latest water level implementation techniques for effective implementation [21, 22].

Implementation of Water Management System

The dataset generated by LabVIEW is used to trained The Linear Regression model and use to predict two critical parameters:

A. Available Hours of Water Supply: Equation 3 is used for predicting the number of hours the current water level will last based on the flow out rate.

B. Tank Refill Duration: Equation 4 is used for predicting the time required to refill the tank to its full capacity based on the flow in rate and the current water level.

These predictive capabilities into the water management system help, the model to provides real-time insights, enabling proactive management of water resources. This approach helps to minimizes water wastage and also ensures a consistent and reliable water supply.

DEVELOPED SYSTEM

The system works as follows:

- i. Flow rate controllers set the flow in and flow out rates. The flow out rate from the underground tank is always equal to the flow in rate of the overhead tank.
- ii. Depending on the flow rates, the water level in the tank increases or decreases. The current water level is displayed on an indicator in the front panel.
- iii. If the current water level falls below a predefined threshold, the lower indicator turns red; otherwise, it remains green.
- iv. The system calculates the projected available hours of water supply based on flow rates and the current water level.
- v. Users can inquire about the estimated tank refill duration for the tank to be full based on the current water level.
- vi. Mathematical formulae used for the forecasting of Available Hours of water supply and Tank refill duration are as follows:

$$\text{Available Hours of Water Supply (hrs)} = \frac{\text{Current water level(l)}(\text{height from bottom})}{\text{Flow out rate}(\frac{\text{l}}{\text{s}})} \times \frac{1}{3600} \quad (3)$$

In equation 3 the parameters used are as follow.

- A. Current Water Level (l):** This parameter shows how much water is in the tank right now, expressed in litres. It is important because it establishes how much water is still usable.
- B. Flow Out Rate (l/s):** This indicates, how quickly water is being used up or drained from the tank, in litres/second.
- C. Conversion Factor (1/3600):** This factor is used to converts the time hours from second.

By using Equation 3 the number of hours the water supply will last based on the current water level and the rate at which water is being used is calculated. For example, if the tank has 10,000 litres of water and the flow out rate is 2 litres per second, the calculation helps in determining how long the water will last before the tank is empty. This information is important for managing water usage and during periods of shortage.

$$\text{Tank Refill Duration (hrs)} = \frac{\text{Total height of tank(l)} - \text{current water level (l)}}{\text{Flow in rate}(\frac{\text{l}}{\text{s}})} \times \frac{1}{3600} \quad (4)$$

Equation 4 can be used to calculate the time required to fill a tank from its current level to its maximum capacity. For example, if a tank has a capacity of 5,000 liters and currently holds 1,000 liters at a flow rate of 3 liters per second, this formula will help determine time required to fill the remaining 4000 liters in the tank. This estimation is required to plan activity of filling and stoping water for refilling and avoiding overflow of tank respectively.

- vii. Readings of flow in rate, flow out rate, current water level, available hours of water supply, and tank refilled

Date	Time	Flow out rate (l/s)	Flow in rate (l/s)	Current water level(l)	Cutoff Activation	Avail. hrs. of water supply	Time required to fill the tank(hrs)
30-10-2023	Monday 11:41:37 AM	0	2.12	0	1	0	1.31
30-10-2023	Monday 11:41:38 AM	0	2.12	2.12	1	0	1.31
30-10-2023	Monday 11:41:39 AM	0	2.12	4.24	1	0	1.31
30-10-2023	Monday 11:41:40 AM	0	2.12	6.36	1	0	1.31
30-10-2023	Monday 11:41:41 AM	0	2.12	8.47	1	0	1.31
30-10-2023	Monday 11:41:42 AM	0	2.12	10.59	1	0	1.31
30-10-2023	Monday 11:41:43 AM	0	2.12	12.71	1	0	1.31
30-10-2023	Monday 11:41:44 AM	0	2.12	14.82	1	0	1.31
30-10-2023	Monday 11:41:45 AM	0	2.12	16.95	1	0	1.31
30-10-2023	Monday 11:41:46 AM	0	2.12	19.07	1	0	1.31
30-10-2023	Monday 11:41:47 AM	0	2.12	21.19	1	0	1.31
30-10-2023	Monday 11:41:48 AM	0	2.12	23.32	1	0	1.31
30-10-2023	Monday 11:41:49 AM	0	2.12	25.42	1	0	1.31
30-10-2023	Monday 11:41:50 AM	0	2.12	27.54	1	0	1.31
30-10-2023	Monday 11:41:51 AM	0	2.12	29.66	1	0	1.31

Figure 9. Dataset obtained from LabVIEW simulation.

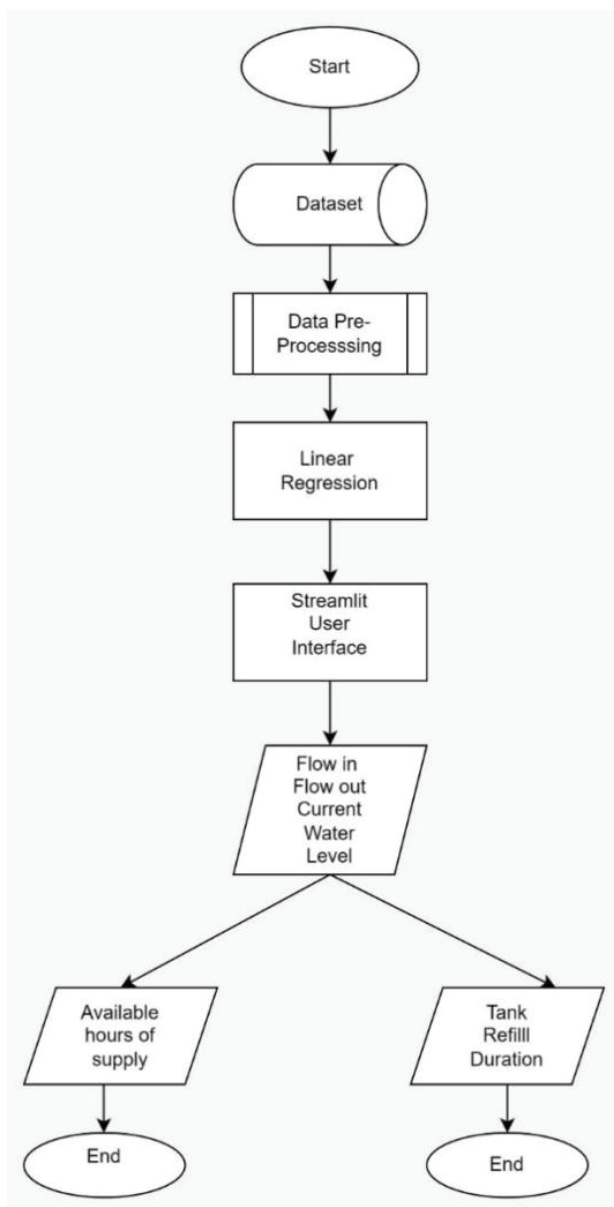


Figure 10. Flow chart of machine learning algorithm (linear Regression).

time, are continuously logged and shown in Figure 9 for obtained data set.

viii. The logged data is transferred to Excel sheets at regular intervals, creating a comprehensive dataset. Sample dataset is as follows:

- ix. The dataset serves as training data for a machine learning model, particularly using the linear regression algorithm as shown in figure 10.
- x. The trained model predicts values such as Available Hours of Water Supply and Tank Refill Duration based on new input data.
- xi. The generated predictions, along with other readings, are logged back into Excel sheets for generating record of daily or weekly patterns.

This process forms a closed loop, continually updating and refining predictions based on real-time data and historical patterns.

RESULTS AND DISCUSSION

A graph is generated that correlates water, volume and time using LabVIEW. This graph helps to predict water consumption patterns by identifying peak periods, facilitating improved resource planning, and forecasting future water availability most effective way. Figure 11 shows a line graph that depicts the trend of Current water level (l) for different Timestamp.

Figures 11 and 12 illustrates fluctuations in the water levels of underground and overhead tanks for different time. This helps to provide information about usage patterns and identifying peak periods of water requirement.

Thus, Figure 13 and 14 helps to analyze the efficiency of the water flow control system and how the system manages the balance between water inflow and outflow. Initially, a certain water flow-in rate and flow-out rate is provided to the underground tank. Once the underground tank is full, the flow-in rate is reduced to nearly zero with the constant flow-out rate. Since the system is developed such that the water flow-out rate from the underground tank becomes the flow-in rate for the water flowing in the overhead tank,

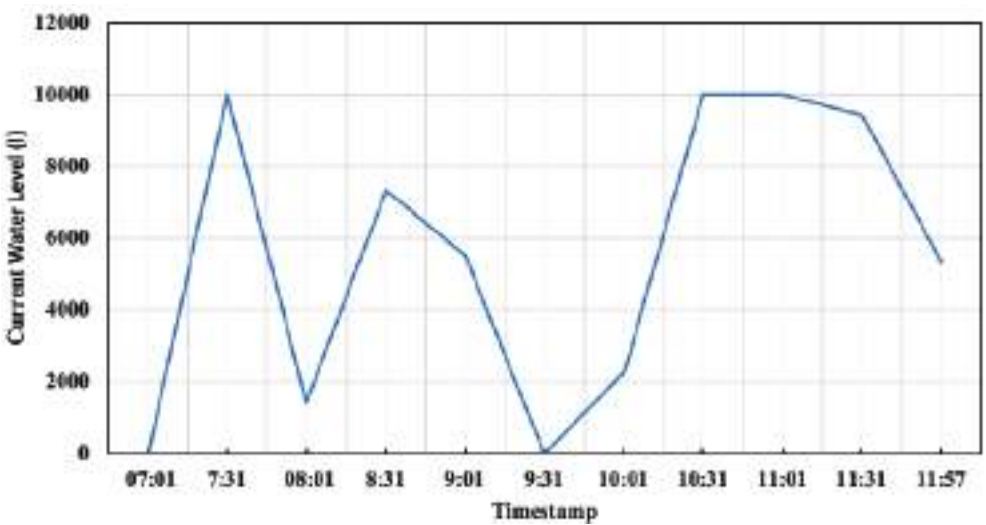


Figure 11. Current water level (liters) in underground tank for different timestamps on tuesday evening.

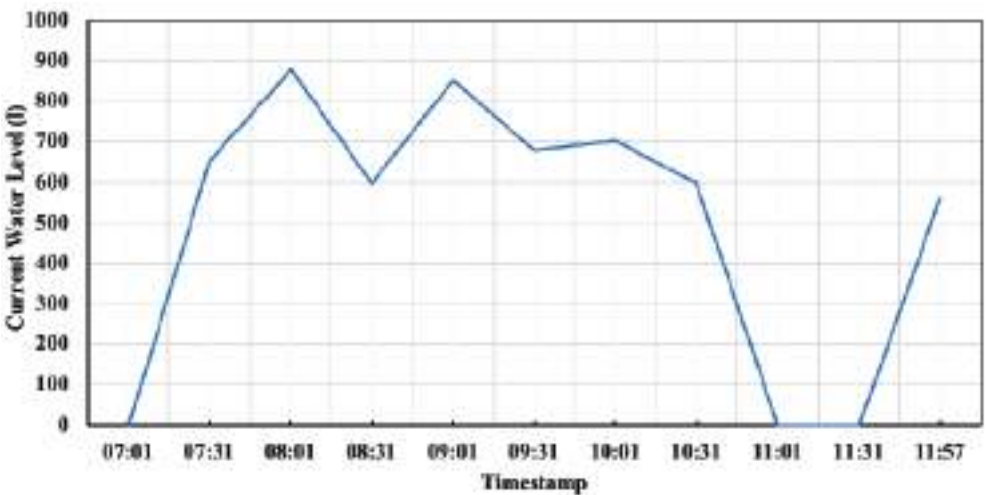


Figure 12. Current water level (liters) in overhead tank for different timestamp on tuesday evening.

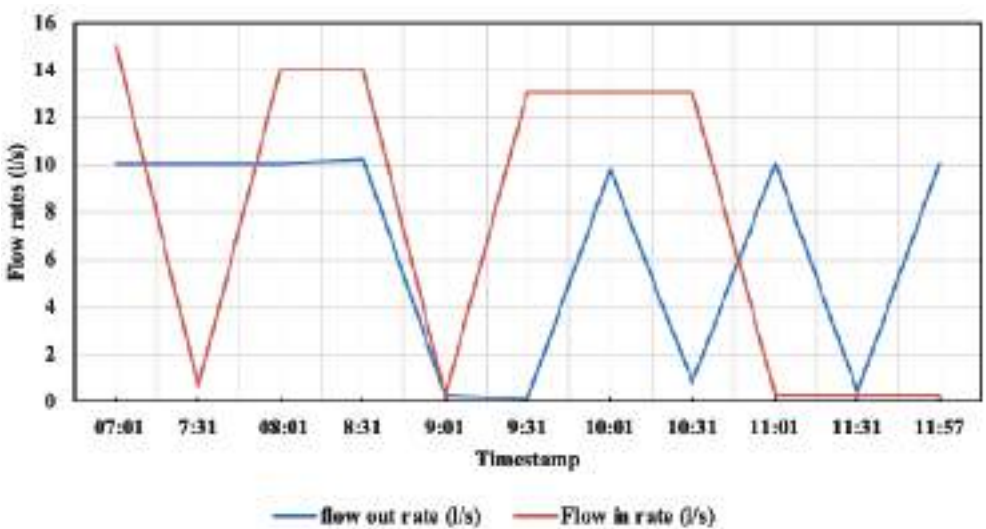


Figure 13. Flow-in and flow-out rates for underground tank for different timestamp on tuesday evening.

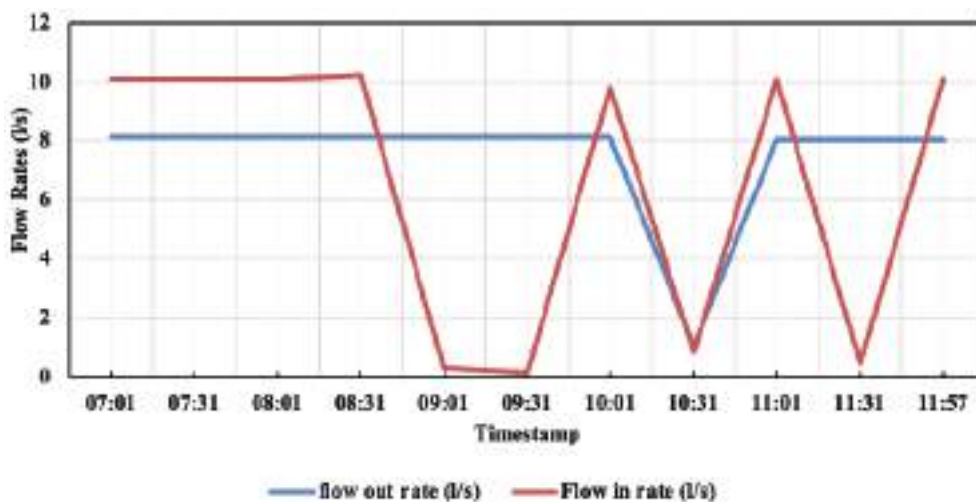


Figure 14. Flow-in and flow-out rates for overhead tank for different timestamp on tuesday evening.

there is a peak in the flow-in rates of the overhead tank in the Fig. 14 at the point at which the flow-out rates for the underground tank in the Fig. 13 is incremented. The flow-out rate of the underground tank is kept at a constant fluctuating rate between nearing 0 and 10.03 l/s considering that water will be pulled out of the tank with the help of a motor and thus will remain constant. The flow-out rate for the overhead tank is kept at constant keeping in mind that water flows out from the overhead tank with the help of gravity and thus, will remain constant all the time.

Figure 15 and 16 helps with water distribution schedule planning and prediction by showing the change in water supply hours over time. In Fig. 15, the trend of the available hours of water supply is similar to that of current water level as the more the water level, the more is the available hours

of water supply. It highly depends on the current water level and the flow-out rate of the respective tank. Figure 16 also shows a similar trend.

Figure 17 and 18 illustrates how long it takes to refill the tanks and provides information on how well the system works to restore water levels. At a given instant of time, for the underground tank, the flow-in and flow-out rates both are provided leading to the depiction of the above mentioned line graph. In this graph, the tank refill duration is highly dependent on the current water level and the flow-in rate of the respective tanks. Line graphs are chosen for their suitability in depicting continuous variation over time. These graphs offer a thorough understanding of the operation of the water supply system, supporting measures for conservation and improvement.

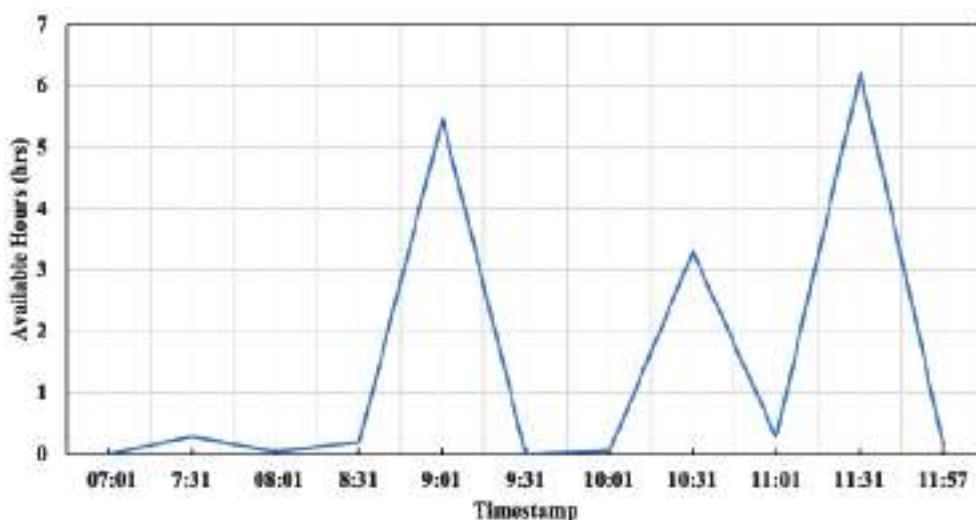


Figure 15. Available hours of supply (hrs) for different timestamp in underground tank on tuesday evening.

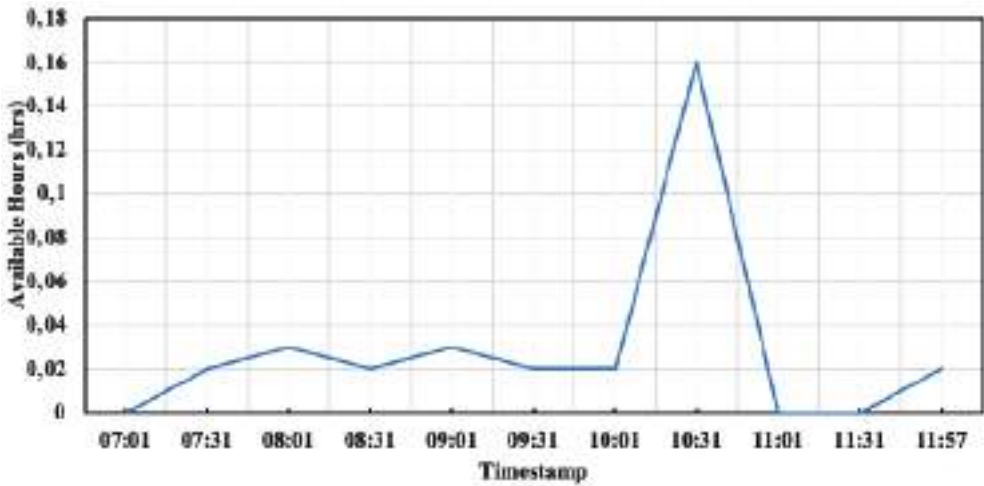


Figure 16. Available hours of supply (hrs) for different timestamp in overhead tank on Tuesday evening.

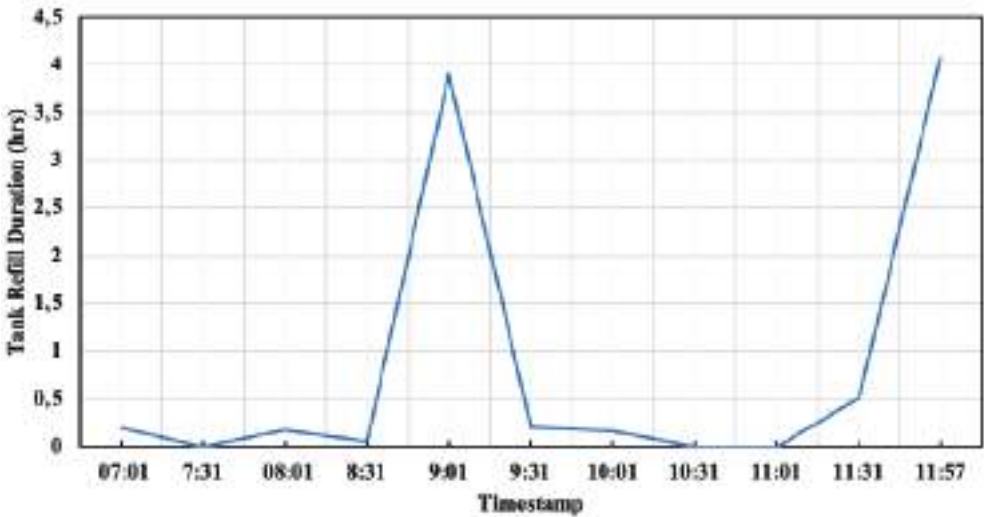


Figure 17. Tank Refill Duration (hrs) of underground tank for different timestamp on Tuesday evening.

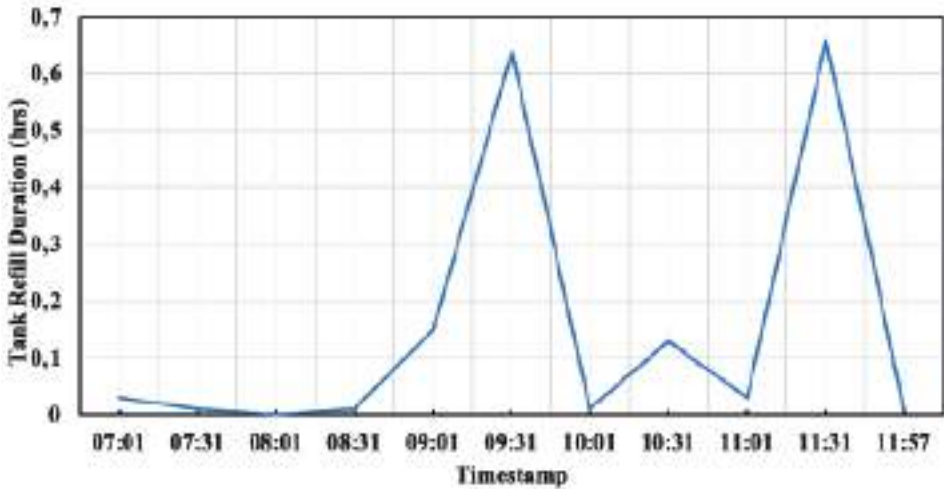


Figure 18. Tank refill duration (hrs) of overhead tank for different timestamps on Tuesday evening.

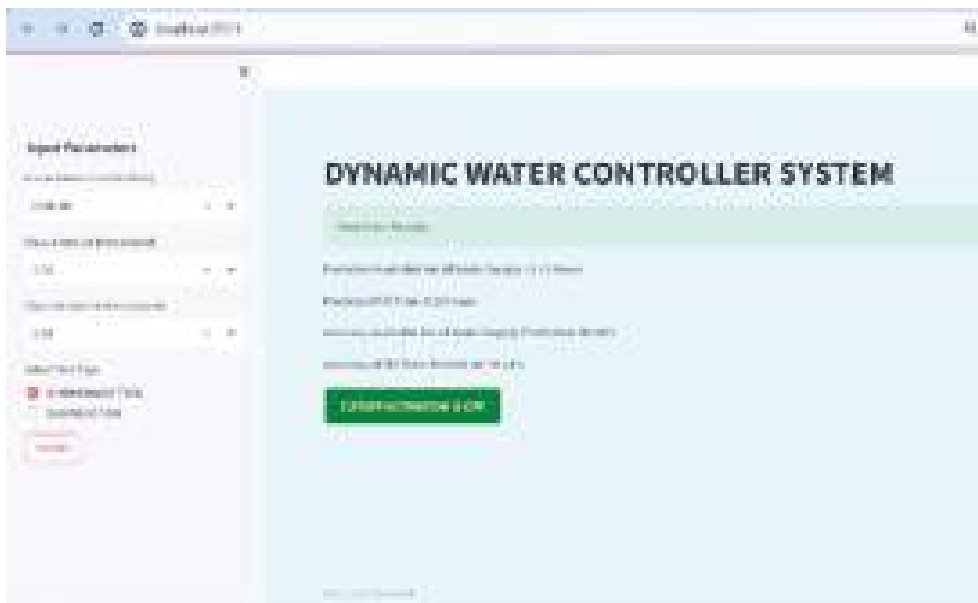


Figure 19. User interface for decision making.

The linear regression model was trained from the dataset obtained from LabVIEW. After that, a User Interface using streamlit was generated as shown in figure 18, which takes flow in, flow out rate, and current water level as input. Streamlit is an open-source Python Library that allows you to create web applications in an efficient manner for Machine Learning and Data Science projects. It allows you to visualize the data and also make the system more user friendly with its simple and easy user interface. It can be generated easily using familiar python syntax and scripts. With the provided inputs from LabVIEW and by using linear regression it will predict Available hours of water supply and Tank refill duration which will further help us to efficiently control water usage and save water from being wasted.

With the help of the obtained Dataset, the LabVIEW and Machine Learning results were analyzed. It was found that LabVIEW estimated values and Machine Learning predicted values are very close.

- i. **Dependence on Accurate Data Simulation:** The LabVIEW-simulated data is significantly used for developing the model. Though LabVIEW offers a robust structure for developing realistic datasets, it may affect accuracy of prediction in actual circumstances. The forecasts accuracy is depends on any differences between the simulated and real situations data sets.
- ii. **Limited Size of Dataset:** - Large size datasets are required to enhanced machine learning model performance, particularly for regression analysis models. In spite of extensive, the dataset created by LabVIEW for this project may be have limitations. This may have restriction on the model's capacity to generalize and

make correct predictions in a different scenario. For limited data set Linear Regression is the best option compare to sophisticated models like PROPHET and ARIMA.

- iii. **Real-time Data Integration:** - The machine learning models is trained by data produced by LabVIEW. While integrating real-time data from actual water management systems, technical difficulties could arise such as data transfer delay, contineous updatation, and possible differences between real-time and historical data. It implies that certain hardware and software setups are required when using LabVIEW as the main tool for data production for training model. The developd system's capacity to scale and adapt to other contexts or systems that do not support LabVIEW may be limited by this dependence.
- iv. **External and Environmental Factors:** The developed system's prediction algorithms may not consider all other factors that could affect water levels and flow rates, like variations in weather patterns, user habits, or unforeseen interruptions to the water supply. This may increase uncertainty and decrease prediction accuracy.
- v. **Simplified Assumptions:** In this dynamics system, assumptions, serve as the foundation for the mathematical formulas used to predict the number of hours available for water delivery and the time needed to refill the tank. In Real-world water systems different tank designs, pressure variations, and sporadic supply disruptions will might behave more complicatedly.
- vi. **Maintenance and Calibration:** Regular maintenance, calibration of sensors and other hardware components are required for the system to sustain its accuracy over

time. Any maintenance gaps will result in inaccurate data collection, which would affect the accuracy of the system's forecasts.

- vii. **Requirement for User Expertise:** - To maintain this developed system, certain amount of expertise with both LabVIEW and machine learning is required. Users without prior experience in these fields may find this requirement difficult which could restrict the system's usability and accessibility.
- viii. **Problems with Scalability:** - The developed model is a scale down model, there may be difficulties when trying to integrate it with other tanks and distribution systems or scale it up to handle larger water networks. For accurate implementation, more complex data management, and the requirement for more advanced prediction models are some of these problems.

CONCLUSION

The developed Dynamic Water Management System, which integrates LabVIEW for dataset collection, a Linear Regression model for forecasting, and a user-friendly interface for real-time interaction, offers an efficient solution for water resource management. LabVIEW is an excellent platform for accurate data collection, while machine learning ensures accuracy in prediction of time requirement to tank refill duration. The intuitive user interface allows easy input of critical parameters such as flow-in, flow-out, and current water level, making the system accessible to various users. An accuracy of 94.56% and 99.24% for predicting Available Hours of Water Supply and Tank Refill Duration respectively using Linear Regression Machine Learning model. For implementing with actual sensors for flow rate and water level detection will enable real-time data acquisition, leading to a dynamic and efficient automated water control system utilizing IoT. This integration highlights the system's scalability and adaptability for larger applications like municipal water systems and large-scale industrial use. The use of machine learning for predictive analytics highlights data-driven decision-making in water management, motivating further research into advanced models. Additionally, optimizing water usage and minimizing waste contributes to sustainable resource management, addressing global environmental challenges. The Dynamic Water Management System satisfies current water management needs and also establishes a new benchmark for intelligent, efficient, and sustainable practices, paving the way for future research and development.

ACKNOWLEDGEMENTS

This work was supported by Fr. C. Rodrigues Institute of Technology, Vashi. The authors would like to thank for this 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] Durga SN, Ramakrishna M, Dayanandam G. Autonomous water tank filling system using IoT. *Int J Comput Sci Eng* 2018;6:1–4. [\[CrossRef\]](#)
- [2] Johari A, Abd Wahab MH, Abdul Latif NS, Ayob ME, Ayob MI, Ayob MA, et al. Tank Water Level Monitoring System using GSM Network. *Int J Comput Sci Inf Technol* 2011;2:1114–1120.
- [3] Rao KR, Srinija S, Bindu KH, Kumar DS. IOT based water level and quality monitoring system in overhead tanks. *Int J Eng Technol* 2018;7:379–383. [\[CrossRef\]](#)
- [4] Ahmed AN, Yafouz A, Brima AH, Kisi O, Huang YF, Sherif M. Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia. *Eng Appl Comput Fluid Mech* 2022;16:422–440. [\[CrossRef\]](#)
- [5] Herath M, Jayathilaka T, Hoshino Y, Rathnayake U. Deep machine learning-based water level prediction model for colombo flood detention area. *Appl Sci* 2023;13. [\[CrossRef\]](#)
- [6] Barbade GM, Chandurkar SN, Shounak VS, Nimkar VR, Patil UB. Automatic water tank filling system with water level Indicator. *Indian J Microprocess Microcontr* 2021;1:1–7. [\[CrossRef\]](#)
- [7] Olisa SC, Asiegbu CN, Olisa JE, Ekengwu BO, Shittu AA, Eze MC. Smart two-tank water quality and level detection system via IoT. *Heliyon* 2021;7. [\[CrossRef\]](#)

- [8] De Souza EV, Da Silva MAC. Management system for improving the efficiency of use water systems water supply. *Procedia Eng* 2014;89:458–466. [\[CrossRef\]](#)
- [9] Higa ML, Tawy DM, Lord SM. An introduction to LabVIEW exercise for an electronics class. *Front Educ Conf* 2002;32:T1D–T1D. [\[CrossRef\]](#)
- [10] Shen J, Valagolam D, McCalla S. Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea. *PeerJ* 2020;8. [\[CrossRef\]](#)
- [11] Atashi V, Gorji HT, Shahabi SM, Kardan R, Lim YH. Water Level Forecasting Using Deep Learning Time-Series Analysis: A Case Study of Red River of the North. *Water* 2022;14. [\[CrossRef\]](#)
- [12] Guarnaccia C, Longobardi A, Mancini S, Viccione G. Drinking water tank level analysis with ARIMA models: A case study. *Environ Sci Proc* 2020;33. [\[CrossRef\]](#)
- [13] Viccione G, Guarnaccia C, Mancini S, Quartieri J. On the use of ARIMA models for short-term water tank levels forecasting. *Water Sci Technol Water Supply* 2020;20:787–799. [\[CrossRef\]](#)
- [14] Mancini S, Francavilla AB, Longobardi A, Viccione G, Guarnaccia C. Predicting daily water tank level fluctuations by using ARIMA model. A case study. *J Phys Conf Ser* 2022;2162:012007. [\[CrossRef\]](#)
- [15] Bata M, Carriveau R, Ting DS-K. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water* 2020;5. [\[CrossRef\]](#)
- [16] Yao W, Li L. A new regression model: Modal linear regression. *Scand J Stat* 2014;41:656–671. [\[CrossRef\]](#)
- [17] Gianey HK, Choudhary R. Comprehensive review on supervised machine learning algorithms. *Proc Int Conf Mach Learn Data Sci* 2017;2017:38–43. [\[CrossRef\]](#)
- [18] Satu MS, Howlader KC, Mahmud M, Kaiser MS, Shariful Islam SM, Quinn JMW, et al. Short-term prediction of covid-19 cases using machine learning models. *Appl Sci* 2021;11. [\[CrossRef\]](#)
- [19] Liu G, Savic D, Fu G. Short-term water demand forecasting using data-centric machine learning approaches. *J Hydroinform* 2023;25:895–911. [\[CrossRef\]](#)
- [20] Bellfield SL. Short-term domestic water demand: estimation, forecasting and management [Master thesis]. Leeds, UK: University of Leeds; 2001.
- [21] Shahabi A, Tahvildari N. A deep-learning model for rapid spatiotemporal prediction of coastal water levels. *Coast Eng* 2024;190:104504. [\[CrossRef\]](#)
- [22] Mendonça F, Mostafa SS, Morgado-Dias F, Azevedo JA, Ravelo-García AG, Navarro-Mesa JL. Noncontact automatic water-level assessment and prediction in an urban water stream channel of a volcanic island using deep learning. *Electronics* 2024;13:1145. [\[CrossRef\]](#)



Research Article

An abstract approach to sequences and series using soft complex numbers

Atiqe UR RAHMAN^{1,*}, Aram NOORI QADIR², Pishtiwan OTHMAN SABIR³

¹Department of Mathematics, University of Management and Technology, Lahore 54000, Pakistan

²Department of Mathematics, College of Education, University of Garmian, Kalar 46021, Iraq

³Department of Mathematics, College of Science, University of Sulaimani, Sulaymaniyah 46001, Iraq

ARTICLE INFO

Article history

Received: 29 January 2024

Revised: 26 March 2024

Accepted: 30 March 2024

Keywords:

Soft Cauchy Sequence; Soft Complex Convergence; Soft Complex Number; Soft Complex Sequence; Soft Complex Series; Soft Set

ABSTRACT

The idea of a soft set offers a consistent structure for combining multiple data types and supporting a range of sources and representations. Because of its adaptability, it can be used in a variety of industries where precise and ambiguous information supervision is crucial for reliable and productive evaluation. This article delves into the interplay between sequences and series within the domain of soft complex numbers. It establishes a thorough understanding of soft complex boundedness, offering precise definitions for soft complex convergent sequences, soft complex limits, and soft complex series. Our exploration lays the groundwork with a broad perspective, paving the way for a detailed examination of specific properties embedded in the Bolzano-Weierstrass theorems. Additionally, we unravel the intricacies of soft complex Cauchy sequences and scrutinize the soft complex limits associated with both convergent sequences and series. This comprehensive discussion sheds light on the nuanced aspects of these mathematical concepts, providing a deeper insight into their implications.

Cite this article as: Ur Rahman A, Noori Qadir A, Othman Sabir P. An abstract approach to sequences and series using soft complex numbers. Sigma J Eng Nat Sci 2025;43(5):1751–1759.

INTRODUCTION

An adaptable scheme for handling ambiguity and uncertainty in a variety of everyday circumstances is provided by fuzzy set [1]. One benefit is that it may depict the differentiation between membership and non-membership, enabling more complex conceptual depictions. Because of its versatility, complex systems with ill-defined constraints can be better modeled. This theory has successfully been applied in various fields of study like differential equations [2–5], and decision-making [6,7]. To equip fuzzy

set with parameterization mode, Molodtsov [8] introduced a groundbreaking mathematical approach known as soft set theory (SST), designed to address uncertainties. It has multiple applications in the visualization of knowledge, data mining, and governance. Its capacity to manage the confusion and unpredictability present in practical problem data is a significant benefit. A more adaptable and truthful depiction of intricate systems is made possible by soft sets (SSETs), which permit the inclusion of elements with incomplete memberships. In decision support systems,

*Corresponding author.

*E-mail address: aurkhhb@gmail.com

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



where precise data may be omitted or it may be challenging to collect, adaptability is especially essential. Subsequent research by Maji et al. [9,10] expanded the scope of SST, presenting various operations applied to decision-making problems. Rahman et al. [11] employed parameterized approach to decision making using extension of fuzzy SSETs. Vimala et al. [12] investigated some algebraic structures and their relating results using multi aspects of fuzzy SSETs. Further contributions by Shabir and Naz [13] introduced soft topological spaces, while Majumdar and Samanta [14] explored mappings between SSETs. The study by Feng et al. [15] investigated the intersection of SSETs with fuzzy and rough sets. Researchers in [16,17] introduced soft real sets (SRSs), soft real numbers (SRNs), soft complex sets (SCSs), and soft complex numbers (SCNs), demonstrating their properties and practical applications. Saeed et al. [18] introduced the notions of soft elements and members. Later on, they [19] introduced the notions of soft algebraic structures using soft elements and members. A subsequent study in 2016, focused on fundamental operations in a soft context, discussing soft functions and including results like the Bolzano theorem [20]. Irfan [21] studied various properties of convexity and its inverse in soft and fuzzy SSETs environments. Salih and Sabir [22] explored specific characteristics related to convexity and concavity within the context of SSETs. The researchers [23–27] extended the previously defined knowledge of convexity to some of its other variants in soft and fuzzy SSET environments. Further study related to ambiguous focus on fuzzy complex numbers [28] that extend traditional complex numbers to accommodate uncertainty and imprecision, enabling degrees of membership and uncertainty in both real and imaginary components. Demir [29] discussed the idea of soft complex valued metric spaces and looked at a few topological facets of them. Next, for a variety of soft mappings on soft complex valued metric spaces, certain fixed soft element theorems were developed. Selvachandran et al. [30,31] studied the relations and distance measures between complex vague SSETs. Sabir [32] studied the extension of complex fuzzy set. This extension enhances the representation of fuzzy information within mathematical contexts, particularly in systems where imprecision and ambiguity play a significant role. Wardowski [33] demonstrated the natural relationship between soft operations and soft objects in soft topological spaces and proposed a novel concept of soft element of a SSET. Öztürk [34] examined the connection between bipolar SSETs and bipolar soft points and originated the idea of bipolar soft points. A novel interpretation of soft point was introduced by Senel [35], which allows one to construct all soft points that vary with any parameter that occurs in a SSET. Taşköprü and Altıntaş [36] provided a topology whose members are sets of soft elements and explained how this topology relates to basic soft topology. Allam et al. [37] first proposed the idea of a soft element and then modified the idea of a soft point to eliminate any connected issues with the soft point. Polat et al. [38] introduced a number of

properties on soft topological spaces, based on the notion of soft elements, which provide us with an alternative viewpoint for the advancement of soft set theory. These qualities include the neighbourhood structure of a soft element, soft interior, and among others. Hameed and Khalil [39] proved some new theorems on the equality of infimum soft sequences. Asma et al. [40] and Kirişçi [41] discussed decision making applications based on uncertain environments.

Novelty and Motivation

The scientific literature reviewed above makes it evident that while certain scholars have studied soft elements, soft members, and soft complex numbers, there needs to be more research on the description of soft complex sequences and series in the literature. The SCNs provide an effective tool for handling equivocal and imperfect data in an organized way. Their capacity to express indeterminate values with real and imaginary aspects is one of its advantages since it enables a more thorough explanation of intricate situations. Because of its adaptability, it may be utilized for analyzing systems with multifaceted unpredictability, which is prevalent in a wide range of practical uses. Furthermore, while maintaining uncertain knowledge, SCNs offer an adaptable structure for carrying out computations. The SCNs and soft complex sequence (SCSQs) are useful in many domains where handling ambiguous or imprecise data is crucial. They can be employed to simulate systems with ambiguous or fuzzy parameters in engineering, enabling more reliable and flexible control schemes. The SCNs, which take into account erratic characteristics and unclear market data, can help in risk evaluation and investment optimization in the financial sector. The SCSQs can also be used in medical diagnosis to examine patient information as well as medical records, taking into account the inherent uncertainty in diagnosis and treatment results. Additionally, SCNs can also improve the processing of ambiguous or partial data in artificial intelligence and machine learning, resulting in more trustworthy and precise decisions. All things considered, SCNs have implications in a variety of domains, such as financial services, engineering, and artificial intelligence, where managing uncertainty is essential to solving problems successfully. Thus, keeping in mind above mentioned advantages and adaptability of SCNs, this study is aimed to introduce the novel concept of a SCSQ, examining its convergent components using SCNs and soft real sequences (SRSQ). Additionally, it establishes a convergent series on SCNs, exploring their relationship.

Section 2 reviews some basic concepts related to soft sets, soft elements, soft members, soft complex numbers, etc., and this is how the rest of the article is structured. In Section 3, the soft complex limit and associated theorems have been characterized and described using SCNs and SCSQs. Section 4 has a discussion of the boundedness of SCSQs, soft complex Cauchy sequences using SCNs, and their corresponding results. In Section 5, the convergence of these SCNs soft complex series and the corresponding

outcomes have been examined. Lastly, Section 6 presents the overview, future scope, and implications of the anticipated effort

PRELIMINARY KNOWLEDGE

The following section is meant to review fundamental definitions and operations for SSETs, SRSs, and SCSs, as outlined in [13,16,17,42].

Definition 1 Let $\tilde{U}$ be an introduction to the universe of discussion and $\tilde{\theta}$ be a set of parameters. Let $2^{(\tilde{U})}$ shows the set of power of $\tilde{U}$ and $\tilde{\theta}$ be a subset that is not empty of $\tilde{\theta}$. A pair $(\tilde{\zeta}, \tilde{\theta})$ is called a SSET over $\tilde{U}$, where $\tilde{\zeta}$ is a mapping given by $\tilde{\zeta}: \tilde{\theta} \rightarrow 2^{(\tilde{U})}$.

Put otherwise, a set above $\tilde{U}$ is a family of parameterized discourse universe subsets $\tilde{U}$. For $\tilde{q} \in \tilde{\theta}$, $\tilde{\zeta}(\tilde{q})$ might be seen as a set of $\tilde{q}$ approximation SSETs elements $(\tilde{\zeta}, \tilde{\theta})$ and if $\tilde{q} \notin \tilde{\theta}$ then $\tilde{\zeta}(\tilde{q}) = \emptyset$, that is $\tilde{\zeta}_{\tilde{\theta}} = \tilde{\zeta}(\tilde{q}): \tilde{q} \in \tilde{\theta} \subseteq \tilde{\theta}$, $\tilde{\zeta}: \tilde{\theta} \rightarrow 2^{(\tilde{U})}$. The family of all these SSETs over $\tilde{U}$ denoted by $\tilde{\Omega}_{\tilde{\theta}}(\tilde{U})$.

Definition 2 Let $\tilde{\zeta}_{\tilde{\theta}}, \tilde{f}_{\tilde{\theta}} \in \tilde{\Omega}_{\tilde{\theta}}(\tilde{U})$ then $\tilde{\zeta}_{\tilde{\theta}}$ is a soft subset of $\tilde{f}_{\tilde{\theta}}$, symbolized by $\tilde{\zeta}_{\tilde{\theta}} \subseteq \tilde{f}_{\tilde{\theta}}$, if $\tilde{\theta} \subseteq \tilde{\theta}$ and $\tilde{\zeta}_{\tilde{\theta}} \subseteq \tilde{f}_{\tilde{\theta}} \forall \tilde{q} \in \tilde{\theta}$. In this case $\tilde{\zeta}_{\tilde{\theta}}$ is described as a soft subset of $\tilde{f}_{\tilde{\theta}}$, and $\tilde{f}_{\tilde{\theta}}$ is described as a soft super set, $\tilde{\zeta}_{\tilde{\theta}}, \tilde{f}_{\tilde{\theta}} \supseteq \tilde{\zeta}_{\tilde{\theta}}$.

Definition 3 A pair of soft subsets $\tilde{\zeta}_{\tilde{\theta}}$ and $\tilde{f}_{\tilde{\theta}}$ over $\tilde{U}$ are considered equal if $\tilde{\zeta}_{\tilde{\theta}}$ is a soft subset of $\tilde{f}_{\tilde{\theta}}$ and $\tilde{f}_{\tilde{\theta}}$ is a soft subset of $\tilde{\zeta}_{\tilde{\theta}}$.

Definition 4 The soft subsets complement $(\tilde{\zeta}, \tilde{\theta})^c$ denoted by $(\tilde{\zeta}, \tilde{\theta})^c$ is determined by $(\tilde{\zeta}, \tilde{\theta})^c = (\tilde{\zeta}^c, \tilde{\theta})$: $\tilde{\zeta}^c: \tilde{\theta} \rightarrow 2^{(\tilde{U})}$ is a mapping provided by $\tilde{\zeta}^c(\tilde{q}) = \tilde{U} - \tilde{\zeta}(\tilde{q}), \forall \tilde{q} \in \tilde{\theta}$ and $\tilde{\zeta}^c$ and $\tilde{\zeta}$ is said the soft complement function is equivalent to $\tilde{\zeta}$. Clearly $(\tilde{\zeta})$ is similar to $\tilde{\zeta}$ and $((\tilde{\zeta}, \tilde{\theta})^c)^c = (\tilde{\zeta}, \tilde{\theta})$.

Definition 5 The difference between SSETs $(\tilde{\zeta}, \tilde{\theta})$ and $(\tilde{f}, \tilde{\theta})$ over $\tilde{U}$ displayed by $(\tilde{\zeta}, \tilde{\theta}) - (\tilde{f}, \tilde{\theta})$ is the SSET $(D, \tilde{\theta})$ such that $D(\tilde{q}) = \tilde{\zeta}(\tilde{q}) \setminus \tilde{f}(\tilde{q}), \forall \tilde{q} \in \tilde{\theta}$.

Definition 6 The union of SSETs and $(\tilde{f}, \tilde{\theta})$ Over $\tilde{U}$ is the SSET $(D, \tilde{C})$, where, $\tilde{C} = \tilde{\theta} \cup \tilde{\theta}$ for all $\tilde{q} \in \tilde{C}$,

$$D(\tilde{q}) = \begin{cases} \tilde{\zeta}(\tilde{q}) & \text{if } \tilde{q} \in \alpha \\ \tilde{f}(\tilde{q}) & \text{if } \tilde{q} \in \beta \\ \tilde{\zeta}(\tilde{q}) & \text{if } \tilde{q} \in \gamma \end{cases} \quad (1)$$

where, $\alpha = \tilde{\theta} - \tilde{\theta}, \beta = \tilde{\theta} - \tilde{\theta}$ and $\gamma = \tilde{\theta} \cap \tilde{\theta}$ and written as $(\tilde{\zeta}, \tilde{\theta}) \cup (\tilde{f}, \tilde{\theta}) = (D, \tilde{C})$.

Definition 7 The intersection $(D, \tilde{C})$ between two SSETs $(\tilde{\zeta}, \tilde{\theta})$ and $(\tilde{f}, \tilde{\theta})$ over $\tilde{U}$, denoted $(\tilde{\zeta}, \tilde{\theta}) \cap (\tilde{f}, \tilde{\theta})$, $(\tilde{\zeta}, \tilde{\theta})$ and $(\tilde{f}, \tilde{\theta})$ can be defined as $\tilde{C} = \tilde{\theta} \cap \tilde{\theta}$ and $D(\tilde{q}) = \tilde{\zeta}(\tilde{q}) \cap \tilde{f}(\tilde{q}), \forall \tilde{q} \in \tilde{C}$.

Definition 8 Given a non-empty parameter set $\tilde{\theta}$ and a non-empty set $\tilde{U}$. Then a function $\tau: \tilde{\theta} \rightarrow \tilde{U}$ is defined as a soft element in $\tilde{U}$. A soft element τ of $\tilde{U}$ belongs to a SSET $(\tilde{\zeta}, \tilde{\theta})$ of $\tilde{U}$ which is symbolized by $\tau \in (\tilde{\zeta}, \tilde{\theta})$ if $\tau(\tilde{q}) \in \tilde{\zeta}(\tilde{q}), \forall \tilde{q} \in \tilde{\theta}$. Thus for a SSET $(\tilde{\zeta}, \tilde{\theta})$ of $\tilde{U}$ concerning the index set $\tilde{\theta}$, it gives $\tilde{\zeta}(\tilde{q}) = \{\tau(\tilde{q}): \tau \in (\tilde{\zeta}, \tilde{\theta}), \forall \tilde{q} \in \tilde{\theta}\}$. In that case, τ is said to be a soft element of the SSET $(\tilde{\zeta}, \tilde{\theta})$.

For SSETs $(\tilde{\zeta}, \tilde{\theta})$ and $(\tilde{f}, \tilde{\theta})$. $(\tilde{f}, \tilde{\theta})$ within $\tilde{\Omega}_{\tilde{\theta}}(\tilde{U})$ the relation $(\tilde{\zeta}, \tilde{\theta}) \subseteq (\tilde{f}, \tilde{\theta})$, holds iff each soft element belonging to $(\tilde{\zeta}, \tilde{\theta})$ is also a soft element in $(\tilde{f}, \tilde{\theta})$.

Definition 9 Consider R as the collection of real numbers, let $2^{(R)}$ be the set of all bounded, non-empty subsets of R , and $\tilde{\theta}$ be a set of parameters. In this context, a mapping $\tilde{\zeta}: \tilde{\theta} \rightarrow 2^{(R)}$ is termed a SRS, denoted as $(\tilde{\zeta}, \tilde{\theta})$.

In particular if $(\tilde{\zeta}, \tilde{\theta})$ is a singleton, then recognizing $(\tilde{\zeta}, \tilde{\theta})$ together with the equivalent soft element, it will be defined as a SRN.

Definition 10 The sequence $\{\tilde{\mu}_n\}$ it is claimed that a set of SRNs is convergent to the soft real limit $\tilde{\mu}$ in $R(\tilde{\theta})$, when it has the soft real limit $\tilde{\mu}$, otherwise it is claimed as divergent.

Lemma 1 [42] There is a soft convergent sub sequence for every SRSQ.

This outcome confirms clearly that there is a convergent sub sequence for any bounded SRSQ.

Definition 11 Let C denote the collection of complex numbers, $\Lambda(C)$ be the collection of all bounded non-empty subsets of the complex number set and $\tilde{\theta}$ be a set of parameters. Then a mapping $\tilde{\zeta}: \tilde{\theta} \rightarrow 2^{(C)}$ is said to be a SCS. It's shown by $(\tilde{\zeta}, \tilde{\theta})$.

Definition 12 In particular, if $(\tilde{\zeta}, \tilde{\theta})$ a SSET with a singleton, then identifying $(\tilde{\zeta}, \tilde{\theta})$ it will be defined as a SCN together with the corresponding soft element. We used the symbols $\tilde{\omega}_1, \tilde{\omega}_2$ and $\tilde{\omega}_3$, to denote SCNs.

Example 1 Let C be the set of complex numbers. The $\tilde{\theta}$ be a set of parameters given by $\tilde{\theta} = \{\tilde{q}_1, \tilde{q}_2, \tilde{q}_3, \tilde{q}_4, \tilde{q}_5\}$. Then $(\tilde{\zeta}, \tilde{\theta})$ can be considered as a SCN where

$$(\tilde{\zeta}, \tilde{\theta}) = \{(\tilde{q}_1, \{2 + i\}), (\tilde{q}_2, \{2i\}), (\tilde{q}_3, \{2 - i\}), (\tilde{q}_4, \{4i\}), (\tilde{q}_5, \{5 - 6i\})\}.$$

Definition 13 A SCS $(\tilde{\zeta}, \tilde{\theta})$ is claimed as soft pure imaginary set if $\tilde{\zeta}(\lambda)$ is a subset of the set of pure imaginary numbers for each $\lambda \in \tilde{\theta}$.

Let us denote the set of all SCSs by $X_{(\tilde{\theta})}$ and the set of all SCNs by $X^{(\tilde{\theta})}$.

Definition 14 Assume that $\tilde{\zeta}$ is a SCS or number. Subsequently, $Re \tilde{\zeta}$ is a real part and $Im \tilde{\zeta}$ is an imaginary part of $\tilde{\omega}$ and are stated as follows:

$$Re \tilde{\zeta}(\tilde{q}) = \{Re(\tilde{\omega}): \tilde{\omega} \in \tilde{\zeta}(\tilde{q})\}, Re \tilde{\zeta}(\tilde{q}) = Re(\tilde{\zeta}(\tilde{q})) \quad (2)$$

and

$$Im \tilde{\zeta}(\tilde{q}) = \{Im(\tilde{\omega}): \tilde{\omega} \in \tilde{\zeta}(\tilde{q})\}, Im \tilde{\zeta}(\tilde{q}) = Im(\tilde{\zeta}(\tilde{q})) \quad (3)$$

For every $\tilde{q} \in \tilde{\theta}$. It is shown that $Re \tilde{\zeta}$ and $Im \tilde{\zeta}$ are SRSs or numbers.

Definition 15 Let $(\tilde{\zeta}, \tilde{\theta})$ be a collection of SCS. Then the modulus of $(\tilde{\zeta}, \tilde{\theta})$ is symbolized by $|\tilde{\zeta}, \tilde{\theta}|$ and is defined by $|\tilde{\zeta}|(\lambda) = \{|\tilde{\omega}|: \tilde{\omega} \in \tilde{\zeta}(\lambda); \lambda \in \tilde{\theta}\}$.

Limits of a Sequence in SCNs

Definition 16 If $\tilde{\mu}_n$ and $\tilde{v}_n$ be any two SRSQs then $\tilde{\omega}_n = \tilde{\mu}_n + i\tilde{v}_n$ is a SCSQ denoted by

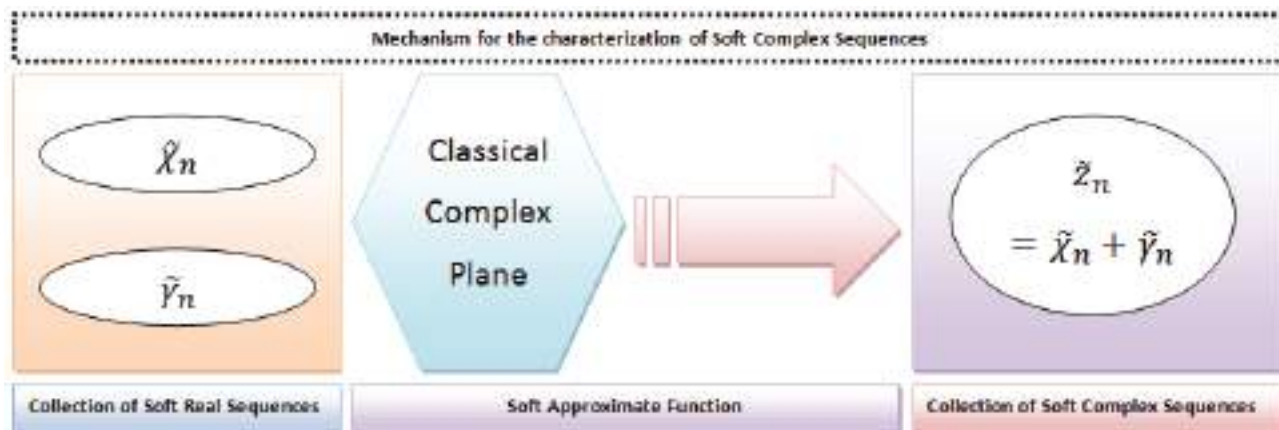


Figure 1. Mechanism for the characterization of SCSQs.

$\{\tilde{\omega}_n\}_{n=1}^{\infty} = \{\tilde{\omega}_1, \tilde{\omega}_2, \tilde{\omega}_3, \dots, \tilde{\omega}_n, \dots\}, \tilde{\omega}_n \in X^{(\tilde{\theta})}$ for all $n = \{1, 2, 3, \dots\}$.

Mechanism for the characterization of SCSQs can be viewed from Figure 1.

Definition 17 The sequence $\{\tilde{\omega}_n\}$ of SCNs is called convergent to soft complex limit $\tilde{\varrho} \in X^{(\tilde{\theta})}$, when it has the soft complex limit $\tilde{\varrho}$, otherwise it is claimed as divergent.

Definition 18 If $\{\tilde{\omega}_n\}$ represents a sequence of SCNs. We conclude that $\tilde{\omega}_n$ approaches the soft complex limit $\tilde{\varrho}$ (as $n \rightarrow \infty$) whenever for every SRN $\tilde{\eta} > \tilde{\theta} \exists \tilde{h}(\tilde{\eta})$, a soft positive integer (SPI) depending on $\tilde{\eta}$ such that

$$|\tilde{\omega}_n - \tilde{\varrho}| < \tilde{\eta}, \text{ for all } n \geq \tilde{h}(\tilde{\eta}) \quad (4)$$

If $\tilde{\omega}_n$ approaching the limit of the soft complex $\tilde{\varrho}$, we write $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \tilde{\varrho}$ or $\tilde{\omega}_n \rightarrow \tilde{\varrho}$ as $n \rightarrow \infty$.

Example 2 Consider $\tilde{\omega}_n = \tilde{\mu}_n + i\tilde{\nu}_n \in X^{(\tilde{\theta})}$ with $\tilde{\theta}$ is a finite set of parameters where $\tilde{\mu}_n = \tilde{\theta}$ and $\tilde{\nu}_n = \left(\frac{1}{2n}\right), n \in \mathbb{N}$.

Then, we claim that $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \tilde{\theta} = \tilde{\theta} + i\tilde{0}$ is the number of the soft complex, where $\tilde{0}(\tilde{q}) = \tilde{\theta}(\tilde{q}) + i\tilde{0}(\tilde{q}) = \tilde{\theta} + i\tilde{0}$. In fact, given $\tilde{\eta} > \tilde{\theta}$ it gives that $k = \min\{\tilde{\eta}(\tilde{q}) : \tilde{q} \in \tilde{\theta}\} > 0$. Since $\tilde{\theta}$ is a finite set of parameters. Because of the Archimedean characteristic of real numbers, $n_0 \in \mathbb{N}$, such that $\frac{1}{n_0} < \tilde{h}$. Given the fact that

$$\begin{aligned} |\tilde{\omega}_n - \tilde{0}(\tilde{q})| &= \left| \left(\frac{1}{2n}\right) - \tilde{0}(\tilde{q}) \right| = \left| \left(\frac{1}{2n}\right)(\tilde{q}) - \tilde{0}(\tilde{q}) \right| \\ &= \left| \frac{1}{2n} - 0 \right| = \frac{1}{2n} \leq \frac{1}{n_0} < k \leq \tilde{\eta}(\tilde{q}), \end{aligned}$$

for every $\tilde{q} \in \tilde{\theta}$ and $n \geq n_0$. It means that $|\tilde{\omega}_n - \tilde{0}| < \tilde{\eta}$.

Theorem 1 Let $\{\tilde{\omega}_n\}$ be a sequence in the SCNs. Write $\tilde{\omega}_n = \tilde{\mu}_n + i\tilde{\nu}_n$ where $\tilde{\mu}_n$ and $\tilde{\nu}_n$ are sequences of SRNs. Then the given sequence convergent in $X^{(\tilde{\theta})}$ iff $\{\tilde{\mu}_n\}$ and $\{\tilde{\nu}_n\}$ converge in $\mathbb{R}(\tilde{\theta})$.

Proof. Suppose $\tilde{\mu}_n \rightarrow \tilde{\mu}$ and $\tilde{\nu}_n \rightarrow \tilde{\nu}$. We now prove that $\tilde{\omega}_n = \tilde{\mu} + i\tilde{\nu}$. Indeed, fix $\tilde{\eta} > \tilde{\theta}$ and find, by the stated convergence, there exists $\tilde{h}$ such that $n > \tilde{h}$. It given that $|\tilde{\mu}_n - \tilde{\mu}| < \tilde{\eta}/2$ and $|\tilde{\nu}_n - \tilde{\nu}| < \tilde{\eta}/2$. Then, for $n > \tilde{h}$, it gives

$$\begin{aligned} |\tilde{\omega}_n - (\tilde{\mu} + i\tilde{\nu})| &= |(\tilde{\mu}_n - \tilde{\mu}) + i(\tilde{\nu}_n - \tilde{\nu})| \leq |\tilde{\mu}_n - \tilde{\mu}| \\ &\quad + |\tilde{\nu}_n - \tilde{\nu}| < \tilde{\eta}/2 + \tilde{\eta}/2 = \tilde{\eta}. \end{aligned}$$

Conversely, suppose that $\tilde{\omega}_n = \tilde{\mu} + i\tilde{\nu}$. Then, given $\tilde{\eta} > \tilde{\theta}$, we find SPI $\tilde{h}$ such that for $n > \tilde{h}$, it holds that $|\tilde{\omega}_n - (\tilde{\mu} + i\tilde{\nu})| < \tilde{\eta}$. So clearly

$$|\tilde{\mu}_n - \tilde{\mu}| \leq |\tilde{x}_n - \tilde{\mu}| + |\tilde{y}_n - \tilde{\nu}| = |\tilde{\omega}_n - (\tilde{\mu}_n + i\tilde{\nu}_n)| < \tilde{\eta}, \quad (5)$$

and

$$|\tilde{\nu}_n - \tilde{\nu}| \leq |\tilde{y}_n - \tilde{\nu}| + |\tilde{\mu}_n - \tilde{\mu}| = |\tilde{\omega}_n - (\tilde{\mu}_n + i\tilde{\nu}_n)| < \tilde{\eta}. \quad (6)$$

Proposition 1 Let $\{\tilde{\omega}_{1n}\}$ and $\{\tilde{\omega}_{2n}\}$ be convergent infinite sequences of SCNs. Then the sequences $\{\tilde{\omega}_{1n} + \tilde{\omega}_{2n}\}$, $\{\tilde{\omega}_{1n} - \tilde{\omega}_{2n}\}$, and $\{\tilde{\omega}_{1n} \cdot \tilde{\omega}_{2n}\}$ are convergent, and

$$\begin{aligned} \lim_{n \rightarrow \infty} (\tilde{\omega}_{1n} + \tilde{\omega}_{2n}) &= \lim_{n \rightarrow \infty} \tilde{\omega}_{1n} + \lim_{n \rightarrow \infty} \tilde{\omega}_{2n}, \\ \lim_{n \rightarrow \infty} (\tilde{\omega}_{1n} - \tilde{\omega}_{2n}) &= \lim_{n \rightarrow \infty} \tilde{\omega}_{1n} - \lim_{n \rightarrow \infty} \tilde{\omega}_{2n}, \\ \lim_{n \rightarrow \infty} (\tilde{\omega}_{1n} \cdot \tilde{\omega}_{2n}) &= \left(\lim_{n \rightarrow \infty} \tilde{\omega}_{1n} \right) \left(\lim_{n \rightarrow \infty} \tilde{\omega}_{2n} \right). \end{aligned}$$

In addition, if $\tilde{\omega}_{2n} = 0$, for all $n \in \mathbb{N}$ and $\lim_{n \rightarrow \infty} \tilde{\omega}_{1n} \neq \tilde{\theta}$, then the sequence $\{\tilde{\omega}_{1n} / \tilde{\omega}_{2n}\}$ of SCNs is convergent, and

$$\lim_{n \rightarrow \infty} \left(\frac{\tilde{\omega}_{1n}}{\tilde{\omega}_{2n}} \right) = \frac{\lim_{n \rightarrow \infty} \tilde{\omega}_{1n}}{\lim_{n \rightarrow \infty} \tilde{\omega}_{2n}}.$$

Proof. Let $\lim_{n \rightarrow \infty} \tilde{\omega}_{1n} = \tilde{\varrho}_1$ and $\lim_{n \rightarrow \infty} \tilde{\omega}_{2n} = \tilde{\varrho}_2$. Suppose some SRN $\tilde{\eta} > \tilde{\theta}$ be given. It follows from the idea of

limits sequence that there are SPI $\check{h}_1(\check{\eta})$ and $\check{h}_2(\check{\eta})$ depending on $\check{\eta}$ such that $|\check{\omega}_{1n} - \check{\varrho}_1| < \check{\eta}/2$ whenever $n \geq \check{h}_1$ and $|\check{\omega}_{2n} - \check{\varrho}_2| < \check{\eta}/2$ whenever $n \geq \check{h}_2$. If $n > \check{h}$, where $\check{h}$ is the maximum of $\check{h}_1$ and $\check{h}_2$, then

$$|\check{\omega}_{1n} + \check{\omega}_{2n} - (\check{\varrho}_1 - \check{\varrho}_2)| \leq |\check{\omega}_{1n} - \check{\varrho}_1| + |\check{\omega}_{2n} - \check{\varrho}_2| < \check{\eta}/2 + \check{\eta}/2 = \check{\eta}.$$

Thus, $\lim_{n \rightarrow \infty} (\check{\omega}_{1n} + \check{\omega}_{2n}) = \lim_{n \rightarrow \infty} \check{\omega}_{1n} + \lim_{n \rightarrow \infty} \check{\omega}_{2n}$.

Let $\check{\vartheta}$ be some complex number. We prove that $\lim_{n \rightarrow \infty} \check{\vartheta} \check{\omega}_{2n} = \check{\vartheta} \check{\varrho}_2$. Now, given any SPI $\check{h}$, we can choose a soft positive $\check{\delta}$ small enough to ensure that $|\check{\vartheta}| \check{\delta} < \check{\eta}$. Now $\lim_{n \rightarrow \infty} \check{\omega}_{2n} = \check{\varrho}_2$, and therefore, there exists some SPI $\check{h}$ such that $|\check{\omega}_{2n} - \check{\varrho}_2| < \check{\delta}$ whenever $n \geq \check{h}$. But then $|\check{\vartheta} \check{\omega}_{2n} - \check{\vartheta} \check{\varrho}_2| = |\check{\vartheta}| |\check{\omega}_{2n} - \check{\varrho}_2| < \check{\delta} |\check{\vartheta}|$. This gives the fact that $\lim_{n \rightarrow \infty} \check{\vartheta} \check{\omega}_{2n} = \check{\vartheta} \check{\varrho}_2$.

On apply this result with $\check{\vartheta} = -1$, we see that $\lim_{n \rightarrow \infty} -\check{\omega}_{2n} = -\check{\varrho}_2$. It follows that $\lim_{n \rightarrow \infty} \check{\omega}_{1n} - \check{\omega}_{2n} = \check{\varrho}_1 - \check{\varrho}_2$. Next we prove that if $\{\check{\omega}_{3n}\}$ and $\{\check{\omega}_{4n}\}$ are infinite sequences of SCNs, and if $\lim_{n \rightarrow \infty} \check{\omega}_{3n} = \bar{0}$ and $\lim_{n \rightarrow \infty} \check{\omega}_{4n} = \bar{0}$, then $\lim_{n \rightarrow \infty} (\check{\omega}_{3n} \check{\omega}_{4n}) = \bar{0}$. Suppose some SRN $\check{\eta} > \bar{0}$ be given, it follows from the definition of limits sequence that there exists SPIs $\check{h}_3$ and $\check{h}_4$ such that $|\check{\omega}_{3n}| < \sqrt{\check{\eta}}$ whenever $n \geq \check{h}_3$ and $|\check{\omega}_{4n}| < \sqrt{\check{\eta}}$ whenever $n \geq \check{h}_4$. Suppose $\check{h}$ be the maximum of $\check{h}_3$ and $\check{h}_4$, if $n \geq \check{h}$ then that $|\check{\omega}_{3n}| < \sqrt{\check{\eta}}$ and $|\check{\omega}_{4n}| < \sqrt{\check{\eta}}$. Therefore, $|\check{\omega}_{3n} \check{\omega}_{4n}| < \sqrt{\check{\eta}}$. It shows that $\lim_{n \rightarrow \infty} (\check{\omega}_{3n} \check{\omega}_{4n}) = \bar{0}$.

We can apply this result with $\check{\omega}_{3n} = \check{\omega}_{1n} - \check{\varrho}_1$ and $\check{\omega}_{4n} = \check{\omega}_{2n} - \check{\varrho}_2$ for all positive integer n , where $\check{\varrho}_1 = \lim_{n \rightarrow \infty} \check{\omega}_{1n}$ and $\check{\varrho}_2 = \lim_{n \rightarrow \infty} \check{\omega}_{2n}$. Now $\lim_{n \rightarrow \infty} \check{\omega}_{3n} = \bar{0}$ and $\lim_{n \rightarrow \infty} \check{\omega}_{4n} = \bar{0}$ follows that

$$\begin{aligned} \bar{0} &= \lim_{n \rightarrow \infty} (\check{\omega}_{3n} \check{\omega}_{4n}) = \lim_{n \rightarrow \infty} ((\check{\omega}_{1n} - \check{\varrho}_1)(\check{\omega}_{2n} - \check{\varrho}_2)) \\ &= \lim_{n \rightarrow \infty} (\check{\omega}_{1n} \check{\omega}_{2n} - \check{\varrho}_2 \check{\omega}_{1n} - \check{\varrho}_1 \check{\omega}_{2n} + \check{\varrho}_1 \check{\varrho}_2) \\ &= \lim_{n \rightarrow \infty} (\check{\omega}_{1n} \check{\omega}_{2n}) - \check{\varrho}_2 \lim_{n \rightarrow \infty} \check{\omega}_{1n} - \check{\varrho}_1 \lim_{n \rightarrow \infty} \check{\omega}_{2n} + \lim_{n \rightarrow \infty} (\check{\varrho}_1 \check{\varrho}_2) \\ &= \lim_{n \rightarrow \infty} \check{\omega}_{1n} \check{\omega}_{2n} - \check{\varrho}_1 \check{\varrho}_2 \end{aligned}$$

Therefore $\lim_{n \rightarrow \infty} \check{\omega}_{1n} \check{\omega}_{2n} = \check{\varrho}_1 \check{\varrho}_2$.

Finally, suppose that $\check{\omega}_{2n} \neq \bar{0}$ for every positive integers n , and that $\check{\varrho}_2 \neq \bar{0}$. Then for any given SRN $\check{\eta} > \bar{0}$, there exists some SPIs $\check{h}_5$ such that

$$|\check{\omega}_{2n} - \check{\varrho}_2| < \frac{1}{2} |\check{\varrho}_2|^2 \check{\eta} \text{ and } |\check{\omega}_{2n} - \check{\varrho}_2| < \frac{1}{2} |\check{\varrho}_2|,$$

whenever $n \geq \check{h}_5$. For $n \geq \check{h}_5$, then $|\check{\omega}_{2n}| \geq |\check{\varrho}_2| - |\check{\omega}_{2n} - \check{\varrho}_2| > \frac{1}{2} |\check{\varrho}_2|$, and therefore,

$$\left| \frac{1}{\check{\omega}_{2n}} - \frac{1}{\check{\varrho}_2} \right| = \left| \frac{\check{\varrho}_2 - \check{\omega}_{2n}}{\check{\varrho}_2 \check{\omega}_{2n}} \right| \leq \frac{2}{|\check{\varrho}_2|^2} |\check{\omega}_{2n} - \check{\varrho}_2| < \check{\eta}$$

Thus, $\lim_{n \rightarrow \infty} \frac{1}{\check{\omega}_{2n}} = \frac{1}{\check{\varrho}_2}$. It given that $\lim_{n \rightarrow \infty} \frac{\check{\omega}_{1n}}{\check{\omega}_{2n}} = \frac{\check{\varrho}_1}{\check{\varrho}_2}$, as required.

Theorem 2 If $\check{\varrho}$ is the convergent soft complex limit for the SCSQ $\check{\omega}_n$, then it is also the convergent soft complex limit for $\check{\omega}_{n+s}$ ($s \in \mathbb{N}$).

Proof. Since $\{\check{\omega}_n\}$ converges to soft complex limit $\check{\varrho}$ then for any SRN $\check{\eta} > \bar{0}$, there exists SPI $\check{h}(\check{\eta})$ such that $|\check{\omega}_n - \check{\varrho}| < \check{\eta}$, for all $n \geq \check{h}(\check{\eta})$. If we take $m = n + s$, for all $s \in \mathbb{N}$, then $m > n > \check{h}(\check{\eta})$ implies $m = n + s$ and $|\check{\omega}_{n+s} - \check{\varrho}| < \check{\eta}$. Hence $\check{\varrho}$ is the convergent soft complex limit for the SCSQ $\check{\omega}_{n+s}$.

Theorem 3 A convergent sequence of SCNs has a unique soft complex limit.

Proof. Let $\{\check{\omega}_n\}$ be a soft complex convergent sequence. Suppose that $\lim_{n \rightarrow \infty} \check{\omega}_n = \check{\varrho}_1$ and $\lim_{n \rightarrow \infty} \check{\omega}_n = \check{\varrho}_2$. We wish to prove $\check{\varrho}_1 = \check{\varrho}_2$. For given $\check{\eta} > \bar{0}$, by the Definition 18, there exists a soft real positive integer $\check{h}_1$ such that $n \geq \check{h}_1$ implies $|\check{\omega}_n - \check{\varrho}_1| < \check{\eta}/2$. Moreover, there exists a soft real positive integer $\check{h}_2$ such that $n > \check{h}_2$ implies $|\check{\omega}_n - \check{\varrho}_2| < \check{\eta}/2$. For $n > \max\{\check{h}_1, \check{h}_2\}$, by the triangle inequality it gives

$$|\check{\varrho}_1 - \check{\varrho}_2| = |\check{\varrho}_1 + \check{\omega}_n - \check{\omega}_n - \check{\varrho}_2| \leq |\check{\omega}_n - \check{\varrho}_1| + |\check{\omega}_n - \check{\varrho}_2| < \frac{\check{\eta}}{2} + \frac{\check{\eta}}{2} = \check{\eta}.$$

This shows that $|\check{\varrho}_1 - \check{\varrho}_2| < \check{\eta}$ for all $\check{\eta} > \bar{0}$. It follows that $|\check{\varrho}_1 - \check{\varrho}_2| = \bar{0}$ and hence, $\check{\varrho}_1 = \check{\varrho}_2$.

Theorem 4 The SCSQ $\{\check{\omega}_n\}$ is convergent iff the SRSQ $\{|\check{\omega}_n|\}$ is also convergent.

Proof. Let $\{\check{\omega}_n\}$ be a soft complex convergent sequence, and let $\lim_{n \rightarrow \infty} \check{\omega}_n = \check{\varrho}$. For given $\check{\eta} > \bar{0}$, by Definition 18, there exists a soft real positive integer $\check{h}$ such that $n > \check{h}$ implies $|\check{\omega}_n - \check{\varrho}| < \check{\eta}$. So by the triangle inequality we can show that $||\check{\omega}_n| - |\check{\varrho}|| \leq |\check{\omega}_n - \check{\varrho}| < \check{\eta}$. Conversely, can be shown similarly.

Remark 1 The SCSQ $\{\check{\omega}_n\}$ converges to zero iff the SRSQ $\{|\check{\omega}_n|\}$ converges to zero.

Bounded by the Sequence of SCNs

Definition 19 Soft complex number sequence $\{\check{\omega}_n\}$ is claimed as bounded if a SRN $\check{M} \in \check{P}$ ($\check{M} > 0$) exists such that $|\check{\omega}_n| \leq \check{M} \forall n \in \mathbb{N}$. A sequence is unbounded if it is not bounded.

Example 3 Consider $\check{\omega}_n = \iota \left(\frac{1}{n} \right) \in X^{(\check{\theta})} (n \in \mathbb{N})$ where $\check{\theta}$ is a finite set of parameters. Then, it gives that $\check{\omega}_n$ is bounded since clearly there exist $\check{1} \in \check{P}$ and $|\check{\omega}_n|(\check{q}) = \left| \iota \left(\frac{1}{n} \right) \right|(\check{q}) = \left| \left(\frac{1}{n} \right) \right|(\check{q}) = \left| \frac{1}{n} \right| \leq \check{1}$ for every $\check{q} \in \check{\theta}$.

Definition 20 Let $\{\check{\omega}_n\}$ be a SCSQ. We say that $\{\check{\omega}_n\}$ is a soft complex Cauchy sequence if for all SRNs $\check{\eta} > \bar{0}$, there exists a positive soft integer $\check{h}$ such that $|\check{\omega}_n - \check{\omega}_m| < \check{\eta}$, for all $n, m \geq \check{h}$.

For $\check{\eta}$ to be less than the difference between any two soft elements that occur before and following the term's $\check{h}$.

Proposition 2 Let $\{\check{\omega}_n\}$ be a sequence of SCNs that converges. Then it is bounded.

Proof. Suppose that the SCSQ $\{\check{\omega}_n\}$ converge to a soft complex limit $\check{\varrho} \in X^{(\check{\theta})}$. Then for $\check{\eta} = \check{1} > \bar{0} \exists \check{h} \in \mathbb{N}$

such that $|\tilde{\omega}_n - \tilde{\varrho}| < \tilde{1}$ for all $n \geq \tilde{h}$. It follows that $|\tilde{\omega}_n| < |\tilde{\varrho}| + |\tilde{\omega}_n - \tilde{\varrho}| < |\tilde{\varrho}| + \tilde{1}$.

If we set $\tilde{m} = \max\{|\tilde{\varrho}|, |\tilde{\omega}_1|, |\tilde{\omega}_2|, |\tilde{\omega}_3|, \dots, |\tilde{\omega}_{n-1}|\}$ then $|\tilde{\omega}_n| < \tilde{m} + \tilde{1}$ for all $n \in \mathbb{N}$. This implies that the SCSQ $\{\tilde{\omega}_n\}$ is bounded.

Theorem 5 Let $\{\tilde{\omega}_n\}$ be a bounded sequence of SCNs. Then $\{\tilde{\omega}_n\}$ has a convergent subsequence.

Proof. Since $\{\tilde{\omega}_n\}$ is a bounded sequence of SCNs, then $\{Re \tilde{\omega}_n\}$ and $\{Im \tilde{\omega}_n\}$ are bounded sequences of SRNs. By Lemma 1, $\{Re \tilde{\omega}_n\}$ and $\{Im \tilde{\omega}_n\}$ has a convergent subsequences $\{Re \tilde{\omega}_{n_k}\}$ and $\{Im \tilde{\omega}_{n_k}\}$, respectively. Now $\{Re \tilde{\omega}_{n_k}\}$ is a subsequence of the convergent sequence $\{Re \tilde{\omega}_{n_k}\}$, and so converges to the same soft complex limit. Hence, $\{\tilde{\omega}_{n_k}\}$ is a convergent subsequence of $\{\tilde{\omega}_n\}$ as its real and imaginary parts are both convergent sequence of SCNs.

Theorem 6 If $\{\tilde{\omega}_n\}$ is a SCSQ bounded and converges to soft complex limit $\tilde{\varrho} \in X^{(\tilde{\varrho})}$. Then $\tilde{\omega}_n^2 = \tilde{\varrho}^2$.

Proof. Suppose $\{\tilde{\omega}_n\}$ be a SCSQ bounded and it is converge to $\tilde{\varrho} \in X^{(\tilde{\varrho})}$. So, by Definition 18, for a SRN $\tilde{\eta} > \tilde{0}$, there is a soft natural integer $\tilde{h}$ in which $|\tilde{\omega}_n - \tilde{\varrho}| \leq \frac{\tilde{\eta}}{\tilde{M} + |\tilde{\varrho}|}$, $n \geq \tilde{h}$, also since $\{\tilde{\omega}_n\}$ is bounded, therefore, $|\{\tilde{\omega}_n\}| < \tilde{M}$ ($n \in \mathbb{N}$), this means that

$$|\tilde{\omega}_n^2 - \tilde{\varrho}^2| = |(\tilde{\omega}_n - \tilde{\varrho})(\tilde{\omega}_n + \tilde{\varrho})| = |\tilde{\omega}_n - \tilde{\varrho}| |\tilde{\omega}_n + \tilde{\varrho}|.$$

The triangular inequality has given us $|\tilde{\omega}_n + \tilde{\varrho}| \leq |\tilde{\omega}_n| + |\tilde{\varrho}|$, this implies that $|\tilde{\omega}_n + \tilde{\varrho}| \leq \tilde{M} + |\tilde{\varrho}|$, $n \in \mathbb{N}$. Now it gives

$$|\tilde{\omega}_n^2 - \tilde{\varrho}^2| \leq |\tilde{\omega}_n - \tilde{\varrho}| (\tilde{M} + |\tilde{\varrho}|), n \geq \tilde{h} \quad (7)$$

implies that

$|\tilde{\omega}_n^2 - \tilde{\varrho}^2| \leq \frac{\tilde{\eta}}{\tilde{M} + |\tilde{\varrho}|} (\tilde{M} + |\tilde{\varrho}|) = \tilde{\eta}$, $n \geq \tilde{h}$. This completes the proof.

Theorem 7 If $\{\tilde{\omega}_n\} = \{\tilde{\mu} + i\tilde{\nu}\}$ is a sequence of SCNs then $\{\tilde{\omega}_n\}$ is a Cauchy sequence iff both sequences of SRNs $\{\tilde{\mu}_n\}$ and $\{\tilde{\nu}_n\}$ are Cauchy sequence of SRNs.

Proof. If we suppose $\{\tilde{\omega}_n\}$ is a Cauchy sequence, then for every SRN $\tilde{\eta} > \tilde{0} \exists$ a positive soft integer $\tilde{h}$ such that $|\tilde{\omega}_n - \tilde{\omega}_m| \leq \tilde{\eta}$ ($\tilde{n}, \tilde{m} \geq \tilde{h}$). Then for $n \geq \tilde{h}$, it gives that

$$\begin{aligned} |\tilde{\mu}_n - \tilde{\mu}_m| &= \sqrt{(\tilde{\mu}_n - \tilde{\mu}_m)^2} \leq \sqrt{(\tilde{\mu}_n - \tilde{\mu}_m)^2 + (\tilde{\nu}_n - \tilde{\nu}_m)^2} \\ &= |\tilde{\omega}_n - \tilde{\omega}_m| \leq \tilde{\eta} \end{aligned}$$

and

$$\begin{aligned} |\tilde{\nu}_n - \tilde{\nu}_m| &= \sqrt{(\tilde{\nu}_n - \tilde{\nu}_m)^2} \leq \sqrt{(\tilde{\mu}_n - \tilde{\mu}_m)^2 + (\tilde{\nu}_n - \tilde{\nu}_m)^2} \\ &= |\tilde{\omega}_n - \tilde{\omega}_m| \leq \tilde{\eta}. \end{aligned}$$

So both $\{\tilde{\mu}_n\}$ and $\{\tilde{\nu}_n\}$ are Cauchy sequence of SRNs.

Conversely, suppose that both $\{\tilde{\mu}_n\}$ and $\{\tilde{\nu}_n\}$ are Cauchy sequence of SRNs. Then for every SRN $\tilde{\eta} > \tilde{0}$, since $\{\tilde{\mu}_n\}$ is a Cauchy sequence, there exists a positive soft integer $\tilde{h}_1$ such that for all $n, m \geq \tilde{h}_1$, then

$|\tilde{\mu}_n - \tilde{\mu}_m| \leq \tilde{\eta}/2$. Similarly, since $\{\tilde{\nu}_n\}$ is Cauchy then there exists an positive soft integer $\tilde{h}_2$ such that for all $n, m \geq \tilde{h}_2$. Then $|\tilde{\nu}_n - \tilde{\nu}_m| \leq \tilde{\eta}/2$. If $\tilde{h} = \max\{\tilde{h}_1, \tilde{h}_2\}$, then for $n, m \geq \tilde{h}$, it gives

$$\begin{aligned} |\tilde{\omega}_n - \tilde{\omega}_m| &= |(\tilde{\mu}_n + i\tilde{\nu}_n) - (\tilde{\mu}_m + i\tilde{\nu}_m)| \\ &= |(\tilde{\mu}_n - \tilde{\mu}_m) + i(\tilde{\nu}_n - \tilde{\nu}_m)| \leq |\tilde{\mu}_n - \tilde{\mu}_m| + |\tilde{\nu}_n - \tilde{\nu}_m| < \tilde{\eta}. \end{aligned}$$

So $\{\tilde{\omega}_n\}$ is a Cauchy sequence of SCNs.

Theorem 8 Every soft Complex Cauchy sequence has a bound.

Proof. Suppose that $\{\tilde{\omega}_n\}$ is a soft Complex Cauchy sequence. Then by Definition 18, for every SRN $\tilde{\eta} = 1 > 0$, there exists a positive integer $\tilde{h}$ such that $|\tilde{\omega}_n - \tilde{\omega}_m| < 1$, for all $n, m \geq \tilde{h}$. So for $n \geq \tilde{h}$, it gives $|\tilde{\omega}_n| \leq 1 + |\tilde{\omega}_{\tilde{h}}|$.

Theorem 9 Every convergent SCSQ is a soft complex Cauchy sequence.

Proof. Let $\{\tilde{\omega}_n\}$ be a SCSQ and convergent to $\tilde{\varrho} \in X^{(\tilde{\varrho})}$. For $\tilde{\eta} > \tilde{0}$, there exists a positive integer $\tilde{h}$ such that

$$|\tilde{\omega}_s - \tilde{\varrho}| \leq \frac{\tilde{\eta}}{2} (s \geq \tilde{h}).$$

This implies that for $n, m \geq \tilde{h}$, it gives

$$|\tilde{\omega}_n - \tilde{\omega}_m| \leq |\tilde{\omega}_n - \tilde{\varrho}| + |\tilde{\omega}_m - \tilde{\varrho}| < \tilde{\eta}.$$

Theorem 10 Every soft complex Cauchy sequence is convergent.

Proof. Let $\{\tilde{\omega}_n\} = \{\tilde{\mu}_n + i\tilde{\nu}_n\}$ be any soft complex Cauchy sequence, where $\tilde{\mu}_n$ and $\tilde{\nu}_n$ are Cauchy sequences of SRNs. Then $\{\tilde{\mu}_n\}$ and $\{\tilde{\nu}_n\}$ are convergent sequences of SRNs to $\tilde{\mu}_0$ and $\tilde{\nu}_0$, respectively.

Convergence of Series in SCNs

Let $\{\tilde{\omega}_n\}$ be a sequence of SCNs, we make a new sequence that is specified by $\tilde{\omega}_1 = \tilde{\omega}_1, \tilde{\omega}_2 = \tilde{\omega}_1 + \tilde{\omega}_2, \tilde{\omega}_3 = \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3, \dots, \tilde{\omega}_n = \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 + \dots + \tilde{\omega}_n$, where $\tilde{\omega}_n$ is called the sequence of n^{th} partial sums of sequence of SCNs $\{\tilde{\omega}_n\}$. The sequence $\{\tilde{\omega}_n\}$ is symbolized by $\tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 + \dots = \sum_{n=1}^{\infty} \tilde{\omega}_n$ called an infinite series of SCNs. If $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \tilde{\omega}$ exists then the series is claimed as convergent and $\tilde{\omega}$ is its sum, i.e.

$$\sum_{n=1}^{\infty} \tilde{\omega}_n = \lim_{n \rightarrow \infty} \sum_{k=1}^n \tilde{\omega}_k = \lim_{n \rightarrow \infty} \tilde{\omega}_n = \tilde{\omega}.$$

If a series of SCNs is not convergent, it is called divergent.

Example 4 The soft complex series $\sum_{n=0}^{\infty} i^n$ does not convergent. Because the SCSQ of its parial sums is periodic and therefore doses not have a limit.

Theorem 11 If the infinite series of SCNs $\sum_{n=1}^{\infty} \tilde{\omega}_n$ is convergent then $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{0}$.

Proof. Let $\tilde{\omega}_n = \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 + \dots + \tilde{\omega}_{n-1} + \tilde{\omega}_n$ be the n^{th} partial sum of the series of SCNs. Suppose that the series is convergent and $\bar{\omega}$ be the sum $\sum_{n=1}^{\infty} \tilde{\omega}_n$, this implies that $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{\omega}$, we get that $\tilde{\omega}_n = \tilde{\omega}_n - \tilde{\omega}_{n-1}$, by $\tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 + \dots + \tilde{\omega}_{n-1} = \tilde{\omega}_{n-1}$, taking limit on both sides, it gives that

$$\lim_{n \rightarrow \infty} \tilde{\omega}_n = \lim_{n \rightarrow \infty} \tilde{\omega}_n - \lim_{n \rightarrow \infty} \tilde{\omega}_{n-1} = \bar{\omega} - \bar{\omega} = 0.$$

Remark 2 Consider the infinite series of SCNs $\sum_{n=0}^{\infty} \tilde{\omega}_n = \tilde{\omega}_0 + \tilde{\omega}_1 + \tilde{\omega}_2 + \dots$ if $\tilde{r}_n = \tilde{\omega}_n + \tilde{\omega}_{n+1} + \tilde{\omega}_{n+2} + \dots$ then $\tilde{r}_n$ is called remainder of the infinite series of SCNs. If $\bar{\omega}$ is the sum of infinite series of SCNs then $\bar{\omega} = \tilde{\omega}_n - \tilde{r}_n$.

Theorem 12 A series $\sum_{n=1}^{\infty} \tilde{\omega}_n$ of terms with SCNs is convergent iff for every SRN $\bar{\eta} > \bar{0}$, there exists a positive integer $\bar{h}$ such that $|\tilde{\omega}_n + \tilde{\omega}_{n+1} + \dots + \tilde{\omega}_{n+p}| < \bar{\eta}$, for all $n \geq \bar{h}$ and $p \geq 0$.

Proof. Suppose $\sum_{n=1}^{\infty} \tilde{\omega}_n$ is convergent, and let $\tilde{\omega}_n = \tilde{\omega}_1 + \tilde{\omega}_2 + \dots + \tilde{\omega}_{n-1}$ be the n^{th} partial sum of the series of SCNs. Suppose that the series is convergent and $\bar{\omega}$ be a sum of series, it means that $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{\omega}$, for every SRN $\bar{\eta} > \bar{0}$ there exists a positive integer $\bar{h}$ such that $|\tilde{\omega}_n - \bar{\omega}| < \bar{\eta}$, for all $n \geq \bar{h}$.

Let $\tilde{r}_n = \tilde{\omega}_n + \tilde{\omega}_{n+1} + \tilde{\omega}_{n+2} + \dots$ be the remainder of the infinite series of SCNs, we get that $\bar{\omega} = \tilde{\omega}_n - \tilde{r}_n$, so that $|\tilde{\omega}_n - \bar{\omega}| = |\tilde{\omega}_n - \tilde{\omega}_n| = |\tilde{r}_n| < \bar{\eta}$, for all $n \geq \bar{h}$, or $|\tilde{\omega}_n + \tilde{\omega}_{n+1} + \tilde{\omega}_{n+2} + \dots| < \bar{\eta}$, for all $n \geq \bar{h}$. So that, $|\tilde{\omega}_n + \tilde{\omega}_{n+1} + \tilde{\omega}_{n+2} + \dots + \tilde{\omega}_{n+p}| < \bar{\eta}$, for all $n \geq \bar{h}$ and $p \geq 0$.

Conversely, we know that $\sum_{n=1}^{\infty} \tilde{\omega}_n = \tilde{\omega}_1 + \tilde{\omega}_2 + \dots + \tilde{\omega}_n + \dots$ and if $\bar{\omega}$ is its partial sum then, $\bar{\omega} = \tilde{\omega}_n - \tilde{r}_n$ and $|\tilde{r}_n| < |\tilde{\omega}_n - \bar{\omega}|$. From $|\tilde{r}_n| = |\tilde{\omega}_n + \tilde{\omega}_{n+1} + \tilde{\omega}_{n+2} + \dots + \tilde{\omega}_{n+p}| < \bar{\eta}$ for all $n \geq \bar{h}$ and $p \geq 0$ so that $|\tilde{\omega}_n - \bar{\omega}| < \bar{\eta}$, it means that $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{\omega}$.

Definition 21 Let $\tilde{\omega}_n \in X^{(\bar{\theta})}$ ($n \geq \bar{0}$), a series $\sum_{n=1}^{\infty} \tilde{\omega}_n$ converges to $\bar{\omega}_0$, if for every SRN $\bar{\eta} > \bar{0}$, there exists a positive integer $\bar{h}$ such that $|\sum_{k=1}^{\infty} \tilde{\omega}_k - \bar{\omega}_0| < \bar{\eta}$, for all $n \geq \bar{h}$.

Definition 22 A SCNs series $\sum_{n=1}^{\infty} \tilde{\omega}_n$ converges absolutely if $\sum_{n=1}^{\infty} |\tilde{\omega}_n|$ converges.

Proposition 3 If the series $\sum_{n=1}^{\infty} \tilde{\omega}_n$ converges absolutely then, $\sum_{n=1}^{\infty} \tilde{\omega}_n$ converges.

Proof. Let $\tilde{\omega}_n = \tilde{\omega}_1 + \tilde{\omega}_2 + \tilde{\omega}_3 + \dots + \tilde{\omega}_{n-1} + \tilde{\omega}_n$ be the series of partial sums in the context of its convergence. Then for every SRN $\bar{\eta} > \bar{0}$, there exists a positive integer $\bar{h}$ such that $|\sum_{j=m+1}^{\infty} \tilde{\omega}_j| < \bar{\eta}$, $\forall n \geq \bar{h}$. If $n, m \geq \bar{h}$ then,

$$\begin{aligned} |\tilde{\omega}_n - \tilde{\omega}_m| &= |\tilde{\omega}_{m+1} + \tilde{\omega}_{m+2} + \dots + \tilde{\omega}_n| \\ &= \left| \sum_{j=m+1}^{\infty} \tilde{\omega}_j \right| \leq \sum_{j=m+1}^{\infty} |\tilde{\omega}_j| \leq \sum_{j=m+1}^{\infty} |\tilde{\omega}_j| < \bar{\eta}. \end{aligned}$$

Finally, it gives that $\{\tilde{\omega}_n\}$ is a Cauchy sequence in SCNs so by Theorem 9 it is a convergent sequence in SCNs and $\sum_{n=1}^{\infty} \tilde{\omega}_n$ is convergent.

Theorem 13 Let $\tilde{\omega}_n = \tilde{\mu}_n + i\tilde{\nu}_n$ ($n \in \mathbb{N}$) be a SCSQ and $\bar{\omega} = \bar{\mu} + i\bar{\nu} \in X^{(\bar{\theta})}$. Then the soft complex series

$\sum_{n=1}^{\infty} \tilde{\omega}_n = \bar{\omega}$ iff the soft real series $\sum_{n=1}^{\infty} \tilde{\mu}_n = \bar{\mu}$ and $\sum_{n=1}^{\infty} \tilde{\nu}_n = \bar{\nu}$.

Proof. First we write the sequence of partial sums of the series $\sum_{n=1}^{\infty} \tilde{\omega}_n$ as $\tilde{\omega}_n = \tilde{\omega}_{1n} + i\tilde{\omega}_{2n}$, where $\tilde{\omega}_{1n} = \sum_{n=1}^{\infty} \tilde{\mu}_n$ and $\tilde{\omega}_{2n} = \sum_{n=1}^{\infty} \tilde{\nu}_n$. This implies that $\sum_{n=1}^{\infty} \tilde{\omega}_n = \bar{\omega}$ iff $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{\omega}$. Now by Theorem 1 on convergent SCSQs we get that $\lim_{n \rightarrow \infty} \tilde{\omega}_n = \bar{\omega}$ iff $\lim_{n \rightarrow \infty} \tilde{\omega}_{1n} = \bar{\mu}$ and $\lim_{n \rightarrow \infty} \tilde{\omega}_{2n} = \bar{\nu}$. Therefore, $\sum_{n=1}^{\infty} \tilde{\omega}_n = \bar{\omega}$.

Conversely, since $\tilde{\omega}_{1n} = \bar{\mu}$ and $\tilde{\omega}_{2n} = \bar{\nu}$ are partial sums of the real soft series $\sum_{n=1}^{\infty} \tilde{\mu}_n$ and $\sum_{n=1}^{\infty} \tilde{\nu}_n$, respectively, so the result is obtained directly.

CONCLUSION

The SCNs offer a useful tool for organizing and managing ambiguous and incomplete data. One of its benefits is that they may articulate imprecise values with both real and imaginary features, which allows for a more in-depth description of complex circumstances. Due to its flexibility, it can be applied to the analysis of systems with complex randomness, which is frequent in many real-world applications. In light of the aforementioned benefits and flexibility of SCNs, by using SCNs and SRSQ, the novel notion of a SCSQ has been presented, and its link with SCNs has been demonstrated by a convergent series analysis. Further, the soft complex limit of SCSQs has been introduced, which is an initial route for the convergence of SCSQs. Some famous soft convergence theorems are presented. Important theorems such as the Bolzano-Weierstrass theorem have also been explored. Soft complex boundedness of SCSQs is also well described. There are also some related soft complex theorems present. Cauchy Sequences in SCNs have also been reflected, as is their relationship with bounded sequences in SCNs. Furthermore, by using SCSQs with specific characteristics, we introduce the infinite series and its convergent in SCNs. The presented work can be used in order to characterize the various notions of differential and integral calculus in SSET environment. Moreover, the various theorems and ideas in fuzzy measure theory and fuzzy functional analysis can also be modified for SSET environment using the SCNs, and SCSQs. The suggested approach has enormous potential in a number of different fields for its future possibilities. Scholars could further explore ways to improve the suggested approaches to better manage situations in the actual world when specific and vague knowledge coexist. Additionally, there is an exciting opportunity for combining the suggested idea with modern innovations like machine learning and artificial intelligence, where the structure's versatility may help create stronger and more effective algorithms for pattern recognition and data analysis.

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 LA. Fuzzy sets. *Inf Control* 1965;8:338–353. [\[CrossRef\]](#)
- [2] Abu Arqub O. Adaptation of reproducing kernel algorithm for solving fuzzy Fredholm-Volterra integrodifferential equations. *Neural Comput Appl* 2017;28:1591–1610. [\[CrossRef\]](#)
- [3] Abu Arqub O, Singh J, Maayah B, Alhodaly M. Reproducing kernel approach for numerical solutions of fuzzy fractional initial value problems under the Mittag-Leffler kernel differential operator. *Math Methods Appl Sci* 2023;46:7965–7986. [\[CrossRef\]](#)
- [4] Abu Arqub O, Singh J, Alhodaly M. Adaptation of kernel functions-based approach with Atangana-Baleanu-Caputo distributed order derivative for solutions of fuzzy fractional Volterra and Fredholm integrodifferential equations. *Math Methods Appl Sci* 2023;46:7807–7834. [\[CrossRef\]](#)
- [5] Alshammari M, Al-Smadi M, Arqub OA, Hashim I, Alias MA. Residual series representation algorithm for solving fuzzy Duffing oscillator equations. *Symmetry* 2020;12:572. [\[CrossRef\]](#)
- [6] Blanco-Mesa F, Merigó JM, Gil-Lafuente AM. Fuzzy decision making: a bibliometric-based review. *J Intell Fuzzy Syst* 2017;32:2033–2050. [\[CrossRef\]](#)
- [7] Kahraman C, Onar SC, Oztaysi B. Fuzzy multi-criteria decision-making: a literature review. *Int J Comput Intell Syst* 2015;8:637–666. [\[CrossRef\]](#)
- [8] Molodtsov D. Soft set theory: first results. *Comput Math Appl* 1999;37:19–31. [\[CrossRef\]](#)
- [9] Maji PK, Roy AR, Biswas R. An application of soft sets in a decision-making problem. *Comput Math Appl* 2002;44:1077–1083. [\[CrossRef\]](#)
- [10] Maji PK, Biswas R, Roy AR. Soft set theory. *Comput Math Appl* 2003;45:555–562. [\[CrossRef\]](#)
- [11] Rahman AU, Alballa T, Alqahtani H, Khalifa HAEW. A fuzzy parameterized multiattribute decision-making framework for supplier chain management based on picture fuzzy soft information. *Symmetry* 2023;15:1872. [\[CrossRef\]](#)
- [12] Vimala J, Begam SS, Saeed M, Khan KA, Rahman AU. An abstract context to lattice-based ideals (filters) with multi-fuzzy soft settings. *New Math Nat Comput* 2023;1–15. [\[CrossRef\]](#)
- [13] Shabir M, Naz M. On soft topological spaces. *Comput Math Appl* 2011;61:1786–1799. [\[CrossRef\]](#)
- [14] Majumdar P, Samanta SK. On soft mappings. *Comput Math Appl*. 2010;60:2666–2672. [\[CrossRef\]](#)
- [15] Feng F, Li C, Davvaz B, Ali MI. Soft sets combined with fuzzy sets and rough sets: a tentative approach. *Soft Comput* 2010;14:899–911. [\[CrossRef\]](#)
- [16] Das S, Samanta SK. Soft real sets, soft real numbers and their properties. *J Fuzzy Math* 2012;20:551–576.
- [17] Das S, Samanta SK. On soft complex sets and soft complex numbers. *J Fuzzy Math* 2013;21:195–216.
- [18] Saeed M, Hussain M, Mughal AA. A study of soft sets with soft members and soft elements: a new approach. *Punjab Univ J Math* 2020;52:1–15.
- [19] Saeed M, Shafique I, Rahman AU, El-Morsy S. Soft algebraic structures embedded with soft members and soft elements: an abstract approach. *J Math Comput Sci* 2024;33:250–263. [\[CrossRef\]](#)
- [20] Tantawy OA, Hassan RM. Soft real analysis. *J Prog Res Math* 2016;8:120–129.
- [21] Irfan I. Convex and concave sets based on soft sets and fuzzy soft sets. *J New Theory* 2019;29:101–110.
- [22] Salih HM, Sabir PO. Some new properties of convex and concave soft sets. *Diyala J Pure Sci* 2019;15:87–98. [\[CrossRef\]](#)
- [23] Majeed SN. Some notions on convex soft sets. *Ann Fuzzy Math Inform* 2016;12:517–526.
- [24] Rahman AU, Saeed M, Arshad M, Ihsan M, Ahmad MR. (m; n)-Convexity-cum-Concavity on fuzzy soft set with applications in first and second sense. *Punjab Univ J Math* 2021;53:19–33. [\[CrossRef\]](#)
- [25] Rahman AU, Saeed M, Ihsan M, Arshad M, Ayaz S. A conceptual framework of m-convex and m-concave sets under soft set environment with properties. *Trans Math Comput Sci* 2021;1:40–60.
- [26] Ihsan M, Rahman AU, Saeed M, Khalifa HAEW. Convexity-cum-concavity on fuzzy soft expert set with certain properties. *Int J Fuzzy Log Intell Syst* 2021;21:233–242. [\[CrossRef\]](#)
- [27] Ihsan M, Saeed M, Rahman AU. A rudimentary approach to develop context for convexity cum concavity on soft expert set with some generalized results. *Punjab Univ J Math* 2021;53:621–629. [\[CrossRef\]](#)
- [28] Buckley J. Fuzzy complex numbers. *Fuzzy Sets Syst* 1989;33:333–345. [\[CrossRef\]](#)

- [29] Demir I. Some soft topological properties and fixed soft element results in soft complex valued metric spaces. *Turk J Math* 2021;45:971–987. [\[CrossRef\]](#)
- [30] Selvachandran G, Maji PK, Abed IE, Salleh AR. Relations between complex vague soft sets. *Appl Soft Comput* 2016;47:438–448. [\[CrossRef\]](#)
- [31] Selvachandran G, Maji PK, Abed IE, Salleh AR. Complex vague soft sets and its distance measures. *J Intell Fuzzy Syst* 2016;31:55–68. [\[CrossRef\]](#)
- [32] Sabir PO. Notes on extension of fuzzy complex sets. *Tikrit J Pure Sci* 2017;22:88–93. [\[CrossRef\]](#)
- [33] Wardowski D. On a soft mapping and its fixed points. *Fixed Point Theory Appl* 2013;2013:1–11. [\[CrossRef\]](#)
- [34] Öztürk TY. On bipolar soft points. *TWMS J Appl Eng Math* 2020;10:877–885.
- [35] Senel G. A comparative research on the definition of soft point. *Int J Comput Appl* 2017;163:1–5. [\[CrossRef\]](#)
- [36] Taşköprü K, Altıntaş I. A new approach for soft topology and soft function via soft element. *Math Methods Appl Sci* 2021;44:7556–7570. [\[CrossRef\]](#)
- [37] Allam A, Ismail TH, Muhammed R. A new approach to soft belonging. *Ann Fuzzy Math Inform* 2017;13:145–152. [\[CrossRef\]](#)
- [38] Polat NÇ, Yaylalı G, Tanay B. Some results on soft element and soft topological space. *Math Methods Appl Sci* 2019;42:5607–5614. [\[CrossRef\]](#)
- [39] Hameed RA, Khalil MM. On equality of infimum soft sequences. *Int J Math Trends Technol* 2018;55. [\[CrossRef\]](#)
- [40] Mahmood A, Abbas M, Murtaza G. Observation of experts, attitudes through multi-criteria decision-making. *Sigma J Eng Nat Sci* 2023;41:926–937. [\[CrossRef\]](#)
- [41] Kirişçi M. Generalized Pythagorean fuzzy sets and new decision-making method. *Sigma J Eng Nat Sci* 2022;40:806–813. [\[CrossRef\]](#)
- [42] Khattak AM, Abdullah S, Khattak ZA. A study on some fundamental characteristics of sequences of soft real numbers. *Acta Sci Malays* 2018;2:8–13. [\[CrossRef\]](#)



Research Article

Electrokinetic effect on flow of two-phase fluid in a circular tube

Sangeetha J^{1,*}, Ponalagusamy R¹

¹Department of Mathematics, National Institute of Technology, Tiruchirappalli, Tamil Nadu, 620015, India

ARTICLE INFO

Article history

Received: 29 February 2024

Revised: 12 June 2024

Accepted: 24 September 2024

Keywords:

Electric Field; K-L Fluid;
Newtonian Fluid, Two-Phase
Model, Wall Shear Stress

ABSTRACT

In this article, we investigated the steady-state motion of two-phase immiscible fluids through a rigid wall tube. The fluid confined in the mid-region is represented as Kuang-Luo (K-L) fluid and the outer, region is occupied by Newtonian fluid. Both the fluid is considered as an electrically conducting fluid and the fluid undergoes the electroosmotic effect when an electric field is applied to it, externally. Based on the circumstances, with an appropriate boundary condition the analytical formulation for the velocity profiles of the fluid, flow flux, motion in the plug core region, plug core radius and wall shear stress was done. The numerical analysis of simulation effects was made on the K-L parameters, electro-osmotic effect and yield stress. This case study reveals the fact that whenever the K-L parameter (θ_2) and the yield stress (τ_0) increases, the velocity profiles of the fluid decrease and the opposite behaviour is identified by increasing the electrokinetic term and K-L parameter (θ_1) found in dealing with the wall shear stress properties. The significance of formulating the K-L constitutive equation using an empirical formula obtained from the experimental values of apparent viscosity, shear rate and yield stress of the human and canine blood samples provides the best approximate result. Aspired from this accuracy, the model has been constructed and analysed with appreciable values under the external electric field in the rigid circular conduit and the graphical results were displayed. A comparison of the various existing models with the present model was made and the applicability of the prescribed model provides a generalized case of non-Newtonian fluids like Casson fluid, Bingham Fluid and Newtonian fluid. Additionally, an analysis of the proposed mathematical model's significance on physiological applications has been discussed. For the first time, it is reported that for any value of the yield stress, the flow of two-phase immiscible fluids exhibits a realistic flow phenomenon when $\theta_1 \leq \theta_2$. The coarse-grained consequential amount of declined percentage has been seen from the two-layered Newtonian-Newtonian model is approximately higher than the Casson-Newtonian (6.74%), Bingham-Newtonian (17.92%) and K-L-Newtonian (22.77%) in velocity profiles. In the absence of an electric field ($U_e = 0$), Casson fluid is approximately 1.02% higher than the presence of an electric field ($U_e = 0.5$) while dealing with wall shear stress and similarly for the K-L model ($U_e = 0$) is appropriately greater than 0.078 % for $U_e = 0.5$. The pressure drop (Δp) for the K-L parameter $\theta_1 = 0.1, 0.2$ is approximately 0.21 % higher than the $\theta_1 = 0.3$ for the fixed viscosity index $\theta_2 = 0.3$. The results of the prescribed model are highly beneficial for simulating electrically conductive lubricants as applications of the engineering field.

Cite this article as: Sangeetha J, Ponalagusamy R. Electrokinetic effect on flow of two-phase fluid in a circular tube. Sigma J Eng Nat Sci 2025;43(5):1760–1784.

*Corresponding author.

*E-mail address: sangeethajagan009@gmail.com

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



INTRODUCTION

The flow pattern of fluid motion will vary depending upon the conduit and the environs which it presents. If the fluid is subjected to the multiphase fluid with the nature of immiscibility, its flow behaviour depends on the adjacent fluid also. Many researchers were involved in studying the immiscible characteristics of the fluids which undergone certain circumstances. Immiscible fluids are two or more fluids that are incapable of forming a homogeneous mixture or solution when combined. Because, immiscible properties are possessed by the strong intermolecular force of one fluid over another, different densities, viscosity, molecular structures and polarities. So, it is restricted to providing an adherence nature. These kinds of fluids are seen in our day-to-day life. In the biological system, the flow of blood in blood vessels, and food through the alimentary canal and in Industrial applications, emulsion formation, the extraction process, separation techniques, lubrications, fuel and cleaning and degreasing. The experimental work on the two-phase viscous flow in the case of oil pipelines was investigated by Michael Bentwich [1]. He performed this experiment in order to reduce the pumping costs. This research has concluded that the potentiality of the solution was made a valuable acceptance and this method produces the analogue to various problems in the fields of elasticity, heat and mass transfer, mechanics and the electric potential distribution. Following the validation of experimental work, Packham and Shail [2] were inspired to study the flow of two immiscible viscous fluids in the long horizontal pipe which is independent of time. He made a statement that the influence of the interface region between two fluids was very effective during the flow in a fully developed region because of the symmetricity of the conduit. On behalf of bio-rheological applications, Chaturani and Ponalagusamy [3] studied the behaviour of blood flow in the stenosed arteries by pondering blood vessels as a two-layer model. In the presence of a porous region, Srinivasan and Vafai [4] started the pioneer studies on the theoretical analysis of the two immiscible fluids. Furthermore, they made a statement about the physical importance of the porous material and flow effects in the presence of porous material. Following them, Chamkha [5] distinguished the difference between flow traits of electrically conducting two-phase immiscible fluids in the presence of porous and non-porous regions in the symmetric channel. In addition to that, the impact of electrically conducting fluids was more efficient in the porous region than in the non-porous region. Umavathi et al. [6] investigated the two-phase immiscible viscous fluids under the unsteady oscillatory motion along with heat transfer through a horizontal channel in the presence of an isothermal permeable wall. Meanwhile, the laminar condition on the fluid flow was very important to investigate in the case of two-phase immiscible fluids in order to utilize the bio-engineering fields like microchips, biochips, Lab-on-Chip etc... Based on this statement, Umavathi et

al. [7] intervened in the motion of two immiscible viscous fluids in the long vertical wavy channel under the laminar condition. Prathap Kumar et al. [8] formulated the analytical solution of free convective two-layer immiscible micropolar and viscous fluids through the vertical channel. Ponalagusamy and Tamil Selvi [9] intervened in the stenosed arteries in the form of two-layered (Casson-Newtonian) blood flow along with variable slip wall conditions by assuming the axisymmetric cylindrical tube. Using a perturbation technique, Umavathi and Shekar [10] the mixed convective of fully developed laminar flow was interpreted by assuming different conductivity and viscosity in the two-layer viscous immiscible fluids. Along with the different aspects of immiscible fluids, it becomes necessary to analyse the immiscible combination of Newtonian and non-Newtonian. Such studies made a huge impact on industrial applications. Devakar and Ramgopal [11] performed an analytical study on immiscible fluids like Couple-stress fluid and Newtonian fluid through the circular cylinder in the presence of both porous and non-porous regions by assuming the condition axisymmetric nature. In the presence of particles or contamination, the fluids are assumed to be micropolar and the immiscible nature of micropolar-Newtonian in the porous region through the circular conduit in the axisymmetric condition has been analysed by Yadav and Verma [12]. Considering an optimal amount of density difference, the miscible nature of high-viscosity fluid with low-viscosity fluid was analysed by Akbari et al. [13]. Gemma Houston et al. [14] investigated the microfluidic devices that have dealt with miscible characteristics of the fluids and the experiments were also carried out in well-acceptable optimized value. The magnetohydrodynamics flow of variable viscosity in the two-phase model was analysed in the configuration of blood flow through the stenosed arteries analytically using the Frobenius method by Ponalagusamy [15].

Initially, several authors deal with Newtonian fluid under various circumstances. Newtonian fluid with two different viscosities possessing an immiscible nature was investigated by Rao et al. [16] under peristaltic motion. Vajravelu et al. [17] discussed the unsteady immiscible Newtonian fluid under a magnetic field between the permeable beds. The corrugated vertical channel was analysed using the perturbation technique by Umavathi et al. [18]. Using MHD flow in the isothermal wall, Dragiša D. Nikodijević et al. [19] interpreted the three immiscible fluids in a channel subjected to the horizontal position. In the periodically constricted tube with constant pressure, the immiscible Newtonian fluids were studied by Fraggedakis et al. [20]. Abd Elmaboud [21] analysed the impact of the magnetic field in a vertical semi-corrugated channel for two immiscible fluids. The combined effect of non-Newtonian and Newtonian fluids was considered an immiscible property and the interpretation was carried out in various circumstances. Devakar and Ramgopal [22] studied the immiscible flow of couple stress- Newtonian fluid in the

presence of porous and non-porous medium through a tube. The discussion of multiphase fluid immiscible models like micropolar-Newtonian fluid was interpreted under the assumption of unsteady motion and heat transfer by Devakar and Ankush Raje [23]. From the perspective of various liquid combination models subjected to immiscible nature, viscoelastic effects were studied by Seng Hoe Hue et al. [24]. Shan Fang et al. [25] made discussion in a pore doublet model, the immiscibility was interpreted by considering the effects of geometric structure also. The investigation on the unsteady flow of two immiscible electromagnetic effects in the porous medium by Sangeetha and Ponalagusamy [26].

Owing to this fact, the real-life application is probably based on non-Newtonian fluids. Non-Newtonian fluids contain variable viscosity, yield stress, non-linear shear rate components and other parameters. Depending on the fluid nature, a suitable non-Newtonian fluid will be taken to model. In both the biomedical and industrial fields, the fluids possessed non-Newtonian behaviour. Because the immiscible nature only occurs in the presence of highly complex molecular structures. Some of the non-Newtonian fluids are Power-law fluid, K-L model, Sutterby fluid, Casson fluid, Quemada model fluid, Jeffrey fluid, Thixotropic fluid, Hershey-Bulkley fluid, second-order fluid, Bingham fluid, etc., were made simulation by various researchers using various techniques. Akbari et al., intervened in the flow of micropolar fluid on the linearly stretched porous sheet using a homotopy perturbation technique [27]. Applying Akbari-Ganji's Method on the complex two-phase nanofluids on the revolving system in the presence of a uniform magnetic field was investigated by Gholinia et al. [28]. Implanting the highly precision numerical method known as a variational iterative method, convective 2D micropolar fluid flow in the permeable channel was probed by Gholinia et al. [29]. On a theoretical basis, Dewangan et al., studied the superficial velocity and the temperature ratio of two-phase stratified fluid flow through the horizontal channel using the ANSYS fluent [30]. Özkan et al., experimentally computed the utilization of nanofluids on the conventional reinforced materials to analyse the mechanical efficiency of the wind turbine [31]. Employing the spectral collocation method, the computation analysis of the thermal conductance of the fluid properties has been carried out on the Casson nanofluid flow in the presence of variable magnetic field Akaje et al. [32]. Hattab et al. [33] interpreted the flow of convective magnetohydrodynamics of carbon nanotube and water as base fluid with the isothermal fins under the uniform magnetic field using a finite control volume method in the square enclosed conduit.

Among all other non-Newtonian models, the K-L model equation consists of three parameters which were experimentally verified by Luo and Kuang [34]. The model was constructed by analysing the shear-thinning behaviour of human blood. By using the weighted least square method, they formulated the constitutive equations

with experimental data and this model satisfies with shear thinning characteristics of blood with high and low shear rates. Based on the human blood component, the variables present in this model are the yield stress and functions of blood components like hematocrit, the viscosity of plasma component and other additional chemical variables. The comparative analysis of the Casson fluid was made by Zhang and Kuang [35] experimentally by making certain modifications in the K-L model proving that it is far better than providing the pseudoplastic properties than the Casson fluid. Mostly, the K-L model has been dealing with blood flow behaviour. Somchai Sriyab [36] investigated the K-L blood flow in the stenosis and the narrow arteries region. In the presence of a non-symmetrical stenosed narrow artery, characteristics of the blood (K-L model) flow were analysed by Rekha Bali and Nivedita Gupta [37]. Ponalagusamy and Ramakrishna Machi [38-40], dealt with the two-layered (K-L- Newtonian) blood flow of mild stenosed arteries in six various types, in the presence of porous walls with stenosed arteries and pulsatile flow in the tapered stenosis as single-layer K-L model. Using Aris's method of moments, Shalini Singh and Murthy [41] have investigated the unsteady behaviour of solute dispersion in the pulsatile K-L fluid motion in the rigid circular conduit. In this examination, they computed the higher order moments like the 3rd and 4th central moments for the mean concentration and found that the dispersion of the solute provides precise information for the higher order moments. By considering blood as a K-L model, the dispersion of soluble matter in blood flow has been analysed by Ponalagusamy et al. [42]. In these works, the impact of the electric field on the flow is not investigated. One can derive other types of fluids by making some components in appropriate measured values which will be discussed later.

As so far, in the prior prescribed discussion, the effect of an electrokinetic factor was least discussed. While dealing with the electrically conducting fluid, it is possible to make a single-phase or multi-phase fluid undergone for the external electric field. Electrokinetic effects have a wide range of applications. Mechanism-based on the electroosmotic phenomena is when the wall of the tube is charged with positive (or negative), the fluid present in the tube can move counter ions near the wall. It tends to form a strong bond near the wall. Due to the external electric field, it possesses a strong electric field near the wall and tends to form the second layer. These layers are called the electric double layer.

Based on this electrostatic phenomenon, many authors interpret based on various kinds of fluid. In the existence of an external electric field, the motion of two immiscible fluids in the microchannel was examined by Gao et al. [43]. Zhao and Yang [44] discussed the behaviour of electrokinetic phenomena applied to the Newtonian fluid in detail. Afonso et al. [45] investigated analytically the two-phase electro-osmotic flow of stratified fluids based on the behaviour of viscoelastic nature. In the formation of

electroosmotic flow with multilayer fluid in the existence of a peristaltic effect, many authors discussed. In the existence of peristaltic motion in the wall, two-layered: Power Law and Newtonian models were discussed by Goswami et al. [46]. Considering the two immiscible fluid flows consisting of the electrically conducting fluid and non-conducting fluids subjected to Newtonian fluid was discussed by Harshad Gaikwad et al. [47] presented their work analytically. Mithilesh et al. [48] proposed the bio-rheological factors of micropolar fluids through the microfluidic channels with peristaltic wave walls under the electro-osmotic flow. Based on the peristaltic motion, the Ellis-Newtonian model was investigated by Ali et al. [49] and Hussain and Ali [50] analysed the Multilayer Power-law. Ponalagusamy and Sangeetha [51] interpreted the flow characteristic of the two-phase (Thixotropic- Newtonian) model in the presence of electroosmotic force. Furthermore, the influence of the electro-osmotic effect on several characteristic assumptions was carried out by many authors. The pulsatile nature of nanofluid in the microchannel was analysed by Mukherjee et al. [52]. Tanoy Kahali et al. [53], interpreted the trajectory of a compound drop in a microfluidic channel and numerical investigations were performed by Mid Ismaveel et al [54], by assuming the Electrodiffusio-osmotic transport in nanochannels under various parameters. Followed by fellow researchers, the effective motion of electrically conducting-non-conducting immiscible fluids has been computed analytically using periodic motion on electric under the external impact of the magnetic field was formulated by Ponalagusamy and Sangeetha [55]. Sangeetha et al. [56], computed the electric field impact on the time-dependent peristaltic motion of two-phase immiscible fluids in the elastic tube using the perturbation technique under the magnetic field through the porous medium. In this case, they analyse the influence of the electroosmotic effect at the interface region. The impact of electroosmotic force in electrically conducting fluids is worth noticing. In this case, the electroosmotic effect on the peristaltic motion of immiscible flow in the circular conduit along with the variable viscosity coefficient has been studied analytically by Sangeetha et al. [57].

In the aforementioned studies, the interpretation of the immiscible fluids becomes very complex when dealing with the combined effects of non-Newtonian and Newtonian fluids, particularly the non-Newtonian fluids since it consists of highly complex structures in nature. On behalf of dealing with other non-Newtonian fluids which possess a viscoelastic nature, the K-L model proposed by Kuang and Luo was purely done by the experimental process based on the human and canine blood and its properties formulating the equations using the empirical formula using the values of apparent viscosity, shear rate and yield stress. So far, the studies have not described the limitations while dealing with the K-L parameters. In our research, the formulation of the limiting case was carried on with the suitable mathematical condition. Moreover, the impact of external body

forces like an electrokinetic body force made an efficient flow in the prescribed conduit which is also considered in this work. Inspired by these quantities, the model was constructed in a possible way which suggests dealing with the industrial-based application under different environments. Therefore, the analysis is mathematically modelled by two-layered immiscible flow in the existence of the external electric field. The mid region represents K-L fluid and the outer region is Newtonian fluid. The mathematical formulation of velocity profiles, fluid flow rate, the plug core radius and the wall shear stress are expressed analytically with properly prescribed boundary conditions. Our present investigation became the generalized model of certain non-Newtonian fluids like Casson fluid, Bingham fluid and Newtonian fluid.

MATHEMATICAL FORMULATION

Geometrical Description

Consider the steady-state, axisymmetric, incompressible, laminar and two-phase motion of immiscible fluids subjected to the electrically conducting properties through the cylindrical coordinate system $(r_1^*, \theta_1^*, z_1^*)$, where r_1^* represents the direction towards the radial distance, θ_1^* indicates the direction over the circumferential rotation and z_1^* denotes the direction of the axial distance. Meanwhile, the flow towards θ_1^* Direction dissipated because of the axisymmetric nature of the model. Based on the above assumptions, the motion in the present model projects the unidirectional way in the axial distance. The suggested model contains two sections: the mid-area named as K-L fluid model and the outer sector named Newtonian fluid. In Figure 1, the configured shape of the presently studied two-phase model is exhibited in a well-manner form.

Charge Distribution in the Pipe

In order to analyse the charge distribution, certain assumptions need to be taken into account. The surface of the tube should be homogenous. In heterogeneous material used, the charge distribution over a concern region will vary and it might cause variation in flow pattern. Owing to the phenomena of electrostatic theory, the formulation to construct the potential charge distribution which leads to the disposition of an Electric Double Layer (EDL) at the surface of the wall which obeys the linearized form of Poisson-Boltzmann equation,

$$\varepsilon^* \nabla^2 \psi^* = -\rho_e^* \quad (1)$$

where ψ^* denoted the potential distribution of electric force, ε^* represents the permittivity of the medium and ρ_e^* indicates the net charge density which formulated as $\rho_e^* = ez^\pm(n^+ - n^-)$. Here, the components present in the net charge density were composed of the electric charges (e), the valence of the respective charged ions ($z^\pm$) and the

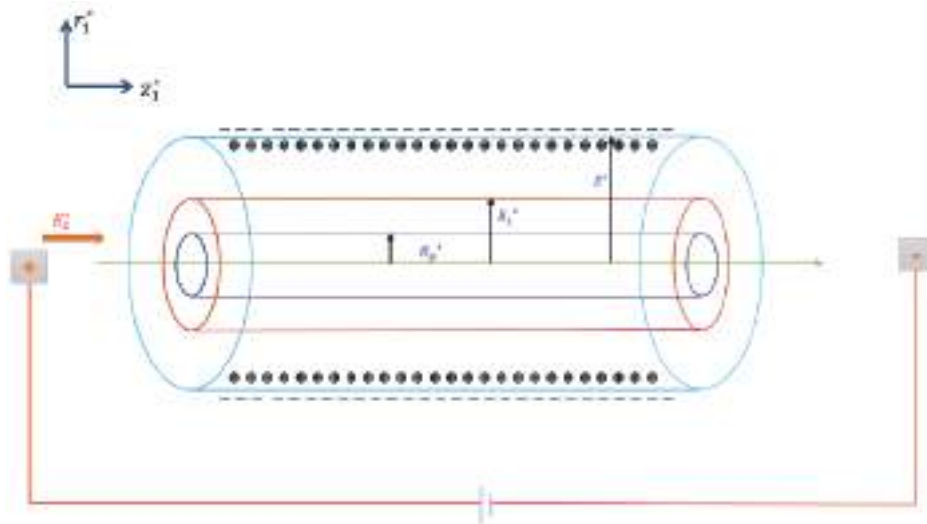


Figure 1. A visual depiction of the prescribed mathematical model in schematic form. In the current study, the wall of the tube assumed as negatively charged and the fluid area has a net positive charge which stabilises the surface charge.

number of the density of ions and counterions (n^*). Since, the fluid flow acquaintance with the additional body force term namely electrokinetic force because of the external electric field E_z^* [55-57].

$$F_{EK}^* = -\rho_e^* E_z^* \quad (2)$$

Fluid Flow for Velocity Distribution

The corresponding continuity and the momentum equations present in both the regions are displayed as follows:

- i. For core region (K-L fluid): $0 \leq r_1^* \leq R_1^*$

$$\frac{\partial w_{11}^*}{\partial z_1^*} = 0 \quad (3)$$

$$-\frac{\partial p_0^*}{\partial z_1^*} - \frac{1}{r_1^*} \frac{\partial}{\partial r_1^*} \left(r_1^* \tau_{r_1^* z_{11}^*}^* \right) + F_{EK}^* = 0 \quad (4)$$

- ii. For peripheral region: $R_1^* \leq r_1^* \leq R^*$

$$\frac{\partial w_{12}^*}{\partial z_1^*} = 0 \quad (5)$$

$$-\frac{\partial p_0^*}{\partial z_1^*} + \frac{1}{r_1^*} \frac{\partial}{\partial r_1^*} \left(r_1^* \tau_{r_1^* z_{12}^*}^* \right) + F_{EK}^* = 0 \quad (6)$$

where w_{11}^* , w_{12}^* represents the velocity distribution occupies the areas of the core and the peripheral layers in the axial direction, r_1^* is the length of the radial direction and $-\frac{\partial p_0^*}{\partial z_1^*}$

indicated the pressure gradient. “*” implies the dimensional quantities.

The constitutive equation for the corresponding fluids is given as

$$\tau_{r_1^* z_{11}^*}^* = \tau_0^* + \eta_2^* \gamma^* + \eta_1^* \sqrt{\gamma^*}; \quad \tau_{r_1^* z_{11}^*}^* > \tau_0^* \quad (7)$$

$$\frac{\partial w_{11}^*}{\partial z_1^*} = 0; \quad \tau_{r_1^* z_{11}^*}^* \leq \tau_0^* \quad (8)$$

$$\tau_{r_1^* z_{12}^*}^* = \mu_n^* \frac{\partial w_{12}^*}{\partial r_1^*}; \quad (9)$$

where the stress components act in the $r_1^* z_1^*$ plane, τ_0^* , η_1^* , η_2^* , γ^* and μ_n^* are yield stress and functions of the blood components like hematocrit, the viscosity of plasma components, other additional chemical variables, shear rate and viscosity in the Newtonian region respectively.

The relative boundary condition in the dimensional form is

$$\text{i) at } r_1^* = 0; \tau_{r_1^* z_{11}^*}^* \text{ is finite} \quad (10)$$

$$\text{ii) at } r_1^* = R_p^*; \frac{\partial w_{11}^*}{\partial z_1^*} = 0 \quad (11)$$

$$\text{iii) at } r_1^* = R_1^*; w_{11}^* = w_{12}^*; \tau_{r_1^* z_{11}^*}^* = \tau_{r_1^* z_{12}^*}^* \quad (12)$$

$$\text{iv) at } r_1^* = R^*; w_{12}^* = 0 \quad (13)$$

where R_p^* , R_1^* and R^* are plug core radius, interfacial thickness and height of the tube from the axial direction to the outer wall. Solving the momentum equations using the respective boundary conditions, the following dimensional

quantities have been implemented to make the non-dimensional form,

$$\begin{aligned} r_1 &= \frac{r_1^*}{R^*}; z_1 = \frac{z_1^*}{R^*}; R_1 = \frac{R_1^*}{R^*}; w_{11} = \frac{w_{11}^*}{U^*}; \\ w_{12} &= \frac{w_{12}^*}{U^*}; p_0 = \frac{R^*}{U^* \mu_n^*} p_0^*; \tau_{r_1 z_{11}} = \frac{R^*}{U^* \mu_n^*} \tau_{r_1 z_{11}}^*; \\ \tau_0 &= \frac{R^*}{U^* \mu_n^*} \tau_0^*; F_{EK} = \frac{R^{*2}}{\varepsilon^* \zeta^* \mu_n^*} F_{EK}^*; \theta_1 = \frac{\eta_1^*}{\mu_n^*} \sqrt{\frac{R^*}{U^*}}; \theta_2 = \frac{\eta_2^*}{\mu_n^*} \end{aligned}$$

where U^* is reference velocity.

The dimensionless governing equations are as follows,

i. For core region: $0 \leq r_1 \leq R_1$

$$\frac{\partial w_{11}}{\partial z_1} = 0 \quad (14)$$

$$-\frac{\partial p_0}{\partial z_1} - \frac{1}{r_1} \frac{\partial}{\partial r_1} (r_1 |\tau_{r_1 z_{11}}|) + U_e \rho_e = 0 \quad (15)$$

ii. For peripheral region: $R_1 \leq r_1 \leq 1$

$$\frac{\partial w_{12}}{\partial z_1} = 0 \quad (16)$$

$$-\frac{\partial p_0}{\partial z_1} + \frac{1}{r_1} \frac{\partial}{\partial r_1} \left(r_1 \frac{\partial w_{12}}{\partial r_1} \right) + U_e \rho_e = 0 \quad (17)$$

where $U_e = -\frac{\varepsilon^* \zeta^* E_z^*}{\mu_n^*}$ is an electrokinetic slip velocity [39,42-44,50].

The constitutive equation (non-dimensional form) is given as

$$\tau_{r_1 z_1} = \tau_0 + \theta_1 \sqrt{-\frac{\partial w_{11}}{\partial r_1} - \theta_2 \frac{\partial w_{11}}{\partial r_1}}; \quad \tau_{r_1 z_1} > \tau_0 \quad (18)$$

$$\frac{\partial w_{11}}{\partial r_1} = 0; \quad \tau_{r_1 z_1} \leq \tau_0 \quad (19)$$

where θ_1 , θ_2 and τ_0 are the K-L parameters and yield stress. The respective boundary conditions are

$$\text{i) at } r_1 = 0; \tau_{r_1 z_{11}} \text{ is finite} \quad (20)$$

$$\text{ii) at } r_1 = R_p; \frac{\partial w_{11}}{\partial r_1} = 0 \quad (21)$$

$$\text{iii) at } r_1 = R_1; w_{11} = w_{11}; \tau_{r_1 z_{11}} = \tau_{r_1 z_{12}} \quad (22)$$

$$\text{iv) at } r_1 = 1; w_{12} = 0 \quad (23)$$

To solve the aforementioned non-linear momentum equations, two kinds of techniques have been invoked i.e. either the solution of the equation can be obtained by

momentum equation in addition to the electro-kinetic body force term or to transfer its participation in the form of suitable boundary conditions at the outer wall as electro-kinetic slip parameter [46,49-51,57]. In this study, we opt for the second type by shifting the electro-kinetic body force term to the no-slip boundary condition at the conduit wall, and then boundary condition (23) becomes

$$\text{at } r_1 = 1; w_{12} = U_e \quad (24)$$

After transferring the slip parameter, the momentum equations become

$$-\frac{\partial p_0}{\partial z_1} - \frac{1}{r_1} \frac{\partial}{\partial r_1} (r_1 |\tau_{r_1 z_{11}}|) = 0 \quad (25)$$

$$-\frac{\partial p_0}{\partial z_1} + \frac{1}{r_1} \frac{\partial}{\partial r_1} \left(r_1 \frac{\partial w_{12}}{\partial r_1} \right) = 0 \quad (26)$$

The Analytical Expression for the Solution

Solving the momentum equations (25) and (26), applying the boundary conditions (20)-(22) and (24), the mathematical expression of the velocity distribution in the two sectors is obtained as

In $R_p \leq r_1 \leq R_1$.

$$\begin{aligned} w_{11} &= \frac{P}{4\theta_2} (\theta_2 + R_1^2(1 - \theta_2) - r_1^2) + \frac{1}{2\theta_2^2} (2\tau_0\theta_2 - \theta_1^2)(r_1 - R_1) \\ &+ \frac{\theta_1}{6P\theta_2^3} ((\theta_1^2 + 2\theta_2PR_1 - 4\theta_2\tau_0)^{3/2} - (\theta_1^2 + 2\theta_2PR_1 - 4\theta_2\tau_0)^{3/2}) + U_e \end{aligned} \quad (27)$$

In $R_1 \leq r_1 \leq 1$:

$$w_{12} = \frac{P}{4} (1 - r_1^2) + U_e \quad (28)$$

The rate at which fluid is moving in the plug-core region is

In $0 \leq r_1 \leq R_p$:

$$\begin{aligned} w_{11} &= \frac{P}{4\theta_2} (\theta_2 + R_1^2(1 - \theta_2) - R_p^2) + \frac{1}{2\theta_2^2} (2\tau_0\theta_2 - \theta_1^2)(R_p - R_1) \\ &+ \frac{\theta_1}{6P\theta_2^3} ((\theta_1^2 + 2\theta_2PR_p - 4\theta_2\tau_0)^{3/2} - (\theta_1^2 + 2\theta_2PR_1 - 4\theta_2\tau_0)^{3/2}) + U_e \end{aligned} \quad (29)$$

where $P \left(= -\frac{\partial p_0}{\partial z_1} \right)$ is pressure gradient and R_p is the radius of the plug-core region formulated as $R_p = \frac{2\tau_0}{P}$. The flow flux of a given region, Q (in non-dimensional form) is

$$\begin{aligned}
Q &= 2 \int_0^{R_p} r_1 w_{pc} dr + 2 \int_{R_p}^{R_1} r_1 w_{11} dr + 2 \int_{R_1}^1 r_1 w_{12} dr \\
Q &= -\frac{PR_p^4}{8\theta_2} - \frac{2\tau_0\theta_2 - \theta_1^2}{3\theta_2^2} R_1 R_p^2 + \frac{P}{8\theta_2} (R_1^4(1 - \theta_2) + \theta_2) + \frac{2\tau_0\theta_2 - \theta_1^2}{6\theta_2^2} R_1^3 \\
&\quad - \frac{\theta_1^6}{105P^3\theta_2^5} (14\tau_0\theta_2 P - \theta_1^2) + (A + 2\theta_2 P R_p)^{\frac{3}{2}} (R_p^2 - R_1^2) \quad (30) \\
&\quad + \frac{2R_1}{5\theta_2 P} (A + 2\theta_2 P R_1)^{\frac{5}{2}} - \frac{2R_p}{5\theta_2 P} (A + 2\theta_2 P R_p)^{\frac{5}{2}} \\
&\quad + \frac{2}{35\theta_2^2 P^2} (A + 2\theta_2 P R_p)^{\frac{7}{2}} - \frac{2}{35\theta_2^2 P^2} (A + 2\theta_2 P R_1)^{\frac{7}{2}} + U_e
\end{aligned}$$

where $A = \theta_1^2 - 4\theta_2\tau_0$.

By assuming the flow flux (Q) value, the corresponding pressure gradient (P) values have been calculated with suitable values of the other parameters using MATLAB programming. By using the equations (31) and (28), one can obtain the expression for τ_w ,

$$\tau_w = \left(-\frac{dw_{12}}{dr_1} \right)_{r_1=1} \quad (31)$$

VALIDATION RESULTS

The flow of the two-phase immiscible fluids through the circular tube which was mentioned above has a tendency to perform like a generalized form of other fluid models, viz Newtonian-Newtonian, Bingham Plastic-Newtonian and Casson-Newtonian models with certain restrictions of the provided parameters.

Newtonian-Newtonian

The flow flux of the Newtonian-Newtonian model can be obtained by vanishing the yield stress (τ_0) and K-L parameter (θ_1) and the constitutive equation of the core region (K-L Fluid) behaves like a Newtonian fluid and its flow rate given as follows.

$$Q = \frac{P}{8\theta_2} (R_1^4(1 - \theta_2) + \theta_2) + U_e \quad (32)$$

Bingham-Newtonian

The flow flux of the Bingham-Newtonian model can be obtained by vanishing the K-L parameter (θ_1) and the constitutive equation of the core region (K-L Fluid) behaves like a Bingham fluid and its flow rate given as follows.

$$Q = \frac{P}{8\theta_2} (R_1^4 - R_p^4) - \frac{\tau_0}{3\theta_2} (R_1^3 - R_p^3) + \frac{P}{8} (1 - R_1^4) + U_e \quad (33)$$

Casson-Newtonian

The flow flux of the Casson-Newtonian model can be obtained by formulating the K-L parameter (θ_1) as $2\sqrt{\tau_0\theta_2}$ (i.e.) $\theta_1 = 2\sqrt{\tau_0\theta_2}$ and the constitutive equation of the core region (K-L Fluid) behaves like a Casson fluid and its flow rate is given as follows

$$\begin{aligned}
Q &= -\frac{PR_p^4}{8\theta_2} + \frac{\tau_0}{3\theta_2} (R_1^3 + R_p^3) + \frac{P}{8} \left(1 - R_1^4 + \frac{R_1^4}{\theta_2} \right) \\
&\quad + \frac{\sqrt{8\tau_0 P}}{3\theta_2} R_p^{7/2} - \frac{7\sqrt{8\tau_0 P}}{21\theta_2} R_p^{7/2} - \frac{6\sqrt{8\tau_0 P}}{42\theta_2} R_1^{7/2} + U_e \quad (34)
\end{aligned}$$

Graphical Validation

The graphical outcome of the present model was compared with the existing analytical model in three different cases:

- Velocity profile of different values of hematocrit with no electric osmotic flow [38].
- Two-phase flow of K-L-Newtonian model along with the absence of electric field [38].
- The one-layer flow of the K-L-Newtonian model with no electrokinetic effect [40].

Ramakrishna Manchi and Ponalagusamy [38] derived the analytical solution for steady-state two-phase K-L-Newtonian fluid in the human stenotic artery region with no-slip boundary condition. Ramakrishna Manchi and Ponalagusamy [40] discussed the pulsatile flow of K-L fluid through the stenotic region as a one-layer model analytically. Validation of the velocity profiles for the current study was evaluated from three different perspectives. First, the non-stenotic region of the hematocrit velocity in the absence of an electric field by Ramakrishna Manchi and Ponalagusamy [38] is displayed in Figure 2. Second, the two-phase velocity profiles of the K-L-Newtonian model (with no electroosmotic effect) are presented in Figure 3. Third, Figure 4 shows the validation of the one-layer model which was studied by Ramakrishna Manchi and Ponalagusamy [40] in the non-stenotic region with zero electroosmotic force. Moreover, the velocity profile values of our present work for the one-layer model were in good agreement with the existing work which is shown in Table 1. It is noticed that our results were well matched with the analysis carried out earlier by Ramakrishna Manchi and Ponalagusamy [40].

Observation and Outcomes

In the current analysis, the flow of immiscible fluids is constructed in the two-layered model with K-L fluid occupying the inner region and the outer region confined by the Newtonian fluid. The investigation was made on the flow characteristics of two electrically conducting fluids actuated by the combined effects of electro-osmotic and pressure forces under the influence of the Electric double layer (EDL) effect formed at the wall. The flow behaviour of the fluids in the form of velocity profile, flow flux, plug core radius and skin friction components have been investigated. The numerical computation of the velocity, flux, wall shear stress and plug core radius equations can be wisely calculated using the Matlab software package which will be useful for the graphical representation of the flow pattern. Yield stress, peripheral layer thickness, and electric field will have a preponderance of impacts across the tube, and

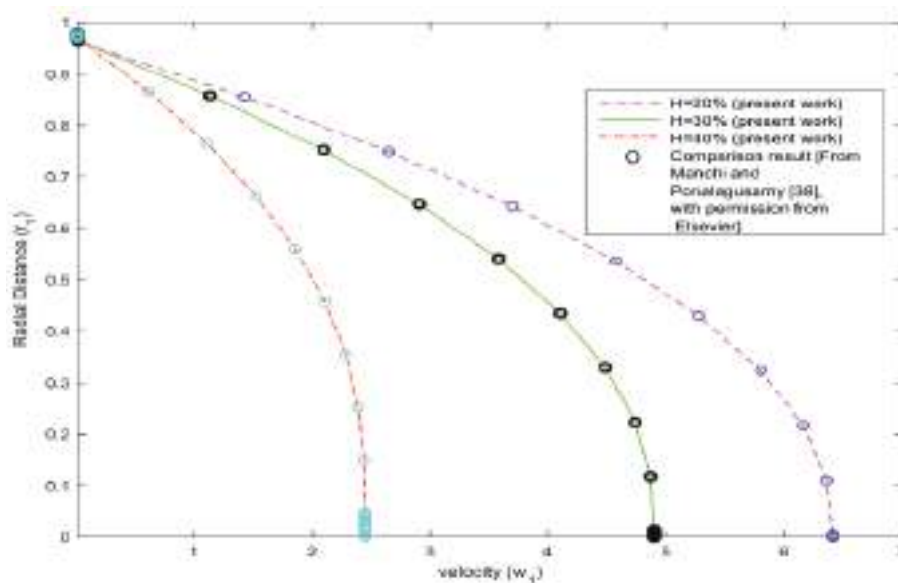


Figure 2. Velocity profiles for different values of hematocrit of the current model compared with the velocity profiles of the non-stenotic region of the two-layer model. [From Manchi and Ponalagusamy [38], with permission from Elsevier].

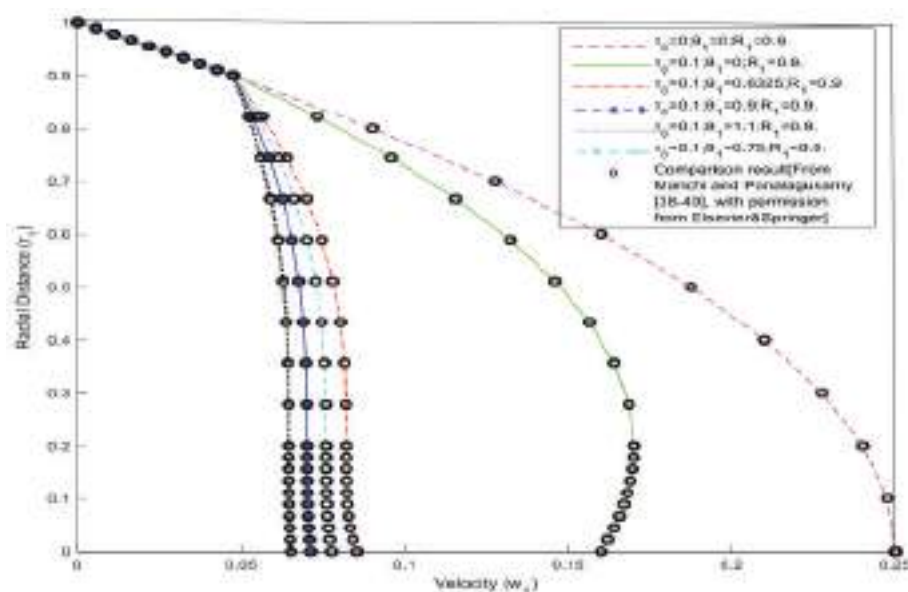


Figure 3. Velocity profiles of the current model compared with the velocity profiles of the non-stenotic region of the two-layer model. [From Manchi and Ponalagusamy [38], with permission from Elsevier].

we bring out the eminence effects of the flow dynamics, which depend on K-L parameters, and the discussion will be carried out in upcoming figures along with the values prescribed in Table 2.

The components present in the K-L constitutive equation are constructed using the empirical formula of experimental values obtained from the human and canine blood samples like the Quemada model, Bi-Exponent model, Casson model and Power-law model [34]. The apparent viscosity,

shear rate and yield stress of the hemorheology provide the better approximate result in K-L Blood Constitutive equations (BCEs) compared to the Casson model. Here, the parameters present in the K-L model, θ_1 represents the relative viscosity that may be related to the aggregative level of Red Blood Cells and θ_2 denotes the viscosity index that may be related to the aggregation of Red Blood Cells. Since there are many BCEs but that cannot be directly related to the aggregation and deformability of RBC it is necessary to

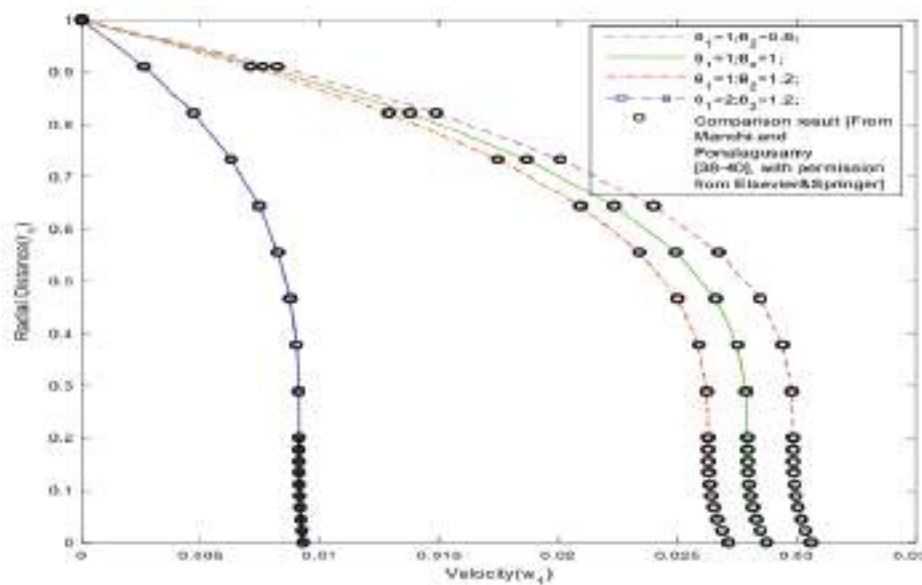


Figure 4. Velocity profiles of the current model compared with the velocity profiles of a non-stenotic region of a one-layer model.[From Manchi and Ponalagusamy [38-40], with permission from Elsevier & Springer].

Table 1. Velocity profiles of the current model compared with the velocity profiles of the non-stenotic region of the one-layer model. [From Manchi and Ponalagusamy [40], with permission from Springer]

$\theta_1 = 2; \theta_2 = 1.2$		$\theta_1 = 1; \theta_2 = 1.2$		$\theta_1 = 1; \theta_2 = 1$		$\theta_1 = 1; \theta_2 = 0.8$	
Manchi & Ponalagusamy [40]	Present work	Manchi & Ponalagusamy [40]	Present work	Manchi & Ponalagusamy [40]	Present work	Manchi & Ponalagusamy [40]	Present work
0	0	0	0	0	0	0	0
0.0051	0.0051253	0.0141	0.014130	0.0151	0.0151255	0.0163	0.01630061
0.0079	0.0078824	0.0223	0.0222899	0.0238	0.02376323	0.0255	0.02548836
0.0089	0.0088688	0.0257	0.025732	0.0274	0.0273547	0.0292	0.02292448
0.0091	0.0091084	0.0263	0.0263022	0.0279	0.02793766	0.0298	0.02984228
0.0091	0.00911110	0.0263	0.02631317	0.0279	0.02794848	0.0299	0.02985302
0.0091	0.0091297	0.0264	0.0263941	0.028	0.0280279	0.0299	0.02993105
0.0092	0.0091812	0.0266	0.0266301	0.0283	0.02825633	0.0302	0.03015238
0.0093	0.0092831	0.0271	0.0271265	0.0287	0.02873066	0.0306	0.03060477

Table 2. The magnitude of dimensionless parameters for the suggested model are displayed as follows

Parameters	Range	References
Electric field, U_e	0-1	Goswami et al. [46], Ali et al. [49], Hussain and Ali [50], Ponalagusamy and Sangeetha [51, 57].
Yield stress, τ_0	0-0.1	Ponalagusamy and Machi [38-40], Ponalagusamy and Sangeetha [51, 57].
K-L parameter, θ_1	0-1	Ponalagusamy and Machi [38-40].
K-L parameter, θ_2	0.01-1.5	Ponalagusamy and Machi [38-40].
Interface thickness, R_1	0.7-0.9	Ponalagusamy and Sangeetha [51].

find out a BCEs which can be related to RBC aggregation and deformability. Based on the human and canine blood samples, the root-mean-square differences between BCEs and the measured results are calculated. Among the BCEs, K-L equations provide more satisfactory results because of lesser RMS values [34],[35]. Inspired by this rheological model, these studies may provide an improvement on the research work-based industrial applications under the external electric field [41].

At this juncture, it is pertinent to observe from equation (27) that when $\theta_2 = 0$, the velocity of the core fluid becomes infinity which fails to represent the real flow phenomenon. Hence, the non-zero value of θ_2 should be considered for investigating its effect on the flow characteristics. The pressure force plays a significant role in dealing with the fluid motion. The pressure force was most prominently influenced by the yield stress (τ_0) and K-L parameters θ_1 and θ_2 . It is computationally noticed that using equation (30) and for a given value of flow rate (Q), we cannot independently assign a value to τ_0 , θ_1 and θ_2 while computing the numerical value of pressure gradient (P), otherwise, it is impossible to obtain acceptance value (real and positive) of the pressure gradient (P). It is observed that for a given value of the yield stress (τ_0), the acceptable pressure gradient values were obtained when $\theta_1 \leq \theta_2$, otherwise the existence of real roots vanished. This is mathematically attributed to the fact that the resultant polynomial equation in the form of a pressure gradient (P) obtained from equation (30) is of the fourth-degree equation if $\left| \frac{2\theta_2 P R_1}{\theta_1^2 - 4\theta_2 \tau_0} \right| \leq 1$.

By considering the condition $\theta_1 > \theta_2$, the discriminant value of the fourth-degree polynomial equation was positive for the appropriate numerical values, and the corresponding roots obtained for this value are non-real roots, which were well matched with the discriminant condition. Without considering the constraint mentioned earlier, we computationally observed that when $\theta_1 > \theta_2$, the realistic value of pressure gradient for the prescribed model is impossible to obtain.

The axial velocity profiles in the radial direction are computed over the concerned parameters. Due to symmetry, the investigation was carried out on the half region and the velocity profiles of both one-layered and two-layered models were scrutinized. The graphical representation of the velocity profiles was analysed under the various parameters shown in Figure 5- 8. In Figures 5 and 6, under the influence of the electric field, K-L parameter (θ_1) and yield stress an investigation of the velocity profile of two-layered and one-layered models was performed. In Figure 5, we analysed the four types of two-fluid models, such as Newtonian-Newtonian, Bingham-Newtonian, Casson-Newtonian ($\theta_2 = 1$; $\theta_1 = 2\sqrt{\theta_2 \tau_0}$) and K-L-Newtonian. The Casson -Newtonian fluid model can also be derived from any assigned value of the K-L parameter (θ_2). Compared to all the two-fluid models, Newtonian fluids are much faster than other fluids. When the yield stress increases it reduces the flow nature of the fluid. This fact is easily verified in the given figure. Moreover, the K-L parameter (θ_2) also possessed a greater influence on the flow nature. Witnessed, it has proven that the flow decreases, whenever

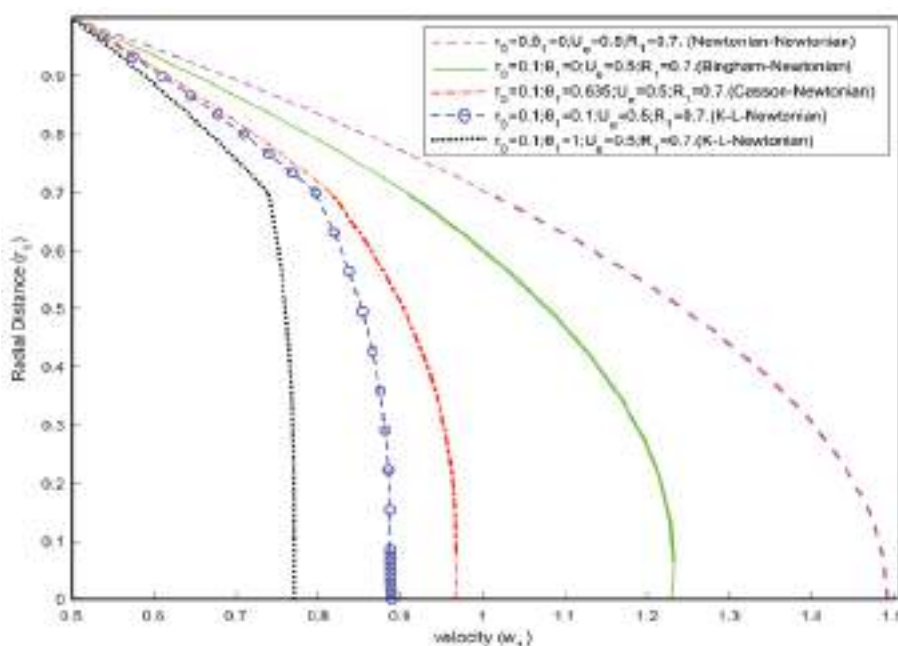


Figure 5. Velocity distribution along with radial direction for distinct values of yield stress (τ_0) and K-L parameter (θ_1). [two-layered model with $\theta_2 = 1$; $\theta_1 \leq \theta_2$].

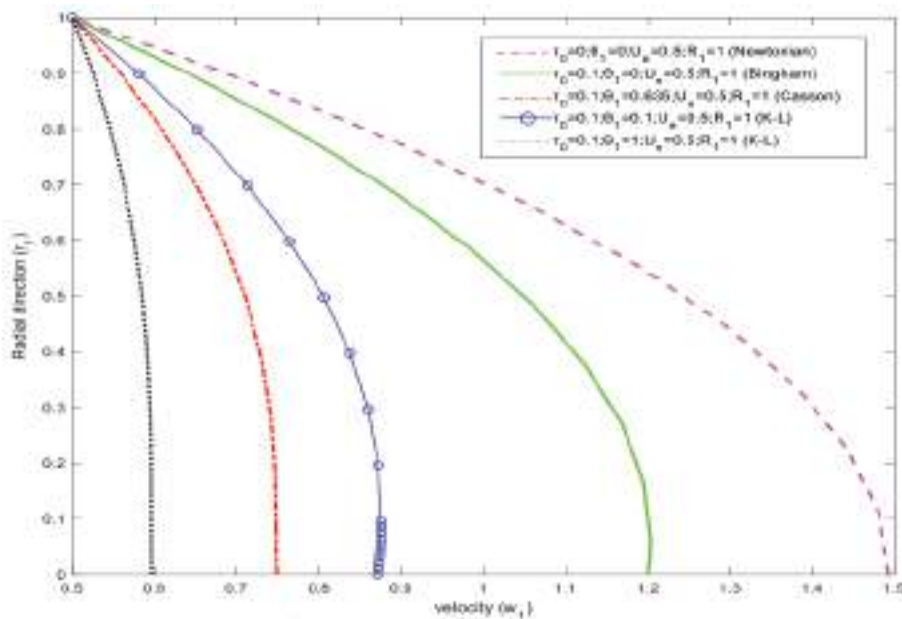


Figure 6. Velocity distribution along with radial direction for distinct values of yield stress (τ_0) and K-L parameter (θ_1). [one-layered model with $\theta_2=1$; $\theta_1 \leq \theta_2$].

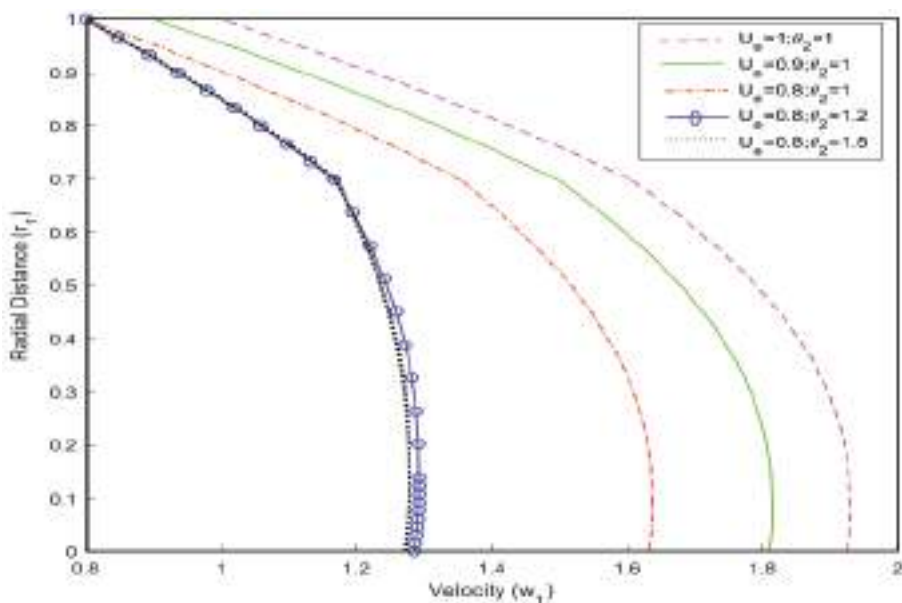


Figure 7. Velocity distribution along with radial direction for distinct values of K-L parameter (θ_2) and electric field (U_e).

θ_2 increases. Furthermore, θ_1 is the coefficient of the shear thinning component and it invokes the behaviour of the core fluid model. The higher the value of θ_1 , the shear rate of pseudo-plasticity increases and hence velocity decreases. Moreover, it analysed that the two-layered Newtonian-Newtonian model is approximately higher than the Casson-Newtonian (6.74%), Bingham-Newtonian (17.92%) and

K-L-Newtonian (22.77%) in velocity profiles along with the data points in Table 3.

In Figure 6, the distribution of velocity along the radial distance with the one-layer model is displayed. If the intermediate layer thickness (R_1) is enlarged or approaches the outer layer of the wall (R), then the fluid flow model is probably considered to be a one-layered model. The discussion made in Figure 6, is based on the yield stress (τ_0), K-L

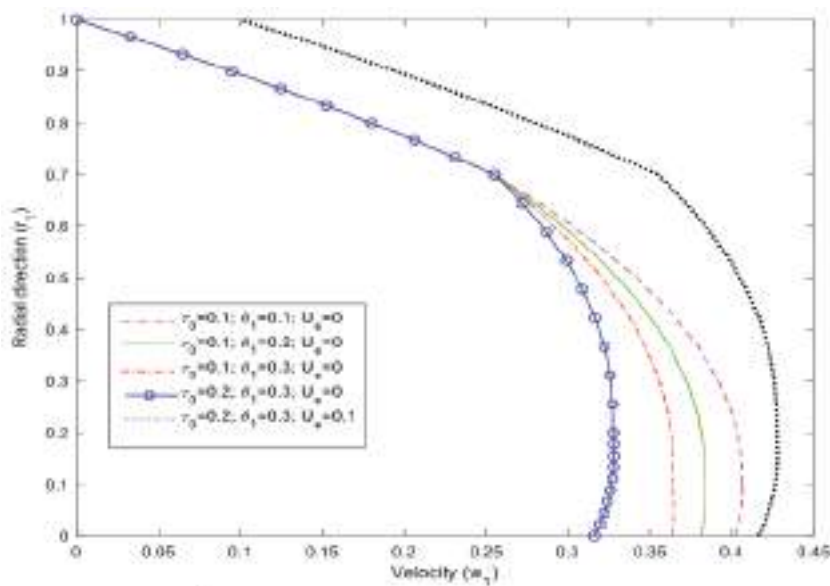


Figure 8. Velocity distribution along with radial direction for distinct values of yield stress (τ_0), K-L parameter (θ_1) and Electric field (U_e).

Table 3. Values for the velocity distribution along with radial direction for distinct values of yield stress (τ_0) and K-L parameter (θ_1) provided in Figure 5. [two-layered model with $\theta_2=1$; $\theta_1 \leq \theta_2$]

Newtonian-Newtonian		Bingham-Newtonian		Casson-Newtonian		K-L-Newtonian ($\theta_1=1$)		K-L-Newtonian ($\theta_1=0.1$)	
y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis
1	0.5	1	0.5	1	0.5	1	0.5	1	0.5
0.96	0.5651	0.96	0.55	0.96	0.5405	0.96	0.53	0.96	0.5381
0.93	0.628	0.93	0.603	0.93	0.5797	0.93	0.56	0.93	0.5749
0.9	0.688	0.9	0.651	0.9	0.6175	0.9	0.58	0.9	0.6105
0.86	0.747	0.86	0.699	0.86	0.6539	0.86	0.616	0.86	0.6447
0.83	0.803	0.83	0.744	0.83	0.6889	0.83	0.6432	0.83	0.677
0.8	0.857	0.8	0.7878	0.8	0.7226	0.8	0.6688	0.8	0.7093
0.76	0.909	0.76	0.8296	0.76	0.7548	0.76	0.6932	0.76	0.7397
0.73	0.959	0.73	0.8695	0.73	0.7857	0.73	0.7162	0.73	0.7688
0.7	1.0064	0.7	0.9077	0.7	0.8153	0.7	0.7391	0.7	0.7966
0.62	1.1085	0.62	0.9759	0.631	0.8503	0.63	0.7476	0.63	0.8191
0.54	1.198	0.55	1.0361	0.56	0.8805	0.56	0.7546	0.563	0.8381
0.46	1.2767	0.48	1.088	0.49	0.906	0.502	0.7601	0.495	0.8537
0.38	1.34	0.41	1.1324	0.42	0.9268	0.436	0.7643	0.427	0.866
0.311	1.3969	0.34	1.1685	0.35	0.9431	0.37	0.7672	0.3588	0.8753
0.23	1.4389	0.27	1.1965	0.28	0.955	0.304	0.7692	0.290	0.8817
0.15	1.469	0.20	1.2166	0.218	0.9627	0.238	0.7702	0.222	0.8856
0.07	1.487	0.13	1.2286	0.149	0.966	0.172	0.7706	0.154	0.8873
0	1.493	0.0625	1.2326	0.080	0.9675	0.106	0.7707	0.0859	0.8876
0	1.493	0.055	1.2326	0.071	0.9675	0.0948	0.07707	0.0764	0.8876
0	1.493	0.048	1.2325	0.062	0.9676	0.082	0.07707	0.0668	0.8876
0	1.493	0.041	1.2323	0.053	0.9677	0.071	0.7707	0.057	0.8876
0	1.493	0.034	1.232	0.044	0.9678	0.059	0.7708	0.0477	0.8877
0	1.493	0.027	1.2317	0.035	0.9677	0.047	0.7708	0.0382	0.8878
0	1.493	0.02	1.2312	0.026	0.9676	0.035	0.7709	0.02866	0.888
0	1.493	0.01	1.2307	0.017	0.9673	0.023	0.7709	0.019	0.8882
0	1.493	0.0069	1.2302	0.0089	0.967	0.01185	0.7710	0.0095	0.8887
0	1.493	0	1.2295	0	0.9666	0	0.7710	0	0.8895

parameter (θ_1) and electric field (U_e). Here the fact was verified that the Newtonian model possesses higher velocity when compared to the Casson ($\theta_2=1$; $\theta_1=2\sqrt{\theta_2\tau_0}$) and K-L fluid model ($\theta_2 = 0.1$ and 1). Significantly, it is noted that the results based on the two-layered model share the same characteristics of the one-layer model concerned with the values of parameters considered. Moreover, the set of data points has been displayed in Table 4.

In Figure 7, the impact of the electric field and K-L parameter (θ_2) was interpreted in the velocity profiles. In the vicinity of the EDL effect, it had a high effect on the flow. The decreasing value of the electric field shows that velocity was reduced and the increasing value of θ_2 tends to reduce more. This shows that the K-L parameter (θ_2), can reduce the fluid flow. This is attributed to the fact that as the K-L parameter (θ_2) increases the apparent viscosity of the fluid slows down the motion of the fluid.

The distribution of velocity along the radial distance with the absence and the existence of an electric field is discussed in Figure 8. The variation of the K-L parameter and yield stress was studied and found that in the absence of an electric field, the higher the value of yield stress (τ_0) and θ_1 tends to slow down the velocity profiles whereas the existence of an electric field made a sudden upsurge in the fluid motion and the data sets have been displayed in Table 5.

Generally, the velocity of the fluid forms merely a parabolic form (depending on the fluid model). This shows that the velocity however higher at the midpoint, whereas it tends to be slower when it approaches the wall. This is because the fluid layer present at the wall made a friction effect with the layer of the wall. This sort of friction was caused due to the tangential force imposed by the fluid motion. This kind of friction is called Skin friction or wall shear stress. The rate at which fluid velocity increases as it travels from the conduit wall to the conduit's middle defines the magnitude of the wall shear stress.

In Figures 9-12, the wall shear stress was discussed with respect to the K-L parameters (θ_2). Here, the new findings have been analysed. Compared to the velocity profiles, the behaviour of K-L parameters (θ_2) and the Electric field projects its behaviour in the opposite nature. In Figure 9, the influence of the electric field is induced to reduce the wall shear stress whenever it increases. The required amount of friction force needed less because of the external electric field. Similarly, as the K-L parameter (θ_1) increases, the tendency of the friction force is also increased. Moreover, as the K-L parameter (θ_1) is held fixed, the magnitude of the skin friction is increased with the increase in (θ_2). Interestingly, the fact was found that the increasing nature of the K-L parameter (θ_2), tends to upsurge the magnitude of the skin friction. In wall shear stress, the Casson fluid

Table 4. Values for the velocity distribution along with radial direction for distinct values of yield stress (τ_0) and K-L parameter (θ_1) provided in Figure 6. [one-layered model with $\theta_2=1$; $\theta_1 \leq \theta_2$]

K-L-Newtonian ($\theta_1=1$)		K-L-Newtonian ($\theta_1=0.1$)		Casson-Newtonian		Bingham-Newtonian		Newtonian-Newtonian	
y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis
1	0.5	1	0.5	1	0.5	1	0.5	1	0.5
0.9012	0.525	0.8995	0.5812	0.8979	0.56	0.8958	0.6475	0.8889	0.7084
0.8025	0.5456	0.7989	0.6523	0.7958	0.6109	0.7917	0.7776	0.7778	0.8923
0.7037	0.5622	0.6984	0.7134	0.6937	0.653	0.6875	0.8903	0.6667	1.0517
0.6049	0.575	0.5979	0.7646	0.5915	0.6869	0.5834	0.9858	0.5556	1.1865
0.5062	0.5843	0.4974	0.8059	0.4894	0.7125	0.4792	1.0638	0.444	1.2969
0.4074	0.5905	0.3968	0.8375	0.3873	0.7305	0.3750	1.1246	0.333	1.3827
0.3086	0.5939	0.2963	0.8593	0.2852	0.7415	0.2709	1.1679	0.2222	1.444
0.2099	0.5954	0.1958	0.8717	0.1831	0.7466	0.1667	1.194	0.1111	1.4807
0.11	0.5956	0.0952	0.8753	0.081	0.7476	0.0625	1.2026	0	1.493
0.0988	0.5956	0.0847	0.8753	0.072	0.7476	0.0556	1.2026	0	1.493
0.0864	0.5956	0.0741	0.8751	0.063	0.7476	0.0486	1.2025	0	1.493
0.0741	0.5956	0.0635	0.8749	0.054	0.7476	0.0417	1.2023	0	1.493
0.0617	0.5957	0.0529	0.8745	0.045	0.7476	0.0347	1.2020	0	1.493
0.0494	0.5957	0.0423	0.8741	0.036	0.7477	0.0278	1.2017	0	1.493
0.037	0.5958	0.0317	0.8735	0.027	0.7478	0.0208	1.2012	0	1.493
0.0247	0.5959	0.0212	0.8728	0.018	0.7481	0.0139	1.2007	0	1.493
0.0123	0.5960	0.0106	0.8719	0.009	0.7484	0.0069	1.2002	0	1.493
0	0.5960	0	0.871	0	0.7489	0	1.1995	0	1.493

Table 5. Values of velocity distribution along with radial direction for distinct values of yield stress (τ_0), K-L parameter (θ_1) and Electric field (U_e) provided for Figure 8.

$\tau_0=0.1; \theta_1=0.1; U_e=0$		$\tau_0=0.1; \theta_1=0.2; U_e=0$		$\tau_0=0.1; \theta_1=0.3; U_e=0$		$\tau_0=0.2; \theta_1=0.3; U_e=0$		$\tau_0=0.2; \theta_1=0.3; U_e=0.1$	
y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis
1	0	1	0	1	0	1	0	1	0.1
0.96	0.0328	0.96	0.0328	0.96	0.0328	0.96	0.0328	0.96	0.1328
0.93	0.0644	0.93	0.0644	0.93	0.0644	0.93	0.0644	0.93	0.164
0.9	0.0950	0.9	0.095	0.9	0.095	0.9	0.095	0.9	0.1950
0.8667	0.1244	0.8667	0.1244	0.8667	0.1244	0.8667	0.1244	0.8667	0.2244
0.833	0.1528	0.833	0.1528	0.833	0.1528	0.833	0.1528	0.833	0.2528
0.8	0.18	0.8	0.18	0.8	0.18	0.8	0.18	0.8	0.28
0.7667	0.2061	0.7667	0.2061	0.7667	0.2061	0.7667	0.2061	0.7667	0.3061
0.733	0.2311	0.733	0.2311	0.733	0.2311	0.733	0.2311	0.733	0.3311
0.7	0.255	0.7	0.255	0.7	0.255	0.7	0.255	0.7	0.355
0.633	0.2881	0.633	0.2840	0.633	0.2804	0.644	0.2720	0.644	0.372
0.5667	0.3170	0.5667	0.3091	0.5667	0.3023	0.588	0.2866	0.588	0.3866
0.5	0.3418	0.5	0.3305	0.5	0.3207	0.533	0.2989	0.533	0.3989
0.43	0.3626	0.43	0.3481	0.43	0.3357	0.477	0.3088	0.477	0.4088
0.366	0.3792	0.366	0.362	0.366	0.3473	0.422	0.3165	0.422	0.4165
0.3	0.3919	0.3	0.3723	0.3	0.3558	0.3667	0.322	0.3667	0.422
0.233	0.4006	0.233	0.3792	0.233	0.3612	0.311	0.3256	0.311	0.4256
0.166	0.4055	0.166	0.3828	0.166	0.3639	0.2556	0.3273	0.2556	0.4273
0.1	0.4068	0.1	0.3836	0.1	0.3644	0.2	0.3276	0.2	0.4276
0.0889	0.4068	0.0889	0.3836	0.0889	0.3645	0.1778	0.3277	0.1778	0.4277
0.0778	0.4067	0.0778	0.3837	0.0778	0.3645	0.1556	0.328	0.1556	0.428
0.0667	0.4064	0.0667	0.3836	0.0667	0.3647	0.133	0.3277	0.133	0.4277
0.0556	0.4061	0.0556	0.384	0.0556	0.3648	0.11	0.3270	0.11	0.427
0.0444	0.4056	0.0444	0.383	0.0444	0.3647	0.0889	0.3258	0.0889	0.4258
0.0333	0.4049	0.0333	0.3826	0.0333	0.3645	0.0667	0.3241	0.0667	0.4241
0.0222	0.4042	0.0222	0.382	0.0222	0.3642	0.0444	0.3219	0.0444	0.4219
0.0111	0.4033	0.0111	0.3813	0.0111	0.3638	0.0222	0.3192	0.0222	0.4192
0	0.4023	0	0.3805	0	0.3633	0	0.316	0	0.416

along with the value of $U_e=0$ was 0.102 % approximately higher for $U_e=0.5$ and for the K-L model ($U_e=0$) is appropriately greater than 0.078 % for $U_e=0.5$. The one-layered and two-layered models were discussed graphically in the existence and absence of an electric field as seen in Figures 10-12. In Figure 10, Casson-Newtonian model ($\theta_1=2\sqrt{\theta_2\tau_0}$) and the K-L- Newtonian model (θ_1) were made with relevant values. The Newtonian-Newtonian model and Bingham-Newtonian model are discussed in Figure 11. The Newtonian-Newtonian fluid has a lesser magnitude of skin friction than the Bingham-Newtonian fluid. The skin friction coefficient of the K-L-Newtonian model was higher friction than the Casson-Newtonian model. In both Figures 10 and 11, it is noticed that the electric field reduces the skin friction. Similarly, the characteristics analysed in Figure 12 are the one-layered model (K-L fluid and Casson

fluid model) which shares the same behaviour exposed by the two-layered model. The comparison of the one-layer and two-layer models was analysed graphically and verified one-layered model features with the two-layered model.

In Figure 13, the interpretation of the pressure drop over the axial direction (z) was made. The prescribed model was designed with two additional body forces: pressure force and electro-osmotic force. Electric force always enhances fluid motion. If suppose this force is absent, the pressure force tends to play a vital role in making fluid flow. Owing to this fact, the concerned graph was studied. Scrutinizing the K-L parameters (θ_1 & θ_2), the analysis of the pressure drop was discussed in Figure 13. It shows that the pressure drop exists only when $\theta_1 \leq \theta_2$ and the Δp increment is found for the increasing values of the K-L parameters. The pressure drop (Δp) for the K-L parameter $\theta_1=0.1, 0.2$ is

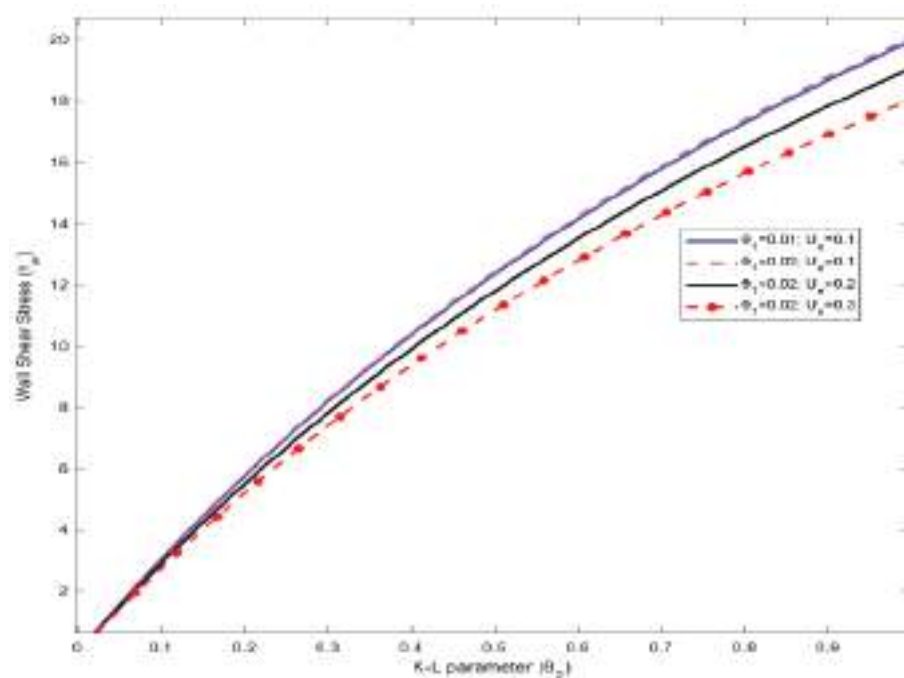


Figure 9. Variation of wall shear stress (τ_w) along with the K-L parameter (θ_2) for different values of K-L parameter (θ_1) and Electric field (U_e).

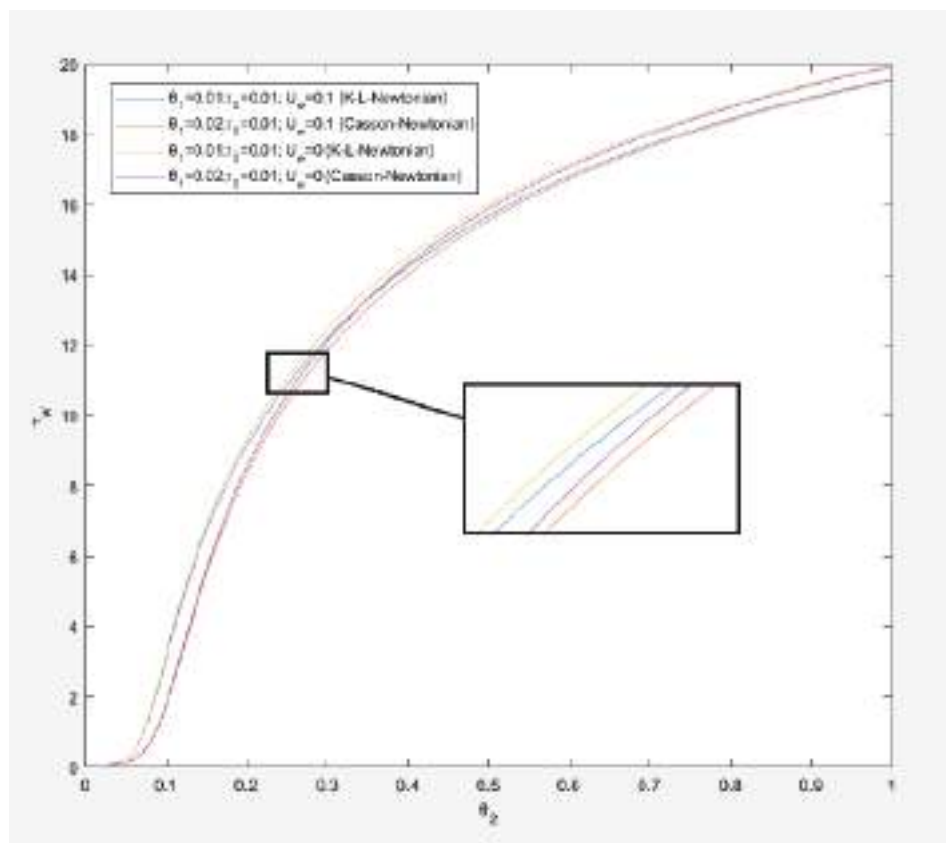


Figure 10. Variation of wall shear stress (τ_w) along with the K-L parameter (θ_2) for different values of K-L parameter (θ_1) and Electric field (U_e).

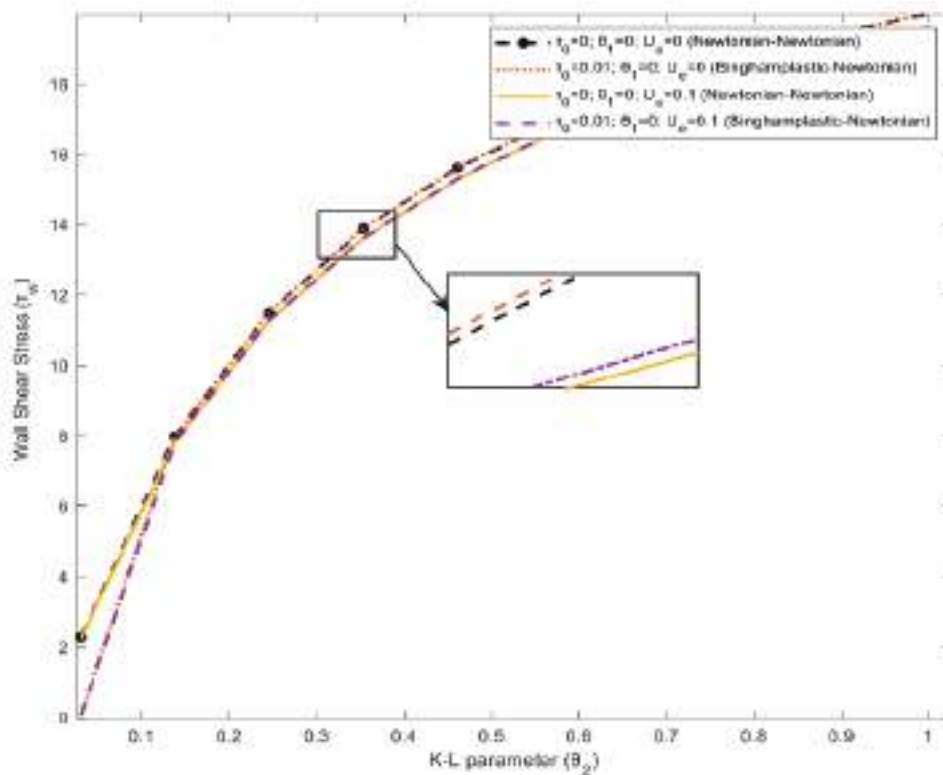


Figure 11. Variation of wall shear stress (τ_w) along with the K-L parameter (θ_2) for different values of K-L parameter (θ_1) and Electric field (U_e).

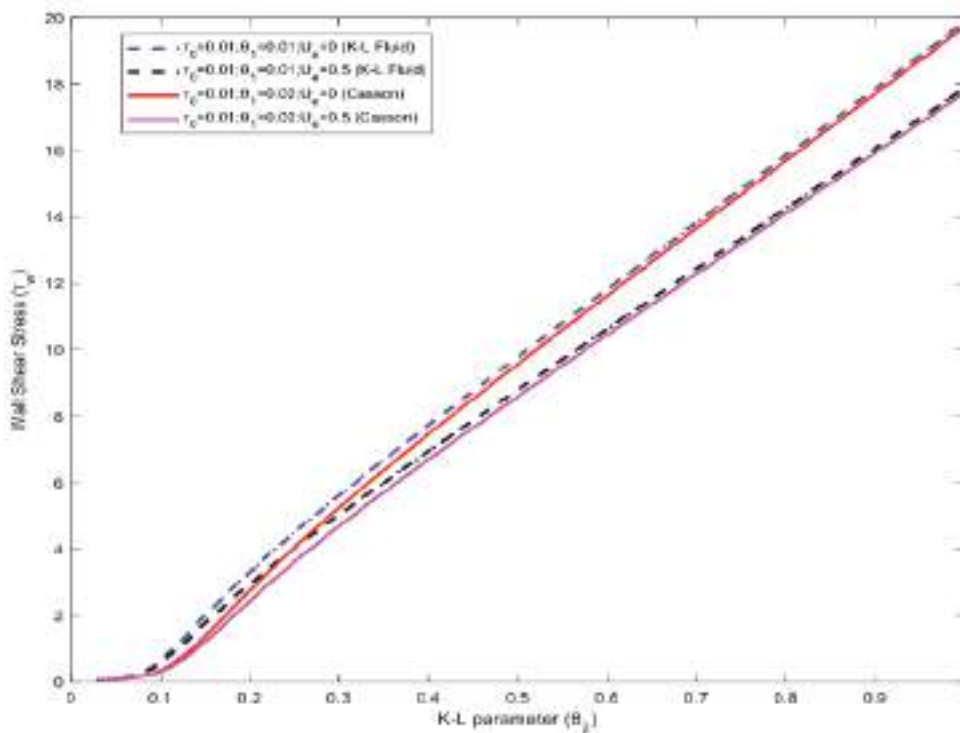


Figure 12. Variation of wall shear stress (τ_w) along with the K-L parameter (θ_2) for different values of K-L parameter (θ_1) and Electric field (U_e) for one-layered model.

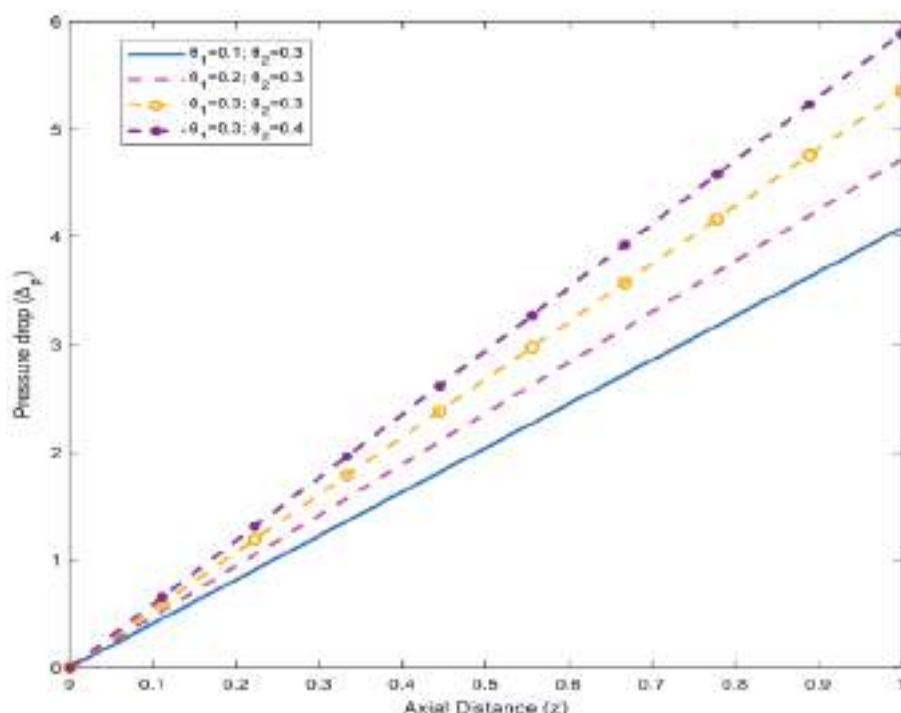


Figure 13. Variation of pressure drop (Δp) along with the axial direction (z) for different values of K-L parameters (θ_1 & θ_2).

Table 6. Values for the variation of pressure drop (Δp) along with the axial direction (z) for different values of K-L parameters (θ_1 & θ_2) provided in Figure 13.

$\theta_1=0.3; \theta_2=0.4$		$\theta_1=0.3; \theta_2=0.3$		$\theta_1=0.2; \theta_2=0.3$		$\theta_1=0.1; \theta_2=0.3$	
x-axis (z)	y-axis (Δp)	x-axis (z)	y-axis (Δp)	x-axis (z)	y-axis (Δp)	x-axis (z)	y-axis (Δp)
0	0	0	0	0	0	0	0
0.1	0.654	0.1	0.5949	0.1	0.5247	0.1	.4535
0.2	1.308	0.2	1.1898	0.2	1.0493	0.2	.9069
0.3	1.962	0.3	1.7847	0.3	1.574	0.3	1.3604
0.4	2.616	0.4	2.3797	0.4	2.0986	0.4	1.8138
0.5	3.27	0.5	2.9746	0.5	2.6233	0.5	2.2673
0.6	3.924	0.6	3.5695	0.6	3.1479	0.6	2.7207
0.7	4.578	0.7	4.1644	0.7	3.6726	0.7	3.1742
0.8	5.232	0.8	4.7593	0.8	4.1972	0.8	3.6276
1	5.886	1	5.342	1	4.7219	1	4.0811

approximately 0.21 % higher than the $\theta_1=0.3$ for the fixed viscosity index $\theta_2=0.3$ and the set of data points has been displayed in Table 6.

In Figures 14 and 15, the impact of the yield stress over the plug core radius region was computed numerically. Depending on the various factors, the analysis was done. The plug core radius is dependent on the pressure gradient and the yield stress of the given fluid model. Plug core radius possesses a visco-plastic nature and the velocity of the fluid in this section is almost constant. This proves that the growth

of the radius in the plug core area is linear with respect to the yield stress. Owing to this fact, the assumed parameters are K-L parameters, Constant interfacial region and electric field. The variation of the Interfacial region and the electric field was discussed in Figure 14. The potential of the electric field has a tendency to make the fluid region move forward and its enhancement has occurred in a small amount. Furthermore, the interfacial region also plays a vital role in the plug core region. It validates the increasing volume of core fluid radius (i.e., interfacial radius), and its influence

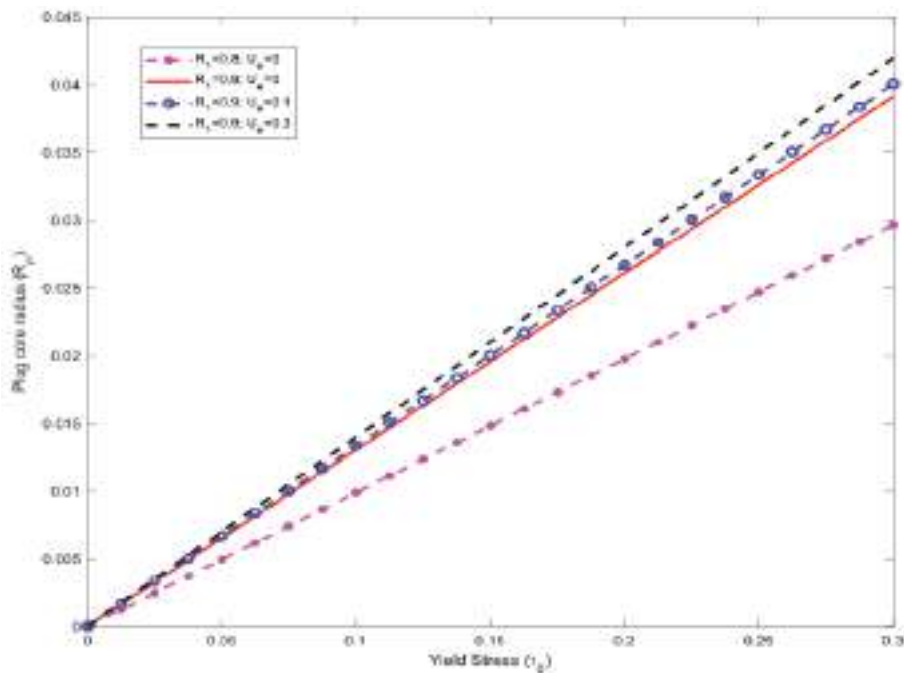


Figure 14. Variation of plug core radius (R_p) along with yield stress (τ_0) for different values of interfacial Region (R_1) and Electric field (U_e).

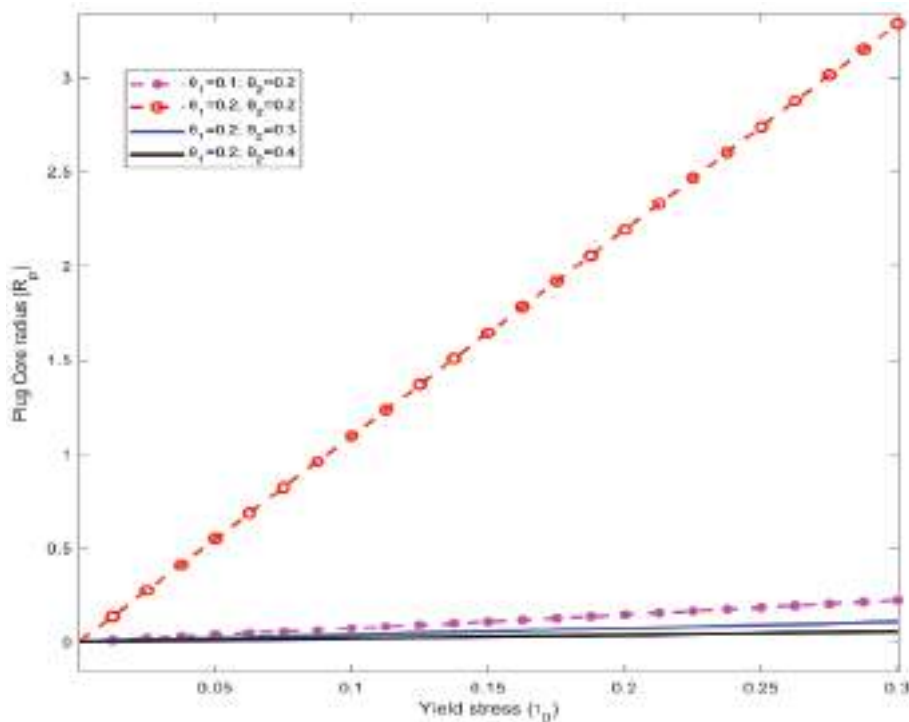


Figure 15. Variation of plug core radius (R_p) along with yield stress (τ_0) for different values of K-L parameters (θ_1 & θ_2).

causes the increase in the region in the plug flow region. In Figure 15, the variation of the K-L parameters (θ_1 , θ_2) was interpreted with appropriate values. The influence of the K-L parameter θ_1 was higher than the θ_2 . The enhancement of

the θ_2 value suppresses the plug core radius and the opposite behaviour was identified in the θ_1 parameter.

The variation of the flow flux was discussed in Figures 16-18 with respect K-L parameter (θ_2). In Figure 16, we

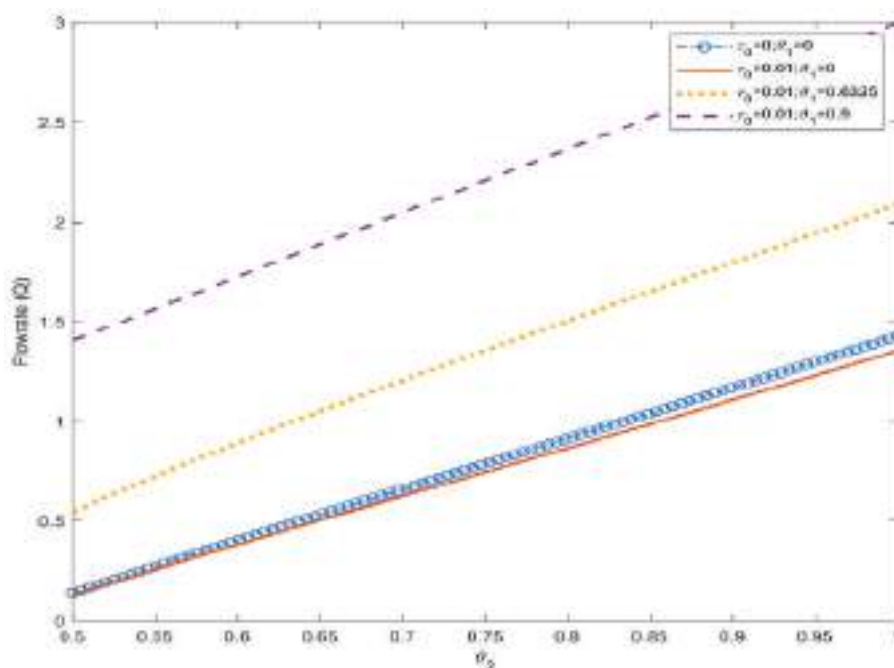


Figure 16. Variation of flow rate with respect to the K-L parameter (θ_2) for the distinct values of yield stress (τ_0) and K-L parameter (θ_1).

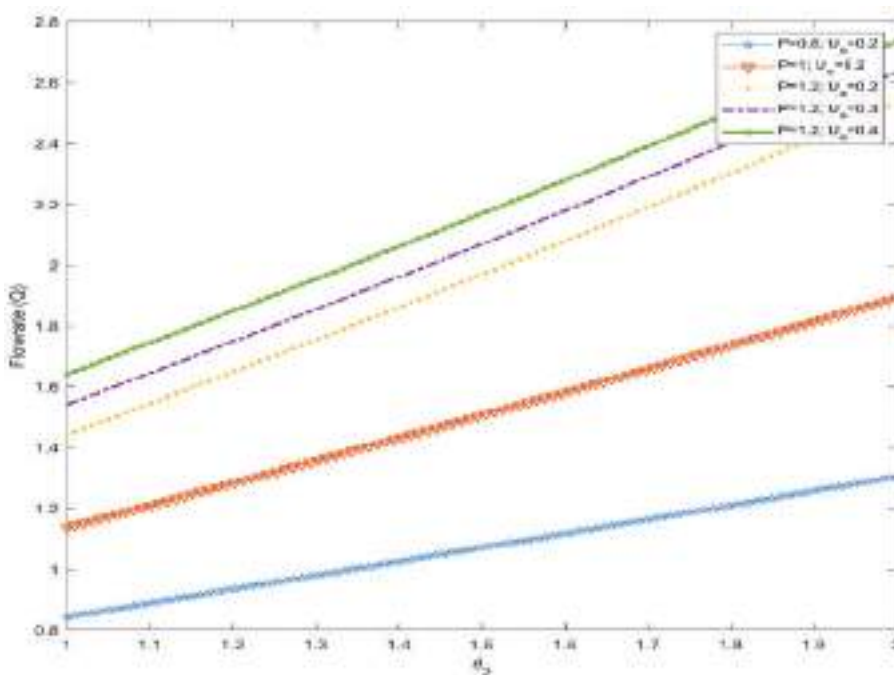


Figure 17. Variation of flow rate with respect to the K-L parameter (θ_2) for the distinct values of pressure gradient (P) and Electric field (U_e).

analysed the variation of flow rate for the distinct values of yield stress (τ_0) and K-L parameter (θ_1). Here, the observation was made that the flow rate was diminished for the increment value of yield stress. This is due to the fact that

the component yield stress represents the solid behaviour of the fluid. If the fluid accumulates the yield stress component, the extra additional force is required to make a fluid move in the conduit. Owing to this physiological fact,

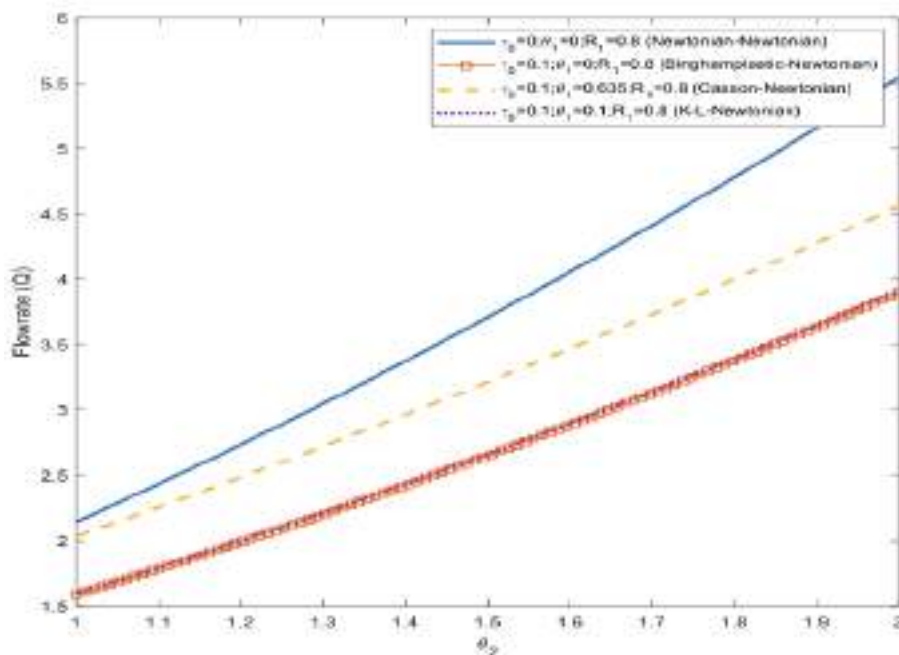


Figure 18. Variation of flow rate with respect to the K-L parameter (θ_2) for the different types of fluids in limiting cases.

our model portrayed these behaviours very well. For the enhanced values of the K-L parameter (θ_1), the flow rate was increased which shows that the component θ_1 provides the favourable quality to ingrown the motion of the flux. The variation of the flow flux was studied for the distinct values of pressure gradient and the electroosmotic flow in Figure 17. This figure shows that the enhancement of flow flux occurs for the increasing values of the pressure gradient and the electroosmotic force. Figure 18 portrays

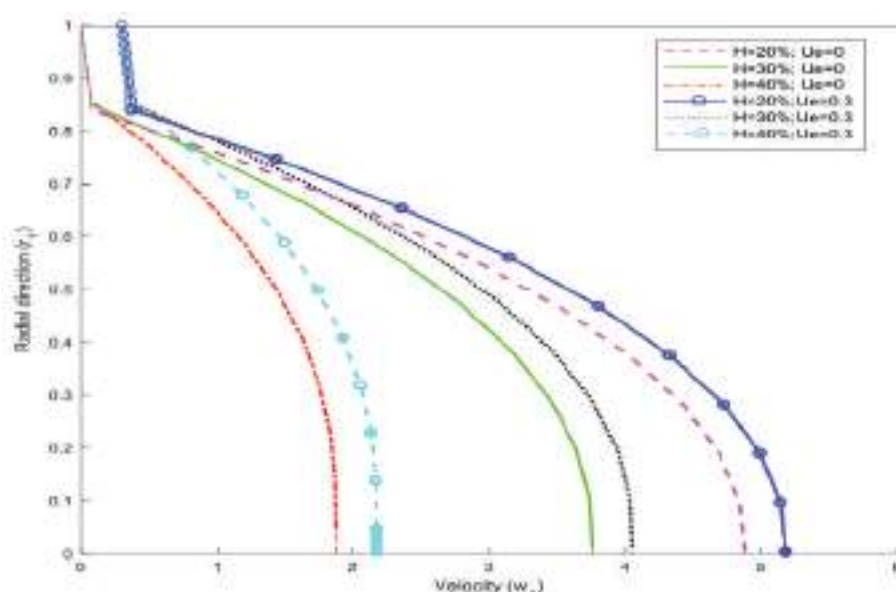
the flow rate of different kinds of fluids with respect to the K-L parameter (θ_2) with certain limiting cases. It shows that Newtonian-Newtonian fluids possess high flow flux compared to the Bingham plastic-Newtonian, Casson-Newtonian and K-L-Newtonian. This is due to the fact that the presence of yield stress makes the fluid decrease its flow wherever it exists and the corresponding data sets for this respective figure are shown in Table 7 to indicate the values of each kind of fluid flow.

Table 7. Variation of flow rate with respect to the K-L parameter (θ_2) for the different types of fluids in limiting cases in Figure 18.

Newtonian-Newtonian		Bingham Plastic-Newtonian		Casson - Newtonian		K-L - Newtonian	
x-axis	y-axis	x-axis	y-axis	x-axis	y-axis	x-axis	y-axis
1.0000	1.0410	1.0000	0.6040	1.0000	0.9480	1.0000	0.6137
1.0714	1.1720	1.0714	0.6710	1.0714	1.0300	1.0714	0.6808
1.1428	1.3040	1.1428	0.7380	1.1428	1.1110	1.1428	0.7483
1.2142	1.4380	1.2142	0.8050	1.2142	1.1930	1.2142	0.8164
1.2857	1.5750	1.2857	0.8740	1.2857	1.2740	1.2857	0.8852
1.3571	1.7140	1.3571	0.9430	1.3571	1.3560	1.3571	0.9549
1.4285	1.8540	1.4285	1.0130	1.4285	1.4390	1.4285	1.0255
1.5000	1.9980	1.5000	1.0850	1.5000	1.5220	1.5000	1.0970
1.5714	2.1430	1.5714	1.1570	1.5714	1.6050	1.5714	1.1690
1.6428	2.2920	1.6428	1.2300	1.6428	1.6900	1.6428	1.2430
1.7142	2.4420	1.7142	1.3050	1.7142	1.7750	1.7142	1.3180
1.7857	2.5950	1.7857	1.3800	1.7857	1.8620	1.7857	1.3940
1.8571	2.7510	1.8571	1.4570	1.8571	1.9490	1.8571	1.4710
1.9285	2.9090	1.9285	1.5350	1.9285	2.0370	1.9285	1.5490
2.0000	3.0700	2.0000	1.6140	2.0000	2.1264	2.0000	1.6280

Table 8. Rheological parameter values. [From Manchi and Ponalagusamy [38], with permission from Elsevier]

H (%)	R_1 (cm)	τ_0^* (dyne/cm ²)	η_1^* (dyne-s ^{1/2} / cm ²)	η_2^* (dyne-s ^{1/2} / cm ²)
20	0.84	0.00201	0.01871	0.02912
30	0.85	0.00593	0.03790	0.03102
40	0.86	0.02383	0.08406	0.04026

**Figure 19.** Velocity distribution along with radial direction for distinct values of hematocrit in the presence and absence of an electric field through a rigid tube.

Real-Life Application

The present formulation gives an effective way of studying the K-L- Newtonian fluid model under the impact of an electric field. So far, the application of the concern fluid model has been utilized in the form of blood transport in normal and stenosed arteries. Here, we gave the flow behaviour of common rheology of the immiscible fluids under the impact of the electric field. Moreover, the fluid becomes very effective when the concentration of the suspended particles (maybe nano-particles in fluids or red blood cells in blood flow) in immiscible fluids can be discussed. Inspired by the human blood model, the present model sheds some limelight to make engineering-based applications.

Table 8 provides the experimental values of hematocrit, K-L parameter and yield stress in dimensional form. In addition to that, appropriate values were used and the discussion was made about the existence and absence of an electric field. This exhibits that the velocity decreases with the increasing value of hematocrit. Moreover, the electric field possesses some additional force to enhance the velocity as displayed in Figure 19. This simulation technique will

foresight its application in high polymer structure fluids. For instance, silver conductive grease, copper conductive grease, carbon nanotube conductive grease and graphite conductive grease are electrically conductive. Based on the percentage of silver, copper, graphite and carbon nanotube present in the concerned lubricants is similar to the hematocrit in blood. Depending on these percentages, the motion of the fluid and its resistance can be calculated. These electrically conductive lubricants (presence of additive) will be applicable in connectors and slip rings, switches and potentiometers, printed circuit boards, semiconductor devices, automotive applications, aerospace technology and Defence sectors. The aforementioned idea will spark the production of suitable fluid applications in the engineering industry.

CONCLUSION

The motion of immiscible fluids in the circular tube constrained by the external electric field was discussed. The fluid confined in the core section is Kuang-Luo (K-L) fluid which is surrounded by another fluid occupied in the outer

region the Newtonian fluid. It is presumed that both fluids are electrically conducting fluids. In addition to pressure force, the electric field tends to create an electroosmotic force and forms an EDL effect near the conduit wall. Based on this consideration, studies on the velocity profiles, wall shear stress, plug core radius and pressure drop were made and the following results were obtained.

- In the velocity profiles, yield stress, K-L parameters, intermediate thickness, and electric fields were discussed, and the augmentation of velocity was found when the electric field increased and the depletion of motion occurred when K-L parameters and yield stress were increased.
- The results obtained for the one-layered model possess a similar behaviour to the two-layer model, but they noticed that the velocity of the two-layer model is much higher than the one-layer model. This is due to the interfacial effects of the fluids.
- The depletion nature of the velocity was spotted in the increasing hematocrit value, and its behaviour was similar in the presence and absence of electric potential distribution.
- In the wall shear stress, the nature of the behaviour of the K-L parameter (θ_2) and the electric field was opposite when compared to the velocity profiles. The impact of K-L parameters (θ_1 & θ_2) makes the wall shear stress increase and the enhancement of the electric field makes the wall shear stress decrease.
- The analysis of the pressure drop was discussed over the K-L parameters (θ_1 & θ_2). It shows the increasing effects of the pressure drop caused by the K-L parameter (θ_1 & θ_2).
- The interpretation of plug core radius with corresponding yield stress was discussed under the variation of K-L parameters, electric field, and interfacial region, and the concern effects were noted. The augmentation of plug core radius occurs for increasing values of the electric field, interfacial thickness, and θ_1 , and the opposite nature is shown in θ_2 .
- For the first time, it is reported that for any value of the yield stress, the flow of two-phase immiscible fluids exhibits a realistic flow phenomenon when $\theta_1 \leq \theta_2$.
- A significant amount of declined percentage has been noticed in the two-layered Newtonian-Newtonian model which is approximately higher than the Casson-Newtonian (6.74%), Bingham-Newtonian (17.92%) and K-L-Newtonian (22.77%) in velocity profiles.
- In the case of wall shear stress, the value of Casson fluid for $U_e=0$ is approximately 1.02% higher for $U_e=0.5$ and similarly, the K-L model ($U_e=0$) is approximately 0.078 % greater for $U_e=0.5$.
- The pressure drop (Δp) for the K-L parameter $\theta_1=0.1$, 0.2 is approximately 0.21 % higher than the $\theta_1=0.3$ for the fixed viscosity index $\theta_2=0.3$.
- The flow flux was enhanced for the increased values of pressure gradient, electroosmotic force and θ_1 and the

opposite trend was observed for the higher values of yield stress.

Hence, the present work sheds some light on the outcome of the K-L Newtonian model under the impact of an electric field and on the varying nature of yield stress, K-L parameters, and peripheral thickness. This model can be further extended to various environments, like the existence of a magnetic field, by inducing nanoparticles and the conduit that possesses the peristaltic behaviour.

ACKNOWLEDGEMENTS

The corresponding author (Ms Sangeetha J) is grateful to the Ministry of Human Resource Development (MHRD), the Government of India for the grant of a research fellowship.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The requisite data used for numerical simulations and comparison purposes are evaluated using the aid of MATLAB software and demonstrated graphically.

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.

NOMENCLATURE

Alphabets

w_{11}, w_{12}	Velocity in core and peripheral region.
w_{pc}	Plug core velocity
p_0	Pressure
r_1, z_1	radial and axial direction
$-\frac{dp_0}{dz_1}$	Pressure gradient
R_I	Interfacial radius
R_{pc}	Plug core radius
U_e	Electrokinetic slip velocity
Q	Flow flux

Greek Symbols

$\tau_{r_1 z_1}$	Shear stress in $r_1 z_1$ plane
θ_1 & θ_2	K-L parameters
ϵ	Permittivity of the fluid
μ_n	viscosity of the Newtonian fluid

τ_0	yield stress
ρ_e	net charge density
τ_w	wall shear stress
Δp	pressure drop

REFERENCES

- [1] Bentwich M. Two-phase viscous axial flow in a pipe. *J Basic Eng* 1964;86:669–672. [\[CrossRef\]](#)
- [2] Packham BA, Shail R. Stratified laminar flow of two immiscible fluids. *Proc Camb Philos Soc* 1971;69:443–448. [\[CrossRef\]](#)
- [3] Chaturani P, Ponalagusamy R. A two layered model for blood flow through stenosed arteries. *Proc 11th Natl Conf Fluid Mech Fluid Power* 1982:16–22.
- [4] Vijayaraghavan Srinivasan, Vafai K. Analysis of linear encroachment in two-immiscible fluid systems in a porous medium. *J Basic Eng* 1994;116:135–139. [\[CrossRef\]](#)
- [5] Chamkha AJ. Flow of two-immiscible fluids in porous and nonporous channels. *J Fluids Eng* 2000;122:117–124. [\[CrossRef\]](#)
- [6] Umavathi JC, Chamkha AJ, Mateen A, Al-Mudhaf A. Unsteady two-fluid flow and heat transfer in a horizontal channel. *Heat Mass Transf* 2005;42:81–90. [\[CrossRef\]](#)
- [7] Umavathi JC, Prathap Kumar J, Shekar M. Mixed convective flow of immiscible viscous fluids confined between a long vertical wavy wall and a parallel flat wall. *Int J Eng Sci Technol* 2010;2:256–277. [\[CrossRef\]](#)
- [8] Prathap Kumar J, Umavathi JC, Chamkha AJ, Pop I. Fully-developed free-convective flow of micropolar and viscous fluids in a vertical channel. *Appl Math Model* 2010;34:1175–1186. [\[CrossRef\]](#)
- [9] Ponalagusamy R, Tamil Selvi R. Brief communications: Two-layered model (Casson-Newtonian) for blood flow through an arterial stenosis with axially variable slip velocity at the wall. *Int J Eng Sci Adv Comput Biotechnol* 2013;4:71–74.
- [10] Umavathi JC, Shekar M. Mixed convective flow of immiscible fluids in a vertical corrugated channel with traveling thermal waves. *J King Saud Univ Eng Sci* 2014;26:49–68. [\[CrossRef\]](#)
- [11] Devakar M, Ramgopal NCh. Fully developed flows of two immiscible couple stress and Newtonian fluids through nonporous and porous medium in a horizontal cylinder. *J Porous Media* 2015;18:549–558. [\[CrossRef\]](#)
- [12] Yadav PK, Verma AK. Analysis of immiscible Newtonian and non-Newtonian micropolar fluid flow through porous cylindrical pipe enclosing a cavity. *Eur Phys J Plus* 2020;135:645. [\[CrossRef\]](#)
- [13] Akbari S, Taghavi SM. Buoyant fluid injections at high viscosity contrasts in an inclined closed-end pipe. *Phys Fluids* 2023;35:022102. [\[CrossRef\]](#)
- [14] Houston G, Oliveria MSN. Flow focusing with miscible fluids in microfluidic devices. *Phys Fluids* 2023;35:052015. [\[CrossRef\]](#)
- [15] Ponalagusamy R. Analysis of MHD flow of blood in stenosed arteries with radially variable viscosity and peripheral plasma layer thickness by means of Frobenius method. *ESS Open Arch* 2023;1–21.
- [16] Rao AR, Usha S. Peristaltic transport of two immiscible viscous fluids in a circular tube. *J Fluid Mech* 1995;298:271–285. [\[CrossRef\]](#)
- [17] Vajravelu K, Arunachalam PV, Sreenadh S. Unsteady flow of two immiscible conducting fluids between two permeable beds. *J Math Anal Appl* 1995;196:1105–1116. [\[CrossRef\]](#)
- [18] Umavathi JC, Shekar M. Mixed convective flow of immiscible fluids in a vertical corrugated channel with traveling thermal waves. *J King Saud Univ Eng Sci* 2014;26:49–68. [\[CrossRef\]](#)
- [19] Nikodijević DD, Stamenković ŽM, Jovanović MM, Kocić MM, Nikodijević JD. Flow and heat transfer of three immiscible fluids in the presence of uniform magnetic field. *Therm Sci* 2014;18:1019–1028. [\[CrossRef\]](#)
- [20] Fraggedakis D, Kouris C, Dimakopoulos Y, Tsamopoulos J. Flow of two immiscible fluids in a periodically constricted tube: transitions to stratified, segmented, churn, spray, or segregated flow. *Phys Fluids* 2015;27:082102. [\[CrossRef\]](#)
- [21] Abd Elmaboud Y. Two layers of immiscible fluids in a vertical semi-corrugated channel with heat transfer: impact of nanoparticles. *Results Phys* 2018;9:1643–1655. [\[CrossRef\]](#)
- [22] Devakar M, Ramgopal NCh. Fully developed flows of two immiscible couple stress and Newtonian fluids through nonporous and porous medium in a horizontal cylinder. *J Porous Media* 2015;18:549–558. [\[CrossRef\]](#)
- [23] Devakar M, Raje A. Modelling and analysis of the unsteady flow and heat transfer of immiscible micropolar and Newtonian fluids through a pipe of circular cross section. *J Braz Soc Mech Sci Eng* 2018;40:1–18. [\[CrossRef\]](#)
- [24] Hue SH, Chagot L, Angeli P. Viscoelastic effects of immiscible liquid-liquid displacement in microchannels with bends. *Phys Fluids* 2022;34:073111. [\[CrossRef\]](#)
- [25] Fang S, Zhenhua C, Baochang S, Meng Z. Optimal displacement of immiscible two-phase fluids in a pore doublet. *Phys Fluids* 2023;35:053332. [\[CrossRef\]](#)
- [26] Sangeetha J, Ponalagusamy R. Impacts of electromagnetic and porous medium on time-dependent flow of two immiscible fluids in a channel. *AIP Conf Proc* 2023;2875:020003:1–11. [\[CrossRef\]](#)
- [27] Akbari N, Gholinia M, Gholinia S, Dabbaghian S, Javadi H, Ganji DD. Analytical and numerical study of micropolar fluid flow in a porous plate due to linear stretching. *Sigma J Eng Nat Sci* 2018;36:1181–1196.

- [28] 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.
- [29] Gholinia M, Gholinia S, Javadi H, Ganji DD. Investigation of micropolar fluid flow and heat transfer in a two dimensional permeable channel by analytical and numerical method. *Sigma J Eng Nat Sci* 2019;37:393–413.
- [30] Dewangan SK, Senapati SK, Deshmukh V. CFD prediction of oil-water two phase stratified flow in a horizontal channel coupled level set-VOF approach. *Sigma J Eng Nat Sci* 2020;38:1–19.
- [31] Özkan R, Ergüner A, Koca K, Genç MS. Improvement of mechanical behaviour of wind turbine blade using nanofluid-graphene and/or glass fiber in epoxy resin. *Sigma J Eng Nat Sci* 2021;39:58–69.
- [32] Akaje TW, Olajuwon BI, Raji MT. Computational analysis of the heat and mass transfer in a Casson nanofluid with variable inclined magnetic field. *Sigma J Eng Nat Sci* 2023;41:512–523. [\[CrossRef\]](#)
- [33] 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. [\[CrossRef\]](#)
- [34] Luo XY, Kuang ZB. A study on the constitutive equation of blood. *J Biomech* 1992;25:959. [\[CrossRef\]](#)
- [35] Zhang JB, Kuang ZB. Study on blood constitutive parameter in different blood constitutive equations. *J Biomech* 2000;33:355–360. [\[CrossRef\]](#)
- [36] Sriyab S. Mathematical analysis of non-newtonian blood flow in stenosis narrow arteries. *Comput Math Methods Med* 2014;2014:479152:1–10. [\[CrossRef\]](#)
- [37] Bali R, Gupta N. Study of non-newtonian fluid by K–L model through a non-symmetrical stenosed narrow artery. *Appl Math Comput* 2018;320:358–370. [\[CrossRef\]](#)
- [38] Ponalagusamy R, Machi R. A study on two-layered (K–L–Newtonian) model of blood flow in an artery with six types of mild stenoses. *Appl Math Comput* 2020;367:1–22. [\[CrossRef\]](#)
- [39] Ponalagusamy R, Machi R. Mathematical study on two-fluid model for flow of K–L fluid in a stenosed artery with porous wall. *SN Appl Sci* 2021;3:508. [\[CrossRef\]](#)
- [40] Ponalagusamy R, Machi R. Biorheological model on pulsatile flow of blood (K–L fluid) through flexible stenotic tapered blood vessels. *Int J Appl Comput Math* 2021;7:1–13. [\[CrossRef\]](#)
- [41] Singh S, Murthy PVS. Unsteady dispersion in pulsatile Luo and Kuang blood flow (K–L model) in a tube with wall reactive absorption. *J Non Newtonian Fluid Mech* 2022;310:104928. [\[CrossRef\]](#)
- [42] Ponalagusamy R, Murugan D, Tamil Selvi R. Effects of rheology of non-newtonian fluid and chemical reaction on a dispersion of a solute and implications to blood flow. *Int J Appl Comput Math* 2022;8:109. [\[CrossRef\]](#)
- [43] Gao Y, Wong TN, Yang C, Ooi KT. Two-fluid electroosmotic flow in microchannels. *J Colloid Interface Sci* 2005;284:306–341. [\[CrossRef\]](#)
- [44] Zhao C, Yang C. Electrokinetic of non-newtonian fluids: A review. *Adv Colloid Interface Sci* 2013;201–202:94–108. [\[CrossRef\]](#)
- [45] Afonso AM, Alves MA, Pinbo FT. Analytical solution of two-fluid electro-osmotic flows of viscoelastic fluids. *J Colloid Interface Sci* 2013;395:277–286. [\[CrossRef\]](#)
- [46] Goswami P, Chakraborty J, Bandopadhyay A, Chakraborty S. Electrokinetically modulated peristaltic transport of power-law fluids. *Microvasc Res* 2016;103:41–54. [\[CrossRef\]](#)
- [47] Gaikwad H, Basu DN, Mondal PK. Electroosmotic transport of immiscible binary system with a layer of non-conducting fluid under interfacial slip: The role applied pressure gradient. *Electrophoresis* 2016;37:1998–2009. [\[CrossRef\]](#)
- [48] Chaube MK, Yadav A, Tripathi D, Bég AO. Electroosmotic flow of biorheological micropolar fluids through microfluidic channels. *Korea Aust Rheol J* 2018;30:89–98. [\[CrossRef\]](#)
- [49] Ali N, Hussain S, Ullah K, Bég AO. Mathematical modelling of two fluid electroosmotic peristaltic pumping an Ellis fluid in an axisymmetric tube. *Eur Phys J Plus* 2019;134:1–18. [\[CrossRef\]](#)
- [50] Hussain S, Ali N. Electro-kinetically modulated peristaltic transport of multilayered power-law fluid in an axisymmetric tube. *Eur Phys J Plus* 2020;135:1–24. [\[CrossRef\]](#)
- [51] Ponalagusamy R, Sangeetha J. A study on electro-hydrodynamic flow of two immiscible fluids in a circular tube. *Proc 66th Congr ISTAM Int Conf* 2021.
- [52] Mukherjee S, Shit GC, Vajravelu K. Effects of diffusive Reynolds number on electro-osmotic pulsating nanofluid flow. *Phys Fluids* 2022;34:122004. [\[CrossRef\]](#)
- [53] Kahali T, Santra S, Chakraborty S. Electrically modulated cross-stream migration of a compound drop in micro-confined oscillatory flow. *Phys Fluids* 2022;34:122015. [\[CrossRef\]](#)
- [54] Ismayeel M, Mehta SK, Mondal PK. Prediction of electrodiffusio-osmotic transport of shear thinning fluids in a nanochannel using artificial neural network. *Phys Fluids* 2023;35:012018. [\[CrossRef\]](#)
- [55] Ponalagusamy R, Sangeetha J. Electroosmotic effect on two immiscible (conducting–non-conducting) fluids flowing in the porous channel under magnetic field. *Proc IMechE Part E J Process Mech Eng* 2023;237:2029–2044. [\[CrossRef\]](#)

- [56] Sangeetha J, Ponalagusamy R, Tamil Selvi R. Unsteady peristaltic flow of two immiscible fluids in an elastic tube under the effect of electromagnetic force and outer region with the porous medium. Proc 68th Congr Indian Soc Theor Appl Mech Int Conf 2023.
- [57] Sangeetha J, Ponalagusamy R, Tamil Selvi R. Electroosmotic peristaltic flow of thixotropic-Newtonian fluids in a circular tube: Effect of variable viscosity co-efficient of core fluid. Chin J Phys 2024;92:470–493. [\[CrossRef\]](#)



Research Article

Investigation of seismic behavior of horizontally curved bridges in comparison with equivalent straight bridges according to the AASHTO LRFD specifications

Esra NAMLI^{1,*}, Turgut ÖZTÜRK¹

¹Department of Civil Engineering, İstanbul Technical University, İstanbul, 34467, Türkiye

ARTICLE INFO

Article history

Received: 06 June 2024

Revised: 23 July 2024

Accepted: 10 September 2024

Keywords:

Horizontally Curved Bridges;
Seismic Behavior; Structural
Analysis

ABSTRACT

Horizontally curved bridges have more complex geometries compared to straight bridges. Their design, structural analysis, and construction require specialized knowledge and expertise. The curved geometry of these bridges affects load distribution and structural behavior. Understanding these behaviors is critical for improving the safety and durability of these bridges. For this reason, many studies have been carried out in order to understand the seismic behavior of horizontally curved bridges and to take the necessary measures in order to design resilient structures. In addition, seismic design regulations have set restrictions on the maximum curvature of bridges and permit engineers to use an equivalent straight bridge for their analysis and design. This paper investigates the restrictions and reviews the AASHTO LRFD Specification Design Fundamentals concerning the seismic responses of horizontally curved bridges by using equivalent straight bridges. In this regard, the seismic responses of 27 horizontally curved and 3 straight bridges, for a total of 30 RC bridges with different span numbers and bridge lengths, are investigated. Numerical parametric structural models have been generated for the selected variables such as; bridge length, span number, span length, and subtended angle. By using the structural analysis program, multi-mode response spectral analyses have been performed for the maximum credible earthquake (MCE) excitation level. The modal periods and frequencies, modal mass participating ratios, maximum displacement of the pier, and internal forces of the structural elements are obtained from the structural analyses of the bridges. The analysis results are compared with horizontally curved bridges and equivalent straight bridges to determine the effect of subtended angle on the seismic behavior of the bridges. It was shown in the study that bridge length and span number had a significant effect on the seismic response of horizontally curved bridges compared to straight bridges. Besides, the subtended angle limitations that AASHTO LRFD specifications put forward regarding allowing the curved bridges to use an equivalent straight bridge should be reviewed again. It suggests that a bridge is considered regular if the subtended angle is smaller than 90°. However, according to the analysis results, the dynamic modal quantities, the displacement and rotations of the pier, and the internal forces of the pier columns and the deck of the bridges could reach their maximum values at lower angles of curvature than 90°. Therefore, the limitations of the subtended angle should be reviewed and re-evaluated for several variable

*Corresponding author.

*E-mail address: enamli@itu.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/>).

parameters by using linear and non-linear analysis methods. The aspects that make this research valuable and different from other studies are firstly, that the parametric models had a wide range of different subtended angles of curvature, bridge span numbers, and bridge lengths. Besides, the analysis results evaluated for the wide scope of the determined bridge configurations to realize the curvature effect of the bridges are very important to designing resilient bridges under seismic excitation.

Cite this article as: Namlı E, Öztürk T. Investigation of seismic behavior of horizontally curved bridges in comparison with equivalent straight bridges according to the AASHTO LRFD specifications. Sigma J Eng Nat Sci 2025;43(5):1785–1806.

INTRODUCTION

Transportation systems from the past to the present and the bridges that are a part of them have been one of the most important factors determining the level of civilization of a country, as well as providing the transportation network within the country.

Damages to bridges after major earthquakes not only affect transportation but also indirectly impact the socio-economic conditions of countries, depending on the level of damage. Therefore, it is of great importance to determine the seismic behavior of bridges and to design them according to those criteria. By examining the types of damage that occurred in past earthquakes, a number of studies have been carried out to develop bridge designs in this aspect. With the increasing population, the coincidence of urban and suburban roads has led to the need for different intersection solutions. Therefore, the use of horizontally curved bridges has become widespread for their aesthetic and economic advantages, which exhibit geometric compatibility. The geometric irregularities of these types of bridges cause more complex and destructive behavior compared to straight bridges. The seismic behavior of bridges in a curved alignment plane contains additional potential damages due to their complex geometries, although similar damages were observed for both curved and straight bridges, as examined in past earthquakes. Particularly as a result of the heavy damage to bridges after the 1971 San Fernando earthquake, the behavior of these bridges under the influence of large ground movements became more significant. In the following years, the 1989 Loma Prieta earthquake, the 1992 Petrolia earthquake, the 1994 Northridge earthquake, and the 2008 Wenchuan earthquakes brought more attention to the importance of damage to bridges in many horizontal curve alignments. The main types of damage in bridges were unseating of the superstructure, column shear collapse, foundation collapse, bearing collapse, and insufficient column bending capacity.

For this reason, many studies have been carried out in order to understand the seismic behavior of horizontally curved bridges and to take necessary measures in order to design resilient structures. In addition, seismic design regulations have set restrictions on the maximum curvature of bridges and permit engineers to use an equivalent straight bridge for their analysis and design.

Bridges with significant horizontal curvature do not exhibit the behavior of an idealized, typical

single-degree-of-freedom (SDOF) system. Therefore, the effects of curvature must also be included in the structural analysis. Conversely, the behavior of bridges with relatively less horizontal curvature can be idealized as a single-degree-of-freedom (SDOF) system.

A curved bridge that can be idealized as an SDOF system can be analyzed as if it were straight. In AASHTO LRFD specifications [1], a force-based seismic analysis design philosophy is adopted, whereas in the AASHTO Seismic Guide Specifications [2], a displacement-based design philosophy is emphasized. In AASHTO LRFD [1], for bridges that have horizontal curvature in plan, the condition that allows bridges to be idealized as straight bridges is specified as having a subtended angle of 900 degrees, whereas this limit has been revised to 300 degrees in the AASHTO Guide Specifications [2]. These conservative limitations arise due to uncertainties in the behavior of curved bridges under seismic excitations. So to point out the importance of the subject, the numerical and experimental studies can be summarized as follows:

Most of the studies include the free vibration response of the horizontally curved bridges and the dynamic behavior under the effect of live loading ([3-7]). In the study, by Williams et al. [4] the multi-directional seismic response of curved bridges was investigated using scaled models with shake table testing. They determined the vulnerability levels of the curved bridges under symmetric mode shapes [8]. Therefore, the torsional effect primarily causes damage to the bearings that provide ductility, and the vulnerability level depends on the shear and bending strength of the hinge bearings and the strength of the shear key. Kawashima et al. [9] presented the analytical and experimental correlation of the seismic responses of bridges. The aim of the study was to emphasize the discontinuous behavior of expansion joints during seismic excitation. In addition, they found that the impact of vertical ground motion is relatively smaller under horizontal transverse seismic effects. In the study by Abdel-Salam et al. [10], it was found that response spectrum analysis fails to capture the range of internal forces when considering the high frequencies of vibration modes in curved bridges. They identified the radius of curvature as the most critical geometric parameter affecting the seismic behavior of curved bridges. In the study by Sennah et al. [11], the development of design criteria for straight and curved box girder bridges in existing specifications was examined. The study emphasized the

necessity for advanced research to address the maximum load distribution in both straight and curved bridges. In the study by Mwafy et al. [12], the detailed seismic performance of a complex curved bridge design was compared with that of constructed bridges for local and global behavior. The comparative study showed significant differences in the lateral capacity and dynamic characteristics between the designed and constructed bridges. Past studies have examined the effect of curvature on the seismic behavior of steel bridges.

Abdel-Salam and Heins ([10], Galindo et al. [13], and Seo and Linzell [14] have demonstrated that as the radius of curvature increases, curved bridges become more susceptible to damages such as deck unseating and crushing during seismic events. Seo and Linzell [15] defined parameters such as span number, length, and curvature radius as the most significant factors affecting the displacement of bearings in continuous bridges. Recently, researchers have been evaluating the seismic performance of curved reinforced concrete bridges. They have focused on the pushover analysis of short-span reinforced concrete bridges and explored the effects of various load directions in seismic analysis. Based on their studies, Araújo et al. [16] concluded that curved bridges exhibit multidirectional dynamic responses and are more sensitive to seismic load directions compared to straight bridges. Khan Easa [17] evaluated a large parametric study that consists of curved bridges with subtended angles varying from 00 to 1800 degrees and having the total arc length equal to the length of the equivalent straight bridge in order to investigate the limitations of the AASHTO specifications about the curved bridges. In his PhD thesis, Shirazi [18] studied the seismic fragility of horizontally curved bridges. According to the study, in addition to ground movements, the horizontal curvature of bridge geometries significantly impacts the seismic fragility of bridges. The seismic responses of individual columns have shown a notable correlation with system fragility curves. The dynamic characteristics of bridges are sensitive to curvature. The curvature geometry is particularly important at curve span angles exceeding 300 degrees. In curved bridges, the combination of longitudinal and transverse modes reduces the primary mode in the bridge response and leads to a combination with larger modes. In their study, Sextos and Taskari [19] investigated the effects of seismic load directions on long-span curved bridges using the finite element method. They modeled the soil under the piers and abutments to examine their influence on the bridges. As a result, all uncertainties aside, increasing the seismic load direction has gradually resulted in a gradual change in bridge behavior. Pahlavan et al. [20] investigated the fragility analysis of two-span curved bridges. They found that the key parameter highlighting fragility in curved bridges is the radius of curvature. Abbasi's [21] PhD thesis examined the mid-span expansion joint design in decks and the effects of shear keys on bridge behavior for both regular and irregular bridges. As a result, shear keys have a more

significant effect on the seismic behavior of curved bridges than straight ones.

Reza Siami Kaleybar Payam Tehrani [22] investigated the effects of geometrical irregularities on the curved bridges under seismic excitation and examined the AASHTO Specification limitations for the curved bridges. Therefore, numerical analyses were done for the selected bridges that have different arrangement types of spans, column heights, abutment conditions, and subtended angles to obtain a wide-spectrum solution. The nonlinear analysis results for bridges with varying column heights and subtended angles differ from those obtained through linear analysis.

This study indicates that the AASHTO Specification limitations for the curved bridges, regarding the analysis method and the subtended angle for the use of equivalent straight bridges, should be reassessed. However, many of the studies mentioned above have not examined the limitation on the subtended angle of curved bridges, as suggested by AASHTO specifications, for analyzing the equivalent straight bridge instead of the curved bridge. In addition generally, displacements and the dynamic behavior of the bridges were investigated in the studies rather than the internal forces of the structural members. Besides, the variation range of the subtended angle is insufficient to evaluate and compare the dynamic behavior of the curved bridges and equivalent straight bridges. In this study, parametric models of bridges were generated by varying the number of spans, bridge length, and the subtended angle of a wide range of curvature angles, which is different from the other studies on this topic.

Multi-mode response spectral analyses were performed for both horizontally curved and equivalent straight bridges under the maximum credible earthquake (MCE) excitation level. As a result of the linear elastic analyses, the dynamic behavior mode shapes and related dynamic quantities (period, natural frequency, and modal mass participation ratios vs.) of the parametric bridges were obtained. Additionally, the displacements at the top of the pier columns, as well as the internal forces of the structural elements (pier column and deck internal forces) have been determined to compare the behavior of the horizontally curved bridges and the straight bridges.

Aashto LRFD Bridge Design Specifications (2020) & AASHTO Guide Specifications for LRFD Seismic Bridge Design (2011)

The dye used in this study is Reactive Orange 16. Reactive Orange 16 is an anionic dye and azo belongs to the class. In Table 1 its characteristics are summarized. Its general characteristics are shown. Figure 1.

According to the table, the seismic analysis methods can be as follows: UL is the uniform load elastic method, SM is the single-mode elastic method, MM is the multi-mode elastic mode, and TH is the time history method.

Table 1. Minimum analysis requirement for seismic effects

Seismic zone	Singe span bridges	Multispan bridges					
		Other bridges		Essential bridges		Critical bridges	
		Regular	Irregular	Regular	Irregular	regular	irregular
1	No Seismic Analysis Required	*	*	*	*	*	*
2		SM/UL	SM	SM/UL	MM	MM	MM
3		SM/UL	MM	MM	MM	MM	TH
4		SM/UL	MM	MM	MM	TH	TH

Table 2. Regular bridge requirements

Parameter	Value				
Number of spans	2	3	4	5	6
Maximum subtended angle for a curved bridge	90°	90°	90°	90°	90°
Maximum span length ratio from span to span	3	2	2	1.5	1.5
Maximum bent/pier stiffness ratio from span to span	-	4	4	3	2

In addition, both of the specifications define a bridge as regular depending on the number of spans, maximum subtended angles, maximum span length ratios, and pier stiffness ratios in the AASHTO LRFD (Table 2) [1]. The AASHTO specifications also propose that a curved continuous bridge may be analyzed as if it were straight, provided that all of the following conditions are met: a) the bridge is regular, b) the subtended angle in plan is not greater than 30° (based on the 2011 AASHTO guide specifications) or is not greater than 90° (based on the AASHTO LRFD 2020 design specifications); and c) the span lengths of the equivalent straight bridge are equal to the arc lengths of the curved bridge [1,2].

According to AASHTO LRFD, bridges are also classified for the seismic categories defined due to S_{D1} . Bridge regularity, importance class, and seismic zone are the criteria used to determine the dynamic analysis method. In this study, the bridges' importance class was determined to be critical for those located in the seismic zone 2 region. For examining the regularity conditions subtended angle of the bridges are varied between 0° up to 180°. Therefore, for both of the conditions, regular and irregular, elastic dynamic analysis performed to investigate the dynamic behavior of the bridges and to compare with the limitations defined in AASHTO specifications.

GENERATING STRUCTURAL MODELS OF THE BRIDGES

Configuration of Parametric Bridges

In this study, a total of 30 reinforced concrete bridges, including 3 straight and 27 curved bridges, were studied.

Bridges have respectively three, four, and five spans with a total length of 135m, 180m, and 225m.

The geometries of the bridges were determined by adjusting the radius of curvature and the subtended angle so that the arc lengths of the bridges would be equal to the lengths of equivalent straight-axis bridges. For one of the abutments, three restraint conditions were considered as fixed, and the others were designed to be fixed only in the vertical direction (i.e., fixed or free in both longitudinal and transverse directions and simplified abutment). The piers are monolithically connected to the deck. The three configurations considered in this study for the bridges are shown in Figure 1a.

For each configuration, the subtended angle varies from 00 to 1800 degrees with an interval of 15°-30° degrees (Fig. 1b). The pier heights were determined as 15m for all bridges. The irregularity of the bridges was provided with the subtended angles varying from 0° to 180° whereas for regularity condition equal column heights and equal span lengths were studied in bridges.

STRUCTURAL ANALYSIS AND DESIGN OF BRIDGES

The parametric bridge models were generated, designed, loaded and analyzed in accordance with the AASHTO LRFD Bridge Design Specifications [1] and TBEC [23].

The columns were designed with a rectangular geometry of 2m x 4m (Fig. 2a) and the deck was considered as a post-tensioned concrete box girder with mono-cell (Fig. 2b). The deck had a transverse moment of inertia of 72.041m⁴ and a cross-section area of 7.2792m².

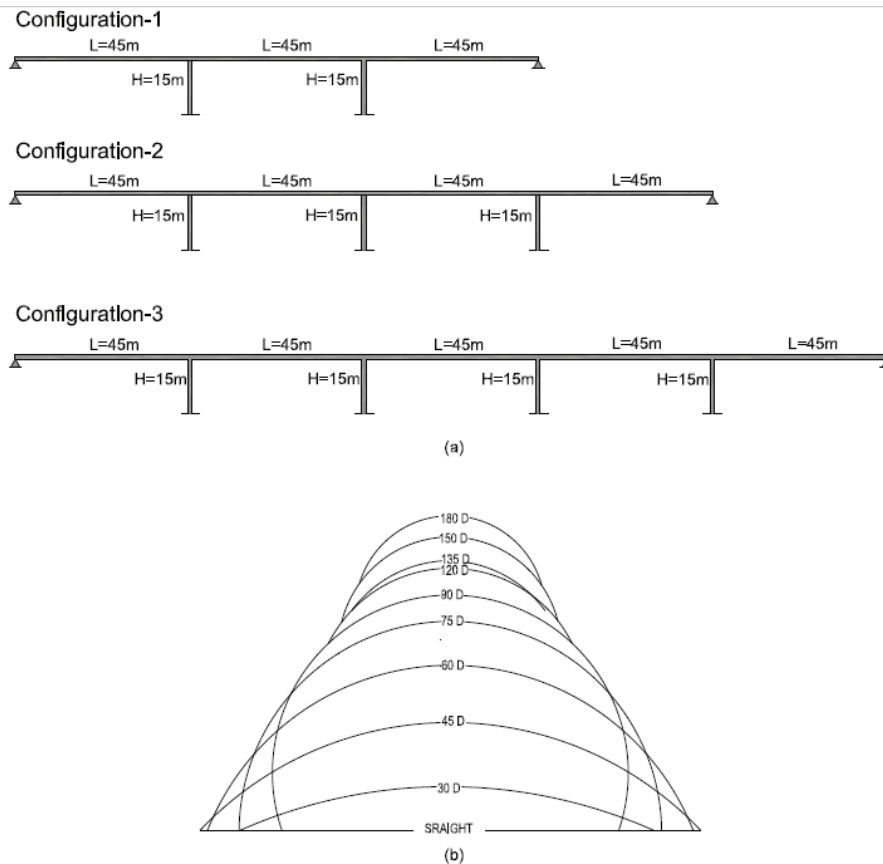


Figure 1. a) Different bridge configurations b) Plane view of straight and curved bridge with different subtended angles.

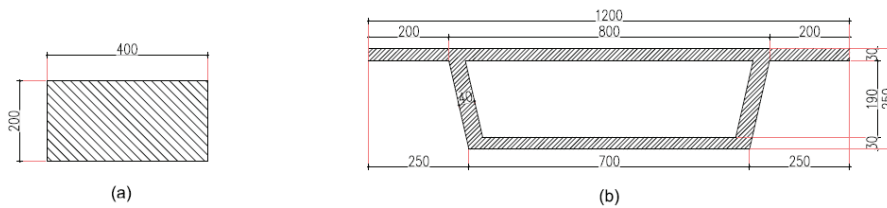


Figure 2. a) Pier column section b) Deck section.

The material of the deck was determined as C45($f_{ck}=45\text{Mpa}$) type concrete and the other structural elements had C30($f_{ck}=30\text{Mpa}$). The selfweight of the structural elements were taken into consideration by the program where concrete unit weight defined as $g_c=25\text{kN/m}^3$ and $g_k=23\text{kN/m}^3$ for asphaltic coating and levelling. In addition, according to the AASHTO specifications, parameters such as percentage of longitudinal bars, limitations of ductility, the minimum concrete cover, the percentage of the axial load, spacing of transverse reinforcement, the ratio of transverse bar volume to the core volume, and the shear capacity design of columns were considered.

The seismic loads acting on the bridges have been taken into account by using multi-mode response spectral analysis method in accordance with AASHTO LRFD Table 2 “seismic design”[1]. Seismic loadings were determined for different seismic ground motion levels (DD-1) and soil class (ZB) as defined in Section 2.2 of the TBEC 2018 Regulation [23]. The design response spectrum has been derived from ‘Turkey Earthquake Hazard Maps’ and the associated website <https://tdth.afad.gov.tr/>. Horizontal elastic design spectrum was created and incorporated into the structural models of the bridges (Fig. 3). The seismic load combinations were defined as follows by adding the

earthquake loads obtained from the multi-mode response spectral analysis to the dead load.

$$1.0 \text{ DL} + 1.0 \text{ EQ}_X + 0.3 \text{ EQ}_Y \quad (\text{COMB1})$$

$$1.0 \text{ DL} + 0.3 \text{ EQ}_X + 1.0 \text{ EQ}_Y \quad (\text{COMB2})$$

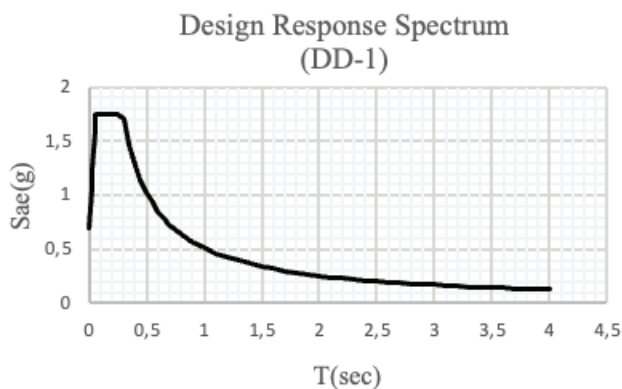


Figure 3. Design response spectrum.

Numerical modeling studies were conducted for the determined variable parameters and bridge configurations. During the structural modelling stage, a general purpose structural analysis program SAP2000 [24] has been used. Bridges with selected configurations; 3-span, 4-span, and 5-span bridge configurations were modelled in 3-D using the SAP2000 structural analysis program (Fig. 4).

In the structural models, the deck and the piers were defined with frame finite elements. The abutments were not modelled exactly instead they were represented with rigid links. Besides, it is important to define the pier-deck connection regions in a manner that accurately reflects the real behavior of the structural elements and geometry. Specifically, for monolithic pier-deck connections, these rigid regions can be represented in the model using fictitious bar elements with infinite rigidity or elastic spring (link) elements with infinite rigidity (Fig. 5).

Elastic spring link elements were selected in the models. The deck were defined with finite shell elements firstly, but analysis results had showed that frame element definitions were preferably used for reliability and the convenience of the analysis results. The additional dead loads of asphaltic cover, guardrails and the pavements were added to the unit weight of the deck definition. "From Element and Additional Masses" option, the program allows the mass distributions from joints and at each node. The program, assigns masses that are compatible with the length of the beam finite elements at that node. Response spectrum data were defined as a function to be used in multi-mode spectral analysis and seismic load cases are defined by these function. The program, correctly analysed the bridge models without any warnings. After the structural analysis the first predominant periods of the bridges were also checked by the excel sheets to verify the result of the program.

For each bridge configuration, numerical models were created and initially analyzed using the linear elastic analysis method. Subsequently, the bridge behaviors in both

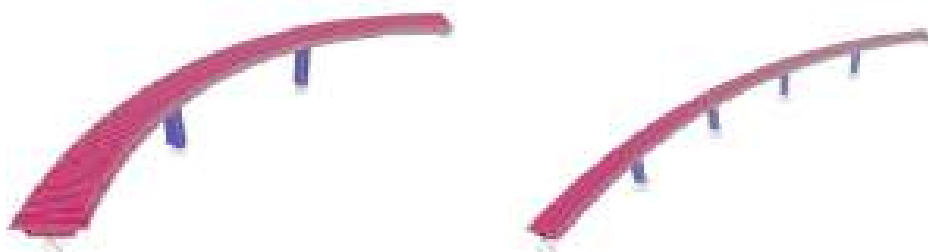


Figure 4. 3-Span, 5-Span Bridges with subtended angles 60°-120° respectively.

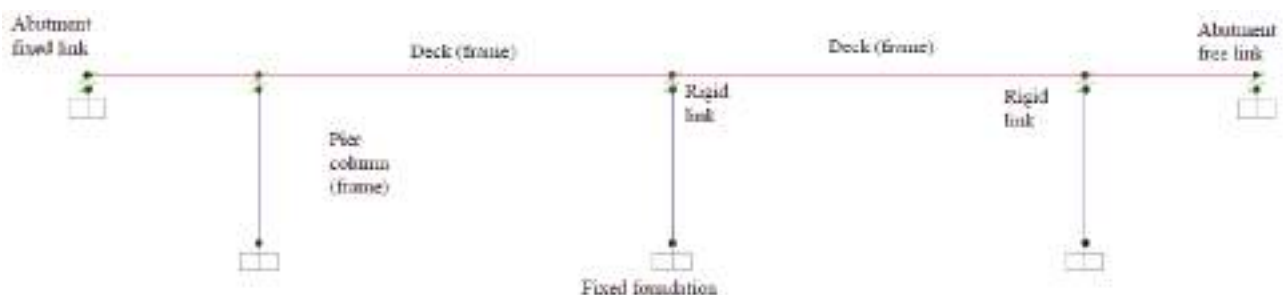


Figure 5. 4-Span Bridge structural model.

straight-axis and curved configurations were examined using the elastic dynamic analysis method defined in AASHTO LRFD [1]. Firstly, a model was prepared for the bridge, and multi-mode response spectral analyses were performed by applying the seismic response spectrum to this model. According to the analysis results, modal quantities of dynamic behavior (mode shapes, periods, frequencies, mass participation ratios), the displacement at the top of the piers, internal forces occurred in the structural elements (columns and deck) were obtained both for equivalent straight bridges and curved formed bridges. The obtained results were graphed to compare the seismic behavior of the bridges and to evaluate the effect of the subtended angle variation due to the radius of curvature. Lastly to get more refined analysis results according to the AASHTO specifications limitations only the subtended angle limit exceedance was taken into consideration in this study.

ANALYSIS RESULTS OF THE STRUCTURAL MODELS OF THE BRIDGES

In this part, the results obtained by using linear modal analyses for different bridge configurations are discussed. For the determined bridge configurations, multi-mode response spectral analyses were performed under the horizontal design response spectrum derived from the 'Turkey Hazard Map' for soil class ZB and the seismic ground motion level DD1 (2475-year return period). AASHTO LRFD specification design criteria were taken into account in the seismic analysis of the bridge (equivalent straight and curved bridges). Earthquake directions due to horizontal curvature were taken into consideration in the structural models. The sign convention for frame internal forces described in the structural program is illustrated in Figure 6 [23]. This sign convention can be described by defining the concept of the positive and negative faces of an object.

The internal forces were obtained from the analysis results according to this sign convention.

Elastic Dynamic Analysis Results of Bridges with 3-Spans of 135m (3x45m) in Length

In this study, to evaluate the limitations and analysis methods suggested by AASHTO Specifications for curved bridges, elastic dynamic analyses were performed, and the results were obtained for both equivalent straight bridges and curved bridges. 3-spans of 135m (3x45m) in length bridges were one of the bridge configurations. As a result of multi-mode response spectral analyses of these bridges, modal quantities of dynamic behavior (i.e., period, natural frequency, modal mass participation) and the pier displacements, and the structural element internal forces were obtained for the equivalent straight-axis bridge of 135m in length and for the curved bridges that had an arc length of 135m with a variable subtended angle of 300-1800 degrees.

The results of the linear modal analyses were obtained, and the variations of the relevant modal quantities with respect to the subtended angle of curvature are presented in Figure 7.

According to the analysis results, the displacements and rotations at the top of the pier columns of the bridges were examined (Fig. 8). The displacement of the pier in the longitudinal direction of the bridges increased as the subtended angle increased, reaching a maximum rate nearly 15 times greater than that of an equivalent straight bridge when compared to a curved bridge with a subtended angle of 180°.

The AASHTO LRFD specification suggests that a bridge is regular if the subtended angle is smaller than 90°; however, according to the analysis results, the displacement of the pier slightly increases for the initial values of the curvature angle.

The displacement of the column in the transverse direction of the bridges did not increase linearly with the

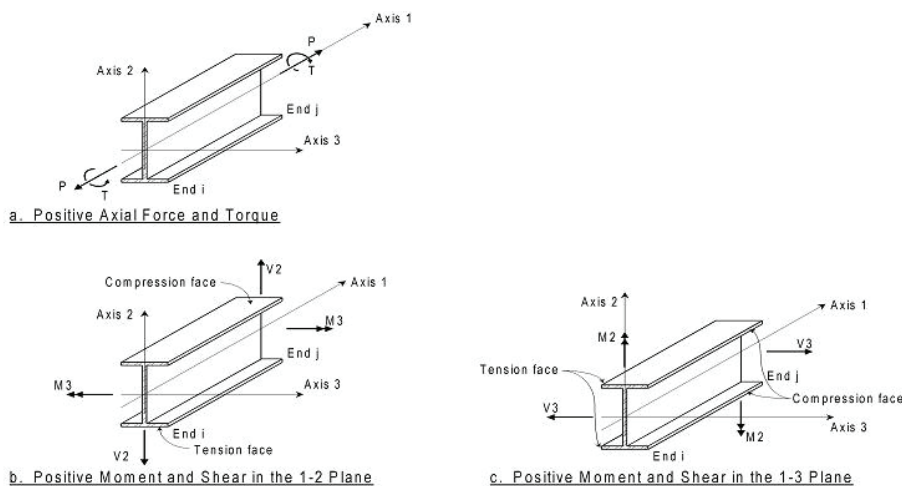


Figure 6. Sign convention for frame internal forces.

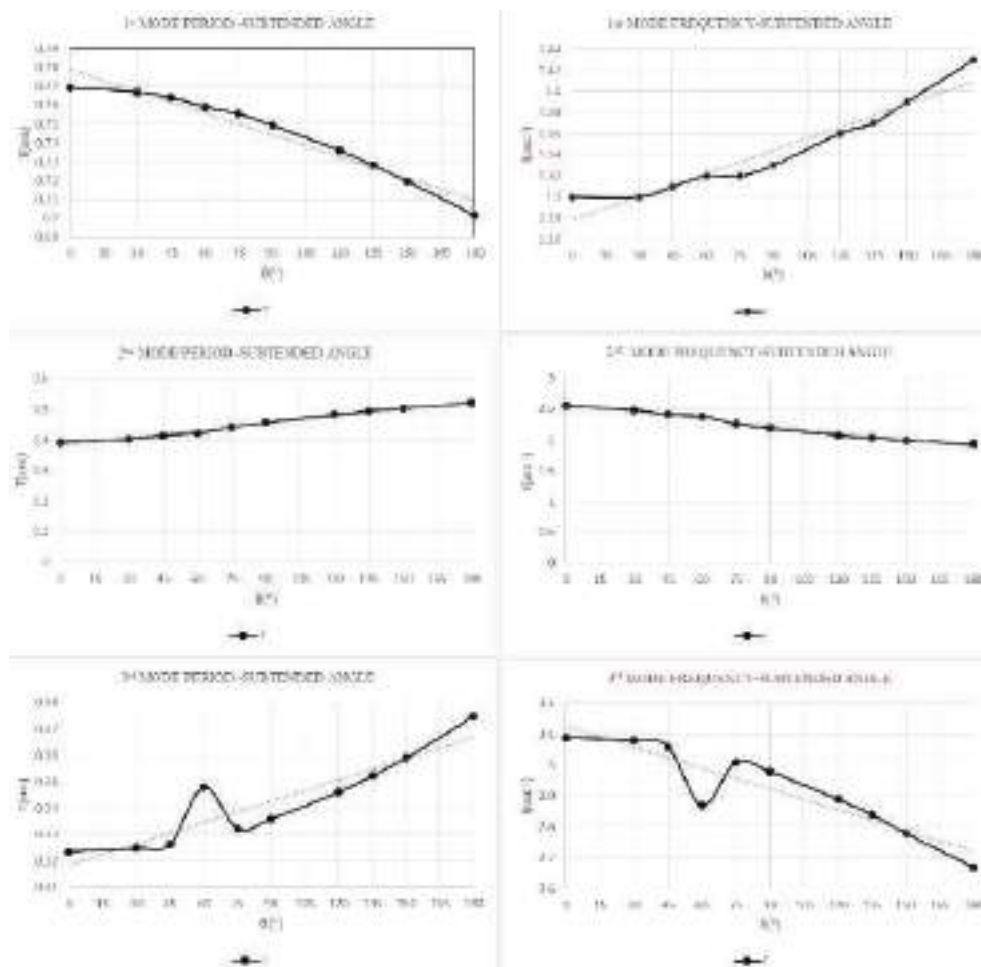


Figure 7. Period and frequency of the first three predominant modes-subtended angle variation.

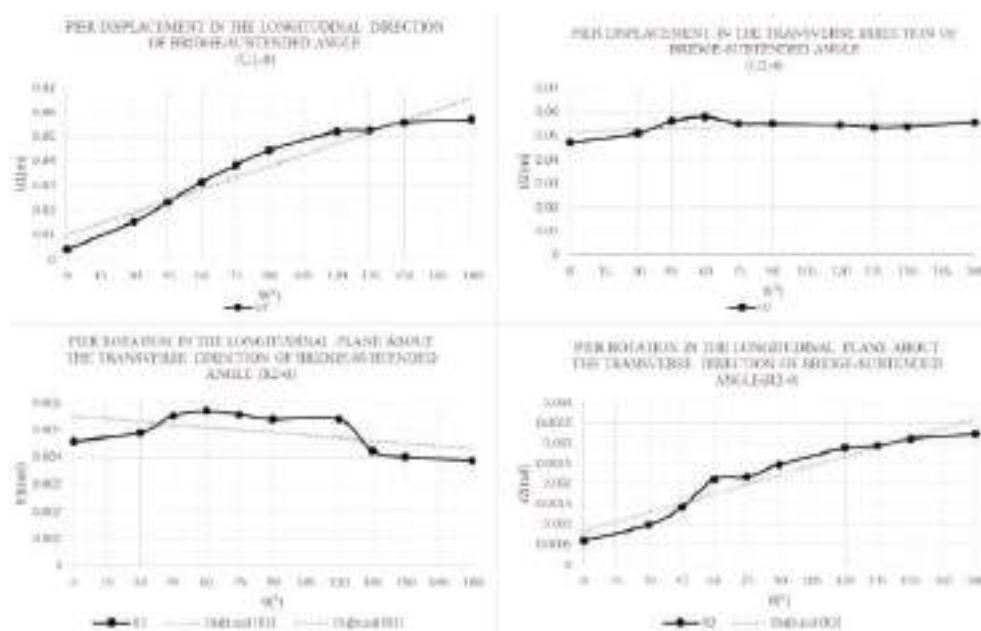


Figure 8. Displacement and rotation at the top of the pier - subtended angle variation.

increase in the subtended angle. The rotation of the pier in the longitudinal direction of the column reached its maximum value for the bridge at a subtended angle of 60° . Lastly, the rotation of the column about the transverse direction of the bridge increased with the variation of the subtended angle. The bridge with a 180° subtended angle had a column rotation nearly six times greater than that of the equivalent straight bridge.

According to the multi-mode response spectral analysis results, the internal forces of the pier columns of the bridges were examined (Fig. 9). The axial force, the shear force in the longitudinal direction of the bridge, the torsional moment, and the bending moment in the longitudinal direction of the bridge increased as the subtended angle increased. Notably, after the subtended angle reached 30° , the shear force and the bending moment in the longitudinal direction increased to nearly five times those of the equivalent straight bridges. The shear force and the bending moment in the transverse direction of the bridge reached their maximum values for the bridge with a subtended angle of 60° ,

which is 37% larger compared to the internal forces of the piers of the equivalent straight bridge.

The internal forces of the bridges' decks were examined using multi-mode response spectral analysis results (Fig. 10). The shear forces, bending moments, and torsional moments of the decks in both directions increased as the subtended angle of the bridges increased, except for the axial force. Particularly notable was the substantial increase in the torsional moment with variations in subtended angle. For instance, the torsional moment of the deck with a subtended angle of 180° was six times greater than that of the equivalent straight bridge. In comparison, the torsional moment of the deck with a subtended angle of 60° was three times greater than the equivalent straight bridge. After the subtended angle surpassed 30° , the shear force and bending moment in the transverse direction increased at a rate nearly three times higher than that of the equivalent straight bridges.

The shear force and bending moment in the transverse direction reached their maximum for the bridge with a subtended angle of 180° . Up to a subtended angle of 90° ,

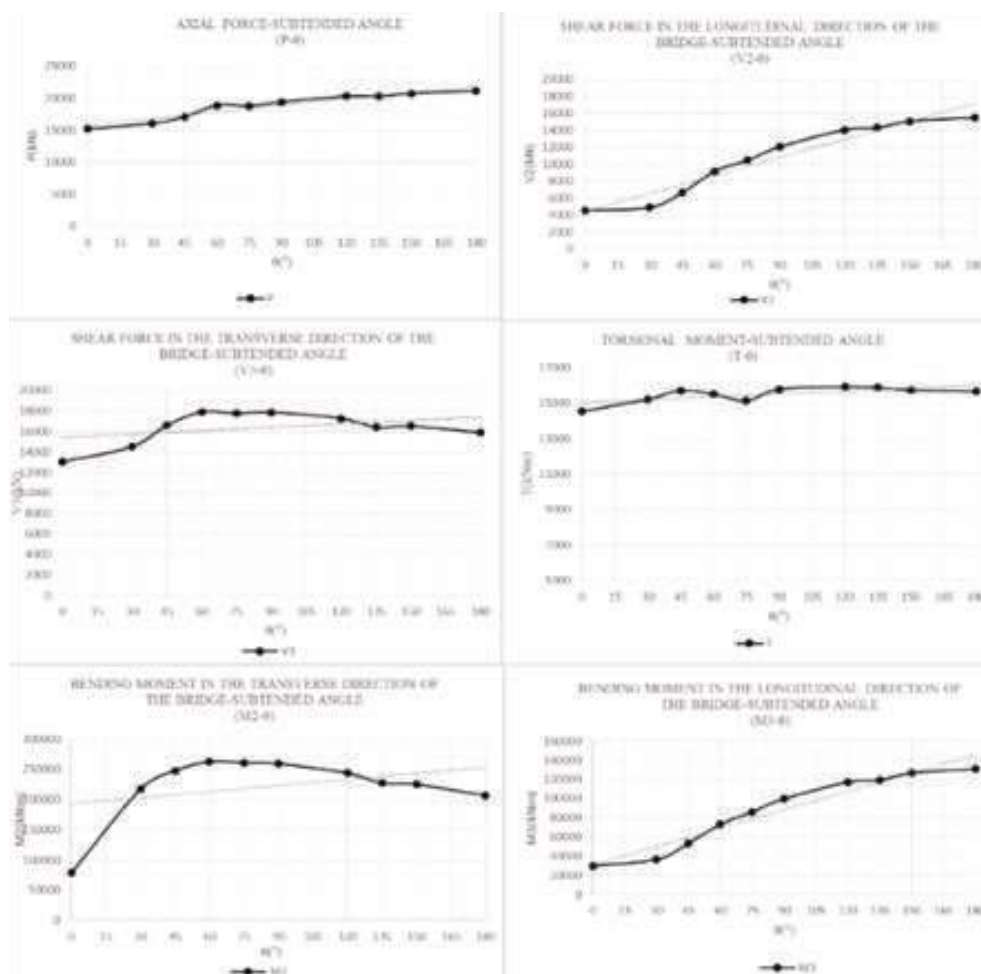


Figure 9. Internal forces of the columns of the bridges - subtended angle variation.

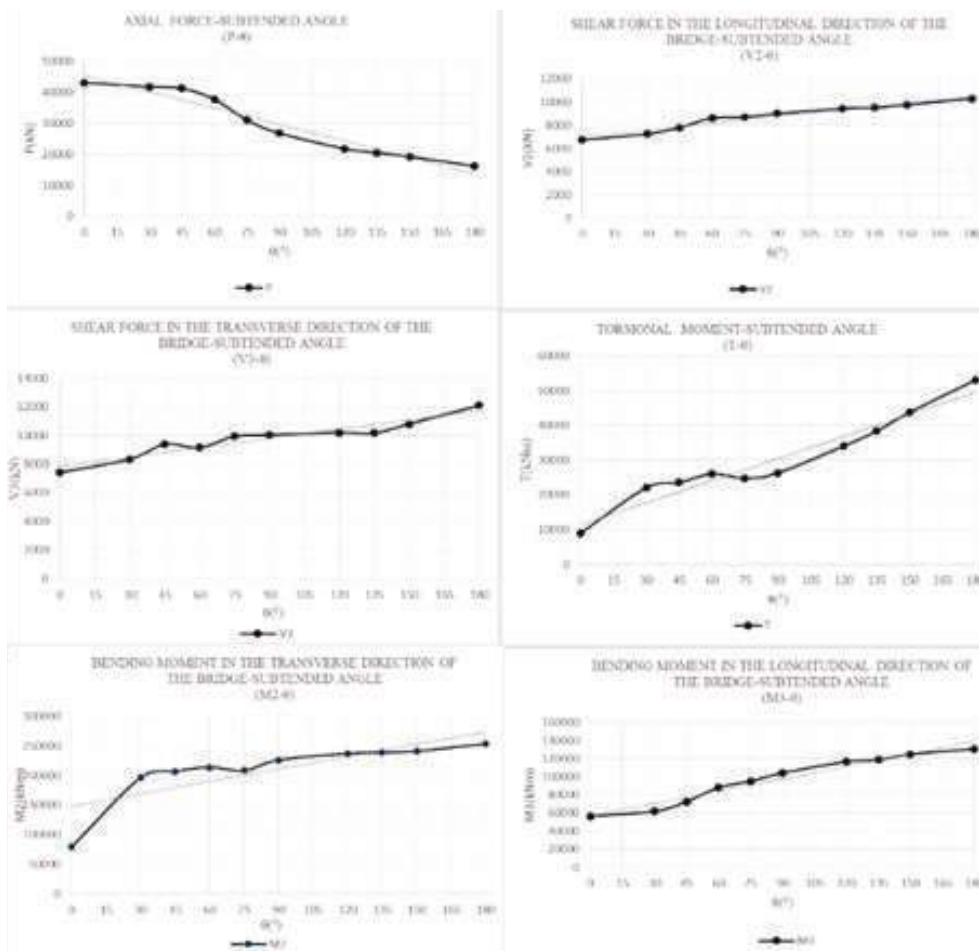


Figure 10. Internal forces of the deck of the bridges - subtended angle variation.

the internal forces increased at a rate 2.7 times higher than those of the equivalent straight bridges.

Elastic Dynamic Analysis Results of Bridges With 4-Spans of 180m (4x45m) in Length

In this study, to evaluate the limitations and analysis methods suggested by AASHTO specifications for curved bridges, elastic dynamic analyses were performed, and the results were obtained for both equivalent straight bridges and curved bridges. One of the bridge configurations analyzed was a 4-span bridge with a total length of 180 m (4x45 m). As a result of the multi-mode response spectral analyses of these bridges, the modal quantities of dynamic behavior (i.e., period, natural frequency, modal mass participation), pier displacements, and structural element internal forces were obtained for an equivalent straight-axis bridge of 180 m in length and for curved bridges with an arc length of 180 m and a variable subtended angle ranging from 30° to 180° degrees.

The results of the linear modal analyses were obtained, and the variations of the relevant modal quantities with

respect to the subtended angle of curvature are presented in Figure 11.

According to the analysis results, the displacements and the rotations at the top of the column of the pier of the bridges were examined (Fig. 12). The displacement of the pier in the longitudinal direction of the bridges increased as the subtended angle increased at a rate of nearly three times more than the equivalent straight bridge compared to a curved bridge that had a subtended angle of 180°. The AASHTO LRFD specification suggests that the bridge is regular in cases where the subtended angle is smaller than 90° but according to the analysis results, the displacement of the pier slightly increased for the initial values of the curvature angle.

The displacement of the column in the transverse direction of the bridges increased with an increase in the subtended angle at a ratio of 1.25. The rotation of the pier in the transverse plane about the longitudinal direction of the column had the maximum value for the bridge with a subtended angle of 60° and decreased at a rate of nearly 44% up to the subtended angle of 180°. Lastly, the rotation of the column in the longitudinal plane about the transverse

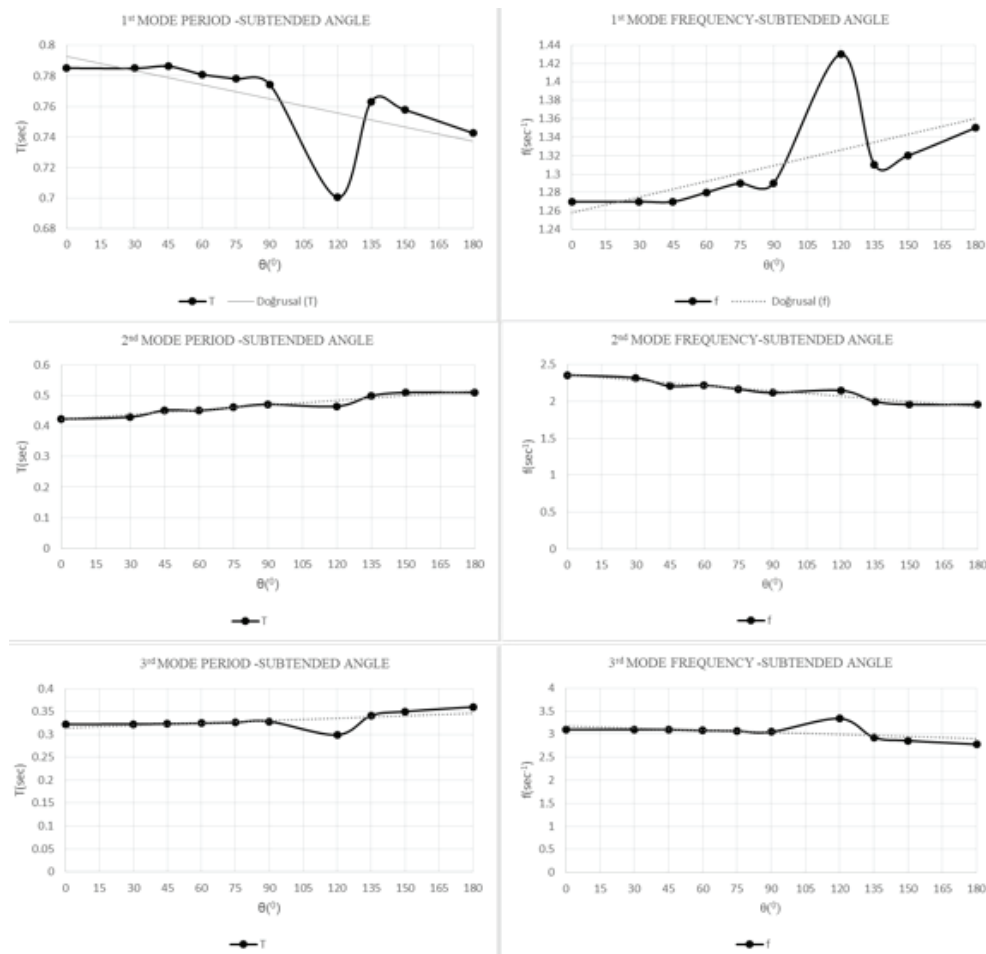


Figure 11. Period and frequency of the first three predominant modes-subtended angle variation.

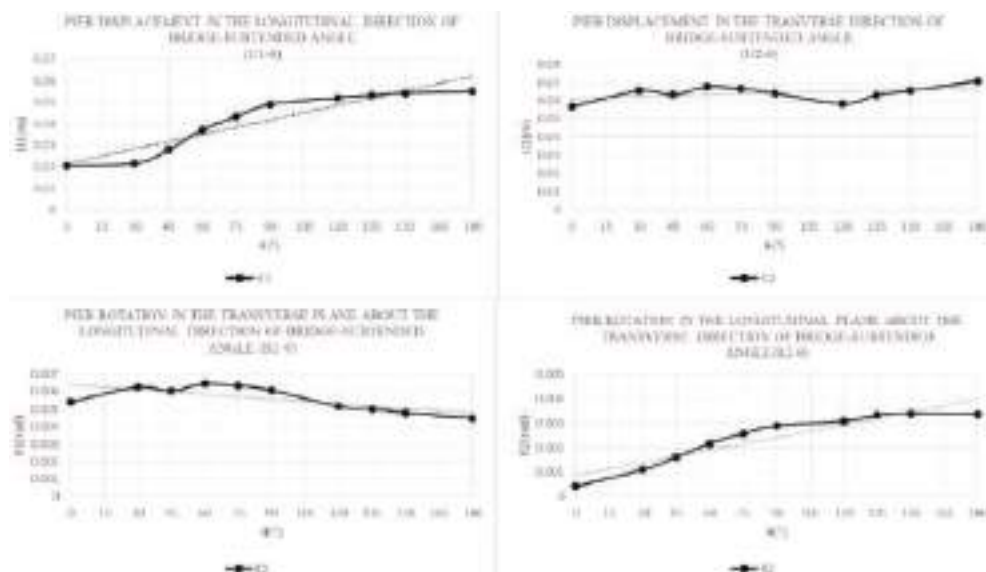


Figure 12. Displacement and rotation at the top of the pier - subtended angle variation.

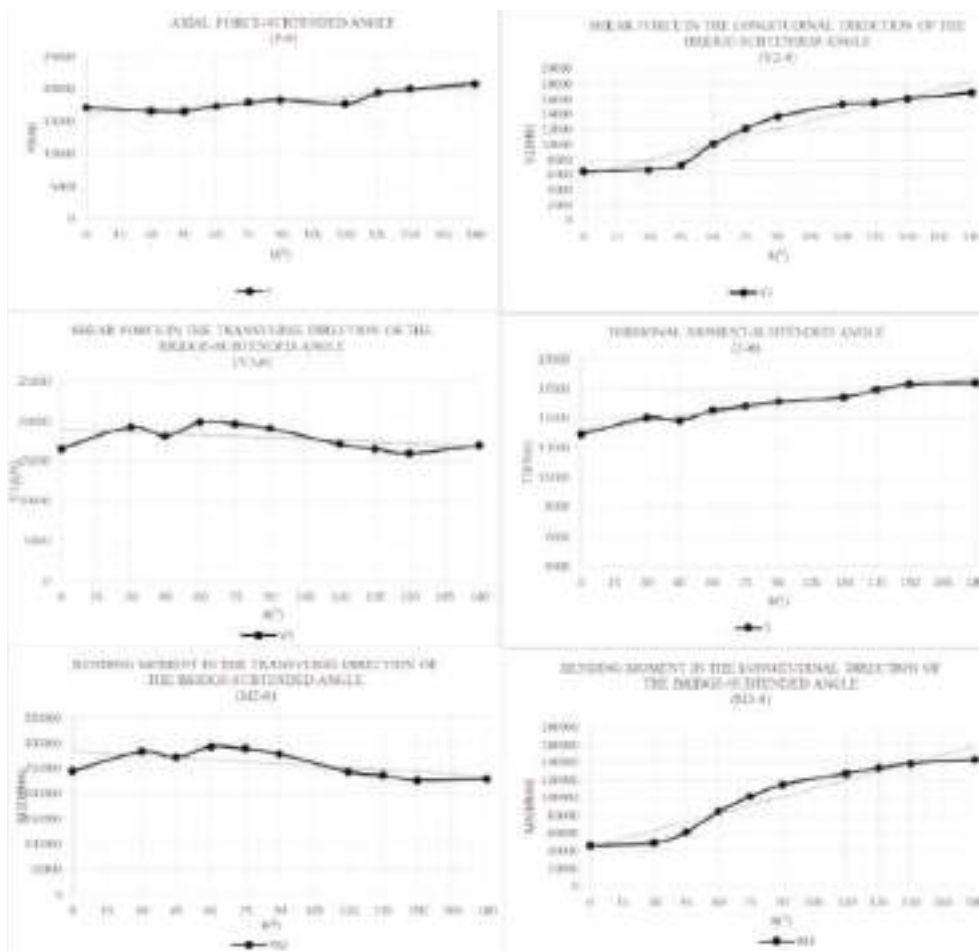


Figure 13. Internal forces of the columns of the bridges - subtended angle variation.

direction of the bridge increased with the variation in the subtended angle.

The bridge with a 180° degree of subtended angle had a column rotation nearly eight times greater than the rotation of the equivalent straight bridge.

According to the multi-mode spectral analysis results, the internal forces of the pier columns of the bridges were examined (Fig. 13). The axial force, shear force in the longitudinal direction of the bridge, torsional moment, and bending moment in the longitudinal direction of the bridge increased as the subtended angle increased.

Particularly after a subtended angle value of 30° , the shear force and bending moment in the longitudinal direction increased to nearly three times more than those of equivalent straight bridges.

The shear force and bending moment in the transverse direction of the bridge reached their maximum values for a bridge with a subtended angle of 60° , which is 20% larger compared to the internal forces of the piers of the equivalent straight bridge.

According to the multi-mode spectral analysis results, the internal forces of the bridge deck were examined (Fig.

14). The shear forces, bending moments, and torsional moments in both directions of the bridge increased as the subtended angle of the bridge increased, except for the axial force. The increment ratio for the torsional moment was particularly high given the variation in the subtended angle. The torsional moment of the deck of the bridge with a subtended angle of 180° was twice that of the equivalent straight bridge. The torsional moment of the deck remained stable between subtended angle values of 30° and 120° and increased up to a subtended angle of 180° . Additionally, the shear force and bending moment in the transverse direction increased by nearly 50%, reaching their maximum values for a bridge with a subtended angle of 180° . The bending moment of the bridge increased by 66% up to a subtended angle of 180° compared to the results of the equivalent straight bridge.

Elastic Dynamic Analysis Results of Bridges With 5-Spans of 225m (5x45m) in Length

5-spans of 225m (5x45m) in length bridges were one of the bridge configurations used to evaluate the curved bridges and straight bridges. As a result of multi-mode

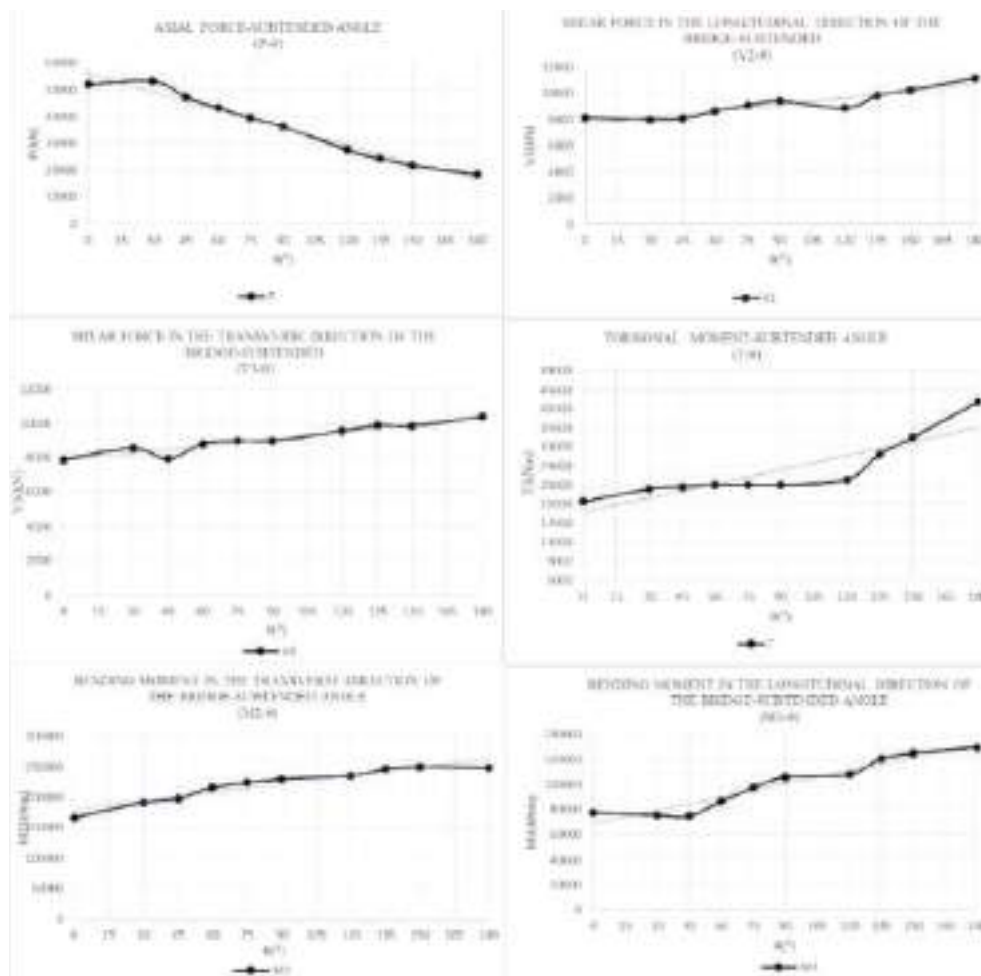


Figure 14. Internal forces of the deck of the bridges - subtended angle variation.

response spectral analyses of these bridges; modal quantities of dynamic behavior (i.e., period, natural frequency, modal mass participation), the pier displacements and the structural element internal forces were obtained for the equivalent straight-axis bridge of 225m in length and for the curved bridges that had an arc length of 225m with a variable subtended angle of 30°-180° degrees.

The results of the linear modal analyses were obtained, and the variations of the relevant modal quantities with respect to the subtended angle of curvature are presented in Figure 15.

According to the analysis results, the displacements and the rotations at the top of the column of the pier of the bridges were examined (Fig. 16). The displacement of the pier in the longitudinal direction of the bridges increased as the subtended angle increased at a rate of nearly two times more than that of an equivalent straight bridge compared to a curved bridge that had a subtended angle of 180°. The displacement of the column in the transverse direction of the bridges increased with the increase of the subtended angle at a ratio of 1.30 times. The rotation of the pier in

the longitudinal direction of the column had the maximum value for the bridge with a subtended angle of 45° and decreased at a rate of nearly 23% up to the subtended angle of 180° for 5-span bridge. Lastly, the rotation of the column about the transverse direction of the bridge increased with the variation in the subtended angle. The bridge with an 180° degree subtended angle had a column rotation nearly four times greater than the rotation of the equivalent straight bridge.

According to the multi-mode spectral analysis results, the internal forces of the pier columns of the bridges were examined (Fig. 17). The axial force, the shear force in the longitudinal direction of the bridge, the torsional moment, and the bending moment in the longitudinal direction of the bridge was increased as the subtended angle increased. Especially after the subtended angle value of 30°, the shear force and the bending moment in the longitudinal direction increased at a ratio of nearly 2.5 times more than the equivalent straight bridges. The shear force and the bending moment in the transverse direction of the bridge reached their maximum value for the bridge, which had a subtended

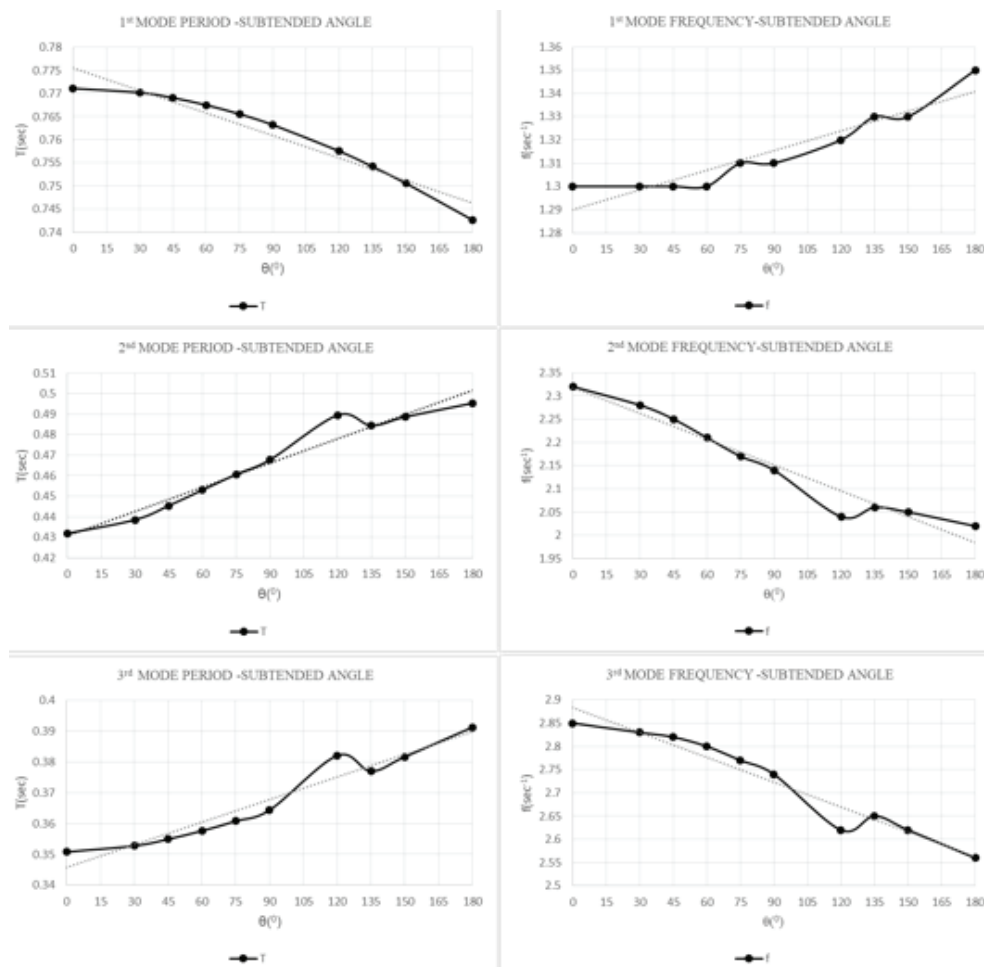


Figure 15. Period and frequency of the first three predominant modes-subtended angle variation.

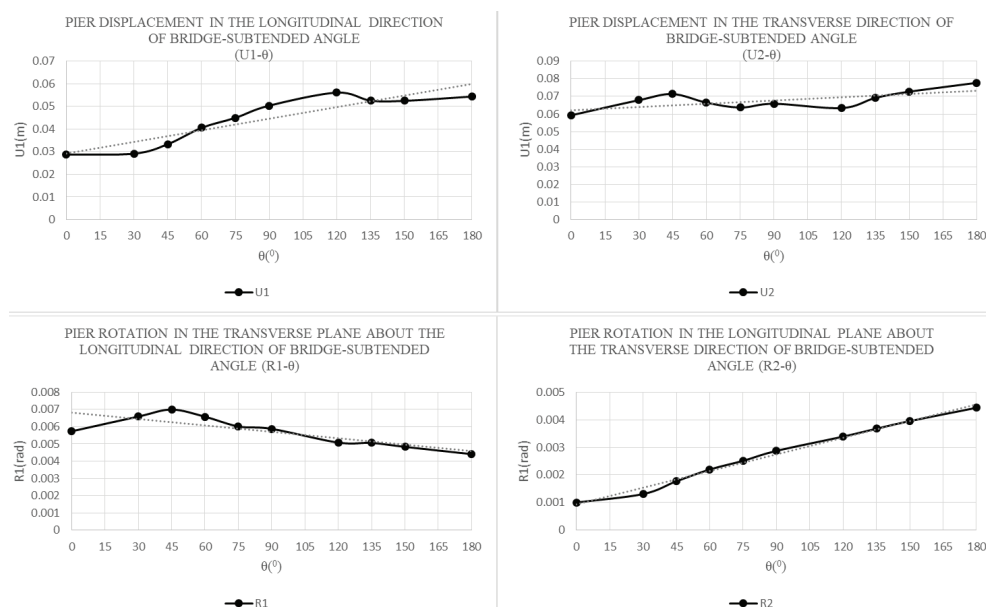


Figure 16. Displacement and rotation at the top of the pier - subtended angle variation.

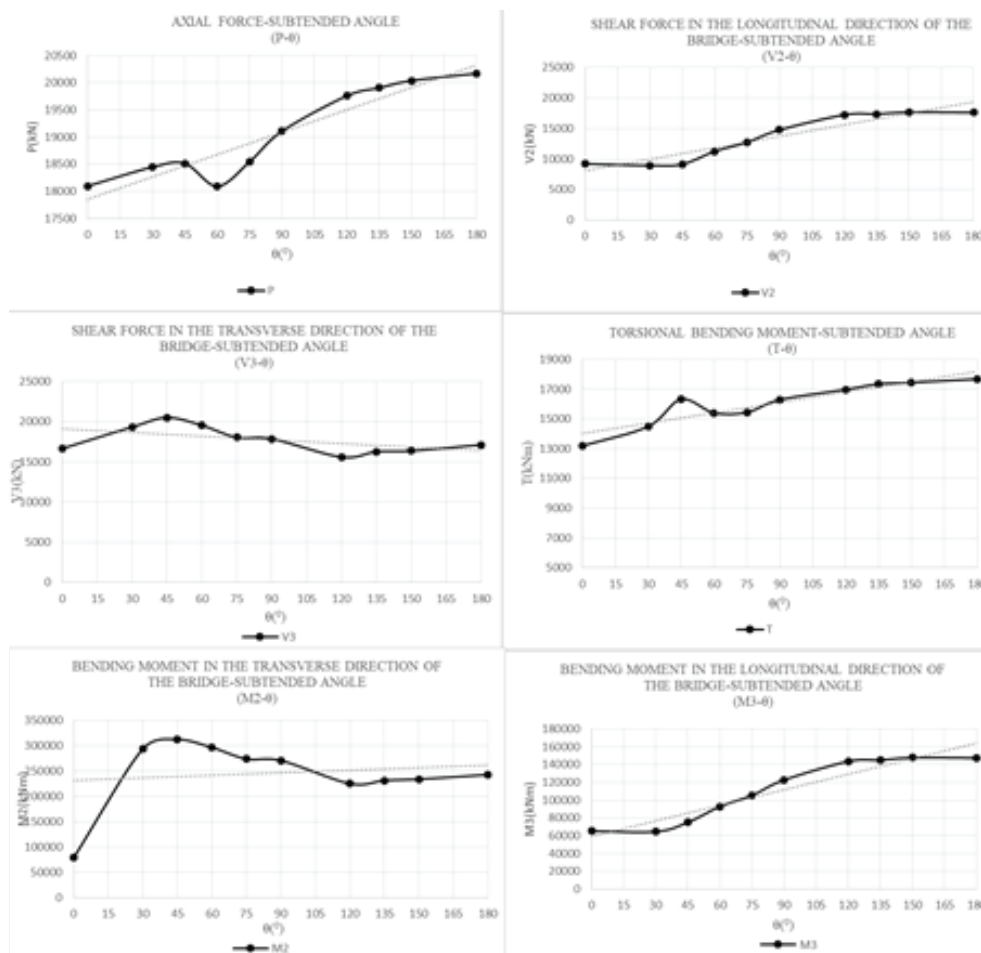


Figure 17. Internal forces of the columns of the bridges - subtended angle variation.

angle of 45° that was four times larger than compared with the internal forces of an equivalent straight bridge.

According to the multi-mode spectral analysis results, the internal forces of the deck of the bridges were examined (Fig. 18). The shear forces, bending moments and the torsional moment of the deck in both directions of the bridge were increased as the subtended angle of the bridge were increased except the axial force. The torsional moment of the deck of the bridge with a subtended angle of 180° was four times greater than that of the torsional moment of the equivalent straight bridge. The torsional moment of the deck of the bridge with a subtended angle of 45° was three times greater than the torsional moment of the equivalent straight bridge. In addition, the shear force and the bending moment in the transverse direction increased at a ratio of nearly three times more than the equivalent straight bridge. In the transverse direction of the bridge reached the maximum value for the bridge that had a subtended angle of 180° . Bending moment of the bridge increased at the rate of %55 up to the subtended angle 180° compared to the results of the equivalent straight bridge.

THE COMPARISON OF ELASTIC DYNAMIC ANALYSIS RESULTS OF BRIDGES WITH 3-4-5-SPANS

In this study, for evaluating the limitations and the analysis methods suggested by AASHTO specifications for curved bridges, elastic dynamic analyses were performed, and the results were obtained for both equivalent straight bridges and curved bridges for the determined configurations of 3-spans, 4-spans, and 5-spans of bridges. As a result of multi-mode response spectral analyses of these bridges; modal quantities of dynamic behaviour (i.e., period, natural frequency, modal mass participation), the pier displacements and the structural element internal forces were obtained for equivalent straight axis bridges and for the curved bridges that had arc lengths of 135m, 180m and 225m with a variable subtended angle 30° - 180° degrees. The analysis results were then compared according to variation of the span number and the bridge length.

The results of the linear modal analyses were obtained, and the variations of the relevant modal quantities with respect to the subtended angle of curvature were presented in Figure 19.

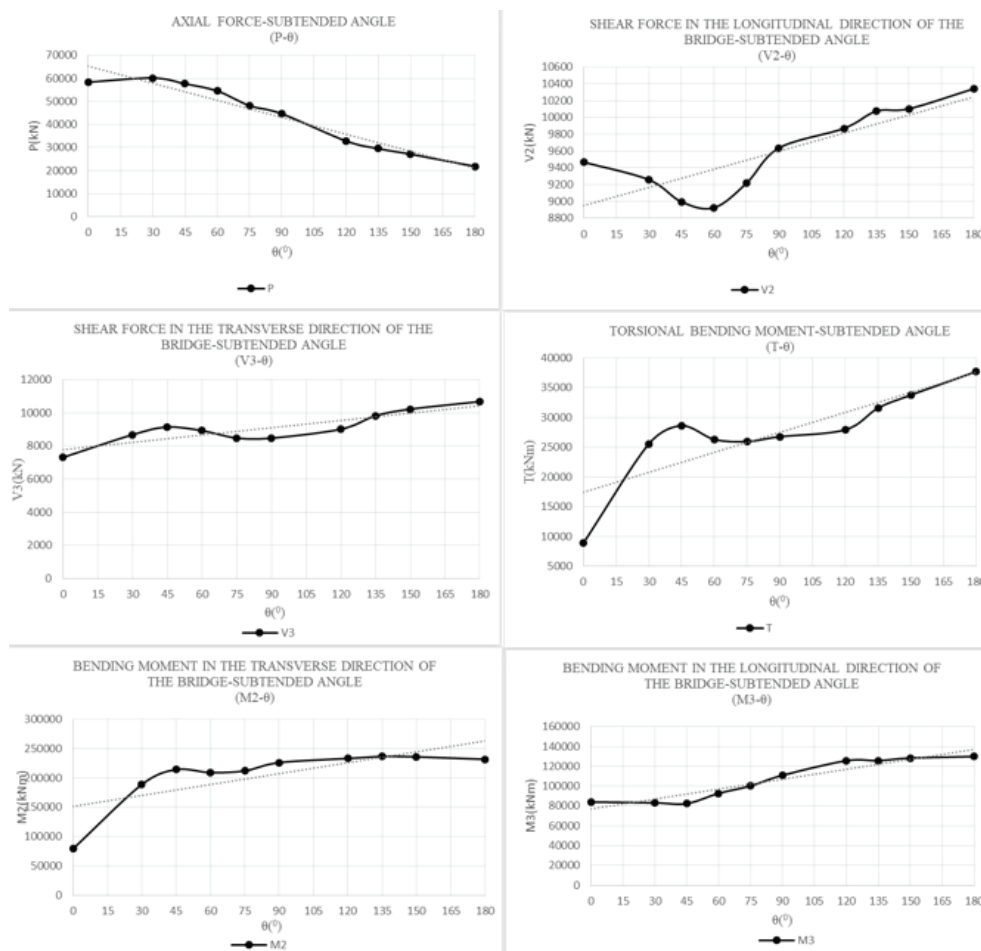


Figure 18. Internal forces of the deck of the bridges - subtended angle variation.

Mode Shape	Subtended Angle (°)	3-Span Bridge					4-Span Bridge					5-Span Bridge				
		T ⁽¹⁾ (sec)	f (sec ⁻¹)	SumUX ⁽¹⁾	SumUY ⁽¹⁾	SumRZ ⁽¹⁾	T ⁽¹⁾ (sec)	f (sec ⁻¹)	SumUX ⁽¹⁾	SumUY ⁽¹⁾	SumRZ ⁽¹⁾	T ⁽¹⁾ (sec)	f (sec ⁻¹)	SumUX ⁽¹⁾	SumUY ⁽¹⁾	SumRZ ⁽¹⁾
1 st Mode shape	0	0.770	1.3	0	0.31368	0.49082	0.785	1.27	0.000	0.201	0.519	0.771	1.3	0.000	0.161	0.432
	30	0.767	1.3	0.003	0.309	0.667	0.785	1.27	0.008	0.193	0.496	0.770	1.3	0.006	0.156	0.412
	45	0.764	1.31	0.007	0.303	0.645	0.786	1.27	0.015	0.215	0.524	0.769	1.3	0.013	0.151	0.389
	60	0.759	1.32	0.013	0.294	0.611	0.781	1.28	0.023	0.186	0.444	0.768	1.3	0.022	0.145	0.359
	75	0.756	1.32	0.021	0.286	0.577	0.778	1.29	0.035	0.178	0.406	0.766	1.31	0.033	0.136	0.325
	90	0.749	1.33	0.031	0.275	0.533	0.774	1.29	0.048	0.167	0.361	0.763	1.31	0.045	0.126	0.284
	120	0.736	1.36	0.056	0.249	0.430	0.701	1.43	0.076	0.159	0.291	0.758	1.32	0.072	0.101	0.199
	135	0.728	1.37	0.070	0.235	0.375	0.763	1.31	0.096	0.137	0.237	0.754	1.33	0.085	0.088	0.158
	150	0.720	1.39	0.086	0.221	0.320	0.758	1.32	0.120	0.130	0.195	0.751	1.33	0.098	0.074	0.119
	180	0.702	1.43	0.115	0.193	0.219	0.743	1.35	0.135	0.085	0.096	0.743	1.35	0.120	0.046	0.056
2 nd Mode shape	0	0.392	2.55	0.000	0.839	0.586	0.423	2.36	0.000	0.836	0.886	0.432	2.32	0.000	0.841	0.852
	30	0.403	2.48	0.086	0.794	0.919	0.430	2.32	0.072	0.799	0.881	0.439	2.28	0.083	0.787	0.846
	45	0.414	2.41	0.169	0.751	0.916	0.452	2.21	0.173	0.819	0.958	0.445	2.25	0.160	0.739	0.844
	60	0.421	2.38	0.260	0.706	0.908	0.451	2.22	0.244	0.694	0.877	0.453	2.21	0.237	0.687	0.845
	75	0.442	2.26	0.336	0.661	0.908	0.462	2.17	0.317	0.647	0.875	0.461	2.17	0.304	0.641	0.852
	90	0.456	2.19	0.405	0.620	0.904	0.471	2.12	0.376	0.602	0.872	0.468	2.14	0.354	0.596	0.855
	120	0.482	2.07	0.506	0.547	0.892	0.465	2.15	0.471	0.536	0.881	0.490	2.04	0.466	0.508	0.849
	135	0.493	2.03	0.539	0.515	0.884	0.499	2	0.486	0.513	0.917	0.485	2.06	0.418	0.499	0.877
	150	0.503	1.99	0.561	0.484	0.874	0.510	1.96	0.531	0.518	0.947	0.489	2.05	0.413	0.474	0.880
	180	0.519	1.93	0.580	0.427	0.844	0.510	1.96	0.475	0.420	0.857	0.495	2.02	0.377	0.437	0.871
3 rd Mode shape	0	0.323	3.09	0.000	0.839	0.586	0.322	3.1	0.003	0.836	0.886	0.351	2.85	0.000	0.844	0.911
	30	0.325	3.08	0.086	0.796	0.920	0.323	3.1	0.073	0.803	0.885	0.353	2.83	0.125	0.788	0.906
	45	0.326	3.06	0.169	0.754	0.917	0.323	3.1	0.176	0.826	0.968	0.355	2.82	0.239	0.739	0.904
	60	0.348	2.87	0.261	0.707	0.909	0.324	3.08	0.244	0.704	0.889	0.358	2.8	0.356	0.691	0.902
	75	0.332	3.01	0.338	0.665	0.911	0.326	3.07	0.318	0.660	0.892	0.361	2.77	0.464	0.649	0.902
	90	0.336	2.98	0.409	0.624	0.908	0.328	3.05	0.378	0.617	0.892	0.364	2.74	0.555	0.610	0.896
	120	0.346	2.89	0.515	0.551	0.899	0.299	3.34	0.592	0.566	0.900	0.382	2.62	0.708	0.560	0.894
	135	0.352	2.84	0.550	0.519	0.893	0.341	2.93	0.562	0.548	0.935	0.377	2.65	0.750	0.540	0.888
	150	0.359	2.78	0.576	0.488	0.885	0.350	2.86	0.704	0.575	0.959	0.382	2.62	0.789	0.526	0.884
	180	0.375	2.67	0.605	0.430	0.864	0.360	2.78	0.651	0.490	0.959	0.391	2.56	0.836	0.510	0.874

Figure 19. Modal quantities obtained from the linear modal analyses for all bridge Configurations.

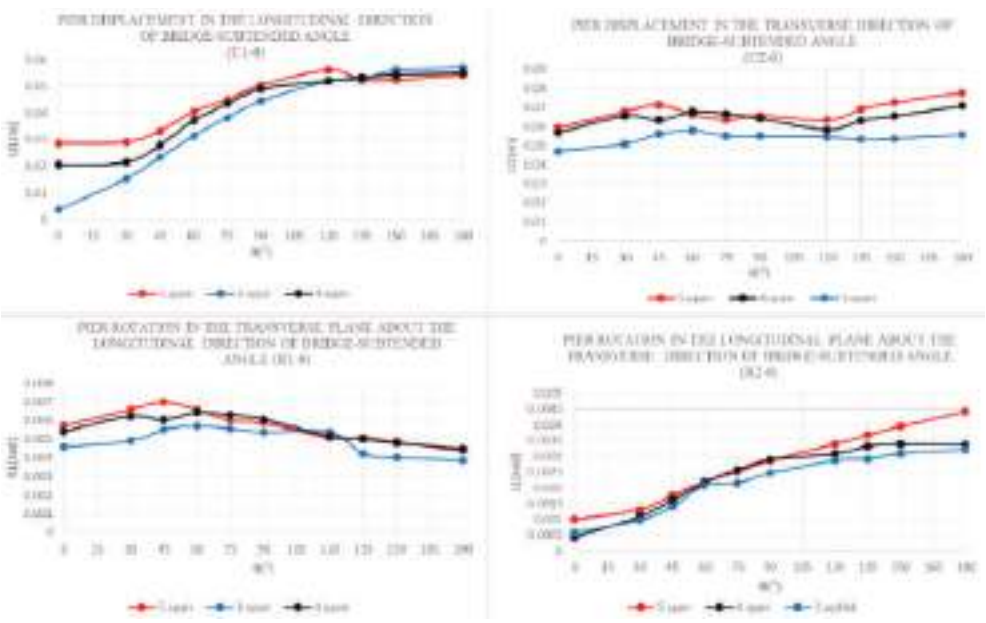


Figure 20. Displacement and rotation at the top of the pier - subtended angle variation.

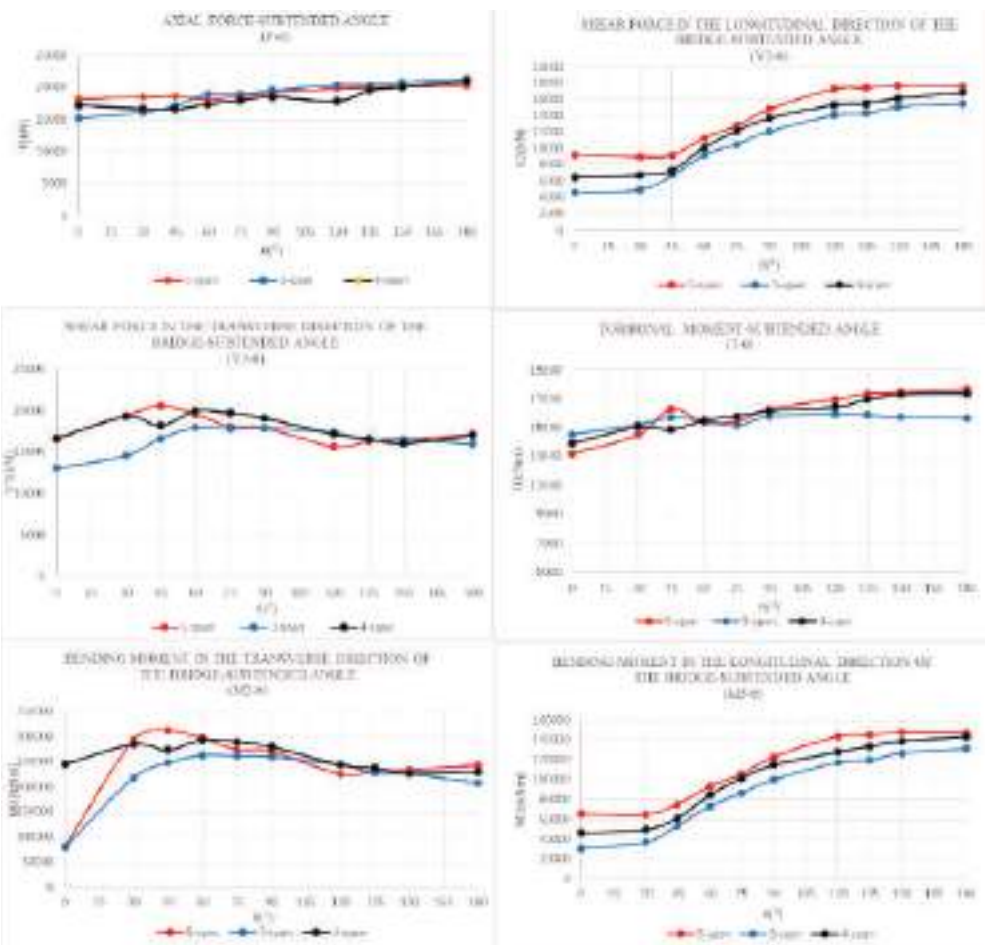


Figure 21. Internal forces of the columns of the bridges - subtended angle variation.

According to the analysis results, the displacements and the rotations at the top of the column of the pier of the bridges were examined (Fig. 20). The displacement of the pier in the longitudinal direction of the bridges increased as the subtended angle increased at a rate of nearly fifteen times more than the equivalent straight bridge compared to a curved bridge that had a subtended angle of 180° for the 3-span bridge configuration. The displacement of the column in the transverse direction of the bridges increased with the increase of the subtended angle at a ratio of 1.375 times more than for the 5-span bridges. The rotation of the pier in the transverse plane about the longitudinal direction of the column had the maximum value for the bridge with a subtended angle of 45° and decreased at the rate of nearly %23 up to the subtended angle of 180° for the 5-span bridge. Lastly, the rotation of the column about the transverse direction of the bridge increased with the variation the subtended angle for all bridge configurations. However, the highest increase rate was for the 5-span bridge with four times greater than the equivalent straight bridge. The bridge with a 180° of subtended angle had the column rotation of nearly four times more than the rotation of the equivalent straight bridge.

According to the multi-mode spectral analysis results, the internal forces of the pier columns of the bridges were examined (Fig. 21). The axial force, shear force in the longitudinal direction of the bridge, torsional moment, and bending moment in the longitudinal direction of the bridge increased as the subtended angle increased. Particularly after a subtended angle value of 30° , the shear force and bending moment in the longitudinal direction increased to nearly three times more than those of equivalent straight bridges for 3-span bridges whereas the maximum values of the internal forces were obtained for 5-span bridge configurations. The shear force and bending moment in the transverse direction of the bridge reached their maximum values for a bridge with a subtended angle of 45° for a 5-span bridge, which is four times greater compared to the internal forces of the piers of the equivalent straight bridge, whereas for 3-span and 4-span bridges the maximum values of those forces were obtained with a subtended angle of 60° .

According to the multi-mode response spectral analysis results, the internal forces of the bridge deck were examined (Fig. 22). The shear forces, bending moments, and torsional moments in both directions of the bridge increased as the subtended angle of the bridge increased, except for

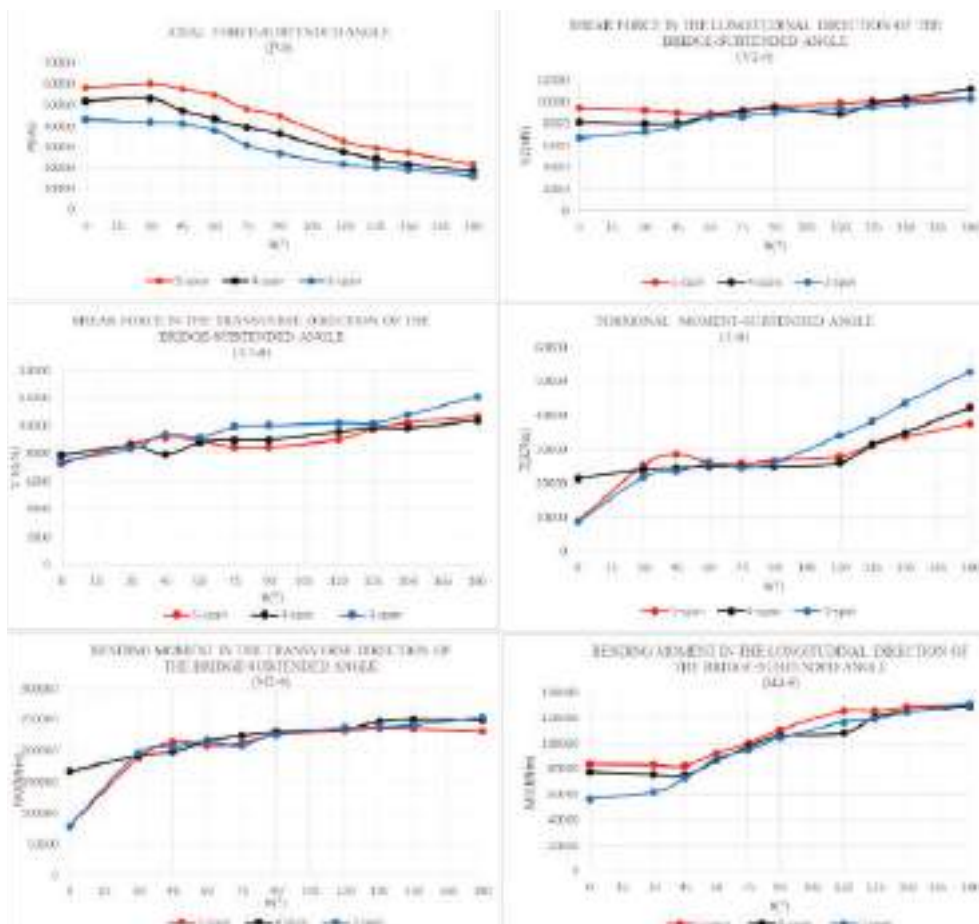


Figure 22. Internal forces of the deck of the bridges - subtended angle variation.

the axial force for all bridge configurations. The increment ratio for the torsional moment was particularly high with the variation in the subtended angle. The torsional moment of the deck of the bridge with a subtended angle of 180° was six times more than that of the equivalent straight bridge for the 3-span bridge configuration, which was the greater value of all bridge configurations.

Furthermore, the transverse shear force and bending moment increased by almost 63% and reached their maximum values for a bridge with a subtended angle of 180° . In comparison to the results of the similar straight bridge for the 3-span structure, the bending moment of the bridge rose by 221% up to a subtended angle of 180° .

CONCLUSION

In this study, the parametric bridge models were generated, designed, loaded, and analyzed in accordance with the AASHTO LRFD Bridge Design Specifications [1] and the TBEC [22] by varying the number of spans, bridge length, and the subtended angle of the curvature. In AASHTO LRFD [1], for bridges that have horizontal curvature in plan, the condition that allows bridges to be idealized as straight bridges is specified as having a subtended angle up to 90° degrees, whereas this limit has been revised to 30° degrees in the AASHTO Guide Specifications [2]. These conservative limitations arise due to uncertainties in the behavior of curved bridges under seismic excitations, so in order to investigate these limitations, numerical analyses were carried out. Multi-mode spectral analyses were performed for both horizontally curved and equivalent straight bridges for investigating the effect of the curvature during seismic excitation.

The dynamic behavior mode shapes and related dynamic quantities (period, natural frequency, and the modal mass participation ratios vs.) of the parametric bridges were obtained by using the linear analysis method. Additionally, the displacements at the top of the pier columns, as well as the internal forces of the structural elements (pier column and deck internal forces), have been determined.

The multi-mode spectral analysis results for defined bridge configurations can be summarised as follows by using the fuzzy technique that introduce fuzzy variables and membership functions to describe the relationship between the subtended angle and the results.

Symbols and notations:

q : Subtended angle

T_1, T_2, T_3 : The periods of the first, second, and third mode shapes, respectively.

$MM_L, MM_T, MM_{Torsion}$: The modal mass participation ratios in the longitudinal, transverse, and torsional directions

D_L, D_T : The longitudinal and transverse displacements at the top of the pier

R_L, R_T : The rotations of the pier about the longitudinal and transverse directions

N : Axial force of the structural elements

V_L, V_T : The longitudinal and transverse shear force of the structural elements respectively

M_L, M_T : The longitudinal and transverse bending moment of the structural elements respectively

T : Torsion of the structural elements

$F(q)$: Fuzzy sets that describes the relationship between the analysis results and the subtended angle q .

$\mu(q)$: Fuzzy membership function that captures the sensitivity of the analysis results with respect to the variation of the subtended angle q .

1. The analysis results indicated that the period of the first mode, denoted; as T_1 decreased as the subtended angle θ increased across all bridge configurations. The membership function $\mu(T_1)$ represents the influence of θ on the period T_1 . For the 4-span bridge configuration, the period of the first predominant mode was higher than that of the other configurations. Specifically, at a subtended angle of 120° degrees, the period T_1 , decreased by approximately 12%. However, as θ approached 180° degrees, the period T_1 showed an increase, as captured by the fuzzy set $F_T(\theta)$. For the 3-span bridge configuration, the period decrease ratio with respect to the subtended angle θ was the highest, with a rate of nearly 10%, indicating a significant sensitivity of T_1 to θ . This sensitivity is described by the fuzzy membership function $\mu(\theta)$. For the second and third mode shapes, denoted as T_2 and T_3 , the periods increased as the subtended angle increased. The fuzzy sets $F_{T_2}(\theta)$ and $F_{T_3}(\theta)$ illustrate this relationship, showing that the behavior of the bridges remained consistent across different configurations during the seismic analysis, as indicated by similar membership functions $\mu(T_2)$ and $\mu(T_3)$.

This fuzzy-based formulation of the mode period analysis highlights the nuanced relationship between the subtended angle and the dynamic behavior of the bridges. The results underscore the importance of incorporating fuzzy logic into the analysis to account for the uncertainties and variations in the seismic response of curved bridges.

As a conclusion, although the modal behavior of bridges is independent of the bridge span number and length, the 3-span bridge configuration showed the most rigid behavior depending on the curvature variation.

2. The analysis showed that the modal mass participation ratio in the longitudinal direction (MM_L) of the bridges increased as the subtended angle θ increased. The fuzzy set $F_{MML}(\theta)$ describes this relationship, indicating that MM_L reached up to 20 times the modal mass participation ratio of the equivalent straight bridge ($\theta=0^\circ$) for the 4-span bridge configuration. The membership function $\mu(MM_L)$ captures the sensitivity of the longitudinal mass participation to changes in θ . In contrast, the mass participation ratios in the transverse direction

(MM_T) decreased as the subtended angle θ increased across all bridge configurations. The fuzzy set $F_{MMT}(\theta)$ represents this trend, with the membership function $\mu(MM_T)$ reflecting the decreasing influence of θ on the transverse mass participation. For the torsional mode, the mass participation ratio ($MM_{Torsion}$) decreased with increasing subtended angle θ for the first predominant mode shape. Among the configurations, the 3-span bridge exhibited the highest torsional mass participation ratio, $MM_{Torsion}$, reaching up to 75%, as characterized by the fuzzy set $F_{MMTorsion}(\theta)$ and its associated membership function $\mu(MM_{Torsion})$. As a conclusion, the variation of modal mass participation ratios of the bridges with the variation of subtended angle were directly effected by the span number and bridge length. 4-span bridge and 3-span bridge had the highest modal mass participation ratio variation effects.

3. The analysis results revealed that the displacement of the pier in the longitudinal direction (D_L) of the bridges increased as the subtended angle θ increased. The fuzzy set $F_{DL}(\theta)$ describes this relationship, indicating that D_L for the 3-span bridge reached a maximum value nearly fifteen times greater than that of the equivalent straight bridge ($\theta=0^\circ$) at a subtended angle of 180° degrees. Similarly, the longitudinal displacement of the pier increased by approximately 12 times for a curved bridge with a subtended angle of 90° degrees. For the 4-span and 5-span bridges, the increase in D_L was nearly 2.5 and 2 times, respectively, as captured by the fuzzy membership functions $\mu(D_L)$ for each configuration. In the transverse direction (D_T), the pier displacement increased with the variation of the subtended angle θ , reaching a ratio of 1.375 for the 5-span bridges, as represented by the fuzzy set $F_{DT}(\theta)$. The rotation of the pier in the transverse plane about the longitudinal direction (R_L) of the column exhibited a maximum value for a bridge with a subtended angle of 45° degrees. As the subtended angle increased to 180° degrees, R_L decreased at a rate of approximately 23%, described by the fuzzy membership function $\mu(R_L)$. Lastly, the rotation of the pier about the transverse direction (R_T) increased with the variation in subtended angle θ across all bridge configurations. The fuzzy set $F(R_T)$ captures this trend, emphasizing the impact of θ on the rotational behavior of the piers.

As a conclusion, the pier displacement in the longitudinal direction variation with the subtended angle decreased by the increase of bridge span number and the bridge length. For the rotation of the pier along the transverse direction of the bridge variation with the subtended angle increased by the increase of bridge span number and bridge length. Lastly, the pier displacements and rotations of the piers arose critical values for the subtended angle less than 90° .

4. The analysis revealed that the internal forces in the bridge namely, axial force (N), shear force (V), torsional

moment (T), and bending moment (M) in the longitudinal direction (L) of the bridge increased as the subtended angle (θ) increased. Specifically, when θ exceeded 30° degrees, the shear force (V_L) and bending moment (M_L) in the longitudinal direction for 3-span bridges were found to be approximately three times higher than those of equivalent straight bridges ($\theta=0^\circ$). For 5-span bridge configurations, the maximum values of the internal forces, represented by the fuzzy set $FL(\theta)$, were obtained. In the transverse direction (T), the shear force (V_T) and bending moment (M_T) reached their peak values for a 5-span bridge at a subtended angle of 45° degrees. The fuzzy set $F_T(\theta)$ corresponding to this configuration indicated that the internal forces were approximately four times greater than those in the piers of the equivalent straight bridge. For 3-span and 4-span bridges, the maximum values of V_T and M_T occurred at a subtended angle of 60° degrees, as determined by the membership functions $\mu(V_T)$ and $\mu(M_T)$, where μ represents the degree of influence of θ on the respective internal forces. The results, formulated through fuzzy logic, suggest that the subtended angle θ significantly affects the internal forces, particularly in the transverse and longitudinal directions.

As a conclusion, the pier column internal forces (The axial force, shear force, torsional moment, and bending moment in the longitudinal direction of the bridge) variation with the subtended angle increased as the bridge length and the span number increased. It should be emphasized that, the AASHTO LRFD specification suggests that a bridge is considered regular if the subtended angle is smaller than 90° . However, according to the analysis results, the internal forces of the pier columns of the bridges could reach their maximum values at lower angles of curvature than 90° .

5. The analysis results demonstrated that the shear forces (V), bending moments (M), and torsional moments (T) in both longitudinal (L) and transverse (T) directions of the bridges increased as the subtended angle θ increased, except for the axial force, which remained relatively unaffected across all bridge configurations. The fuzzy set $F_{TL}(\theta)$ describes the relationship between the subtended angle and the torsional moment in the longitudinal direction, highlighting a particularly high increase ratio with variations in θ . For the 3-span bridge configuration, the torsional moment of the deck reached its maximum value at $\theta=180^\circ$, which was six times greater than that of the equivalent straight bridge ($\theta=0^\circ$), as represented by the fuzzy membership function $\mu(T_L)$. This was the highest value observed among all bridge configurations. For the 5-span bridge configuration, the torsional moment reached its peak at $\theta=45^\circ$ degrees, captured by the fuzzy set $F_{TL}(\theta)$, before decreasing up to $\theta=90^\circ$ degrees. Additionally, the shear force (V_T) and bending moment (M_T) in the transverse direction increased by nearly 63%, reaching

their maximum values for a bridge with a subtended angle of 180° degrees. The fuzzy membership function $\mu(V_T)$ and $\mu(M_T)$ characterize this trend. The bending moment (M_L) of the 3-span bridge increased by 221% up to a subtended angle of 180° degrees compared to the results of the equivalent straight bridge, as captured by the fuzzy set $F_{ML}(\theta)$. This fuzzy logic-based formulation elucidates the complex relationship between the subtended angle θ and the internal forces in curved bridges.

As a conclusion, the deck torsional moment variation with the subtended angle decreased as the bridge length and the span number increased where the torsional behavior of the horizontally curved bridges seriously effected the vulnerability and fragility.

It should be emphasized that, the AASHTO LRFD specification suggests that a bridge is considered regular if the subtended angle is smaller than 90° . However, according to the analysis results, the internal forces of the deck of the bridges could reach their maximum values at lower angles of curvature than 90° .

It was shown in the study that; bridge length and span number had a significant effect on the seismic response of the horizontally curved bridges compared to straight bridges. Besides, the subtended angle limitations that AASHTO LRFD specifications put forward regarding allowing the curved bridges to use an equivalent straight bridge should be reviewed again. It suggests that a bridge is considered regular if the subtended angle is smaller than 90° . However, according to the analysis results, the dynamic modal quantities, the displacement and rotations of the pier, the internal forces of the pier columns and the deck of the bridges could reach their maximum values at lower angles of curvature than 90° . Therefore, the limitations of the subtended angle should be reviewed and re-evaluated for several variable parameters by using linear and non-linear analysis methods.

The aspects that make this research valuable and different from other studies are, firstly, the parametric models had a wide range of different subtended angles of curvature, bridge span numbers, and bridge lengths. Besides, the analysis results evaluated for the wide scope of the determined bridge configurations to realize the curvature effect of the bridges is very important to design resilient bridges under seismic excitation. It is of great importance that critical bridges that provide transportation, communication, and logistics should remain standing without damage in major earthquakes. Therefore, the up-to-date studies that discuss the restrictions of the regulations to get better solutions are very important to advance in the analysis and design fields of civil engineering.

For future studies, to realize the effect of curvature angle, in addition to linear analysis, nonlinear time history and pushover analysis should be done under near-fault seismic excitations. Besides, for long multi-span horizontally curved bridges, different bearing solutions (elastomeric bearings, lead rubber bearings, and spherical bearings)

should be investigated to lighten the mass of the deck in order to reduce the seismic effects. Lastly, according to the results of these researches the limitations of the specifications about the horizontally curved bridges should be reevaluated to design resilient bridges.

REFERENCES

- [1] AASHTO. AASHTO LRFD Bridge Design Specifications-9th Edition. Washington, DC: American Association of State Highway and Transportation Officials; 2020.
- [2] AASHTO. AASHTO Guide Specifications for LRFD Seismic Bridge Design-2nd Edition. Washington, DC: American Association of State Highway and Transportation Officials; 2011.
- [3] Heins C, Bell L. Curved girder bridge analysis. *Comput Struct* 1972;2:785–797. [\[CrossRef\]](#)
- [4] Heins C, Oleinik J. Curved box beam bridge analysis. *Comput Struct* 1976;6:65–73. [\[CrossRef\]](#)
- [5] Li W, Tham L, Cheung Y. Curved box-girder bridges. *J Struct Eng* 1988;114:1324–1338. [\[CrossRef\]](#)
- [6] Senthilvasan J, Brameld G, Thambirantam D. Bridge-vehicle interaction in curved box-girder bridges. *Microcomput Civ Eng* 1997;12:171–181. [\[CrossRef\]](#)
- [7] Huang D. Dynamic analysis of steel curved box-girder bridges. *J Bridge Eng* 2001;6:506–511. [\[CrossRef\]](#)
- [8] Williams D, Godden G. Multidirectional seismic response of a curved highway bridge model. *Bull NZ Soc Earthq Eng* 1976;9:97–114. [\[CrossRef\]](#)
- [9] Kawashima K, Penzien J. Theoretical and experimental dynamic behaviour of a curved model bridge structure. *Earthq Eng Struct Dyn* 1979;7:12–145. [\[CrossRef\]](#)
- [10] Abdel-Salam N, Heins P. Seismic response of curved steel box-girder bridges. *J Struct Eng* 1988;114:2790–2800. [\[CrossRef\]](#)
- [11] Sennah K, Kennedy J. State-of-the-art in design of curved box-girder bridges. *J Bridge Eng* 2001;6:159–167. [\[CrossRef\]](#)
- [12] Mwafy A, Elnashai A. Assessment of seismic integrity of multi-span curved bridges in mid-America. Urbana, IL: University of Illinois, Civil and Environmental Engineering Dept & Mid-America Earthquake Center; 2007.
- [13] Galindo CM, Hayashikawa T, Belda JG. Damage evaluation of curved steel bridges upgraded with isolation bearings and unseating prevention cable restrainers. *World Acad Sci Eng Technol* 2009;35:11–27.
- [14] Seo J, Linzell DG. Nonlinear seismic response and parametric examination of horizontally curved steel bridges using 3-D computational models. *J Bridge Eng* 2011;18:220–231. [\[CrossRef\]](#)

- [15] Seo J, Linzell DG. Influential curved steel bridge fragility analysis parameters. In: Congress on Forensic Engineering, Vol 6. 2012. [\[CrossRef\]](#)
- [16] Araújo M, Marques M, Delgado R. Multidirectional pushover analysis for seismic assessment of irregular-in-plan bridges. *Eng Struct* 2014;79:375–389. [\[CrossRef\]](#)
- [17] Khan E. Direct displacement based seismic design of continuous curved bridges [dissertation]. Raleigh, NC: North Carolina State University; 2015.
- [18] Shirazi R. Seismic response and analytical fragility functions for curved concrete box-girder bridges [dissertation]. Reno, NV: University of Nevada; 2015.
- [19] Sextos AG, Olympia T. Influence of seismic wave angle of incidence over the response of long curved bridges considering soil-structure interaction. In: Dynamic Response of Infrastructure to Environmentally Induced Loads. Lecture Notes in Civil Engineering, vol 2. Springer International Publishing; 2017. p. 1–40. [\[CrossRef\]](#)
- [20] Pahlavan H, Zakeri B, Amiri GG, Shaianfar M. Probabilistic vulnerability assessment of horizontally curved multiframe reinforced concrete box-girder highway bridges. *J Perform Constr Facil* 2015;30. [\[CrossRef\]](#)
- [21] Abbasi M. Seismic vulnerability assessment of as-built and retrofitted multi-frame box-girder bridges [dissertation]. Reno, NV: University of Nevada; 2018.
- [22] Kaleybar RS, Tehrani P. Investigating seismic behavior of horizontally curved reinforced concrete bridges with different types of irregularity in comparison with equivalent straight bridges. *Structure* 2021;33:2570–2586. [\[CrossRef\]](#)
- [23] TBEC. Turkish Building Earthquake Code 2018. Ankara: Disaster and Emergency Management Authority (AFAD); 2018.
- [24] Computers & Structures Inc. SAP2000 Version 14.2.4-15.1.0, Analysis Reference. Berkeley, CA: Computers & Structures Inc; 2002.



Review Article

Recent and significant advances in crypto-biometrics: A detailed review

Priyanka BHATTACHARYYA¹, G.S.G.N. ANJANEYULU^{1,*}

¹Department of Mathematics, SAS, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, India

ARTICLE INFO

Article history

Received: 22 April 2024

Revised: 03 August 2024

Accepted: 11 September 2024

Keywords:

Biometric; Cancelable Biometric;
DNA; Fingerprint; Iris; Multi-
Modal Biometric

ABSTRACT

In the field of information security, Crypto-Biometrics is an advanced technology. Data protection and its secure transmission remain a topmost priority in the current digital era. In recent times, data breaches have become a major issue which affects the organizations across various industrial sectors. In this context, this review on crypto biometrics aims to provide an extensive overview of recent developments in the field of crypto biometrics especially in the area of fingerprints, iris and DNA technologies. The survey provides a comprehensive analysis of the latest development in crypto-biometrics that further focuses on integrating biometrics technology with various cryptographic methods to improve security and privacy. Compared to biometric template, Various algorithms such as advanced encryption standard (AES), triple data encryption standard (3DES), Twofish, Elliptic Curve Cryptography (ECC), Elgamal cryptosystem and Diffie-Hellman key exchange are commonly utilized for encryption, decryption and key generation with enhanced security and privacy. Additionally this review reports various methods and algorithms to improve biometric recognition under various constraints. The study emphasizes the effectiveness and summarises those algorithms with their limitations. Major developments in cancelable biometrics, multi model biometric systems and cypto-bio keys generation are also reviewed in the study. The Survey also addresses several challenges related to biometric template protection, revocability and privacy concern while proposes possible solutions that linked biometrics to cryptography. This comprehensive survey outlines the importance of crypto biometrics for securing sensitive information in the modern digital era. This review also identifies potential areas for further exploration by the researcher.

Cite this article as: Bhattacharyya P, Anjaneyulu GSGN. Recent and significant advances in crypto-biometrics: A detailed review. Sigma J Eng Nat Sci 2025;43(5):1807–1836.

INTRODUCTION

In our era, dominated by information and communication technologies, safeguarding data and ensuring secure transmission have become paramount. Extensive research

has been devoted to fortifying data during both storage and transmission phases. With the rapid evolution of these technologies, it's imperative to shield electronic data from interception and exploitation as it traverses communication channels. The potential for eavesdropping poses a

*Corresponding author.

*E-mail address: anjaneyulu.gsgn@vit.ac.in

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



significant threat, allowing unauthorized individuals to intercept or pilfer sensitive information from computer systems. To mitigate such risks, it's recommended to employ encryption techniques to safeguard data integrity and confidentiality.

Cryptography emerges as a pivotal discipline in meeting these security imperatives. Modern cryptography not only serves to encode messages but also plays a pivotal role in guaranteeing that transmitted information remains accessible exclusively to authorized recipients. Consequently, employing advanced cryptographic methods ensures that any attempt to exploit vulnerabilities in communication channels for unauthorized access to data is thwarted, thereby safeguarding the integrity and confidentiality of information.

Symmetric key cryptography and asymmetric key cryptography are two types in securing the data. In symmetric cryptography such as AES, and DES, the encryption method $C = Enc_k(P)$ and the decryption method $P = Dec_k(C)$ both employ the same key ($k_e = k_d = k$). Two separate keys ($k_e \neq k_d$) are employed in asymmetric cryptosystem (such as the RSA technique), the public key is utilized for encoding a message $C = Enc_k(P)$, while the private key is used to convert the cipher to plaintext $P = Dec_k(C)$.

Ensuring the goals of information security necessitates maintaining the confidentiality of keys, which ultimately relies on the ethical conduct of the individuals involved. The security of the cryptographic technique depends on the reliability of the keys used in the algorithm. If a key is difficult to figure out and impossible to crack [1] in a reasonable amount of time, then it is said to be strong. In this field, constructing hard mathematical problems and generating keys on them are really challenging issues forever. It is highly difficult for a user to remember a key that is particularly large (128, 192, or 256 bits, for instance in the AES method). Additionally, a password-based authentication method is utilized to secure the token or smart-card and limit access to the successful completion of the operation. Conventionally, cryptographic techniques are relied on knowledge-based authentication, such as passwords, or possession-based authentication, such as tokens. However, a password can be misplaced or discovered by brute-force attacks. The non-repudiation property of standard cryptography cannot be ensured by possession or knowledge based authentication schemes.

There is another kind of user authentication system, known as biometrics, which is described as the instinctive identification of individuals based on physiological or behavioral traits. The phrase 'biometrics', which describes the statistical examination of biological data and occurrences, is acquired from the term 'biometry'. An individual's distinctive physical characteristics or other characteristics can be recognized through the procedure of biometrics and computerized devices that record all the characteristics, which is used to verify the individual. Unlike knowledge- or token-based approaches, like passwords or identity cards,

biometric attributes are (specific to each particular) more trustworthy at confirming identity. The physical characteristic of a person is reflected in physiological biometric identifiers, which involved fingerprint, iris, DNA, facial, odor/scent, palm print recognition etc. Voice recognition, keystroke dynamics, signature analysis etc. are examples of behavioral traits that are connected to a person's pattern of activity. Some of the physiological and behavioral characteristics of a person are shown in Figure 1.

A general biometric system is depicted in block diagram in the following Figure 2. The methods of 'Biometric technology' are created with the intention of using biometric information obtained through biometric attributes. The biometric system is automatically processed through a mechanism with a set of steps, which includes, (i) Biometric data is to be fabricated in terms of iris, facial image, palm vein, fingerprints etc., (ii) Data extraction from the stored samples, (iii) Comparison of the real-time data with the source database, (iv) If verification reaches the predefined threshold, then authentication will be successful.

Biometric technology is composed of both hardware and software. Hardware for assembling, studying, and matching biometric information is known as a biometric recognition device. Biometric information is a specimen of a person that is specific to that individual. A biometric engine is a part of the software integrated into biometric technology that processes the biometric information that has been collected. The software often takes place in conjunction with the hardware to run the capturing procedure of the biometric data, extract data, and perform indistinguishability, regarding data matching. An image capture module gathers an individual's fresh biometric information with the use of a sensor during the process of capturing biometric data.

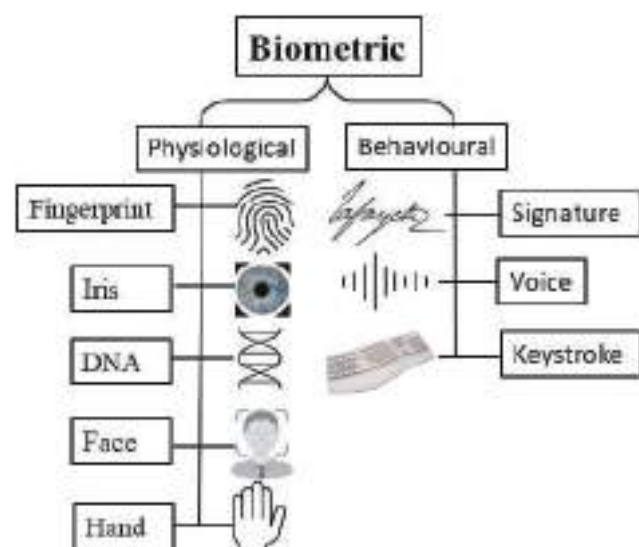


Figure 1. General classification of biometric schemes.

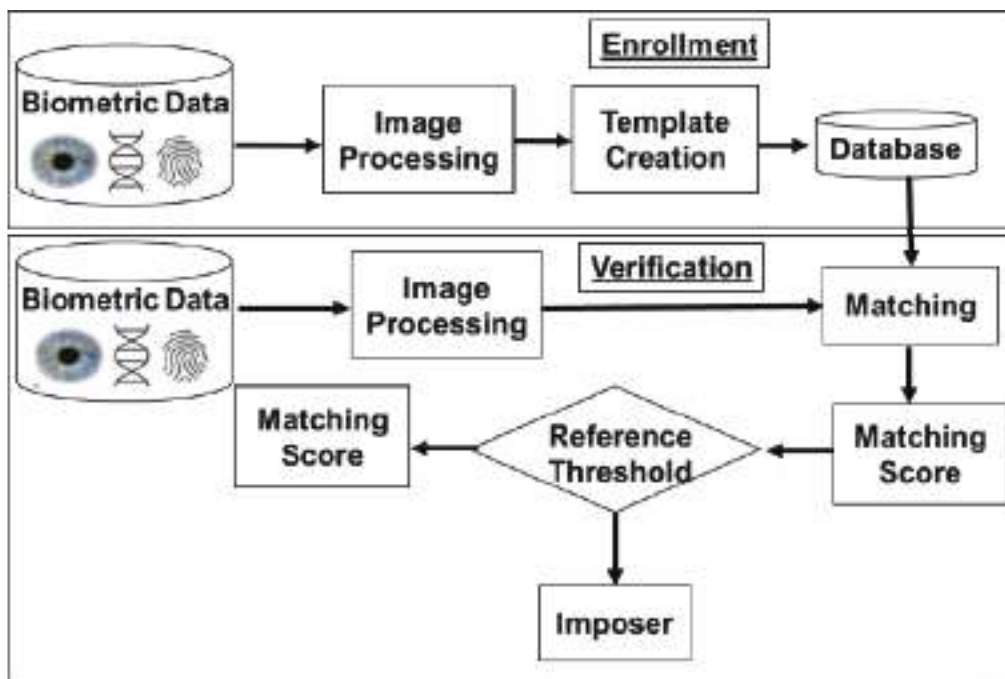


Figure 2. Systematic representation of generic biometric system.

Biometric identification enables a persistent connection between the user's identification and the verifier. In the event of a breach, the biometric information cannot be restored because it is inextricably linked to the user. Furthermore, the usage of biometrics poses a threat to privacy. The personal information of a user may be stolen by connecting the biometric information recorded in several databases.

There are approaches to address the issues that arise from the use of both cryptography and biometrics by integrating both. This composite system, also known as crypto-biometric systems, combines the capacity of biometric technology for individual authentication to boost the safety of cryptographic techniques. Together, these two methods can improve and strengthen the configuration of the system since they have supplementary qualities.

The necessary confidence for encryption can be provided by utilizing the high correlation between the biometric traits and user's identification. Additionally, biometric information can be protected using cryptographic methods while maintaining security. The most straightforward approach is to verify each user using their biometrics and once the verification is successful, free the confidential key relating to the user to the system. Biometric technology can verify only one bit at a time. To achieve a key, a hacker could replace the bit with a relevant outcome (like a Trojan house). The limitation of this method lies in poor setup between cryptographic techniques and biometrics. However, there are certain issues with the crypto-biometric scheme.

A biometric system must include the protection of biometric template to guarantee the confidentiality and the

secrecy of biometric information [2]. Biometric systems have their own issues, including the inability to be revoked, a lack of diversity, and potential privacy breaches. The user authentication system is revocable in nature, which implies that if the validator (e.g., Password) is hacked, a new authenticator can be used instead. In this system, it is not possible for the previous validator to authenticate. Here the user is totally linked to their biometrics characteristics. This makes the biometrics data unchangeable because the biometric qualities are innate and immutable [3]. The templates created in different biometrics systems are remarkably similar under consideration of the identical biometric traits. If a template in a system got compromised, it poses a significant security risk in the system. This is because, biometric traits are permanently associated with specific systems and cannot be replaced with new data. This indicates an inadequacy of variation in templates. As a result, it is possible to log into other systems using a hacked template from one system. The interconnection of templates could jeopardize user confidentiality.

To address the drawback such as irrevocability, non-diversity, and potential invasion of privacy, a new field of study as cancelable biometrics has developed [4–6]. The biometric data in these frameworks is altered by using a one-way transformation. In the altered domain, a matching process takes place between two sample readings from templates. A determination of authentication success or failure results in such a differentiation. Regrettably, because it functions like the traditional biometric authentication system, using it for cryptographic purposes leads to a poor connection between cryptography and biometrics. Furthermore,

biometric information must be transferred across insecure networking channels, to use remotely. Consequently, it is necessary to create a revoking, irreversible encryption key that can be generated from biometric of two independent people while maintaining their confidentiality and secrecy.

There are various methods [7,8] that can establish a solid connection among cryptography and biometrics. The aforementioned system endeavors to acquire biometric data containing encryption keys. These keys, which are achieved using biometric information, are alluded as crypto-bio keys. Key formation [9] and key re-generation are the two main strategies. The key formation system retrieves a reliable bit-sequence known as a crypto-bio key. Several authors on [10–12] provide several instances of key formation systems. Whereas key re-generation systems, which associate a randomly generated key to the source biometric information before regenerating it using a distinct biometric specimen [13–15]. A solitary biometric identifier-based biometric system cannot meet the severe accuracy requirements placed by high-level safety appliances and large-scale recognition systems. By compiling data from multiple biometric traits, multiple biometric systems get beyond this restriction. Multi-biometric systems are more accurate at identification and have a wider community coverage than systems that just use one biometric feature. To make a trustworthy identification determination, the data from diverse resources might be combined at different levels. While discussing several biometric methods, four considerations need to be taken into account which includes the source of the data, the method of operation, the degree of fusion, and the fusion strategy. The different subcategories of multi-biometric systems are multi sample, multi sensor, multi instance, multi algorithm, and multi modal. The verification for the initial four situations is acquired from single biometric attributes, whereas the confirmation for the fifth situation comes from numerous biometric qualities [16]. The combination of multiple biometrics shows that the system's confidentiality and dependability have improved. Different degrees of combination in multiple biometric systems is provided as follows:

- **Sensor level fusion:** Prior to feature extraction, the verification provided by the raw data from many resources of information is combined at this level.
- **Feature level fusion:** Combining data from two independent sets of the same person's biometric attributes.
- **Score level fusion:** Integrating the result of the match scores issued by several biometric matchers to reach a final authentication,
- **Rank level fusion:** The process of combining all the ranks produced by various biometric sub- system to create a consent rank for every identity.
- **Decision level fusion:** When just the person biometric matcher's decision outputs are available, combination is done at the decision level.

These levels in actual execution will be implemented in biometric technology as given in Figure 3.

It is important to choose biometric traits carefully such that at least one of them is exceedingly challenging to take a picture. A secure biometric cryptosystem will be provided by two or more biometric traits combined with an improved cryptographic key generation process.

In the present scenario data secureness has become a big challenging issue in view of network security. Biometric cryptography plays an important role in overcoming the facing challenges. In the year 2019, Kaspersky researchers observed that one-third (37%) of the computers in the company's telemetry that gather, process, and store biometric information were the targets of malware attacks in the third quarter [17]. Recently in the first quarter of 2022, India faced around 18 million cyber- attacks, which are mostly happened during digital transactions as per the data reported by Google's Royal Hansen [18]. Data analysis of cipher shows that approximately 28000 cyber-attacks have come up during the COVID pandemic period [19].

The primary research gap in biometric cryptography is the lack of a thorough analysis that integrates advancements across fingerprint-based, iris-based, and DNA-based cryptography. Despite numerous studies reporting on each approach separately, there is a significant need for a comprehensive comparison and in-depth understanding of

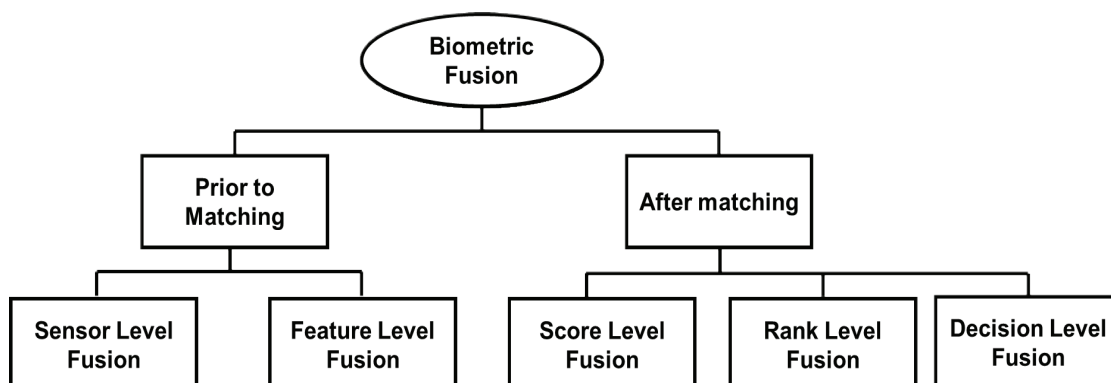


Figure 3. Various levels of fusion.

their performance, limitations, and integration potential. Previous studies have mainly explored biometric approaches in isolation. But they have failed to analyze the benefits and challenges of cross-modal systems and the complexity of scalability, efficiency, and privacy protection. In addition, various issues such as template protection standardization, error correction in fingerprint systems, and the real-time applicability of DNA cryptography are remained unsolved which hinder the progress of this field.

In this context, we present a comprehensive survey on biometric cryptography, which emphasizes the security, privacy, and technical aspects of biometric technologies. This review paper critically evaluates the recent literature by providing a clear overview of the current state of research and identifies its future directions. It also focuses to guide further advancements in biometric cryptography by addressing those overlooked topics. The present study focuses and analyzes the current research and developmental work on crypto- biometrics. In this review, the main attention has been given to fingerprints, iris and DNA-based cryptography.

The initial section of this review provides a comprehensive overview of biometric cryptography, where brief summary of the needs of this research is discussed. In the next section, the details survey focusing on latest advancements in fingerprint, iris, and DNA-based systems have been highlighted. This section qualitatively describes the critical role of integrating biometric data with cryptographic models to improve the security and efficiency of digital communication systems. Then a thorough analysis of various cryptographic models is provided where the performance and effectiveness of various model is compared with existing conventional methods. Finally the review concluded with a brief summary of the key findings and insights from the previous literatures. The conclusion thoroughly discusses the potential advancements in biometric cryptography, addressing current limitations, and an indicative future opportunity for researchers in this field.

DETAILED CURRENT SURVEY ON BIOMETRICS

Review on Fingerprint

Each person has a unique fingerprint, supported by a personal pattern and unique print set. To distinguish one fingerprint from another, it shares three key characteristics that are arch, loop, and whorl. These three characteristics are classified based on the direction of the ridges. Among three characteristics; Arch: has ridges that rise from the center to all directions; Loop: has ridges that have a circular pattern, which enter from one side, flow around the center, and exit on the other side. Whorl: have ridges that form a spiral pattern. The following Figure 4 describes the characteristics of fingerprints. These are the following reviews on fingerprints based on Cryptography.

This section summarizes a series of recently published articles in the area of fingerprints. The integration of fingerprint features into a biometric key-dependent cryptosystem represents an innovative strategy to enhance overall system security. By incorporating fingerprint characteristics as a key component, the system is strengthened against unauthorized access or breaches. The fingerprint-generated key coupled with the system's cryptographic key is shown in Figure 5. In this research, the authors suggested an effective method for extracting minutiae from biometric images to generate irreversible cryptographic keys using fingerprint biometrics. For segmentation, the picture is partitioned into blocks. The gray scale difference is determined for every block. It is relegated to the background if the value is less than the global threshold, otherwise it is allocated to the foreground. The variance threshold distinguishes the foreground and background regions. The ridge structures are present in the segmented foreground sections, while the remaining regions have not been touched. In terms of distinguishing the foreground and background regions, a variance threshold of roughly 100 has been shown to offer the best results. The Normalized image is represented by equation (1),



Figure 4. Characteristics of fingerprints.

$$N(i, j) = M_0 + \sqrt{V_0 \frac{(I(i, j) - M)^2}{V}} \quad \text{if } I(i, j) > M \quad (1)$$

$$= M_0 - \sqrt{V_0 \frac{(I(i, j) - M)^2}{V}} \quad \text{otherwise}$$

For a pixel $I(i, j)$, the estimated mean and variance are denoted by M and V , respectively. M_0 and V_0 represent the intended mean and variance values. As a normalization method, histogram equalization is a mechanism for enhancing the contrast of images by altering their intensity values. The original image's histogram shows that all the intensity values are on the right side of the 0–255 scale and there are no pixels on the left. The normalized image's histogram reveals that the range of intensity values has been modified, resulting in a more equal distribution of dark and light pixels. The suggested approach generates a 128-bit cryptographic key from the system and a 128-bit biometric key from the fingerprint minutiae. A 256-bit secret key is formed by combining those two keys [20].

In conventional cryptosystems, managing and ensuring the confidentiality of keys poses challenges. The Key Distribution Center (KDC) is primarily used for key distribution, which is further secured through enhanced privacy through the integration of biometric data. The KDC facilitates secure communication in symmetric cryptography. In symmetric cryptography, which often uses a special key to share the session key [21].

The fingerprint biometrics of the communicating party serves as a solution, involving the distribution of the session key alongside the user's fingerprint-based key. There is no way to compromise the unique key because it is produced using the user's fingerprint. This eliminates the need for the communication party to memorize the unique key since it is linked to the user's biometric information. The knowledge-based authentication of KDC is converted to biometric authentication using this method. The user's identity is

kept secret even if the KDC is compromised, which safeguards the privacy of fingerprint identity. The KDC is registered with user fingerprint information. Minutiae points are retrieved from the fingerprint image after pre-processing. The fingerprint data is converted into a cancelable template using a one-way Cartesian transformation function [4]. During this transformation process, minutiae points are separated into cells, and each cell can include one or more minutiae points. Minutiae points are moved in accordance with the location of their new cell after the cells are randomly shuffled (using a key as the transformation parameter). The cancelable template created using a user's fingerprint data is registered with the KDC. A Unique Id (Uid) is given to the user, and the cancelable template that corresponds to the Uid is kept in the KDC database. A user's cancellable fingerprint data may be enrolled in one of two methods. The cancelable template can be sent to the KDC over an encrypted channel. Alternately, a user can physically enroll his fingerprint data in a safe position using an offline enrolment method.

A modified minutiae point with (x, y) coordinates values is a component of a cancelable template. First, a minutiae point (let's say, M_i') is chosen, and then the distances (D_{ij} , where $i \neq j$ and $j = i + 1, i = 1$ to n) of this minutiae point from every other minutiae point (i.e., M_I' , where $I = i + 1$ to n) are determined. Each distance can be recorded as a vector which is denoted as $D = (D_{1,j+1}, D_{1,j+2}, \dots, D_{n-1, n})$, where $j = 1$ to n , n = the number of total minutiae in cancelable template. The average distance D_{avg} is calculated using the equation (2),

$$D_{avg} = \frac{2}{n * (n - 1)} \sum_{i=1}^{n-1} D_{i,j} \quad (2)$$

The average distance D_{avg} in each distance D_{ij} is compared and matched, which results in a key in the binary form. If the distance exceeds the average, the bit is 1, and if

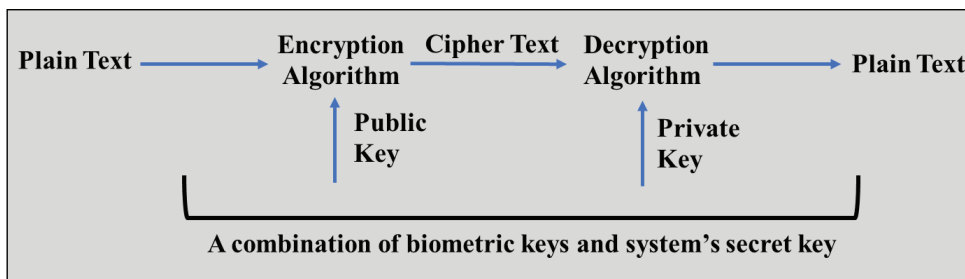


Figure 5. Generation of secret key combining biometric keys and system's secret key.

Table 1. Segment specific time

Execution	Extracting features	Creating cancelable template	Producing key	Overall time to produce a key
Time (sec)	0.05	0.006	0.003	0.059

not, it is 0. This results in the production of a binary vector K_p from all components of the distance vector D_{ij} . To create an everlasting key of N bits, the initial N bits of K_p are required. As a result, KDC creates the fixed keys K_{ps} , K_{pr} from the cancelable templates of the user S and R , respectively, and uses these keys to produce an encrypted communication link with the communication party. Then, for the pair of communicating parties mentioned above, the KDC chooses a confidential key at random. Messages between users S and R will be safely transmitted using this secret key.

For experiment purposes, one fingerprint is used as the input and a genuine key which is everlasting is created using that fingerprint image. The key to the impostor is created using the left fingerprints in the uniform database of fingerprints. In this suggested method, authors have contrasted the genuine key to impostor keys made from different fingerprints to assess the resilience of the genuine key. It is determined that the counterfeiters are unable to produce a key in a manner that is like the real key. The length of the impostor's key is half as long as a genuine key, on average, according to Hamming distance. As a result, to successfully crack the real permanent key, the impersonator must estimate at least half of the bits or 128 bits of the 256 bits. The overall time to construct a key from a fingerprint picture is calculated to assess the overhead time of the proposed method as reported in [21]. The implementation period considers the time spent extracting features, creating cancelable templates, and producing keys. The following Table 1 includes the segment-specific time.

Utilizing a key-pair generated from fingerprint impressions provides a more secure approach to encryption and decryption processes. The ability of the key generation technique to generate the key from the retrieved minutiae points of the fingerprint biometrics is an efficient technique for ensuring the privacy and security of the key creation process. A matrix is used for representing the scanned picture of the fingerprint in the minutiae point extraction procedure. A set of number of furrows and ridges in a small section is represented by the matrix consisting of data regarding the number of ridges and the furrows. In the method proposed by Ankit et al., minutiae points are extracted in a three-step process, and the scanned image is then converted into machine-readable data to analyze the pixel values. In the ridge thinning technique, repetitive erosion is used to identify the ridges of pixel of width one and then the suggested system, which is based on the minutiae point method [22], creates both public and private keys from an individual's fingerprint minutiae points. The time taken for every process is calculated in this work. The suggested method offers the advantage of not requiring the asymmetric key to be kept in a secure location, which drastically reduces security risks. The economic growth of this kind of biometric cryptosystem would be the main emphasis for future development, which would guarantee that any communication channel or network is secure [23].

When two parties communicate over an unsecured channel without prior acquaintance, the main limitation of the Elgamal cryptosystem is the necessity for a pair of encoded messages. The recipient's public key and the sender's private key are used in this computation to encode the message. For highly secret data exchange across the system, instead of using a random number as a private key in the calculation, employing a specific fingerprint ensures better security. The system's data exchange is made more secure by the organizations that uses a specific finger impression as a framework for confirmation [24]. The rough edges of a finger leave a distinctive mark. They are always unique since they are created in person during their final development prior to birth. In the field of criminological study, fingerprints play a significant role. Techniques for coordinating distinctive marks are categorized into three kinds. It consists of edge include based, details based, and connection-based coordinating. The idea put out by Sergey Tulyakov [25] is to protect biometric systems by using symmetric hash capabilities.

Since the user's fingerprint is saved in a database and only pulled out upon authentication, no one may pretend to be the sender using this Elgamal encryption system. By adding a random exponent k , Elgamal makes the Diffie-Hellman key exchange procedure simpler. This exponent takes the place of the receiving party's private exponent. This makes the technique easier to use and allows for one-way encryption without the involvement of the second party. The ability of the algorithm to encrypt electronic messages that are sent via public store and forward services represents a significant advancement of the algorithm in this case. A key component of the encryption process is the randomness of the random exponent k selection, as the ability to guess the random exponent provides the attacker with a large portion of the information needed to decrypt the message. Following the decryption procedure, sender A will select a number arbitrarily and generate a fresh public key, each time A wants to convey a message to the receiver B . In the suggested approach, the sender's (A) produced random number (k) and the receiver's (B) private key serves as each party's unique fingerprint. The decimal values of each fingerprint are obtained. The fingerprint's decimal value was collected through a biometric authentication appliance. With this proposed approach, the sender and the recipient can encode and decode the data while ensuring non-repudiation and verification. It is more safe to communicate information through a system with high confidentiality when a particular finger impression is used as the private key in the computation rather than a random integer.

Integrating fingerprint recognition with cryptographic methods, particularly through the Secure Hash Algorithm SHA-512, combines advanced image processing with SHA-512 to enhance the efficiency and security of fingerprint-based authentication. The study creates a robust identity verification system by extracting distinctive identity

verification system through SHA-512. The hashin process gets optimized through the use of SHA-512 loop unrolling resulting in improved performances along with maintained security feature. This approach reduces the resource allocation and thereby makes the standalone device economical and power efficient. The Look-Up Table (LUT) utilization is reported to be reduced from 63% to 33.07% along with Flip-Flop (FF) utilization decreased from 36% to 13% and Clock Buffer Resources (BUFG) utilization rate stands notably low at 3.13%. This results showcased the improved efficiency in budget allocation, lesser time consumption in hashing. This feature makes the process suitable for time sensitive applications. Future studies can explore the integration of Field Programmable Gate Arrays (FPGA) for further optimization of resource expenditure towards the improvement of system efficiency [26].

Generating symmetric cryptographic key by using fingerprint biometrics offer a unique system in securing communication technology. The system uses a combination of Diffie-Hellman key exchange and fingerprint authentication to create a secure communication channel to encrypt and decrypt between two party's sender and receiver. This system is more secure, convenient and scalable against several attacks, man-in-the-middle attacks, and impersonation attacks. The workflow of the system is as follows: Initially, by using Diffie-Hellman key exchange, the sender and receiver exchange their public keys. Then a fingerprint template is generated by the corresponding sender from their fingerprint images, followed by an encryption of the template with the use of public key of the recipient. After receiving the encrypted template from the sender, the receiver uses their private key to decrypt it. Then both the decrypted template and original fingerprint is compared. If the fingerprint matches, the receiver gets to know that the message is sent by the intended sender.

The system is then examined by using four fingerprint image datasets refer to FVC2002 databases (FVC2002DB1, FVC2002DB2, FVC2002DB3, and FVC2002DB4). Each of this dataset contains 100 subjects, consisting 8 impressions per subject. Performance was assessed based on the FVC protocol, which dictates that all unique impression pairs from the same subjects are used to generate genuine cryptographic keys. The results demonstrated that the system achieved a high level of accuracy [27].

Biometric security is enhanced by integrating Elliptic Curve Cryptography (ECC) with advanced session keys, improving protection against sophisticated cyber threats. By combining ECC with dynamically generated session keys, the approach adds an additional layer of protection, which significantly enhances the security of biometric systems against various cyber threats. The study demonstrates that this innovative method remarkably improves resistance to common cryptographic attacks, such as those targeting static information, by incorporating dynamic session keys. Compared to the traditional static key mechanism, this dynamic session key makes the system more secure and

prevent the attackers from exploiting the system. The simulation results in the proposed system substantially restricts the data breaches and subsequent security threats, resulting in increasing security parameters. Moreover, the approach maintains high system performance while strengthening security, as indicated by the successful simulations that highlight its effectiveness without compromising operational efficiency [28].

Elliptic Curve Cryptography (ECC) parameters for fingerprints are generated through several key processes, including fingerprint preprocessing, ECC parameter extraction, and public key computation [29]. The literature outlines various thresholding methods, such as Global Thresholding and Thresholding with Iterative Application of the Laplacian Operator [30]. Among them, particularly adaptive thresholding is suggested due to its speed and reliability in fingerprint analysis [31,32]. In the complex patterns of fingerprints, where it is important to preserve intricate details, this procedure is effectively adapted.

This work notably highlights the recent advancements in biometric security carried out with the integration of Elliptic Curve Cryptography (ECC) and effectively protects the sensitive dataset as reported in [32]. Through incorporating upgraded session key introduces dynamic features to the biometric security and thereby overcome various limitations as discussed in [33]. This study shows that apart from the parameters 'A', 'B', 'G' and the public key associated to each fingerprint are distinct and stable. This feature reveals the novelty of the ECC parameters derived from a single fingerprint. Fingerprint Quality Index analysis shows that the recovered image after the attack is qualitatively protected. This method potentially tackle the practical challenges indicating its effectiveness against potential threats. The integration of ECC in enhanced session key is found to be advantageous in high security environment including financial sector, defence agencies due to its improved identifications accuracy and security.

An encrypted security model integrating OTP, cryptography, and fingerprint sensor technology was devised to enhance the security of biometric systems. Within this network an OTP generator that can be changed intermittently with variations of inputs. The architecture of this methodology is furnished in Figure 6. The whole system can be implemented through an application in android mobile. The potentiality of the system extended to a pattern identification scheme, where a set of template-based features can be obtained for the enrolled personnel [34].

The system will be accessible after first logging in by the administrator and then by the user. Here admin has the authority to carry out the registration process of a user by taking the fingerprint as well as identification documents of the user. Registered customers can retrieve their own data through biometric as well as OTP verification. The system algorithm consists of (i) Advanced Encryption Standard (AES) Algorithm and (ii) Time-Based One-Time Password (TOTP). The AES algorithm constitutes by three

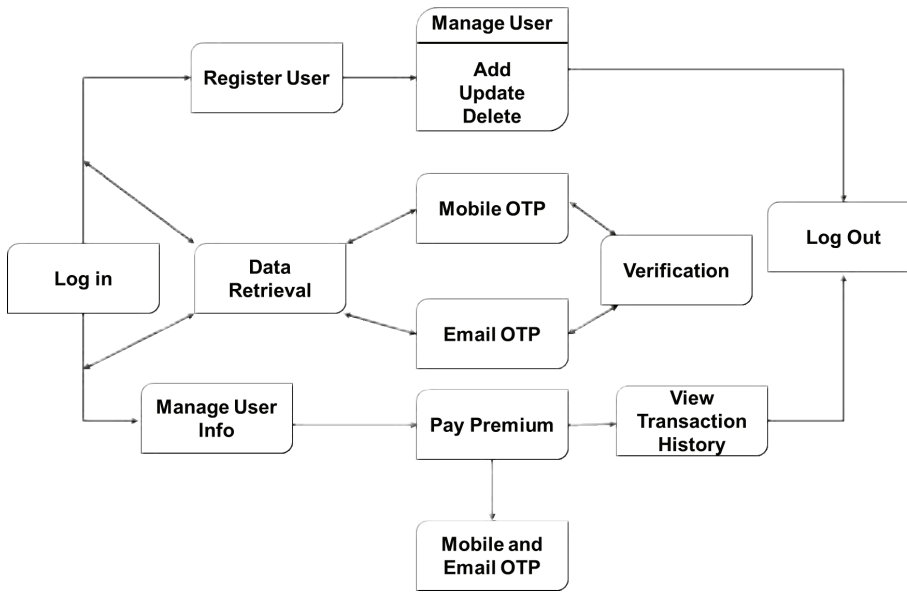


Figure 6. Architecture of the proposed methodology [From Pawar et al. [34], with Permission from Solaris Publication].

block ciphers. By using 128-bit, 192-bit and 256-bit cryptographic keys, all ciphers perform to encrypt and decrypt the block data. Whereas TOTP algorithm executes the OTP in the android-based application in a definite time interval.

As reported in [35], it is not practically possible to generate the same key from an interferer's fingerprint. Neither the user's fingerprint nor the encryption key needs to be stored. This improves the scheme's security. The proposed solution guarantees the confidentiality of a user's biometric data while also addressing the issue of encrypted key storage. Inside the same region of all the fingerprint photos of the same person, a collection of minutiae points is carefully chosen. This assures that all the images of fingerprint of the same person use the same collection of minutiae points. The suggested method only examines ridge ending and ridge bifurcation points. The local pattern of pixels where the ridge terminates or divides is distinguished by analyzing the depleted binary image. Feature points were extracted from the fingerprint image using NIST's NBIS software tool FpMV (Fingerprint Minutiae Viewer) [36]. A fingerprint image is taken by this tool and it outputs minutiae points with x and y coordinates, angles, qualities, and types of minutiae. The keys are generated using x and y coordinate values in the suggested method. Gray coding is an essential component of the presented method since it reduces inconsistencies between keys generated from numerous instances of the same fingerprint. The gray code representation lowers the mistake caused by minor differences in the distances between two minutiae spots in two distinct fingerprints. The practicability of the presented scheme is investigated by calculating the specified difference between the key strings, as a key created using an impostor's fingerprint differs significantly compared

to one created from another occurrence of the same person's fingerprint. The Hamming distance is employed to quantify the difference between two key strings extracted from identical fingerprint images. It calculates the number of differing bits between two binary strings. In the case of legitimate fingerprint, the highest Hamming distance is 20, whereas the average Hamming distance is 12. The impostor fingerprint has maximum and average Hamming distances of 81 and 65 respectively. In this suggested technique, error correction codes are used to fix tiny variations in keys of the same person's fingerprint. The stated scheme becomes much more intriguing as a result of this.

A novel Biometric Cryptosystem is introduced based on fingerprint identification, utilizing an Eigen space algorithm instead of traditional geometric methods. It organizes Eigen space and computes Eigen values and vectors from fingerprint images taken at various angles. In this method (Figure 7), the authors first analyzed fingerprint images from different positions and computed an average image for the training set. Each image $I(x, y)$ in an $N * N$ array is transformed into a $N^2 * 1$ vector. Let $I = [p_1, p_2, p_3, p_4, \dots, p_N]$ represent the pixel values. After normalizing each image, the normalized image is calculated using $\Phi_p = \frac{1}{N} \sum_{n=1}^N \Phi_n$, resulting in a normalized image $I' = [p_1', p_2', p_3', p_4', \dots, p_N']$. After that average image is calculated by, $A = \frac{1}{p} \sum_{i=1}^p I_i'$. Finally, projecting all images onto this eigen space, h_i is calculated by equation (3),

$$[e_1, e_2, e_3, e_4, \dots, e_p]^T \frac{1}{p} \sum_{i=1}^p (I_i' - A)^2 \quad (3)$$

where, $e_1, e_2, e_3, e_4, \dots, e_p$ represent the eigen space vectors, each of which is N-dimensional. In the next step, an unknown input image is projected onto the eigen space

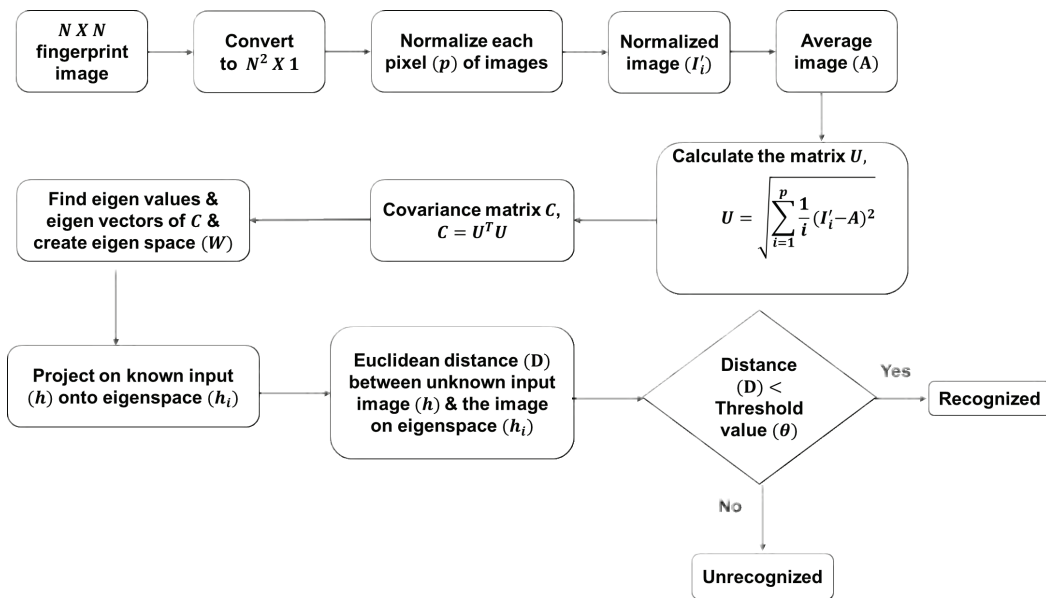


Figure 7. Schematic block diagram of fingerprint recognition model as demonstrated in [37] [Drawn by the author].

which results in a p -dimensional point labelled as h . The Euclidean distance between point h and eigen space h_i is then calculated. If this distance falls below a predetermined threshold, the image is identified; if not, it is considered as unknown.

The effectiveness of this methods across diverse subjects is validated by experimental results. The practical recognition systems customize the precision and efficiency of the system by considering limited sample sizes. This further forms an universal Eigen space through averaging the image and standardizing the pixel value. The time complexity of this method is estimated as $O(p^{3/2})$, demonstrate the simplicity, speed and reliability of the system compared to other methods. This research lacks efforts to improve success rates, especially for larger databases, and to assess real-time applications across different programming languages and hardware platforms [37].

A study aims to demonstrate a prototype of a password-less authentication system integrating cryptographic methods, steganography, and physiological and behavioral biometrics. It shows that the password-less authentication is scalable and can be achievable while balancing both security and usability. This model integrates threat modeling techniques to identify security requirements and appropriately select the proper security protocols. The comparative analysis also shows that the system works as effectively as other password based systems with reasonable authentication time-frames without any additional hardware costs. The system effectively eliminates the risk of data breaches by avoiding the password storage and their hashes. This techniques presents the password-less system as a potential alternative which addresses various user related challenges

[38]. Table 2 summarizes some important results from the recently published articles based on fingerprint.

Review on Iris

Unique Iris pattern recognition can be used in cryptography to create secure passwords and encryption keys and a digital representation of the eye that can be used to authenticate users and encrypt data. The four main regions of the iris, shown in Figure 8, that are used for iris recognition are,

Internal region: Pupil, is the most stable and least susceptible to changes in lighting.

External region: Sclera (the white of the eye), contains many unique features to identify individuals.

Lower region: The region below the pupil which is less susceptible to occlusion by eyelids and eyelashes.

Upper region: The region above the pupil which contains many unique features to identify individuals.

This section summarizes the recently reported articles in the area of Iris.

In 2008, a ground breaking biometric cryptosystem using iris recognition was introduced. It involves during

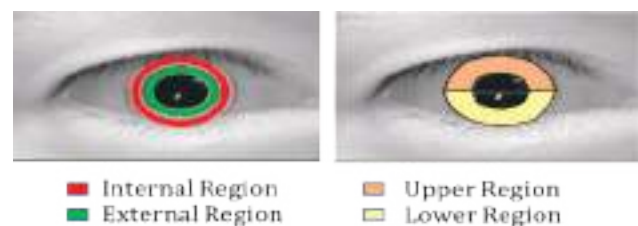


Figure 8. Different regions used for Iris Recognition [39] [Drawn by the author].

Table 2. Significant contributions based on fingerprint

Datasets Used During Performance Analysis	Method	Processor	Advantage	Ref.
-	Generating secret key by combining the fingerprint characteristics as a key and Cryptographic key	-	Fingerprint for enhance security of data	[20]
FVC2004	MINDTCT tools of NBIS software The traditional key distribution center (KDC) is strengthened by the integration of user's fingerprint data, this allows parties who want to connect with one another to share a secret key using fingerprint data.	-	Fingerprint for key distribution center	[21]
http://www.academia.edu/4814389/D01211622 . http://www.yaldex.com/tcpip/0672325659_ch20lev1sec1.html .	An encryption and decryption technique that leverages a key-pair derived from fingerprint impressions.	-	Fingerprint for Network Security	[23]
-	To enhance network security and support this analysis, the ElGamal algorithm has been implemented.	-	Fingerprint for Enhanced Security Mechanism	[24]
FVC2002 databases (i.e., FVC2002DB1, FVC2002DB2, FVC2002DB3 and FVC2002DB4)	A combination of Diffie Hellman key exchange and fingerprint authentication to create a secure communication channel to encrypt and decrypt between two party's sender and receiver.	Commercial software VeriFinger SDK	Fingerprint for Secure communication	[27]
INSURANCE COMPANY's data	Retrieval of data and cryptography AES algorithms and non-repudiation and verification while allowing for encoding and decoding between the sender and the receiver.	TOTP (Time-Based One-Time Password): TOTP in RFC 6238	Fingerprint for Insurance Data Retrieval with Additional OTP Generation and Verification.	[34]
Fingerprint Verification Competition (FVC) 2002 Database DB1	To derive a cryptographic key directly from the unique features obtained from a user's fingerprint.	NIST'sNBIS software tool FpMV(Fingerprint MinutiaeViewer)	Fingerprint for Cryptographic Key Generation.	[35]
-	Eigen space approaches is used. The identity of the individual is known if there is a minimal distance between Eigen space and the unknown picture.	R-software	Fingerprint for user recognition	[37]
-	Combination of cryptographic methods, steganography, fingerprint for password-less authentication. Threat modelling approaches to identify the system's security requirements and choose the best cryptographic protocols.	JAVA	Fingerprint for password less authentication	[38]

encryption process by utilizing a 256-dimensional textural feature vector extracted from the iris image through 2-D Gabor Filters. Reed-Solomon Code generates an Error Correction Code (ECC) simultaneously. Then the feature vector is transformed into an encryption key via a Hash function. This cipher key is employed by encryption algorithms to secure sensitive data. ECC rectified the feature vector obtained from the input iris in the decryption process. The Hash function converts it to the cipher key.

Finally, the data is decrypted using standard decryption techniques with the key [40].

Another study proposes that a unique iris cryptosystem is developed using a modified fuzzy vault algorithm and iris texture features, effective in fingerprint cryptography [41,42]. CRC codes are generated and mixed with the original message, then encoded using Reed-Solomon to produce RS codes of length 256. Grids B and C, each 256 X 3, are created as vaults [43]. RS codes are placed in Grid C

with remaining rows filled with random integers for data obscure. Grid B is filled to minimize dissimilarity between vectors from identical iris images. During decryption, the modified fuzzy vault unlocking algorithm retrieves the feature vector from the decryption iris. RS decoding recovers the message and CRC codes, which are then used to check decryption success. The variations between most of their corresponding elements of two vectors taken from the identical images of the same iris are smaller than the vault tolerance. This difference can be eliminated by determining the order of Grid B by choosing the elements in every row that is closest to the corresponding elements of decryption vector. FAR, FRR, and MC code lengths are 0, 4.6296%, and 128 respectively.

Error-correcting code and Hashing algorithms have been proposed by several authors as the foundations for the iris cryptography system [12,44]. This method assumed that the maximum error rate of the iris templates utilized for encrypting and decrypting is not exceeded 10%. The biometric data is exposed through the secured error-correcting code. Hao [14] rectified the disparities between the encoding and decoding of iris codes using Hadamard and Reed-Solomon code. The weakness in the system lies in its low level of security (Breaking ratio= 2^{-44}). In this paper, an innovative cryptosystem based on the production of iris keys is suggested to address these shortcomings and enhance security. Iris distortion, CCD camera pixel noise, eyelashes, among the other factors, create differences in features retrieved from different images of the same iris. As a result, the iris features cannot be employed as a cipher key directly. This paper employs an error correcting method to eliminate the difference to tackle this challenge. The iris textural feature vector is translated using the Hash function and the Reed- Solomon error-correcting algorithm.

John G Daugman [45] uses the preprocessing procedure to locate and normalize the iris. The inner 3/4 of the bottom half of an iris is defined as a region of interesting (ROI). The ROI is then converted into a (256 * 64) pixels rectangular block. A set of 2-D Gabor filters are applied to the normalized iris image. By converting each iris vector's component to an integer between [0, 15] the feature vector is obtained. The public database CASIA 1.0 [46] is used to test the practicality of the proposed system. Selecting three images of each iris for picture blurriness, occlusion, and other different issues, author conducted (108 * 107)

imposter decryptions and (108 * 1) genuine decryptions. After completing all of the decryptions, the result of FAR and FRR are 0% and 5.5556% respectively. The value of FAR is ideal for preventing data from being effectively decrypted by an imposter. The probability of accessing the message from the vault without the real iris can be determined using the equation (4),

$$p = \frac{\sum_{i=0}^{64} (2^i * C_{256}^i)}{3^{256}} \approx 2^{-137} \quad (4)$$

In other words, the security of this system is equivalent to that of a 137-bit cipher key, which is significantly stronger compared to Hao's method, which has a security level of 44 bits. This method has a FRR of roughly 6%, which indicates that 6% of real users will have to present their iris more than once for decryption, which can still be accepted. For this cryptosystem, a more fruitful technique of extraction can be developed, and the suggested system can be tested on a big database.

A unique and highly secure cryptographic key is created from an iris template by employing AES for encrypting and decrypting identification data. For the genuine identification test, this study utilizes two distance metrics—Hamming distance and Euclidean distance. The Log-Gabor technique is used for feature extraction. The authors evaluate the FAR and FRR at various thresholds using both distance metrics, finding that a threshold of 0.4 yields the optimal FAR and FRR. Additionally, the authors analyze the Equal Error Rate (EER) and Total Success Rate (TSR). The TSR, which measures the verification performance of the iris cryptography system, is calculated as shown in the following equation (5),

$$TSR = \left(1 - \frac{FAR + FRR}{Total\ no.\ of\ accesses}\right) \times 100\% \quad (5)$$

In this case, the Equal Error Rate (EER) is 3.68%, while the computed Total Success Rate is 89%. The comparison results of using Hamming distance and Euclidean distance, are given in Table 3, which we collected from available literature [47].

When compared to the result of the Hamming distance, the result of the measurement of Euclidean distance is substantially weaker and ineffective for the application of iris recognition. Because of its efficiency and simplicity,

Table 3. Comparison Results Using Hamming Distance and Euclidean Distance

Threshold Value (0.4)	Hamming Distance	Euclidean Distance
FAR (%)	0.02	20.5
FRR (%)	11	41.51
FAR (%)	3.68	28.7
FAR (%)	89	56.41

Hamming distance is an appropriate distance metric for computing the difference between one iris template and several templates. But Euclidean distance metric is inadequate for iris template matching. The efficiency of this iris cryptography system can be enhanced by researching and using a hybrid of several error-correcting code techniques to address the burst and backdrop of an iris picture error.

The phase correlation technique is another efficient technique to establish a novel iris recognition system [48]. The iris boundaries in the digital image of the eye are identified using a Canny Edge Detection technique and a Circular Hough Transform. Circle Houghman Detection technique is utilized to build up the speed of the suggested system. In the normalization module, the proposed system employs the Image Registration technique to align a newly acquired picture, $I_a(x, y)$, with a database image, $I_d(x, y)$. The intensity of the new image has been created to be close to points to the reference image, when a mapping function $u(x, y), v(x, y)$ is used to convert the original coordinates. The mapping function at equation (6) must be selected to minimize,

$$\iint (I_d(x, y) - I_a(x - u, y - v))^2 dx dy \quad (6)$$

while being limited in capturing an image coordinate similarity transition from (x, y) to (x', y') , meaning,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} - sR(\emptyset) \begin{pmatrix} x \\ y \end{pmatrix} \quad (7)$$

In this scenario, 's' acts as the scaling factor, while $R(\emptyset)$ is a matrix that corresponds to a rotation by $\emptyset$. When implemented, for a specific pair of iris images I_a and I_d , the warping parameters 's' and ' $\emptyset$ ' are obtained through an iterative process of minimization. For this iris picture registration, a phase correlation base approach is applied. The efficiency of the result offered by the phase correlation method is considerably high.

The normalized iris region is convolved with a 2D Gabor filter to encode the iris region's features. The encoding method creates a bitwise template that contains a lot of bit data, as well as a noise mask that is consistent with the contaminated area of the iris pattern and identifies the template's bits as corrupt. For comparing the quantified vectors and verifying users, Hamming distance measurement is utilized. To enhance security, the Reed-Solomon method is used to encrypt and decrypt data directly. By applying proposed Hadamard and Reed Solomon error correction algorithms, authors were able to reach a 60 percent accuracy rate for a single accurate capture between two different photographs of the same iris. The appropriate acceptance ratio could increase to 93.6% if multiple photos for a single iris are saved for matching and comparison.

By employing visual cryptography, the method [49,50] offers dual-layer security to the biometric template, ensuring the preservation of the iris template's extracted image

within the proposed approach. In addition, each template is given a distinctive number that is encrypted by employing Visual Cryptography. CASIA eye image database [51] has eye images of 756 grayscale with 108 unique eyes or types and seven various images for each individual eye. In this paper, JAVA platform is used to execute the system for the improvement of the security.

The primary flaw with traditional cryptography is that it is incapable of identifying legitimate users. Traditional methods rely on the presence of a token or on knowledge. It is possible to crack them. The human iris is a distinctive structure that can be used for non-invasive biometric analysis. It is reported that encryption keys in software-based cryptography are lengthy bit strings. It is difficult to remember such a big list of random numbers. It is also vulnerable to brute-force attacks or other techniques. Biometric cryptography was introduced to improve the security of encrypted data. For the experiments, the Casia iris database was used to obtain iris images for feature extraction process. A 128-bit secret key is chosen from an iris image which is used as the key in AES encryption. In this paper, Radial edge suppression is used for extracting the image after segmentation. For Normalization process, Dougman's rubber sheet model is applied. Five tests are carried out to ensure that the key is random [52].

Visual Cryptography method is a key tool to create unique key with the use of the iris template for security improvement [53]. Employing the VC module for storing the extracted iris image in the database with the use of random number generator algorithm has proven to be an innovative technique to keep iris templates safe. Visual information (images, texts, etc.) can be encrypted using the cryptographic technique in visual cryptography which enables decryption to be done entirely by sight. Visual cryptography transforms encrypted images based on the principles of human visual perception. The whole system is working based on the two phases: (i) Enrollment, & (ii) Authentication. SNF module is used for segmentation, normalization, and feature extraction process in the enrollment phase. The author employed the VC module for storing the extracted iris image in the database by using the random number generator algorithm. The author has proposed an automatic segmentation algorithm using Circular Hough Transform to analyze the boundary of the iris and the pupil. Daugman's rubber sheet model is used for normalization process. In the authentication phase, a comparison is made between the iris template and the extracted image by sending the image of the eye to the SNF process. In this paper, shares are generated by using the generated random number as input to the VC module only for the monochrome images. The colored eye image will be the focus of future development with the use of color visual cryptography for generating the shares. The proposed visual cryptography method in this paper will solve the problem of the existing method by generating high quality shares [54].

The iris image database pictures are first trained by utilizing segmentation, normalization, Hough transform and feature extraction processes respectively. The text picture is then used to extract the features of the database's matching iris image. The size of the database determines the time it takes to train and text the image [51,55]. For feature extraction, the normalized iris pattern is converted into a 1-D Gabor wavelet. The obtained information is generated and quantized into bit-wise templates that are in adequate sizes for both the real and fictitious responses. The working procedure of the proposed algorithm is described by the following flow chart below (Figure 9). This method is used for the iris template to make it safer from attacks in addition to providing a centralized database to the users [56].

About 10-20 iris database images are used in the suggested technique. Each of the databases has about 20 images of it. This work was carried out by using the MATLAB and the summarized result is given in Table 4, which we collected from [56].

The time of implementation is dependent on the size of the database. As much as the number of images increases, the execution time is also increased.

To develop an efficient method for iris recognition, a study is done by employing advanced techniques such as Hough transform, Daugman rubber-sheet model for iris segmentation, normalization. The equation (8) refers to Log Gabor filter used for iris feature extraction,

$$G(f) = \exp \left[\frac{-\left(\log\left(\frac{f}{f_0}\right)\right)^2}{2\log\left(\frac{a}{f_0}\right)^2} \right] \quad (8)$$

In addition, this equation utilizes 3DES and Twofish algorithms to improve the privacy through the encryption of iris template. It is then concealed within a cover image using a least significant bit technique using steganography, which ultimately creates a stego image. This process makes the changes in master file to be undetectable which addresses the template security in biometric recognition. Twofish, Triple Data Encryption System, and LSB steganography are utilized to emphasize the security Iris template. This technique primarily executes two operations: securing and retrieving. In first operation, Iris template and cover image are taken as inputs whereas stego image along with cover image in the second operation. Over successful operations, the outputs of stego image and iris template assert the strength of the system in the steganography algorithm. This process highly secure the iris template and reducing the risks in information technology environments [57]. The generation of iris biometrics based encryption keys through the extraction and normalization of iris data effectively identify consistent regions within the iris structure. This methods creates feature vectors from local space filling curve descriptors. This encryption keys are generated from the features vectors created from local space. Hybrid feature selection, neighborhood component analysis along with interval-based encoding scheme are combined to express these keys as a distinct feature vector. As shown in Figure 10, the dynamic biometric key generation system is tested with various iris datasets using illumination and rotation invariant descriptors. This results demonstrate that keys in the range of 333 to 1120 bits exhibit high randomness, unlinkability, revocability. This system is reported to be secure against brute-force and JPEG compression attacks. The key generation rate of this system has been achieved up

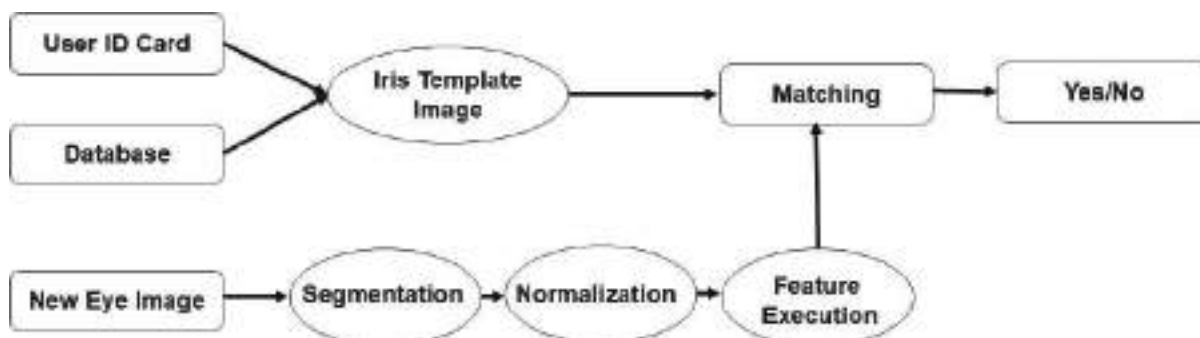


Figure 9. Working Procedure of the proposed algorithm [From Udayini et al. [56], with permission from Blue Eyes Intelligence Engineering and Sciences Publication].

Table 4. Implementation Time

Images in the database (No.)	10	20	30	40	50
Implementation time (sec)	~ 0.602	~ 0.699	~ 0.735	~ 0.806	~ 0.832

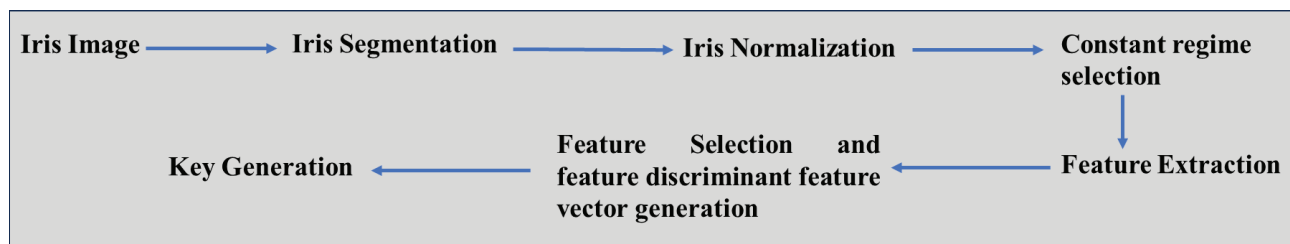


Figure 10. Schematic model for iris-based key generation process as proposed in [58] [Drawn by the author].

to 95.50%, which outperforms existing methods in terms of false acceptance rate (FAR), false rejection rate (FRR), and true acceptance rate (GAR). This results make the system suitable for cryptographic applications in cyber-physical security, IoT, and cloud computing. Future research could explore improving key generation speed and testing with other biometric modalities [58].

The unique iris texture can effectively use to generate secure encryption keys. This provides a secure method for protecting sensitive informations. Using the Histogram of Oriented Gradients (HOG) algorithm, keys of varying lengths (64 to 1024 bits) were produced. The security and reliability of the keys are rigorously tested using chi-square, ENT, and GAP tests. Multiple implementations were performed using high-resolution iris images which shows that the keys consistently maintains high randomness across different samples. The study also suggests that biometric features such as fingerprints can generate similarly secure encryption keys, which offers a wide range of applications for improving cryptographic systems [59].

Firdous Kausar [60] proposed an iris-based cancelable biometric cryptosystem that provides authenticity to the user and safety encryption of user healthcare data which is secured and authenticated. Symmetric key-cryptography is used for encrypting that data and in the encrypted form keeping it on the smart card. The secret encryption key is bound to the patient's cancelable iris template using the fuzzy commitment approach. Using the owner's iris template of the healthcare smart card, the suggested approach allows user verification as well as decryption of healthcare data as needed. The main beneficitation of this paper is the effective combination of a cancelable biometric system with the Fuzzy commitment scheme-based key-binding technique for verifying and safely reserving patient healthcare data on a smart card. The helper data is created with the integration of the secret key's Reed Solomon (RS) encoding and a cancelable transformation of the patient's iris biometric template. The author used symmetric key-cryptography methods to overcome the crucial issue of key-management by authorizing healthcare professionals and patients to share the encoded key, preserving it nowhere. It can be accessed at the running time once the user has successfully completed iris-based biometric authentication. To tie the cancelable iris template with the secret key, author use Reed-Solomon codes. In comparison to [61], where the

maximum key length is 63-bits with FAR of 0% and FRR of 63%, the exploratory findings demonstrate that it can be successfully acquired a key of maximum 252-bits with FAR of 0% and FRR of 7% from the input iris template. According to the security analysis, the adversary is unable to extract the secure key from the stored data on the smart healthcare card. Before applying a secret key to the iris template, a non-invertible cancelable transform function prohibits the attacker from performing a cross-matching attack or obtaining a secret key from the user's stolen iris template.

The One-Time Iris Code (OTIC) encoding scheme effectively addresses the security concerns related to transmitting Iris Codes between clients and servers. As depicted in Figure 11, the model converts the iris code into an 8 X 8 binary matrix and distorts it by rotating the matrix in one of five possible directions. Thus the security of the original code is enhanced. In the next step, the distorted matrix is linearized into a 64-bit sequence, then a 33-bit timestamp is embedded at random positions using four random numbers. This results in 97-bit encoded iris code which is then combined with 3-bit rotation parameter which forms a 100 bit OTIC. This data is then transmitted to the server which decodes this OTIC by reversing the rotations and extracting the embedded timestamp and reconstruct 64-bit Iris Code. The method minimized the risk of the compromisation of the Iris code [62].

The histogram processing of iris images, along with FAR vs. FRR and time complexity analyses, showed that the proposed Iris Recognition System (PIRS) offers greater security and efficiency compared to conventional systems (CIRS). The Equal Error Rate (EER) of PIRS is 0.44, which is nearly similar to CIRS (0.2 to 0.5). PIRS exhibits higher resistance to common security attacks such as MITM, brute-force, and replay attack. The time complexity of PIRS is $O(\log n)$ which is lowered than the time complexity of CIRS ($O(n)$). This leads to faster execution, enhances overall security and minimizes the risk of original iris code compromise.

Cloud storage is relatively economical and flexible business models, but security breaches remain a major concern in various threat landscapes. To enhance data protection in public cloud environments, a biometric-based authentication approach is proposed in the study [63]. The study improves cloud security by combining biometric authentication, encryption, and data placement methods that protects the sensitive information from unauthorized

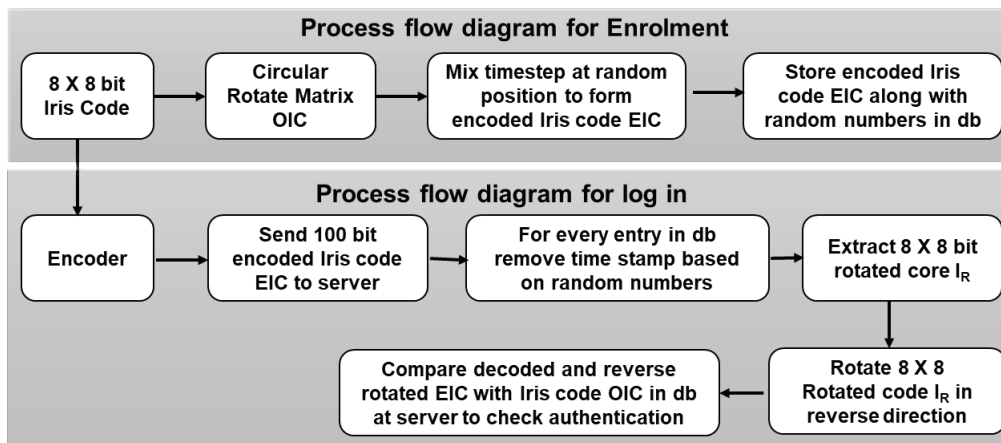


Figure 11. Enrollment and login process [Harikrishnan et al. [62], with permission from Elsevier].

access and potential misuse within cloud environments. This study introduces an integrated crypto-biometric system (ICBS) for secure data access. In addition to this, a data-placement method is introduced which is based on splitting and merging the data into multiple storage containers. This approach ensures accountability, privacy and protection against unauthorised access. This approach further enhances storage efficiency and provides safety assurance even in dispersed across a public cloud environment. Integrated Crypto-Biometric System (ICBS) improves overall data security in cloud environments. This results in strong authentication and secure access control. Furthermore, This approach reduces potential vulnerabilities through the distribution of data pieces across different

locations and thereby reducing the chance of a data breach or unauthorized access to the entire dataset.

This study introduces an efficient data placement strategy, alongside supplementary techniques for data partitioning and merging. The cloud storage system utilizes data partitioning [64] and distributes the divided data across multiple object storage containers within the IBM Bluemix cloud.

To improve performance during the classification process, a more sophisticated learning model like Support Vector Machine (SVM) can be applied. A schematic flow-chart of Iris based recognition in cloud processing is displayed in the Figure 12. The use of the suggested biometric authentication approach helps minimize latency, prevent data loss, and optimize packet handling while processing

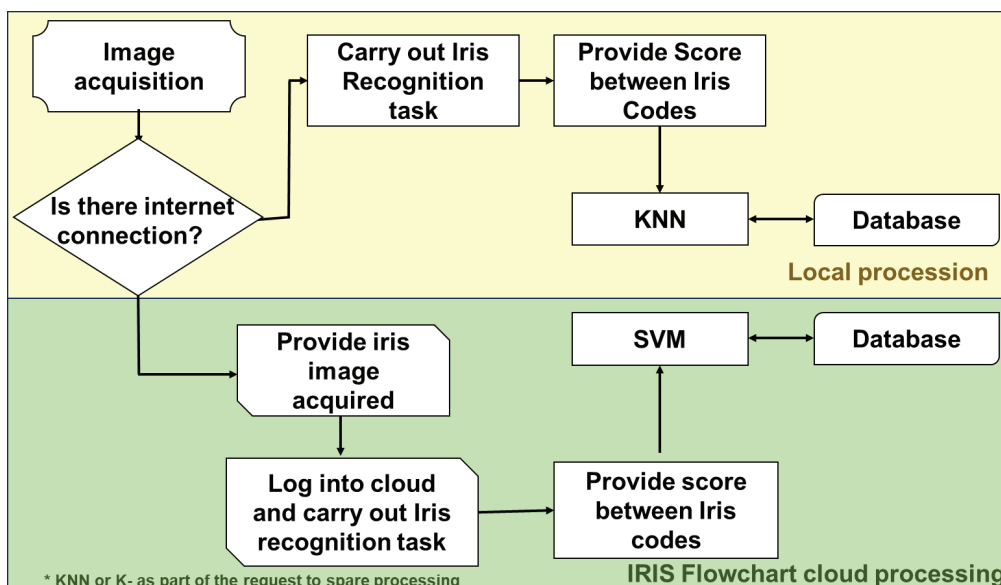


Figure 12. Schematic flowchart of Iris based recognition in cloud processing.

Table 5. Significant contributions based on Iris

Datasets Used During Performance Analysis	Method	Processor	Advantages	Ref
Database CASIA 1.0	To remove the differences caused by iris distortion, eyelashes, eyelids occlusion etc., Reed-Solomon error correcting code is used. The iris textural feature vector is converted to a cipher key for encryption and decryption using the error correcting code and hash functions.	-	Iris Key Generation	[40]
Database CASIA 1.0	The system derived vector features from the iris and utilized an altered fuzzy vault algorithm for encrypting and decrypting messages.	-	Iris for Information Security	[43]
Database CASIA 1.0	To derive a distinctive and reliable cryptographic key from an iris template. The AES cryptography algorithm is utilized for encrypting and decrypting iris image data. The identification process for template matching uses distance metrics like Hamming distance and Euclidean distance.	MATLAB 7.0	Iris for Identity Document	[47]
CASIA 1000 Iris Image Database	A technique for image registration based on phase correlation that converts the iris texture from cartesian to polar coordinates.	-	Iris for Normalization Process	[48]
Database CASIA 1.0	To provide two-fold security to the iris template using Visual Cryptography.	MATLAB & JAVA	Iris for Visual Cryptography	[49]
Database CASIA 1.0	AES is used for both encryption and decryption and some different tests are done to check the randomness of the key.	-	Iris for security improves	[52]
-	To create a distinct robust key using iris template.	-	Iris for Security Purpose	[53]
IIT Delhi Iris Database (Version 1.0)	A proposed algorithm can be used for hiding different types of data and to secure the iris template in data base by using visual cryptography.	MATLAB	Iris for Visual Cryptography	[56]
CASIA- database V3.0	Generation of iris template using cryptography and steganography. To address the issue of exploiting a biometric template hack or assault for malicious reasons.	MATLAB 2013A	Iris template generation	[57]
CASIA-IrisV3-Lamp database	A cancellable biometric cryptosystem utilizing iris recognition is designed to securely store patient healthcare data on a smart card. It uses symmetric key cryptography to encrypt this data before saving it in encrypted form on the smart card.	-	Iris to secure healthcare smart card	[60]

multiple user requests on a single server. Additionally, the data is securely stored using a data placement strategy, with authentication and access control managed through Crypto-Biometric Systems (CBS) in cloud computing, providing strong protection against unauthorized data access. Table 5 summarizes some published articles based on Iris.

Review on DNA

In the recent scenario, DNA cryptography utilizes DNA's encryptic features to address data security challenges. Due to their stability and variability, various segments of DNA are selected for encoding. This approach offers highly

secure data storage and transmission solutions. This unique feature of DNA makes it a promising way to increase information security against cyber threats.

An enhanced key generation proposal based on the concepts of DNA was reported, with reformed encryption and decryption processes. The proposed methods were developed and examined, demonstrating their efficiency in computing, storage, and transmission. Throughout the decryption process, a noticeable increase in dynamism was observed [65]. This study builds upon prior suggested work

Table 6. Datasets during performance analysis

TESTS	.doc	.pdf	.zip	.txt	.gif	.jpeg
File size (kb)	370	524	210	89	2750	3930
Cipher text size (kb)	370	524	210	89	2750	3930
Encrypting time (sec)	7	9	4	3	44	62
Decrypting time (sec)	8	9	4	1	54	62

by incorporating techniques such as key creation, encryption, and decryption.

An algorithm proposed by Roy et al. based on DNA encoding genetic information in codons [66] or symmetric key encryption, along with transmitter and receiver side calculations, is efficiently introduced in this study. The reported key distribution systems presented in the literatures [67–70] are considered to address this issue. In this method, two-level keys are employed, with the first-level key used for converting plaintext to ciphertext, and the second level enhancing the security of the method. This approach proves resilient against brute force attacks due to the utilization of two levels of keys. Based on the datasets in Table 6, the result suggests that the cipher text size is nearly identical in all test situations. It demonstrates that decryption takes substantially less time than encryption. It can also be deduced that even if the file size grows significantly, the encryption and decryption times do not change dramatically.

DNA technology is found to be a source of biometric authentication. Biometrics based on deoxyribonucleic acid (DNA) may be the most precise method of identifying a person [71]. Every human being has their own unique map (blueprint) for every cell they create. Since DNA determines both the physical and mental attributes of humans, it is unlikely that anyone will possess the exact same collection of genes, except in the case of identical twins. Human DNA is shared in 99.7% of cases and 0.3% (1 million nucleotides [72]) are varying. Because this variability is inherited and is consequently particular to every individual. These changeable regions are known as Short Tandem Repeats (or STRs) that may be tested to differentiate one man or woman from another. Extracting, Amplifying, and Sequencing are the three main steps to crack the DNA code. In just a few hours, PCR has enabled the production of hundreds of millions of copies of a single DNA sequence. PCR is an enzyme process that replicates a particular section of DNA over and over again to produce multiple copies of a sequence. This molecular technique comprises heating and cooling samples for 30 cycles in a specific thermal cycling pattern. Every molecule bearing the target sequence receives a copy of the target DNA sequence throughout each cycle. In recent years, it has become possible to PCR amplify 16 STRs in one tube, including the ‘amelogenin’ gender assignment locus. DNA can be amplified in about 20 minutes using

some commercially available devices, such as the Miniature Analytical Thermal Cycling Instrument (MATCI) and on-chip PCR. The suggested DNA cryptography method using XOR techniques is proven to be a user friendly and time saving approach towards network security. In these methods, plain text is chosen to be encrypted, after which a random key is generated. Also, a randomised codon list is generated, which is XORed with the random key. The encrypted message is created using the swap complement procedure. Biotechnology and DNA features are used to create the key. This approach is quick because of the XOR operation, and the time complexity is $O(\log N)$. However, space complexity may improve in the future [73].

A novel DNA cryptography approach was introduced, incorporating the XOR operation and the One-Time Pad for secure data transmission. This approach integrates an arithmetic operation or XOR operation, a One-Time Pad, and a DNA complementary rule, thus providing three layers of security [74]. This method is simple to use and secure against random OTP generation. OTPs are difficult for hackers to guess. Due to several prerequisites being used, this method is not that convenient, therefore, when choosing the OTP, care must be taken to consider the prerequisites.

A pseudo-biological DNA encryption approach was proposed, inspired by the central dogmas of biology. Unlike traditional DNA encryption, this method employs DNA idioms and techniques. Encryption and decryption are executed through transcription, splicing, and translation processes (Figure 13). These encryption steps significantly bolster security measures. Randomly generated keys are utilized to maximize confusion and diffusion, rendering decryption of encrypted text challenging. Robustness analysis is conducted to validate the method's resilience against attacks. This approach advocates the utilization of advanced Biocomputing systems [75].

In DNA-based cryptography, keys can be secured by using hamming code and a block cipher. The use of symmetric cryptography is important for optimizing DNA-based approaches. It compensates for the original algorithm's shortcoming of requiring a match between the length of the plaintext and the length of the secure key. This approach employs an improved matching mechanism with unexpected length to guarantee confusion of the ciphered result. The maximum length matching method has also

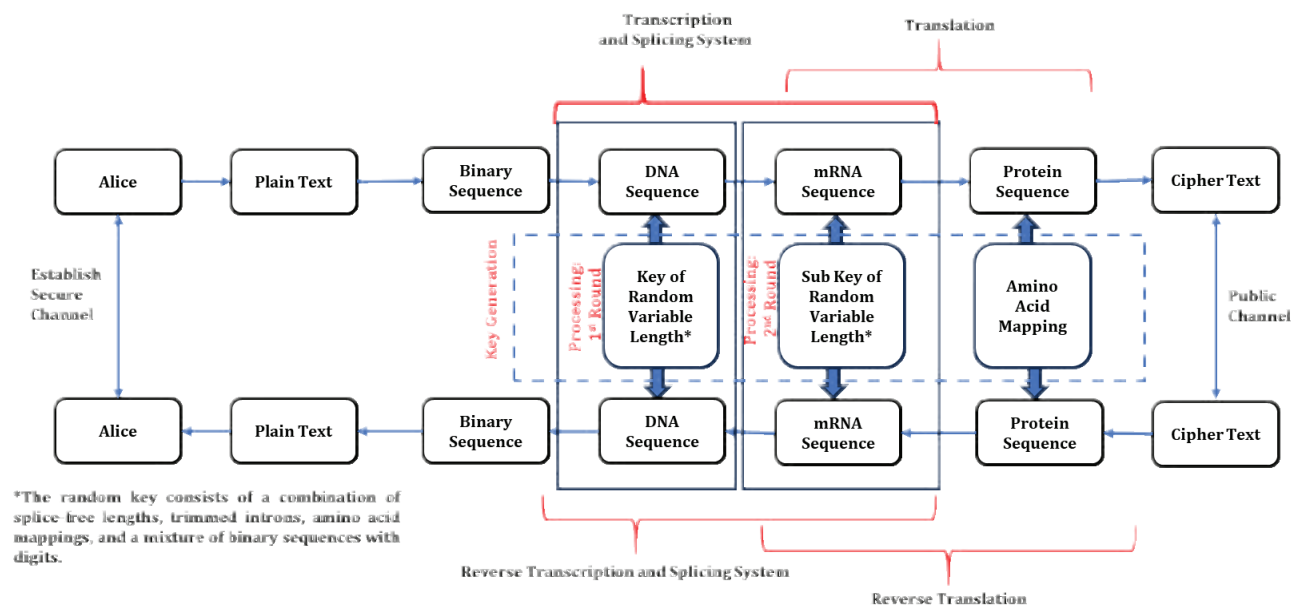


Figure 13. Biotic DNA-based cryptosystem proposed in [75] [Drawn by the author].

been developed in this technique to protect against various attacks. The length restriction of the secret key illustrated by DNA strands is eliminated by the proposed DNA-cryptographic method [76].

The approach of encrypting client-side data before saving in the cloud is an innovative technique based on cryptography methodology. This is a symmetric-key cryptography scheme based on DNA cryptography. The suggested system employs random dynamic encoding tables,

resulting in increased security. The proposed method is CPA-secure, according to the security study. In addition to providing a comprehensive architecture of this methodology (Figure 14) and comparing it to current symmetric-key algorithms (DNA, DES, AES, and Blowfish), the experimental findings show that this method beats the earlier one in terms of cipher text size, encryption time, and throughput. As a result, this new technique is far more efficient and effective than the old one [77].

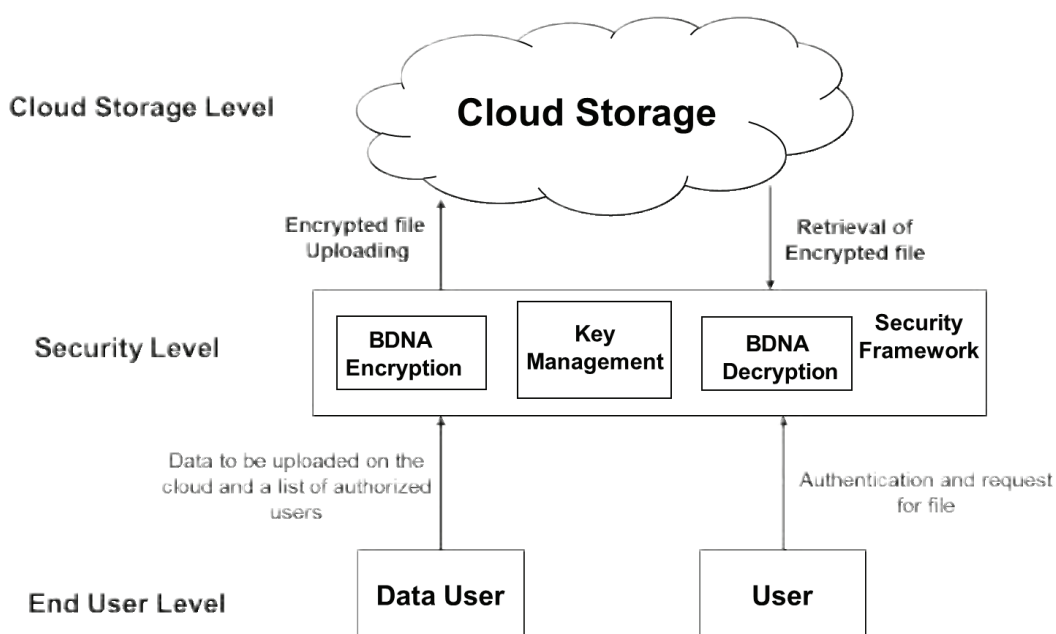


Figure 14. DNA-based methodology to secure cloud computing [From Sohal et al. [77], with permission from Springer].

The strategy for ECC assisted DNA mapping technique generates random DNA codes and the alphabets are assigned to non-repetitive subdivisions. The alphabets are then employed at both ends for encryption and decryption. Standard elliptical curve parameters and techniques were selected for the DNA elliptical cryptography. The robustness of elliptical cryptography against time and simple power analysis attacks (SPA) [77–80] has been validated by existing research. Without needing for a valid authentication key, elliptical cryptography randomly generates data after decoding. Finding the authentication key is the primary goal of attacks against cryptosystems. When useful information is discovered during decryption while utilizing a random key, these attacks are successful. The overall cryptosystem is much more secure from attacks if the encrypted data is itself encoded using a nonrepeating pattern. DNA mapping prior to encoding and DNA remapping after decoding is not a challenging computations according to the linear relation between time consumed and inputs in the described system. Due to the DNA mapping and remapping in the suggested approach, the energy usage is marginally higher compared to the other system. The current elliptical cryptography approach is made more secure by DNA mapping before encryption. This technique is appropriate for real-time IoT applications that have limited resources because of their low timing consumption and high CPU usage. This suggested strategy was tested in an IoT setting utilizing a Raspberry Pi3. According to the security study of the suggested scheme, the existing cryptosystems are far more resistant to SPA attacks and timings when used with the suggested DNA mapping. The referred technology has great capability in cloud-based implementations and mobile applications. The proposed solution holds many promises for future IoT technologies that demand a more compact but effective security framework. The suggested DNA mapping, encryption, decryption, and DNA remapping takes a linear amount of time in relation to the length of the input data. As a result, the suggested DNA

mapping and remapping system (Figure 15) enhances the security of existing elliptical cryptosystems while consuming time [81].

Feature extraction is used to process the fingerprint template, and the features are then encrypted using the DNA cryptographic technique. For the second level of authentication, an XOR technique is used. The feature vector that matches the input feature vector throughout the decryption process is obtained and for the final output the user input image is recovered. There are certain concerns with the algorithms such as AES, DES, IDEA, they've chosen to implement. It is difficult to integrate AES with software. When using DES, the cost of execution rises. The IDEA method is a challenging decision due to the enormous number of weak keys, and management has become complicated. The optical fingerprint imaging technique is utilized here. For capturing fingerprint input, this operates similarly to a digital camera. This study uses a U. are. U 4500 sensor to capture a digital image of a fingerprint. The use of visible light is used to capture this specialized image. MATLAB software is used to process the fingerprints that have been obtained. The safety of the fundamental DNA codon usage is improved by the random generation of DNA sequences that are never used again. It is feasible to authenticate fingerprint templates effectively and safely by integrating DNA cryptographic techniques and a chaotic algorithm. The suggested technique encrypts human fingerprints and iris utilizing DNA characteristics and chaotic algorithms, ensuring that the template's privacy is preserved [82].

As the volume of digital data generated globally continues to grow exponentially, traditional storage mediums are reaching their limits in terms of capacity, durability, and efficiency. DNA storage meets a critical need for more efficient storage solutions with the ability to archive vast amounts of data in an extremely compact form. DNA storage technology represents a groundbreaking advancement in data storage, leveraging artificially synthesized deoxyribonucleic acid (DNA) as a medium for storing information.

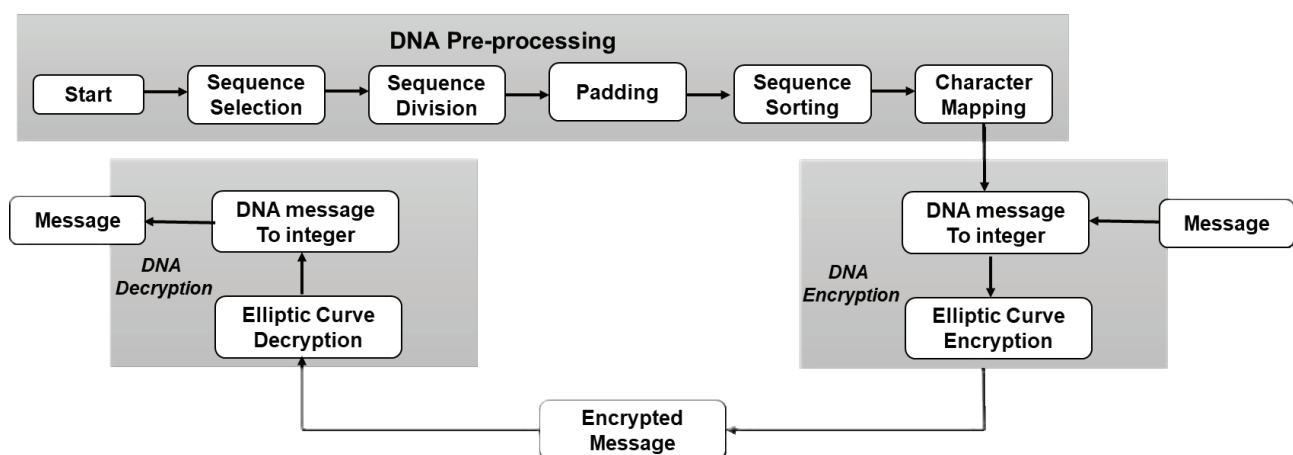


Figure 15. DNA-based elliptic curve cryptosystem for IoT devices.

This technology, characterized by its exceptionally high information density at the molecular scale, offers significant advantages such as enormous storage capacity, long-term durability, ease of access, and maintenance-free operation. Overcoming the current bottlenecks of cost and sequencing speed would propel DNA storage to the forefront of future information technologies, offering a sustainable and scalable solution to the world's growing data storage needs [83].

Encoding data in the hybridization states of DNA oligonucleotides, rather than relying solely on their sequence, has proven to be an advantageous approach for DNA information storage. This method allows for the creation of a metastable information solution that remains stable for long periods at room temperature but can be rapidly and permanently erased by brief heating. The authors demonstrated the effectiveness of this method using eight bitmap images of famous artwork, showing that after over two months of storage at 25 °C, more than 99% of the encoded pixels could be accurately restored, indicating a potential information half-life of at least 15 years. Importantly, the encoded information can be rapidly and permanently erased by heating the solution to 95 °C for just 5 minutes, which disrupts the truth marker pairing, rendering the data unrecoverable. Here bitmap artwork images are used to demonstrate the approach in a visually clear way. However, the method does not depend on long-range patterns within the data. The authors predicted that this approach will be particularly useful for compressed information files with high K-complexity. The approach not only enhances the potential of DNA as a storage medium but also addresses the challenge of securely erasing stored information, which traditional methods struggle to achieve without leaving traces [84].

Another study showed a secure storage scheme that integrates biometric authentication based on the user's genomic short tandem repeat (STR) profile which explore the potential of DNA as a digital data storage medium. The study further combines natural and synthetic DNA for secure data storage. The cryptographic keys of at least 80 bits of entropy is generated by using STR profiles, a form of personal genetic information. With the combination of AES-256 encryption and Reed-Solomon error correction,

these keys enables secure and long-term storage of sensitive data in synthetic DNA. This main advantage of this method is that both the encryption key and the encrypted data are stored in the same DNA medium. The encrypted data can be read simultaneously using sequencing technology. This integration makes the decryption process simple and ensure that the encrypted data can only be deciphered with the correct STR marker data. The analysis confirms that STRs provide sufficient entropy to be used as cryptographic keys and thereby preventing brute-force attacks. The study experimentally generated an 80-bit strong key from human DNA and encrypted 17 kB of information in synthetic DNA. This data is then successfully decrypted which shows the efficacy of the methods in security applications [85].

The use of highly specific, stimuli-responsive DNA materials in DNA cryptography plays a crucial role in information encryption [86]. Leveraging their high information density, low maintenance cost, vast computational parallelism, and ease of synthesis, user-defined DNA materials have been developed for cryptographic applications, including encryption, authentication, and digital signatures [87,88]. DNA based tools are widely used in developing cryptographic methods and biosensors. However, conventional DNA based methods have typically used in controlling enthalpy, which provides unpredictable responses to stimuli and less accurate outcomes due to significant energy fluctuations. Programmable pH-responsive DNA motifs objectively overcome this problem by simultaneously controlling both enthalpy and entropy, which has proven to be an effective method for data storage. These advanced DNA motifs were successfully applied to glucose biosensing and crypto-steganography systems. This technique demonstrates their significant potential in both biosensing technologies and secure information encryption [89].

DNA data storage by combining bioinformatics with the Diffie–Hellman Key exchange method enhances data protection during communication. The study demonstrates a novel cryptosystem that leverages the entire Central Dogma of Molecular Biology (CDMB), encompassing the process by which DNA is translated into proteins (Figure 16).

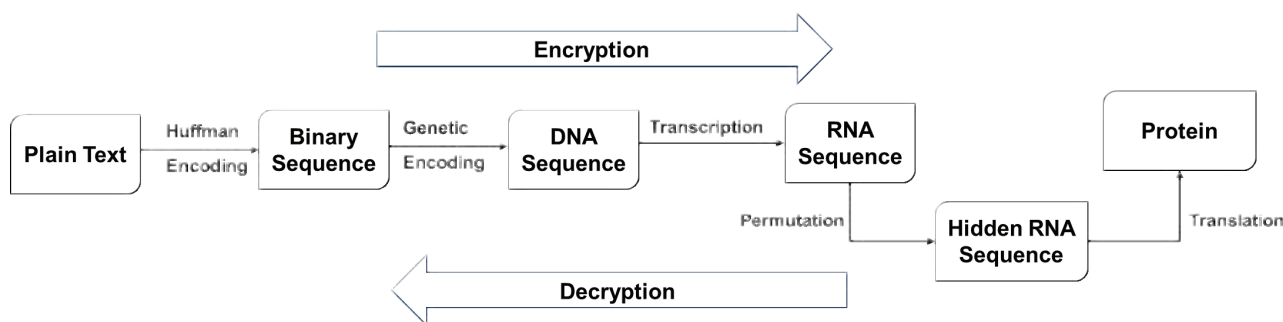


Figure 16. Schematic block diagram of encryption & decryption process of the methodology as reported in [90] [Drawn by the author].

Table 7. Significant contributions based on DNA

Biometric	Method	Processor	Advantage	Ref.
DNA	An enhanced key generation method is proposed using DNA synthesis principles, with improvements made to optimize the encryption and decryption processes.	JAVA	DNA for Enhanced key Generation	[66]
DNA	DNA as a source of biometric authentication	Miniature Analytical Thermal Cycling Instrument (MATCI) and on- chip PCR	DNA for Biometric Authentication	[71]
DNA	XOR encryption for providing DNA based encryption systems.	MATLAB	DNA for Enhancing Asymmetric Encryption	[73]
DNA	A cryptographic mechanism using secret keys derived from biotic DNA.	Pseudo-Biotic DNA cryptography method in C++ and conducted a series of experiments in a network simulator [NS2] to evaluate its effectiveness.	DNA for Adaptive Cryptographic Attacks	[75]
DNA	Study of existing DNA cryptography methods and then proposed a new pseudo-DNA	MATLAB	DNA for Secure Data Transfer.	[74]
DNA	A Hamming code combined with a block cipher mechanism to guarantee the secure transmission of a cryptographic key.	A simulation program, created in C language with VC6.0, has been developed for assessing the algorithm.	DNA for enforced secure key distribution	[76]
DNA	A novel cryptographic method has been proposed that utilizes client-side encryption to secure data before it is uploaded to the cloud. This technique features a multi-faceted symmetric-key encryption approach based on DNA cryptography.	MATLAB	DNA for secure cloud computing	[77]
DNA	A novel hybrid ECC scheme encoded with DNA that offers multiple levels of security.	C and Cortex-A8 running at 200 MHz.	DNA-Based Elliptic Curve Cryptography (ECC) for IoT Devices	[81]
DNA	A fingerprint template encryption scheme, leveraging DNA encoding and genetic algorithms, has been suggested to enhance the security of fingerprint templates.	MATLAB	DNA for Fingerprint Authentication	[82]

To generate key, this approach combines DNA cryptography with Diffie–Hellman Key exchange and includes additional modifications to strengthen the system. The proposed bio-inspired cryptosystem is designed to offer higher cryptographic efficiency and robustness even with large datasets, which provides effective protection against various online threats [90]

Table 7 summarizes a recently published article based on DNA.

PERFORMANCE ANALYSIS

This section offers a comparative analysis of existing Fingerprint, Iris and DNA based cryptographic models, summarizing their performance metrics, results, and effectiveness compared to existing methods.

Fingerprint-Based Cryptographic Model

A comparative analysis of existing fingerprint-based cryptographic models is summarized in Table 8.

Table 8. Performance analysis of existing fingerprint-based cryptographic models

Existing Models	Performance Metric	Performance/ Result Analysis	Comparison Analysis
A New Approach to Symmetric Key Generation Using Combination of Biometrics Key and Cryptographic Key to Enhance Security Of Data [20]	Storage, Security	Generation of 256-bit secret key by combining a 128-bit biometric key from fingerprint minutiae with a 128-bit system-generated key, enhancing security.	Cryptographic key generation process is unique, such key ensures the enhancement of the security and effectiveness.
An approach to cryptographic key distribution through fingerprint based key distribution center [21]	Hamming distance, Time for key generation	Time computed for feature extraction, cancelable template generation, key generation are 0.05 sec, 0.006 sec, 0.003 sec.	- The comparison shows impostors must guess at least half of the bits to break the key due to a 50% average Hamming distance. - Traditional cryptosystems face key sharing and privacy issues. The proposed KDC method addresses these by improving distribution and using biometric data to enhance security and simplify privacy.
Biometrics as a cryptographic method for network security [23]	Time for key generation	- Time calculated for Feature extraction of biometric template, template generation, template encoding and decoding, cryptographic key generation are 0.05 sec, 0.002 sec, 0.15 sec, 0.002 sec respectively - The amount of time needed for the entire process = 0.204 sec	Stronger and offers an improved level of security for methods of encryption and decryption through the employing of a pair of keys generated from fingerprint imprints.
Biometric Based Cryptography for Enhanced Security Mechanism [24]	FAR, FRR, GAR, EER	- The average matching percentage for genuine fingerprint pairs= 89.99%, - Range of Matching percentages= 81.37% to 99.83%, with a standard deviation of 0.043, showing that genuine pairs consistently have a high number of matchable bits for cryptographic keys. - The average Hamming distance is 49.94%, which means there is an average difference of 128 bits between genuine and imposter keys. - Hamming distances range from 37.89% to 61.72%, with a standard deviation of 0.031. Genuine keys differ by 40% to 60% of bits from 99.89% of imposter keys. Only 0.04% of imposter keys have unmatched bits below 40%. An imposter cannot match more than 128 bits out of a 256-bit cryptographic key.	- GAR for the FVC2002 database= 96.49% and GAR for the NIST special databases= 95.89%, indicating proposed method surpasses existing ones. - It effectively counters various attacks, including network, host, replay, and Man-in-the-Middle attacks. - This approach provides an effective solution to secure the session-based communication over unsecured channels.
Designing an Efficient Fingerprint Authentication Algorithm using SHA-512 and its Implementation [26]	Look-Up Table (LUT), Flip-Flop(FF), Clock Buffer Resources (BUFG)	- The reduction of LUT utilization from 63% to 33.07%, results in 47.62% improvement which indicates more efficient resource allocation. - FF utilization dropped from 36% to 13%, reflecting more efficient resource use and speeding up the hashing process, which is especially advantageous for time-sensitive applications. - Utilization rate of BUFG is 3.13% which demonstrates effective clock use. This promotes system reliability and synchronized operation.	- This system uses a secure binary representation and efficient feature extraction, and thereby improving the security and dependability of systems. - Methods like loop unrolling for SHA-512 minimizes the resource usage. This makes the system ideal for cost-effective and low-power standalone electronics.
Fingerprint Based Cryptographic Key Generation [35]	Hamming distance (HD)	- The maximum HD for a genuine case is 20 and the minimum HD is 4; The maximum HD for an imposter case is 81 and the minimum HD is 40. - Impostors cause more mismatches than natural variations in fingerprint images, indicating they cannot replicate an authentic key.	The distinct separation in Hamming distances between genuine and imposter keys demonstrates the effectiveness the proposed algorithm, as imposter-generated keys differ significantly from another fingerprint of the same person.
Finger print recognition based on biometric cryptosystem [37]	Euclidean Distance	- The Euclidean distance of the random image and Eigen space= 43.99. - Time complexity of the proposed method is $O(p^{3/2})$, which ensures the algorithm is efficient, rapid and reliable.	Compared to other approaches, this one is simple, trustworthy and fast.

Table 9. Comparative analysis of existing iris-based cryptographic model

Existing Models	Performance Metric	Performance/Result Analysis	Comparison Analysis
An effective authentication scheme for a secured IRIS recognition system based on a novel encoding technique [62]	Time Complexity, Execution Speed	- Time complexity of proposed iris recognition system (PIRS) is $O(\log n)$, leading to faster execution speeds compared to conventional systems. - The EER for PIRS is 0.44, which comparable to conventional Iris recognition system (CIRS).	- Compared to CIRS (having a time complexity of $O(n)$), PIRS delivers superior security. In addition, it reduces the vulnerability and provides better protection against common threats. - The EER of PIRS is stable and comparable to traditional systems. The system further offer similar accuracy along with improved security features.
Efficient private key generation from iris data for privacy and security applications [58]	Distinctiveness, dissimilarity, randomness, security analysis, segmentation accuracy.	- Key sizes of the system is of 333 to 1120 bits which makes the technique suitable for various cryptographic applications. - The iris key generation rate of the system is 95.50%. This makes the proposed scheme more efficient than other key generation methods in terms of GAR, FAR, and FRR. - Provides high uniqueness, strong inconsistency, passes random tests, resists attacks, supports dynamic and revocable keys, maintains a low mean error rate (MER) under various conditions.	This work outperforms traditional static methods, other biometric modalities like fingerprints, and other dynamic key generation approaches especially in noisy or diverse environments.
Generate and Evaluate Encryption Keys Obtained From Iris Biometric Data [59]	Randomness of generated keys (Chi-square test, ENT test, Block Metric, Gab Metric)	For a 64-bit random key (thresholds: GAB = 3.84, Block = 3.841), the Chi-square test shows the key is random, with entropy of 3.875 bits per byte, GAB and Block metric value 3.781125 and 22 respectively.	-
Iris based cancelable biometric cryptosystem for secure healthcare smart card [60]	False Acceptance Rate (FAR), False Rejection Rate (FRR)	FAR=0%, FRR= 7%, with the ability to generate a robust 252-bit encryption key, indicating high accuracy and security.	- Compare to [61][54], which yields a key length of 63 bits, with a FAR of 0% and an FRR of 63%, proposed model gives better FAR, FRR values. - Compared to traditional encryption methods, the proposed model enhances security by eliminating the need for secure key storage and significantly outperforms in terms of FAR and FRR, making it a superior choice for secure healthcare applications.
A safe and secured iris template using steganography and cryptography [57]	GAR, FAR	GAR=98.70%, FAR=0%	- The proposed approach improves other existing approaches in terms of GAR, FAR values. - Combining 3DES and Steganography enhances performance by improving revocability and distinguishing between genuine and imposter distributions, while reducing sensitivity to outliers compared to existing methods.
Iris image authentication using visual cryptograph [56]	Execution time	As the database size grows, execution time also increases, and conversely, a smaller database size reduces execution time.	Compared to existing methods, proposed visual cryptography method offers greater security by adding an extra layer of authenticity and a centralized user database.
Iris biometric cryptography for identity document [47]	FAR, FRR, EER, TSR	(At threshold value=0.04 using Hamming Distance) FAR=0.02%, FRR=11%, EER=3.68%, TSR=89% (At threshold value=0.04 using Euclidean Distance) FAR=20.5%, FRR=41.51%, EER=28.7%, TSR=56.41%	Comparing to Hamming distance, Euclidean distance is much weaker and not sufficient for the use of iris identification.

Table 10. Comparative analysis of existing DNA-based cryptographic model

Existing Models	Performance Metric	Performance Analysis / Result Analysis	Comparison Analysis
An improved Symmetric key cryptography with DNA based strong cipher [66]	Length of text size, Time.	- The ciphertext size nearly doubles in all test cases, making the process ideal to encrypt a huge file. - Decryption time is much shorter than encryption time, and both encryption and decryption times remain stable even with significantly larger file sizes.	Though this method is efficient and powerful against specific attacks, the partial information contained in the ciphertext makes the method more powerful.
Advancement In DNA As Source of Biometric Authentication [71]	Time	- Time required for extracting DNA= less than 30 seconds - Time required for amplifying DNA=approximately 20 minutes - Time required for sequencing DNA=less than 30 seconds	The high throughput microchips can provide outcomes up to 5 times faster than the existing machinery.
Enhancing Asymmetric Encryption using DNA Based Cryptography [73]	Speed, Time complexity	- The encryption speed of the XOR Cipher technique is extremely high. - A message can be encrypted in polynomial time ($O(\log n)$ time).	-
Pseudo DNA Cryptography Technique using OTP Key for Secure Data Transfer [74]	Computational complexity	Proposed algorithm ensures maximum security during data transfer while minimizing issues related to computational complexity.	Compared to traditional DNA-based cryptography algorithms, proposed pseudo-DNA cryptographic algorithm is much harder to crack due to the additional artificial features.
Inspired Pseudo Biotic DNA Based Cryptographic Mechanism Against Adaptive Cryptographic Attacks [75]	Length of plaintexts and ciphertexts, Key length, Speed, Time, Storage	- Ciphertext length matches plaintext length and varies with key length. - It uses less storage space than plaintext, offering better storage efficiency. - With larger plaintext sizes, the random key length also increases, significantly minimizes the relative size of the key length. - The key and ciphertext transmit faster via safe and public channels, respectively, enhancing the storage and transmission efficiency of the method. - The adversary needs more chosen ciphertexts to obtain the key for a given plaintext, increasing the time required to recover it.	Proposed method is effective against chosen ciphertext and brute force attacks, and it also excels in computation, storage, and transmission efficiency than existing methods.
An optimized DNA based encryption scheme with enforced secure key distribution [76]	Length of the secure key	- Ciphertexts vary significantly because different secure keys are used for each encryption, showcasing the algorithm's effective diffusion property. - Matching positions are below 2%, and identical positions and lengths are under 1.5%, due to the algorithm's unpredictable matching length, satisfying the confusion theory.	The old DNA encryption technique achieves confusion theory but restricts its scope and applications. The proposed enhanced DNA computing, using block cipher symmetric cryptography, overcomes the original algorithm's need for matching plaintext and key lengths.
BDNA-A DNA inspired symmetric key cryptographic technique to secure cloud computing [77]	Ciphertext size, Encryption time	- For plaintext size= 30 kb, ciphertext size= 35.0 - For plaintext size= 30 kb, encryption time= 0.149 - The throughput of BDNA is high.	- The proposed model encrypts plaintext more quickly than AES, DES, DNA, and Blowfish. - BDNA offers higher throughput in comparison to all other approaches. - The recommended approach generates smaller ciphertexts compared to existing encryption techniques.

Iris-Based Cryptographic Models

A summary of the comparative analysis of existing iris-based cryptographic models is provided in Table 9.

DNA-Based Cryptographic Model

A comparative analysis of existing DNA-based cryptographic models is furnished in Table 10.

CONCLUSION

In this paper, we presented recent research contributions in the field of Crypto-Biometrics specially in the areas of fingerprints, Iris, and DNA for cryptographic key generation, template protection. In comparison to conventional systems, the application of this kind of system, which combines biometrics and cryptographic methods, offers significantly greater confidentiality and security. This survey highlights significant advancements in fingerprint-based biometric cryptosystems, demonstrating their enhanced security capabilities through the integration of fingerprint features with cryptographic methods. By integrating biometric data with traditional Key Distribution Centers (KDCs), these systems achieve greater confidentiality and eliminate the need for key memorization. The development of cancelable biometric templates further improves security by transforming and shuffling minutiae to protect against unauthorized access. Performance metrics confirm the system's effectiveness in distinguishing genuine keys from impostor ones, while advanced cryptographic methods such as Secure Hash Algorithm (SHA-512) and Elliptic Curve Cryptography (ECC) contribute to enhanced efficiency and protection. Practical implementation, including mobile applications and encryption algorithms like Advanced Encryption Standard (AES) and Time-Based-One-Time-Password (TOTP), underscores the usability and effectiveness of these systems. The survey covers recent developments and innovations in iris recognition for generating secure key and thereby protecting the data. Methods such as 2-D Gabor filtering, fuzzy vault algorithms and error-correcting codes like Reed-Solomon and Hadamard codes have reported to be a promising approach to enhance the security of biometric cryptosystems. Numerous studies reported the advantage of using visual cryptography and phase correlation to improve security and mitigating potential threats. In Iris recognition, the superiority of Hamming distance over Euclidean distance has been emphasized through the comparison of different distance metrics and encryption algorithms. The combination of steganography, cancelable biometrics, and advanced data placement in cloud environment has made significant progress in this field. The DNA integration improves the security in cryptographic systems while provides a novel solutions for key generation, authentication, encryption and data storage. It is relatively a secure method for encoding data using XOR operations, one-time pads, and symmetric encryption. This methods adopts the stability and variability of DNA.

The current fingerprint-based biometric cryptosystems faces several limitations such as high false rejection rates, issues in practical implications and challenges in integrating fingerprint data with traditional cryptographic methods. System performances and efficiency is affected by the variability in fingerprint quality, computational overhead, and privacy concerns. The main challenges in iris-based cryptography are to maintain image consistence, reduce processing time and ensure key security. It is important to balance false acceptance and rejection rates to improve practical usability. Key length and randomness are the critical parameters. For instance, some methods produce shorter keys with attack vulnerabilities. Commercial implementation is only possible if the system demonstrates a user-friendly, cost-effective solution and ensures robustness against spoofing and attacks. DNA-based cryptography faces various concerns due to its high cost, low scalability and its high complexity. This is mainly due to slow processing and over-reliance on specialized equipment. The practical integration and efficiency is further restricted by several security concerns like vulnerabilities in key management and sequencing exposure along with its high power consumption. For broader adoption in practical, it is important to address these issues.

There is significant potential for several important areas of future research on biometric cryptography. Future work should explore various error correction techniques to increase accuracy in key generation processes. Handling instance-to-instance fluctuations in biometric data such as fingerprints is one of the potential areas that can be explored. In addition, integrating multi-parameter security techniques such as hybrid pairings of DNA with fingerprint or iris could significantly improve protection against highly secure attacks. The use of complex OTPs and consistent fine-point selection can improve the system security. Future works should also be focused on developing more efficient feature extraction methods for iris cryptosystems and the use of color visual cryptography for eye images. Furthermore, incorporating diverse distance metrics and segmentation methods could improve processing outcomes and efficiency. The scope of biometric cryptography can be broadened by exploring multimodal fusion of various biometrics to increase accuracy and key length, as well as optimizing memory usage and execution time across different smartcard types. Implementing advanced encryption techniques like AES and minimizing DNA space complexity present promising paths for future research, supporting the ongoing advancement and improved efficacy of biometric cryptography.

ACKNOWLEDGEMENTS

This research work is supported by Vellore Institute of Technology, Vellore, India.

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] Şolt N, Çalkavur S, Güzeltepe M. Secure encryption over the ring $F_2 + uF_2 + vF_2 + uvF_2$. Sigma J Eng Nat Sci 2024;42:529–533. [\[CrossRef\]](#)
- [2] Jain AK, Nandakumar K, Nagar A. Fingerprint template protection: from theory to practice. In: Campisi P, editor. Security and Privacy in Biometrics. London: Springer; 2013. [\[CrossRef\]](#)
- [3] Jain AK, Flynn P, Ross AA. Handbook of Biometrics. New York: Springer; 2007. 564 p. [\[CrossRef\]](#)
- [4] Ratha NK, Chikkerur S, Connell JH, Bolle RM. Generating cancelable fingerprint templates. IEEE Trans Pattern Anal Mach Intell 2007;29:561–572. [\[CrossRef\]](#)
- [5] Boulton TE, Scheirer WJ, Woodworth R. Revocable fingerprint biotokens: accuracy and security analysis. In: IEEE Conf Comput Vis Pattern Recognit 2007;1–8. [\[CrossRef\]](#)
- [6] Kanade S, Petrovska-Delacretaz D, Dorizzi B. Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data. In: IEEE Conf Comput Vis Pattern Recognit 2009;120–127. [\[CrossRef\]](#)
- [7] Kaur P, Kumar N, Singh M. Biometric cryptosystems: a comprehensive survey. Multimed Tools Appl 2023;82:16635–16690. [\[CrossRef\]](#)
- [8] Sharma S, Saini A, Chaudhury S. A survey on biometric cryptosystems and their applications. Comput Secur 2023;134. [\[CrossRef\]](#)
- [9] Hashem MI, Alibraheemi K. Literature survey: biometric cryptosystems based on fingerprint processing techniques. In: 2022 Int Conf Data Sci Intell Comput (ICDSIC) 2022;198–201. [\[CrossRef\]](#)
- [10] Argyropoulos S, Tzovaras D, Ioannidis D, Strintzis MG. A channel coding approach for human. IEEE Trans Inf Forensics Secur 2009;4:428–440. [\[CrossRef\]](#)
- [11] Monroe F, Reiter MK, Wetzel S. Password hardening based on keystroke dynamics. CCS '99: Proceedings of the 6th ACM conference on Computer and communications security. 1999.
- [12] Davida GI, Frankel Y, Matt BJ. On enabling secure applications through off-line biometric identification. In: 1998 IEEE Symp Secur Priv (Cat No 98CB36186) 1998;148–157. [\[CrossRef\]](#)
- [13] Dodis Y, Reyzin L, Smith A. Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. In: Advances in Cryptology – EUROCRYPT 2004. Berlin: Springer; 2004. p. 523–540. [\[CrossRef\]](#)
- [14] Hao F, Anderson R, Daugman J. Combining cryptography with biometrics effectively. Available at: <https://www.cl.cam.ac.uk/techreports/> Accessed on Sep 01, 2025.
- [15] Kanade S, Camara D, Krichen E, Petrovska-Delacretaz D, Dorizzi B. Three factor scheme for biometric-based cryptographic key regeneration using iris. In: 2008 Biometrics Symp 2008;59–64. [\[CrossRef\]](#)
- [16] Jain AK, Ross AA, Nandakumar K. Introduction to Biometrics. New York Dordrecht Heidelberg London: Springer; 2011. [\[CrossRef\]](#)
- [17] Seals T. ThreatList: a third of biometric systems targeted by malware in Q3. Threat Post. Available at: <https://threatpost.com/threatlist-a-third-of-biometric-systems-targeted-by-malware-in-q3/150778/> Accessed on Sep 01, 2025.
- [18] Sur A. India saw 18 million cyber attacks in first quarter of 2022: Google's Royal Hansen. Available at: <https://www.moneycontrol.com/europe/?url=https://www.moneycontrol.com/news/business/india-saw-18-million-cyber-attacks-in-first-quarter-of-2022-google-executive-royal-hansen-9084911.html> Accessed on Sep 01, 2025.
- [19] Mascellino A. Japan government websites hit by cyber-attacks, Killnet suspected. Whats Hot On. Available at: <https://www.infosecurity-magazine.com/news/japan-govt-websites-killnet/> Accessed on Sep 01, 2025.
- [20] Solanki KH, Sahayak A. A new approach to symmetric key generation using combination of biometrics key and cryptographic key to enhance security of data. Int J Eng Res Technol 2013;2:1–7.
- [21] Barman S, Chattopadhyay S, Samanta D. An approach to cryptographic key distribution through fingerprint based key distribution center. In: 2014 Int Conf Adv Comput Commun Informatics (ICACCI) 2014;1629–1635. [\[CrossRef\]](#)
- [22] Balakumar P, Venkatesan R. Secure biometric key generation scheme for cryptography using combined biometric features of fingerprint and iris. Int J Comput Sci 2011;8:349–356.

- [23] Ankit K, Rekha J. Biometrics as a cryptographic method for network security. *Indian J Sci Technol* 2016;9. [CrossRef]
- [24] Prasanthi KS, Krishna HV, Saranya S. Biometric based cryptography for enhanced security mechanism. *Int J Pure Appl Math* 2017;117:779–787.
- [25] Tulyakov S, Farooq F, Mansukhani P, Govindaraju V. Symmetric hash functions for secure fingerprint biometric systems. *Pattern Recognit Lett* 2007;28:2427–2436. [CrossRef]
- [26] Gupta A, Banakar N, Chetan NK, Aryan M, Purushotham U. Design and implementation of an efficient fingerprint authentication algorithm using SHA-512. In: 2024 Third Int Conf Distrib Comput Electr Circuits Electron (ICDCECE) 2024;1–7. [CrossRef]
- [27] Dwivedi R, Dey S, Sharma MA, Goel A. A fingerprint based crypto-biometric system for secure communication. *J Ambient Intell Humaniz Comput* 2020;11:1495–1509. [CrossRef]
- [28] Gowda VD, Gite P, Rahman M, Kadam KR, Manivasagam G, Prasad KDV. Novel approaches to biometric security using enhanced session keys and elliptic curve cryptography. *J Discrete Math Sci Cryptogr* 2024;27:477–488. [CrossRef]
- [29] Gowda VD, Prasad K, Shekhar R, Srinivas R, Srinivas KNV, Lakineni PK. Development of a real-time location monitoring app with emergency alert features for Android devices. In: 2023 4th IEEE Glob Conf Adv Technol 2023;1–8. [CrossRef]
- [30] Gowda VD, Sharma A, Kumaraswamy S, Sarma P, Hussain N, Dixit SK, et al. A novel approach of unsupervised feature selection using iterative shrinking and expansion algorithm. *J Interdiscip Math* 2023;26:519–530. [CrossRef]
- [31] Gowda VD, Prasad KDV, Gite P, Premkumar S, Hussain N, Chinamuttevi VS. A novel RF-SMOTE model to enhance the definite apprehensions for IoT security attacks. *J Discrete Math Sci Cryptogr* 2023;26:861–873. [CrossRef]
- [32] Gowda VD, Kumar PSVVS, Latha J, Selvakumar C, Shekhar R, Chaturvedi A. Securing networked image transmission using public-key cryptography and identity authentication. *J Discrete Math Sci Cryptogr* 2023;26:779–791. [CrossRef]
- [33] Gowda VD, Sharma A, Raghavendra K, Nagabushanam M, Reddy HGG. Vector space modelling-based intelligent binary image encryption for secure communication. *J Discrete Math Sci Cryptogr* 2022;25:1157–1171. [CrossRef]
- [34] Pawar P, Datar S, Ranade N, Thorat K, Gharu AN. Biometric security using cryptography for insurance data retrieval with additional OTP generation and verification. *Int J Innov Res Technol* 2019;5.
- [35] Suresh K, Pal R, Balasundaram SR. Fingerprint based cryptographic key generation. In: *Intelligent Data Communication Technologies and Internet of Things*. Cham: Springer International Publishing; 2020. p. 704–713. [CrossRef]
- [36] Salamon WJ, Ko K. Fingerprint Minutiae Viewer (FpMV). Available at: <https://www.nist.gov/services-resources/software/fingerprint-minutiae-viewer-fpmv> Accessed on Sep 01, 2025.
- [37] Jalaja V, Gsgn A, Mohan LN. Finger print recognition based on biometric cryptosystem. *J Integr Sci Technol* 2024;763. [CrossRef]
- [38] Oduguwa T, Arabo A. Passwordless authentication using a combination of cryptography, steganography, and biometrics. *J Cybersecr Privacy* 2024;4:278–297. [CrossRef]
- [39] Murugan A, Savithiri G. Feature extraction on half iris for personal identification. In: 2010 Int Conf Signal Image Process 2010;197–200. [CrossRef]
- [40] Wu X, Qi N, Wang K, Zhang D. A novel cryptosystem based on iris key generation. In: *Proc 4th Int Conf Nat Comput (ICNC 2008)* 2008;53–56. [CrossRef]
- [41] Nagar A, Chaudhury S. Biometrics based asymmetric cryptosystem design using modified fuzzy vault scheme. In: 18th Int Conf Pattern Recognit (ICPR'06) 2006;537–540. [CrossRef]
- [42] Juels A, Sudan M. A fuzzy vault scheme. *Des Codes Cryptogr* 2006;38:237–257. [CrossRef]
- [43] Wu X, Qi N, Wang K, Zhang D. An iris cryptosystem for information security. In: 2008 Int Conf Intell Inf Hiding Multimed Signal Process 2008;1533–1536. [CrossRef]
- [44] Davida G, Frankel Y, Matt B, Peralta RE. On the relation of error correction and cryptography to an off line biometric based identification scheme. In: *Proc WCC99 Workshop Coding Cryptogr* 1999;129–138.
- [45] Daugman J. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans Pattern Anal Mach Intell* 1993;15. [CrossRef]
- [46] Service note on CASIA iris image databases. 2012. Available at: <https://www.yumpu.com/en/document/view/3796486/service-note-on-casia-iris-image-databases> Accessed on Sep 01, 2025.
- [47] Moi SH, Rahim NBA, Saad P, Sim PL, Zakaria Z, Ibrahim S. Iris biometric cryptography for identity document. In: 2009 Int Conf Soft Comput Pattern Recognit 2009;736–741. [CrossRef]
- [48] Nithyanandam S, Gayathri KS. A new IRIS normalization process for recognition system with cryptographic techniques. *Int J Comput Sci* 2011;8:342–348.
- [49] Hajare N, Borage A, Kamble N, Shinde S. Biometric template security using visual cryptography. *J Eng Res Appl* 2013;3:1320–1323.
- [50] Ehuil BB, Chen C, Wang S, Guo H, Liu J. A secure mutual authentication protocol based on visual cryptography technique for IoT-cloud. *Chin J Electron* 2024;33:43–57. [CrossRef]

- [51] Kezheng L, Bo F, Hong Z. Visual cryptographic scheme with high image quality. In: 2008 Int Conf Comput Intell Secur 2008;366–370. [\[CrossRef\]](#)
- [52] Asok SB, Karthigaikumar P, Mangai S. Iris based cryptography. Int J Adv Res Comput Commun Eng 2013;2:1310–1313.
- [53] Ajith S, Balaji Ganesh Kumar M, Latha S, Samiappan D, Muthu P. Iris cryptography for security purpose. J Phys Conf Ser 2018;1000:012111. [\[CrossRef\]](#)
- [54] Pawar A, Kumbhare T, Murkute P, Kallapur S. Enhancing iris scanning using visual cryptography. IOSR J Comput Eng 2015;17:54–57.
- [55] Zhang H, Wang X, Cao W, Huang Y. Visual cryptography for general access structure by multi-pixel encoding with variable block size. In: Proc 2008 Int Symp Knowl Acquis Model (KAM) 2008;340–344. [\[CrossRef\]](#)
- [56] Udayini V, Bindu CH, Malleswari PN. Iris image authentication using visual cryptograph. Int J Recent Technol Eng 2019;8:288–291.
- [57] Abikoye OC, Ojo UA, Awotunde JB, Ogundokun RO. A safe and secured iris template using steganography and cryptography. Multimed Tools Appl 2020;79:23483–23506. [\[CrossRef\]](#)
- [58] Dash P, Pandey F, Sarma M, Samanta D. Efficient private key generation from iris data for privacy and security applications. J Inf Secur Appl 2023;75. [\[CrossRef\]](#)
- [59] Abdulraheem AA, Hasso SA. Generate and evaluate encryption keys obtained from iris biometric data. In: 21st Int Multi-Conf Syst Signals Devices (SSD 2024). IEEE; 2024. p. 321–328. [\[CrossRef\]](#)
- [60] Kausar F. Iris based cancelable biometric cryptosystem for secure healthcare smart card. Egypt Inform J 2021;22:447–453. [\[CrossRef\]](#)
- [61] Yang W, Wang S, Hu J, Zheng G, Chaudhry J, Adi E, Valli C. Securing mobile healthcare data: a smart card based cancelable finger-vein bio-cryptosystem. IEEE Access 2018;6:36939–36947. [\[CrossRef\]](#)
- [62] Harikrishnan D, Sunilkumar N, Shelby J, Kishor N, Remya G. An effective authentication scheme for a secured iris recognition system based on a novel encoding technique. Meas Sens 2022;25. [\[CrossRef\]](#)
- [63] Shaleen. A study on integrated crypto-biometric system to protect the unauthorized access of data. J Interdiscip Multidiscip Res 2024;19:593–610.
- [64] Prabu S, Ganapathy G. Secured data storage in public cloud environment through crypto-biometric system. Int J Comput Sci Eng Technol 2017;7:10–14.
- [65] Roy B, Rakshit G, Chakraborty R. Enhanced key generation scheme based cryptography with DNA logic. Comput Sci 2011. Preprint. <https://api.semanticscholar.org/CorpusID:18906078>
- [66] Roy B, Rakshit G, Singha P, Majumder A, Datta D. An improved symmetric key cryptography with DNA based strong cipher. In: Int Conf Devices Commun (ICDeCom-2011) 2011;1–5. [\[CrossRef\]](#)
- [67] Eschenauer L, Gligor VD. A key-management scheme for distributed sensor networks. In: Proc 9th ACM Conf Comput Commun Secur 2002;41–47. [\[CrossRef\]](#)
- [68] Zhu S, Xu S, Setia S, Jajodia S. Establishing pairwise keys for secure communication in ad hoc networks: a probabilistic approach. In: 11th IEEE Int Conf Netw Protocols 2003;326–335.
- [69] Liu D, Ning P, Li R. Establishing pairwise keys in distributed sensor networks. ACM Trans Inf Syst Secur 2005;8:41–77. [\[CrossRef\]](#)
- [70] Du W, Deng J, Han YS, Varshney PK, Katz J, Khalili A. A pairwise key predistribution scheme for wireless sensor networks. ACM Trans Inf Syst Secur 2005;8:228–258. [\[CrossRef\]](#)
- [71] Qaseem IAQM, Asif SAW, Armeen ZF. Advancement in DNA as source of biometric authentication. Int J Eng Res Technol 2013;2:1–8.
- [72] Delibaş E, Arslan A. A new feature vector model for alignment-free DNA sequence similarity analysis. Sigma J Eng Nat Sci 2022;40:610–619. [\[CrossRef\]](#)
- [73] Rani M, Jain S. Enhancing asymmetric encryption using DNA based cryptography. Int J Comput Sci Trends Technol 2014;2:7–11.
- [74] Kalyani S, Gulati N. Pseudo DNA cryptography technique using OTP key for secure data transfer. Int J Eng Sci Comput 2016;5657.
- [75] Suresh Babu E, Naga Raju C, Munaga HM, Prasad K. Inspired pseudo biotic DNA based cryptographic mechanism against adaptive cryptographic attacks. Int J Netw Secur 2016;18:291–303.
- [76] Zhang Y, Liu X, Ma Y, Cheng LC. An optimized DNA based encryption scheme with enforced secure key distribution. Cluster Comput 2017;20:3119–3130. [\[CrossRef\]](#)
- [77] Sohal M, Sharma S. BDNA—a DNA inspired symmetric key cryptographic technique to secure cloud computing. J King Saud Univ Comput Inf Sci 2022;34:1417–1425. [\[CrossRef\]](#)
- [78] Odelu V, Das AK, Choo KKR, Kumar N, Park Y. Efficient and secure time-key based single sign-on authentication for mobile devices. IEEE Access 2017;5:27707–27721. [\[CrossRef\]](#)
- [79] Cheng C, Lu R, Petzoldt A, Takagi T. Securing the Internet of Things in a quantum world. IEEE Commun Mag 2017;55:116–120. [\[CrossRef\]](#)
- [80] Reddy AG, Yoon EJ, Das AK, Odelu V, Yoo KY. Design of mutually authenticated key agreement protocol resistant to impersonation attacks for multi-server environment. IEEE Access 2017;5:3622–3639. [\[CrossRef\]](#)
- [81] Tiwari HD, Kim JH. Novel method for DNA-based elliptic curve cryptography for IoT devices. ETRI J 2018;40:396–409. [\[CrossRef\]](#)
- [82] Vijayakumar S, Jansi M, Lavanya D, Lavanya Sree V. Delay efficient genetic algorithm with DNA based cryptography for fingerprint authentication. Eur J Mol Clin Med 2020;7:2077–2080.

-
- [83] Zhang Z, Zhang Z. DNA information storage and cryptography system. *Acad J Sci Technol* 2024;10:243–249. [\[CrossRef\]](#)
- [84] Kim J, Bae JH, Baym M, Zhang DY. Metastable hybridization-based DNA information storage to allow rapid and permanent erasure. *Nat Commun* 2020;11. [\[CrossRef\]](#)
- [85] Grass RN, Heckel R, Dessimoz C, Stark WJ. Genomic encryption of digital data stored in synthetic DNA. *Angew Chem Int Ed Engl* 2020;59:8476–8480. [\[CrossRef\]](#)
- [86] Zhang C, Ma X, Zheng X, Ke Y, Chen K, Liu D, et al. Programmable allosteric DNA regulations for molecular networks and nanomachines. *Sci Adv* 2022;8. [\[CrossRef\]](#)
- [87] Zhang Y, Wang F, Chao J, Xie M, Liu H, Pan M, et al. DNA origami cryptography for secure communication. *Nat Commun* 2019;10. [\[CrossRef\]](#)
- [88] Fan HJ, Li J, Wang LH, Fan CH, Liu HJ. Constructions of iron atoms arrays based on DNA origami templates for cryptography applications. *Wuli Xuebao Acta Phys Sin* 2021;70. [\[CrossRef\]](#)
- [89] Zheng LL, Li JZ, Wen M, Xi D, Zhu Y, Wei Q, et al. Enthalpy and entropy synergistic regulation-based programmable DNA motifs for biosensing and information encryption. *Sci Adv* 2023;9. [\[CrossRef\]](#)
- [90] Vaishali R, Naik SM. A DNA cryptosystem using Diffie–Hellman key exchange. *SN Comput Sci* 2024;5. [\[CrossRef\]](#)



Review Article

Exploring metal organic frameworks for efficient drug delivery: A bibliometric review

Buddhadeb DUTTA¹, Supratim SUIN^{1,*}

¹Department of Chemistry, Ramakrishna Mission Vivekananda Centenary College, Kolkata, 700118, India

ARTICLE INFO

Article history

Received: 12 May 2024

Revised: 02 July 2024

Accepted: 03 October 2024

Keywords:

Computational Review; Drug Delivery; Metal Organic Frameworks; MOF; Quantitative Review

ABSTRACT

The biomedical application of the metal organic frameworks (MOFs) has become a field of immense interest in the last decade for worldwide research community. The present bibliometric review aims to find quantitatively the global scenario of research involving utilization of MOFs in designing new drug delivery systems. To explore the research direction of MOFs in drug delivery, the research executed in the last decade has been analyzed using sophisticated statistical tools, like, Bibliometrix-Biblioshiny package of R, and VOS viewer, a Java based tool. These software packages facilitate to analyze quantitatively all the research articles published during this time frame to obtain the year wise growth, country wise distribution, mostly used keywords etc. In addition, the collaboration, as well as, the co-occurrences between keywords have been analyzed. Such detailed exploration of the present field utilizing statistical tool can be very much promising to understand the current research position and find grey areas in the research involving utilization MOFs in designing drug delivery systems.

Cite this article as: Dutta B, Suin S. Exploring metal organic frameworks for efficient drug delivery: A bibliometric review. Sigma J Eng Nat Sci 2025;43(5):1837–1851.

INTRODUCTION

The term “drug delivery system” (DDS) refers to a specialized carrier capable of transporting an active pharmaceutical ingredient (API) to a specific target within the body. DDSs serve to enhance the pharmacological effectiveness of an API by addressing factors such as solubility improvement and the optimization of pharmacokinetics and biodistribution. The concept of designing a DDS with precise control over drug release kinetics can be traced back to the 1950s when Smith, Kline, and French introduced a controlled-release formulation for delivering dextroamphetamine over a 12-hour period [1, 2]. Subsequently,

an entire realm of scientific investigation has emerged, offering a multitude of exciting opportunities for enhancing drug performance, safety, and targeted distribution. Organic, inorganic, and hybrid substances like liposomes, micelles, carbon nanotubes, silica, quantum dots, and gold nanoparticles have been explored as the carrier materials in the developments of DDS in the past few decades [3]. The design of a DDS is intrinsically dependent on the specific context, necessitating a tailored selection of material composition, administration routes, and cargo release mechanisms to meet the unique biological requirements. Up to the present day, the materials employed for the synthesis

*Corresponding author.

*E-mail address: supratim1988@rkmvccrahara.org

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



of nanomaterials used in drug delivery encompass organic options such as liposomes, micelles, and dendrimers, inorganic materials like silica and gold nanoparticles, and hybrid materials, including metal-organic frameworks. In the diverse field of biology, these systems can be engineered to respond to external stimuli, such as changes in temperature, pH, acoustic signals, and heat. Peptides and antigens may be added to increase the precision in targeting specific cells. Additionally, they are capable of delivering small and large molecules. In light of the challenges in drug delivery, researchers are exploring alternative systems with enhanced properties and broader applications.

An intriguing option in the area of advanced materials is metal-organic frameworks (MOFs), also referred to as porous coordination polymers (PCPs). MOFs are noteworthy for their remarkable flexibility and adaptability, rendering them exceptionally well-suited for applications in drug delivery. They have characteristics like a generous pore volume, customizable pore dimensions (extending beyond 6 nanometers), substantial surface areas (reaching up to 8000 square meters per gram), and modifiable surface chemistries [4, 5]. Apart from drug delivery, MOFs have found applications in gas storage, energy conversion, chemical sensing, proton conduction, and catalysis. Industrial applications in oleo-chemicals, textiles, transportation, food packaging, and respiratory systems are also noteworthy [6]. Because of their porous structure, MOFs can absorb small molecules and gases within their structure. The application of MOFs in drug delivery was explored in pioneering studies by Horcajada and co-workers [7]. They explored the application of Chromium based MIL and MIL-101 MOFs for absorption of an anti-inflammatory drug molecule ibuprofen. MIL-101 showed unprecedentedly high absorption capacity of 1.4 g of ibuprofen molecule per gram of the MIL-101. Moreover, these MOFs showed excellent release capacity of the absorbed molecule in conditions similar to human body. Bermini et al. performed the Monte Carlo simulation studies and substantiated the adsorption capacity theoretically and predicted the adsorption capacity may be as high as 2 g of ibuprofen in 1 g of the adsorbent for BioMOF-100 [8], establishing strong correlation between pore volume and adsorption capacity of MOFs. These pioneering observations inspired the scientists and a huge surge of research was performed in this field in the last decades. The field expanded from Cr based to Fe based MOFs and other biocompatible metals like zirconium, zinc, and iron were explored for targeted drug delivery materials. In health care systems, application of non-toxic nano-carriers for efficient drug delivery remains a challenge. MOFs have shown remarkable advantage over conventional carriers like polymers, nanoparticles, and bacteria because of their outstanding drug loading capacity and also the ability to precisely release the adsorbed therapeutic agents. The porous structure of the MOFs allows them to encapsulate wide range of therapeutic agents, and also sparingly water soluble cancer treatment agents like curcumin. Gas

transmitter molecules effective as therapeutic agents like nitric oxide (NO), carbon monoxide (CO), and hydrogen sulphide (H_2S) may also be well encapsulated by these MOFs [9-12]. Large biomolecules like nucleic acids and proteins are protected from degradation by encapsulation within MOFs [13]. It is important to have biocompatibility and minimum or no cytotoxicity and immunogenicity for successful application of drug delivery agents. But analysis of cytotoxicity may be a challenging task, because the individual component toxicity may differ in isolation and upon incorporation within MOFs [14], and cytotoxicity may differ depending on the specific cell type under investigation [15]. In addition to the MOF structure itself, it is crucial that the metal ions and ligands used in MOFs exhibit biocompatibility, as there is probability that these components may leach into biological fluids or tissues over time [16-18]

Over the past few decades, there has been much development in different carrier systems for controlled drug release, enhancing the therapeutic benefit and reducing off-target effects [19]. MOFs have been proposed as an exciting new delivery system, with properties such as high loading capacity and tunable pores making them attractive candidates for such an application [20]. There has also been some progress in the characterization of MOF toxicity, behaviour, and uptake in vitro. However, there is still a need to improve and develop an understanding of MOFs in biological systems. The evolution of methods for particle synthesis, surface functionalization techniques, and the scope for developing new composite-hybrid materials has produced nanomaterials with potential for enhanced capabilities such as stimulus-responsive drug release, imaging, cell-specific targeting, and theranostic capabilities. Combining these developments in synthesis, together with a rational selection of biological models, mechanistic characterization and appropriate selection of in vitro assays will offer exciting prospects for the future of this field [21]. In spite of huge number of publications on the utilization of MOFs in developing drug delivery systems, only one article is there to review the biomedical applications of MOFs quantitatively [22]. This computational review aims to explore different bibliometric parameters to analyse the research executed in the last decade in this avenue. The bibliometric parameters include year wise growth, country wise distribution, most cited articles, and the distribution of sources in the MOF drug delivery research. In addition, the dimensional alternation in the research themes during the ten years span can be analysed from the word dynamics and trend topics analysis. The collaborative links between countries, citation of sources, and co-occurrences of keywords can also be found in the present manuscript. As a consequence, this bibliometric review is expected to offer new researchers to understand present research status and explore novel ideas in drug delivery using metal organic frameworks.

MATERIALS AND METHODS

Retrieval of Bibliometric Data

Here, we consider the data revealed from Web of Science (WoS) database, currently administrated by Clarivate Analytics, utilizing a unique search string related to MOF drug delivery. The WoS database was chosen as it gives more relevant results as compared to the other databases available in the present time. The search string considered in the present investigation was “TS = ((MOF OR (Metal organic framework*)) AND (drug delivery))” in the advanced search, selecting the period 01-01-2014 to 31-12-2023, i.e. the last ten complete years. A total of 3384 documents were resulted in the WoS database which includes articles, reviews and other publications such as conference proceedings, book chapters etc.

Bibliometric Parameters Analysis

All the bibliometric parameters were analysed utilizing Bibliometrix-Biblioshiny package of R [23]. The bibliometric data were exported into the web interface of Biblioshiny. After excluding the publications of 2024, total 3014 publications resulted. Our entire analysis was confined into these 3014 articles. The main information related to the present review can be obtained directly in tabular form from this package. The year-wise advancement in the frequency of publications and related citation parameters like, mean total citations per article (MTCPA), mean total citations per year (MTCPY) can be found from this software. Worldwide participation of research community in the MOF drug delivery research can be represented in world map using this software. In addition, most active country according to corresponding author can also be analysed by this tool. This software package also predicts most active authors, sources and related citation parameters like, h-index, average citation per article (ACPP). The most mentioned keywords, in the form of word cloud and data plot can also be generated from this tool. The growth of different words during the span of ten years can also be generated. This software tool also generates trend topics plot to explore the most mentioned topic during different period of time. All the data generated from Bibliometrix-Biblioshiny were plotted in Microsoft Excel.

Collaboration Between Countries, Interlinks Between Sources And Co-Occurrences Between Keywords

A java-based software package, VOS viewer, was utilized to find collaborative links between different countries. This software package has been used by many researchers in the last decade for finding collaborative networks between countries, authors, sources etc. [24] The research on MOF drug delivery research executed in the last decade, as retrieved from Web of Science database, was analysed with VOS viewer. To have visually detectable network paths in studying co-authorship, the minimum number of articles and minimum citations were optimized at 30. Total 84

countries were found to meet this condition, of which only 21 countries were found to possess active links.

The interlinks between the sources publishing articles on MOF drug delivery during the last decade was also analysed in terms of citation utilizing the same VOS viewer package. In this analysis the minimum number of articles and minimum citations were fixed at 40, to have well defined network structure. The VOS viewer package also facilitates the prediction of the co-occurrences of keywords during the timespan considered in this analysis. Here, all keywords, which includes both author keywords and keywords plus, were considered for analysing co-occurrences. The minimum number of occurrences was fixed at 150; only 25 keywords were found to possess interlinks with others. A similar kind of analysis has been reported to explore management science research in Turkey [25].

RESULTS AND DISCUSSION

Analysis of the Bibliometric Parameters

Year-wise growth of publication and citation parameters

The year-wise progression of number of articles published on utilization of metal organic frameworks in drug delivery and related citation parameters, such as, mean total citation per article (MTCPA), and mean total citation per year (MTCPY) has been plotted in Figure 1. As evident from the figure, the number of total publications in each year increases, suggesting growing interest of scientific community towards developing new MOFs for drug delivery. However, the pandemic occurring in 2019-2020 aggravated experimental research and thus resulted in some less article production during this time. In addition, the last five years of research on the said topic published significantly higher number of articles (2345), compared to only 969 articles in the first five years. This result strongly suggests the affinity of research community in MOF drug delivery systems in recent years. On analyzing citation parameters, an irregular trend in MTCPA and MTCPY is evident being maximized in 2016. The decrease in MTCPA and MTCPY in the recent years is very much usual and might be associated with the greater number of articles and less citation received in the last years.

Most active authors: number of publications and citation parameters

Figure 2 illustrates the top ten researchers publishing articles on MOF drug delivery research during 2013-2022 and the related citation parameters like, h-index and average citations per publication (ACPP). As can be seen, Wang Y published highest number of articles (117) on MOF drug delivery, followed by Li Y (111), Liu J (108) Zhang Y (103), and Liu Y (97). It is worthy to mention that, although Liu Z published relatively smaller number of articles (59), the ACPP tops the list, signifying greater acceptability of their findings to the worldwide scientific community. The

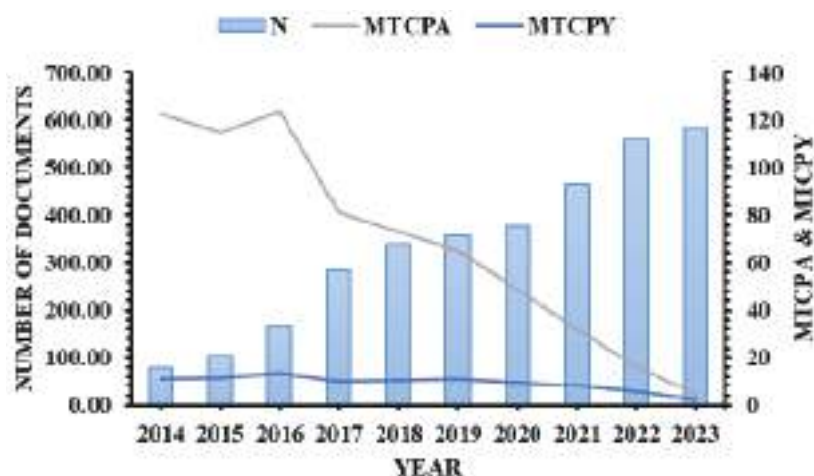


Figure 1. Year-wise growth of number of publications, MTCPA and MTCPY in the period 2014-2023.

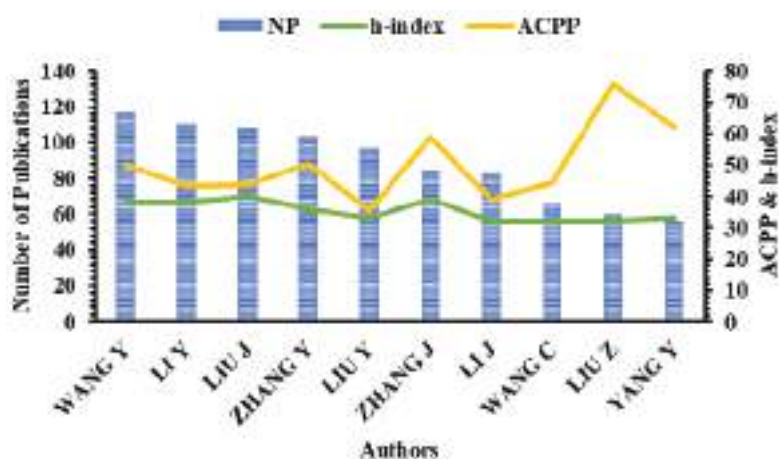


Figure 2. Number of articles, ACPP, and h-index of most active authors (top ten) publishing articles on MOF in drug delivery research in the period 2014-2023.



Figure 3. (a) Countries with most publications in MOF drug delivery in world map; (b) Most relevant countries by corresponding author doing research on MOF in drug delivery research in the period 2014-2023.

other authors receiving appreciably good ACPP are Yang Y, Zhang J, Wang Y, and Wang C. Additionally, h-index,

another important citation parameter, appears as highest for Liu J, followed by Zhang J, Wang Y, and Li Y.

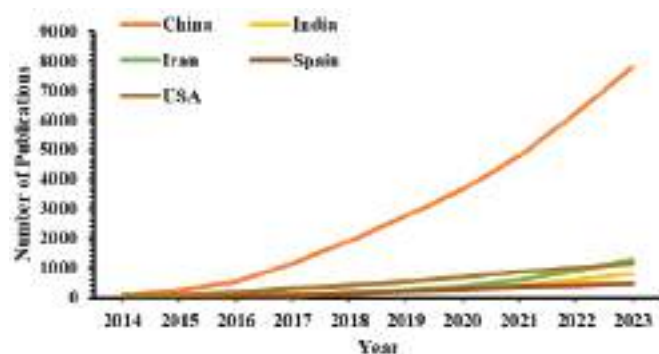


Figure 4. Growth of MOF in drug delivery research executed by top five countries in the period 2014-2023.

Most relevant countries

The most productive ten countries in MOF drug delivery research during 2014-2023 has been designated in the world map in Figure 3a. The deepening of blue colour is directly proportional with the number of publications. As evident, China published the highest number of articles on MOF drug delivery systems followed by Iran, USA, India, and Spain. The designation in the world map also suggests the worldwide participation of countries in the present research field. Figure 3b depicts the most relevant top ten countries by corresponding author in MOF drug delivery research during 2013-2022. The corresponding author data has proved itself to be the most convenient to compare the number of articles published by different countries quantitatively. The plot also predicts the SCP (single country publication) and MCP (multiple country publication), important parameters for inter-country collaboration. As evident from the plot, China published the highest number of articles (related to corresponding author) both as SCP and MCP. This was followed by Iran, USA, India,

and Spain. This data analysis supports highest inter-country collaboration in case of China, followed by Iran, USA, India, and Spain.

Year-wise growth by top five countries

The advancement in the MOF drug delivery research executed by top five countries during 2014-2023 has been represented in Figure 4. As can be seen, China published the highest number of articles in the entire period under consideration. In addition, the progression of MOF drug delivery research executed by China produced highest sloped curve, indicating the greater involvement of researchers of China in the present research topic with time. This trend was followed by countries like, USA, Iran, India and Spain. Although, USA followed China during 2016-2021, in the last couple of years Iran exceeds USA and appears second after China.

Most relevant sources

Figure 5 illustrates the top ten journals and related citation parameters publishing articles on MOF drug

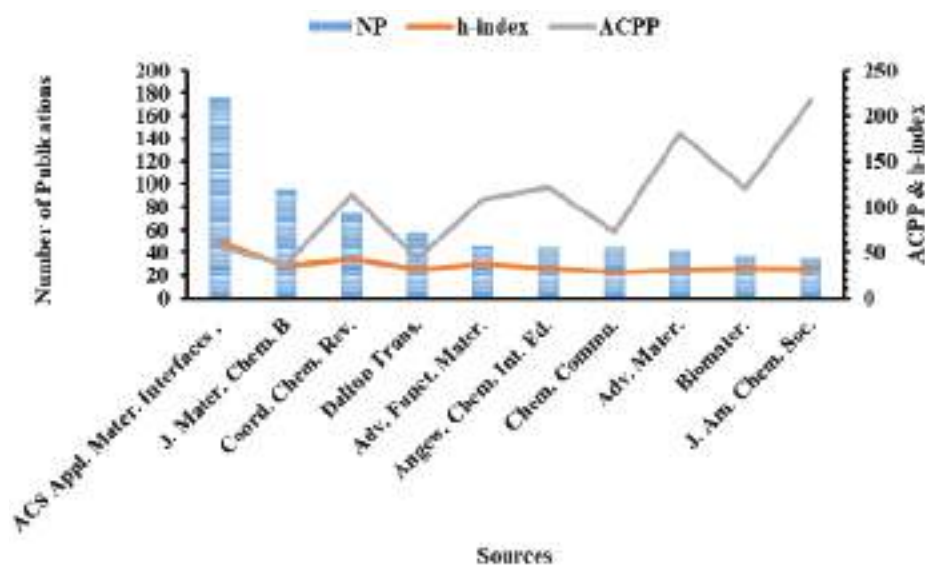


Figure 5. Total number of publications, h-index, and ACPP in the top ten journals.

biomedical applications of MOFs. The other new keywords, such as, 'adsorption', 'release', 'platform', 'design' was found to possess appreciable mention the articles on MOF drug delivery. This study is expected to offer new researchers to find the hot areas, as well as, grey areas of research in the MOF drug delivery systems.

Growth of most mentioned keywords

The growth in the top ten most mentioned keywords during the time frame has been represented in Figure 7. As can be seen, all the keywords considered in the study, gained a momentum after 2015, and the rally continues till the end of the period. The word 'drug delivery' tops the list in the entire period, and is obviously due to its utilization in the search string. The other very much obvious keywords in this line are, 'metal-organic frameworks', 'mof', 'delivery' etc. In addition to the usual keywords, new words like 'nanoparticles' was found to possess impressive growth in the period considered in our study. This trend was followed by 'adsorption', making itself very much growing during the entire period. Another important word 'release' appeared as important after 2016, as it has close relation with the sustained release of drug molecules. This trend was followed by keywords like, 'platform', 'design', and 'cancer'. However, to explore period-wise dominance of keywords, more sophisticated tool like analysis of 'trend topics' may come into play.

Trend topics analysis

To interpret period-wise dominance of topics, trend topics analysis was executed. Figure 8 represents the progressive trends of top twenty-five keywords related to the MOF drug delivery systems during the time frame. In the given figure, the horizontal lines represent the coverage years and the dominating year of the use of keyword is designated by the position of bubble; the size of the bubble

signifies the frequency of the keyword. As can be seen, the early years of research in the MOF drug delivery systems was dominated by keywords like, 'hydrogen storage', 'crystal structure', and 'luminescence' being centralized at 2017. In the later period, 'complexes', 'co-ordination polymer', and 'storage' comes into the picture, being maximized at 2018. So, the early period of research on metal organic frameworks were dominated by applications like gas storage (specifically hydrogen), catalysis etc.

The storage and transportation of gaseous fuels is always challenging associated with their low boiling points, low densities, high diffusivity and high critical pressures. Nowadays these gases are liquefied either by compression at very high pressure (in order of several hundreds of atmosphere) or by lowering the temperature significantly (liquefied hydrogen temperature). Such kind of condition requires a specialized sophisticated facility which is difficult to achieve in ordinary laboratory. MOFs are found to be efficient in storing as well as transporting gaseous fuels at ambient condition. The porosities offered by the 3d engineering of metal and ligands make the storing and transportation of fuels feasible in a cost effective and safe way [33, 34]. The unique structural parameters of the MOFs, such as, high porosities, large surface area, tunable pore size, ease of functionalization makes them efficient as compared to the traditional materials like, zeolites, activated carbon etc. Thus, important gases like, hydrogen, acetylene, methane can easily be stored and transported utilizing MOFs.

The drug delivery utilizing metal organic frameworks gained momentum after 2017, being maximized at 2018. The topics like 'encapsulation', 'adsorption' and 'separation' during 2017-2020 supports this fact. The most extensive work on MOF DDS systems was executed in 2017-2022, being maximized in 2020. The latter period of research on MOF drug delivery was dominated by 'metal organic

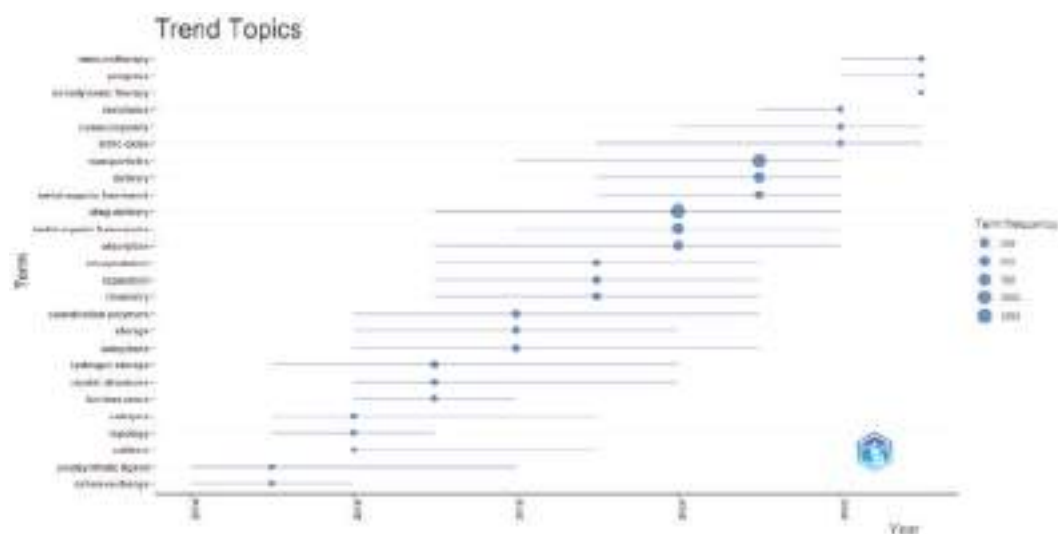


Figure 8. Trend topics analysis for the period 2014-2023.

frameworks, 'nanoparticles', and 'delivery', suggesting utilization of nanoparticles in the latter period of research for delivering of drug molecules. At the more advanced period, topics like 'immunotherapy', 'sonodynamic-therapy' were found to dominate being maximized in 2023, as modern therapeutic treatments of cancer have been inclined towards designing novel systems utilizing new MOFs different therapies were targeted utilizing the MOFs.

Sonodynamic therapy (SDT) is considered as the most advanced therapeutic treatment in cancer treatment owing to its non-invasiveness along with minimum side effects. In SDT, low intensity ultrasound is given in presence of sono-sensitizers. This sort of treatment is employed in treating deep lesions associated with its efficiency to penetrate into the deep tissues. The advantages, such as, large surface area, high porosities, good drug loading capacity and their controlled release, bio-compatibility, and precise targeting capacities make the MOFs efficient in cancer therapeutics [35].

In recent decades, immunotherapy has emerged as a prominent treatment for cancer [36-38]. Immunotherapy involves boosting of natural defence of immune systems in contrary to inhibiting tumor growth in traditional processes of cancer treatment. By alleviating tumour-induced immunosuppression, it triggers an immune response that not only suppresses primary tumour growth but also impedes metastasis and prevents tumour recurrence. MOFs can be well utilized in cancer immunotherapy owing to their large surface area, porous structure and excellent biocompatibility.

Analysis of Collaboration Between Countries, Interlinks Between Sources and Co-Occurrences Between Keywords

Collaboration between countries

The collaboration between different countries, as obtained from VOS viewer, has been represented in Figure

9. In this figure, the frequency of keywords is indicated by the volume of the bubbles, and the intensity of the line is directly related to the number of articles published in that inter-country collaboration. In addition, the distance between the countries defines their relatedness in the present investigation. As can be seen, Peoples R China published the highest number of articles, and found to possess highest link strength (1186). The intensity of the links again suggests collaborative strength between two countries. As evident from the plot, China possesses highest number of links with highest frequency of publication. Countries like, Iran, Australia, India, Singapore are found to publish highest number of articles with Peoples R China. USA followed China in collaborating research with other countries. An appreciably good number of international collaborating articles is well evident from the plot. USA is found to yield good number of articles on collaborating with countries like, China, Iran, India, Australia, England, South Korea etc. This trend was followed by Australia, Iran, and India in publishing articles on MOF drug delivery during 2014-2023. This study reveals the extent of international collaboration by the countries worldwide in the MOF drug delivery research. China tops the list followed by, USA, Australia, Iran, and India.

The density visualization has proved itself to be more promising in evaluating inter-country collaborations. Figure 10 displays the density plot related to inter-country collaboration in MOF drug delivery research. The collaboration between China and USA is well evident from the given density visualization. The density plot also predicts good collaborations between Japan-India, India-South Korea, South Korea- Taiwan, and Australia-Iran. Additionally, a mention-worthy collaboration between USA-Saudi Arabia, Spain-England, and Poland-Brazil is evident from the density visualization.

The development of MOFs utilizing metals and ligands involve researchers from chemistry fraternity, while the

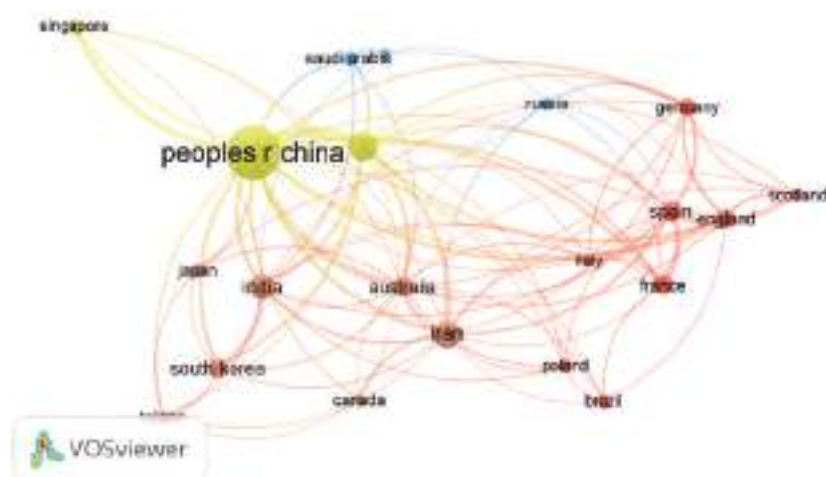


Figure 9. Collaborative networks between different countries in MOFs in drug delivery research during 2014-2023.

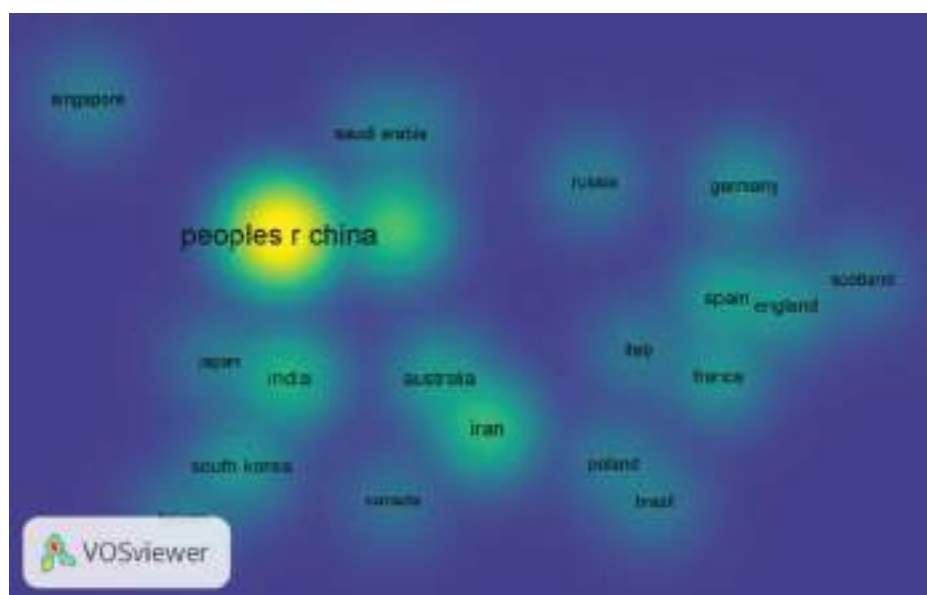


Figure 10. Density visualization of collaborations between different countries in MOFs in drug delivery research during 2014-2023.

applications of the MOFs in drug delivery requires sophisticated biology laboratory. Thus, the present research topic appears as a multidisciplinary field involving chemists, as well as, biologists. A significant number of articles has been published collaborating chemists from structural laboratory with biologists involving in biomedical research. For instance, USA-China collaboration is evident for on command cargo release of zirconium based MOF [39]. Another China-USA joint venture is evident in phototheranostic MOFs for efficient imaging-guided cancer therapy [40].

Thus, the field merges Chemistry with Biology, resulting good collaborations between structural chemists with cell biologists.

Interlinks between sources

Figure 11 depicts the interlinks between different sources publishing articles on MOF drug delivery in the considered time frame. Here, the interlinks suggests the citations by the different sources. As evident from the figure, the size of the bubble is maximum in case of 'ACS Applied Materials and Interfaces', signifying highest number of articles from

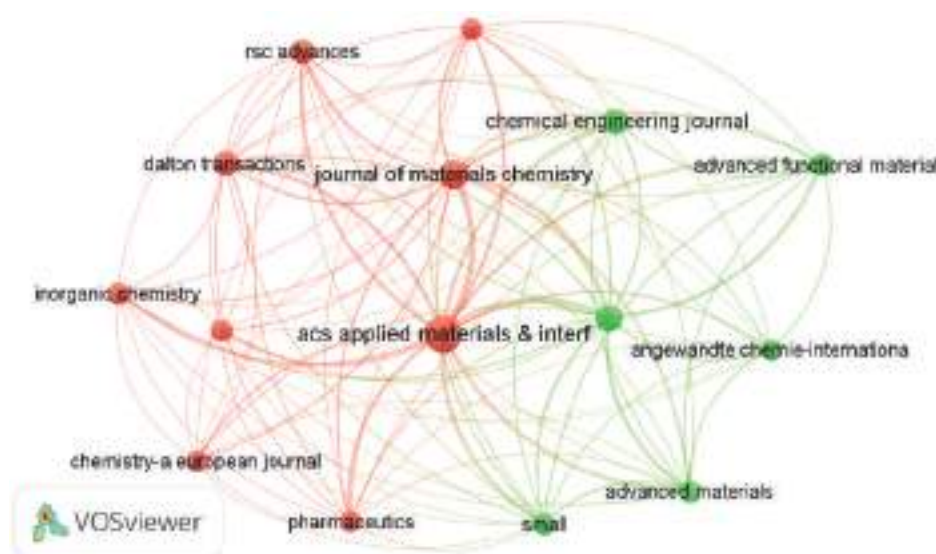


Figure 11. Interconnecting networks between sources publishing articles in MOFs in drug delivery research.

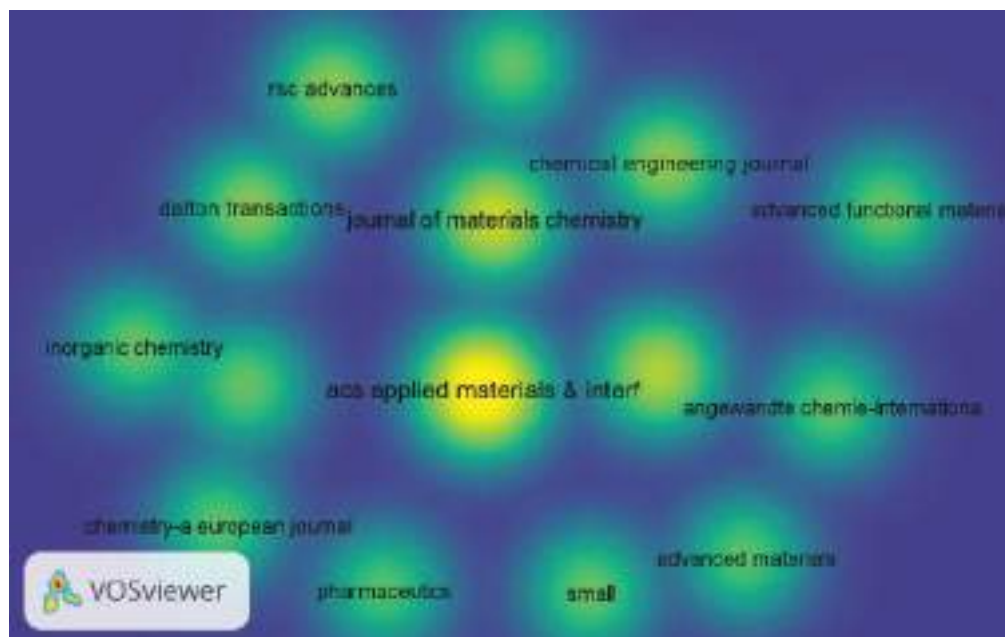


Figure 12. Density visualization between sources publishing articles in MOFs in drug delivery research.

this specific source got cited by other journals. This was followed by sources like, 'Journal of Materials Chemistry B', 'Advanced Materials', 'Advanced Functional Materials', 'Dalton Transactions', and 'Chemical Communications'. As the number of articles were greater in case of these sources the citation links were also found to be high. It is worthy to mention that, although 'Coordination Chemistry Reviews' published significantly high number of articles, the citation links are found to be less and thus has not appeared in the given plot.

Figure 12 represents the density visualization of sources publishing articles on MOF drug delivery systems. As revealed from the density visualization, no strong links between the sources is evident. However, a very weak link is observable between 'Inorganic Chemistry' and 'Microporous and Mesoporous Materials' followed by 'Dalton Transactions' and 'RSC Advances'.

Co-occurrence of keywords

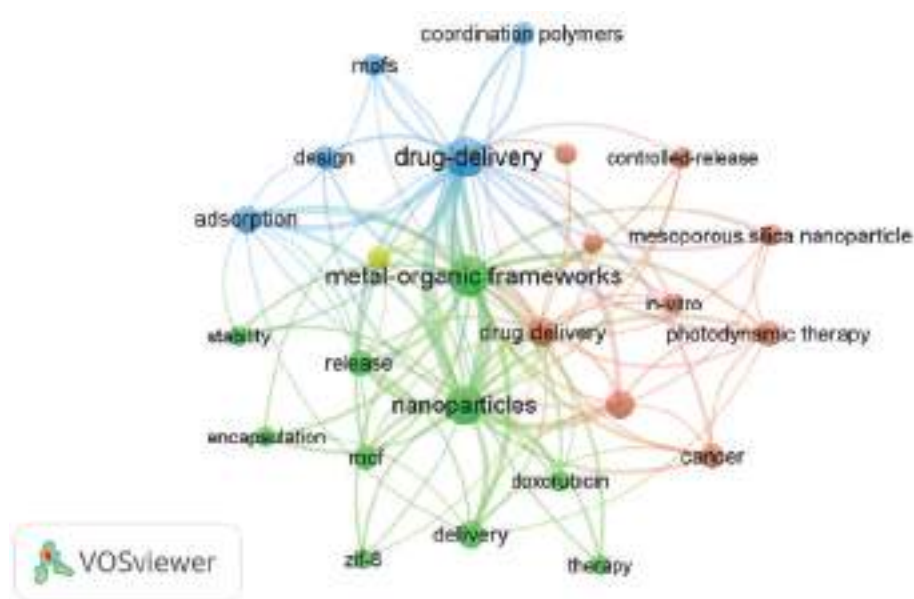
The co-occurrences of different keywords associated with the present study have been plotted in Figure 13. The corresponding parameters related to the co-occurrences of keywords have been summarized in Table 1. The co-occurrences of keywords suggest its association with other keywords. As evident, keywords like, 'metal-organic frameworks', 'drug-delivery', 'nanoparticles', 'drug delivery', 'metal-organic framework', 'delivery' had the most occurrences and total link strength. As the present study involves the quantitative evolution of MOF drug delivery research in the last decade, these keywords are very much obvious and co-occurrences with other keywords. As different MOFs has been developed as the carrier of nanoparticles, as well

as, different MOFs of nano-dimension has been designed for efficient drug delivery, the 'nanoparticles' keyword appears in the topline. This was followed by keywords like, 'adsorption', 'release', 'photodynamic therapy', 'cancer', 'design', 'platform', associated with their close relation with new drug delivery systems for cancer therapeutics. A few materials related keywords, such as, 'mesoporous silica nanoparticles', 'ZIF 8', 'doxorubicin', 'coordination polymers' owing to their extensive investigation in the MOF drug delivery systems. On summary, 'nanoparticles' appears as the most associated keywords in addition to the most MOF drug delivery related keywords. In addition, few materials, as well as, investigation related keywords are found to possess mention worthy occurrences and links.

Recently, there has been significant exploration into utilizing MOFs as host matrices for integrating a variety of functional materials, including nanoparticles, bio entities, and polymers. These combinations result in composites with enhanced or entirely new properties compared to their individual components. Within these MOF composite materials, both the size of the MOF hosts and the guest functional materials can be further reduced to the nanoscale, resulting in metal-organic framework nanocomposites. These MOF nanocomposites exhibit improved compatibility for bio-circulation and serve as efficient nanocarriers for delivering agents in biomedical applications. By incorporating various functional nanoparticles such as Au, Fe_3O_4 , quantum dots, and up conversion nanoparticles into MOF structures, these nanocomposites retain the advantageous properties derived from the crystalline and porous structures of MOFs, while also maintaining the unique biomimetic catalytic, optical-electrical, and

Table 1. Co-occurrence of most relevant keywords and evolution in biodegradable smart and active packaging research during 2014-2023

Keyword	Occurrences	Total link strength
Metal-organic frameworks	1343	2967
Drug-delivery	1317	2862
Nanoparticles	1065	2829
Drug delivery	488	1405
Metal-organic framework	484	1084
Delivery	408	917
Adsorption	387	969
Release	327	975
Photodynamic therapy	295	826
Cancer	259	750
MOF	238	613
Design	227	581
Platform	224	671
Mesoporous silica nanoparticles	207	525
MOFs	200	474
Coordination polymers	197	491
In-vitro	182	520
Encapsulation	168	509
Metal-organic-framework	164	440
Therapy	164	476
Doxorubicin	161	501
Stability	159	466
Controlled-release	158	471
One-pot synthesis	154	476
ZIF-8	151	408

**Figure 13.** Network visualization of co-occurrence of keywords mentioned in the articles involving MOFs in drug delivery research.

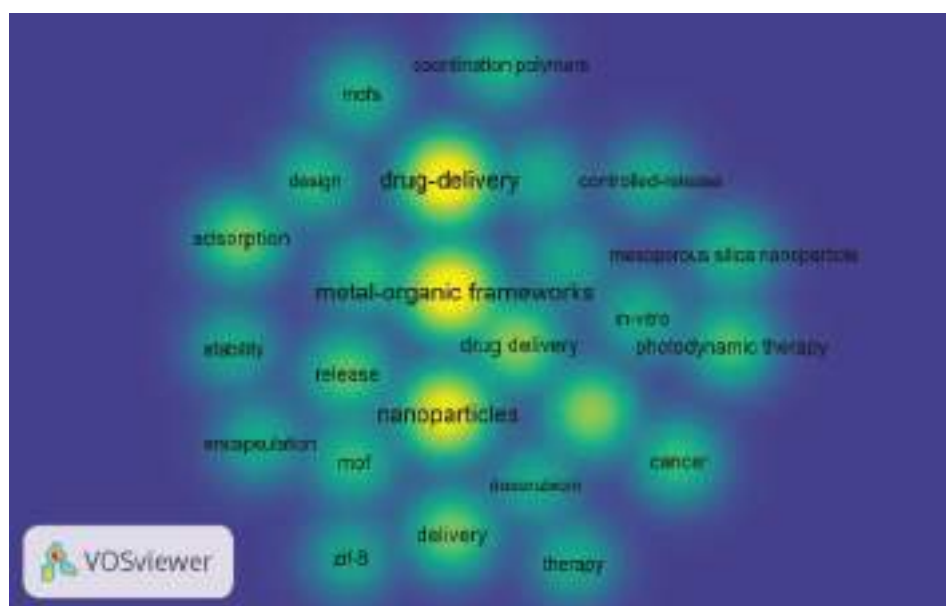


Figure 14. Density visualization of co-occurrence of keywords mentioned in the articles involving MOFs in drug delivery research.

magnetic characteristics of the nanoparticles. Furthermore, the synergistic effects resulting from the integration of these materials can lead to the emergence of novel chemical and physical properties. Many reports have also reviewed about MOF composites for bio-applications such as drug delivery, bio-imaging, and cancer therapy [40–43].

Figure 14 displays the density visualization of co-occurrences of keywords. The close association of ‘metal organic framework’ and ‘drug delivery’ is very much evident from the density plot. This result is quite obvious with the present field of investigation. In addition, ‘release’, ‘mof’, ‘nanoparticles’ are found to have fair overlap, suggesting the exploration of MOFs in nano-dimension, as well as, MOF bound nanoparticles for controlled release of drug molecules. To a very lesser extent, the association of ‘in-vitro’ and ‘photodynamic therapy’ might be associated with the greater in-vitro experiments involving photodynamic therapy employing MOFs.

KEY FINDINGS AND FUTURE SCOPE

On summarizing the results associated with the computational analysis the key findings can be depicted as below:

- The MOF drug delivery research lies in an inclining trend during 2014–2023, suggesting the enhanced interest of worldwide scientific communities in exploring new MOFs for efficient drug delivery
- Worldwide participation in the research involving biomedical application MOF is well evident, China being the highest publishing country with highest international collaboration.

- The country wise production of articles over time displays China being dominating the field in the entire period considered in this study, followed by USA, Iran, India and Spain
- All the research articles were mostly published on journals like, ‘ACS Applied Materials and Interfaces’, ‘Journal of Materials Chemistry B’, ‘Coordination Chemistry Reviews’, ‘Dalton Transactions’, ‘Journal of the American Chemical Society’, ‘Biomaterials’ etc., all highly reputed journals of chemistry.
- The keywords analysis finds ‘Nanoparticles’ as the most mentioned keywords in addition to MOF drug delivery related keywords. The exploration of MOFs as carrier of ‘Nanoparticles’ for efficient drug delivery, as well as, experiment with MOFs with nano dimension results the appearance of ‘Nanoparticles’ in the priority list.
- Immunotherapy, sonodynamic therapy appears as the most recent topic of investigation in the trend topics analysis
- The collaboration study reveals extensive collaboration between structural chemists and cell biologists in developing efficient MOF based drug delivery systems

For efficient drug delivery system, the effect of pH, stability of MOFs in biological medium, loading of drug molecules by controlling porosity must be taken into consideration. Although, there is extensive investigation on the exploration of MOFs in biomedical field, little stress has been given in the toxicity related parameters associated with the drug delivery systems. In addition, there is very less work on the in-vivo experiments, which are very crucial for biomedical applications. In vivo experiments assess biological responses in living organisms, providing

crucial insights into drug absorption, distribution, and safety. They are essential for evaluating the efficacy of therapies before clinical trials, offering information that in vitro studies cannot provide. Moreover, environment friendly green methodologies have gained immense attention in the recent days owing to the sustainability of the environment. This approach often utilizes renewable resources and safer solvents, enhancing sustainability in MOF production. By employing techniques such as microwave-assisted synthesis and hydrothermal methods, researchers can efficiently create MOFs with reduced environmental impact while maintaining high functionality. Environment benign procedures are not much explored and thus have created a lacuna in the field of MOFs for efficient drug delivery.

The incident radiation on a participating media undergoes two phenomena: attenuation and augmentation. Radiation attenuation comes from the effects of absorption and out-scattering (the deviation of the incident radiation to a direction other than the incident direction). On the other hand, radiation augmentation includes the effects of the re-radiation (emission) and in-scattering (increasing the scattering in the direction of the incident radiation). These two phenomena (attenuation and augmentation) are explained by the radiative energy source, which indicates the absorption behavior of a participating media.

Figure 6 shows the radiative energy source which is generated in the volume of participating media (nanoparticulate media). The boundaries of the volume are specified as follows: the upper face is a radiant source (1400 W/m^2), the lower face is adiabatic and reflects the incident radiation inside the volume of the nanosuspension, the side surface is adiabatic surfaces. The main objective is to observe the effect of the nanoparticles on the radiative transfer phenomena and to clarify the effect of particle concentration on the radiative heat generation in the participating media. The generated radiative energy inside a volume of participating media is one of the thermal applications which are desired for different engineering and industrial applications. The spectral properties of the participating media play a significant role in these applications.

CONCLUSION

The utilization of metal organic frameworks in designing novel drug delivery systems has become a field of immense interest towards the worldwide research community in the last decade. This computational review has quantitatively estimated all the research works done during the span of ten years (2014-2023). A total 3014 articles were resulted from the Bibliometrix-Biblioshiny package of R with the data retrieved from Web of Science Database related to the utilization of metal organic frameworks in drug delivery. A year-wise growth in the number of articles was evident, suggesting the growing interest of the research community towards developing new drug delivery systems using metal organic frameworks. In the global participation of the present

research, China was found to produce the highest number of articles, followed by Iran, USA, India, and Spain. In addition, the year-wise growth, as well as, inter-country collaboration was found to be maximized in case of China. On analyzing the sources of articles, 'ACS Applied Materials and Interfaces' was found to publish highest number of articles, followed by, 'Journal of Materials Chemistry B', 'Coordination Chemistry Reviews' and 'Dalton Transactions', all good international journals of chemical sciences. It is worthy to mention that, although 'Journal of American Chemical Society' published a smaller number of articles on MOF drug delivery research, the citation parameters were found to be highest in the articles published in this journal, signifying greater acceptability of their results towards the world-wide scientific community. Additionally, the analysis of top twenty most cited articles revealed that these articles were published in journals like 'Advanced Materials', 'Journal of American Chemical Society', 'Angewandte Chemie International Edition', 'Chemical Communications' etc.; all are high impact journals of international repute. The analysis of the keywords was executed in terms of frequency and year-wise growth. In addition to most obvious keywords, like, 'metal organic frameworks', 'mof', 'drug delivery', 'delivery', 'release' etc., 'nanoparticle' appeared as the most frequent unique keyword in our considered time frame. The utilization MOFs as nano carriers for drug molecules, as well as, designing novel drug delivery systems utilizing the MOFs might appear to be promising to result 'nanoparticles' as highly mentioned keywords. Moreover, the analysis of trend topics revealed new topics like 'immunotherapy', 'sonodynamic therapy', signifying the enhanced interest of modern therapeutic treatments for cancer employing MOFs. On analysing collaborative networks between countries, the highest collaboration was found to be in the articles from Peoples R China was found to possess greatest number of collaborations. The collaborations of Peoples R China with countries like USA, Australia, Iran, India, Singapore etc. is well established from the network structure. ACS Applied Materials and Interfaces, a very renowned journal, was found to publish highest number of articles with mention worthy citation links. The co-occurrences of keywords revealed 'nanoparticles' to be the most mentioned and most linked keywords in addition to the most obvious keywords. In spite of extensive work, little stress has been put in the in-vivo experiments, as well as, the environmental sustainability in synthesizing the MOFs. Thus, this computational analysis is expected to offer researchers a directed path in their early carrier research on metal organic frameworks in biomedical field.

ACKNOWLEDGEMENTS

The authors are thankful to Principal, Ramakrishna Mission Vivekananda Centenary College, Rahara for providing necessary facilities related to this computational analysis.

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] Helfand WH, Cowen DL. Evolution of pharmaceutical oral dosage forms. *Pharm Hist* 1983;25:3–18.
- [2] Park K. Controlled drug delivery systems: past forward and future back. *J Control Release* 2014;190:3–8. [\[CrossRef\]](#)
- [3] Li C, Wang J, Wang Y, Gao H, Wei G, Huang Y, et al. Recent progress in drug delivery. *Acta Pharm Sin B* 2019;9:1145–1162. [\[CrossRef\]](#)
- [4] Horcajada P, Gref R, Baati T, Allan PK, Maurin G, Couvreur P, et al. Metal–organic frameworks in biomedicine. *Chem Rev* 2012;112:1232–1268. [\[CrossRef\]](#)
- [5] Moghadam PZ, Li A, Wiggin SB, Tao A, Maloney AG, Wood PA, et al. Development of a Cambridge structural database subset: a collection of metal–organic frameworks for past, present, and future. *Chem Mater* 2017;29:2618–262. [\[CrossRef\]](#)
- [6] Silva P, Vilela SM, Tome JP, Paz FAA. Multifunctional metal–organic frameworks: from academia to industrial applications. *Chem Soc Rev* 2015;44:6774–6803. [\[CrossRef\]](#)
- [7] Horcajada P, Serre C, Vallet-Regí M, Sebban M, Taulelle F, Férey G. Metal–organic frameworks as efficient materials for drug delivery. *Angew Chem Int Ed* 2006;45:5974–5978. [\[CrossRef\]](#)
- [8] Bernini MC, Fairen-Jimenez D, Pasinetti M, Ramirez-Pastor AJ, Snurr RQ. Screening of biocompatible metal–organic frameworks as potential drug carriers using Monte Carlo simulations. *J Mater Chem B* 2014;2:766–774. [\[CrossRef\]](#)
- [9] Xiao B, Wheatley PS, Zhao X, Fletcher AJ, Fox S, Rossi AG, et al. High-capacity hydrogen and nitric oxide adsorption and storage in a metal–organic framework. *J Am Chem Soc* 2007;129:1203–1209. [\[CrossRef\]](#)
- [10] Lin SX, Pan WL, Niu RJ, Liu Y, Chen JX, Zhang WH, et al. Effective loading of cisplatin into a nanoscale UiO-66 metal–organic framework with preformed defects. *Dalton Trans* 2019;48:5308–5314. [\[CrossRef\]](#)
- [11] He C, Lu K, Liu D, Lin W. Nanoscale metal–organic frameworks for the co-delivery of cisplatin and pooled siRNAs to enhance therapeutic efficacy in drug-resistant ovarian cancer cells. *J Am Chem Soc* 2014;136:5181–5184. [\[CrossRef\]](#)
- [12] Farboudi A, Mahboobnia K, Chogan F, Karimi M, Askari A, Banihashem S, et al. UiO-66 metal–organic framework nanoparticles loaded carboxymethyl chitosan/polyethylene oxide/polyurethane core–shell nanofibers for controlled release of doxorubicin and folic acid. *Int J Biol Macromol* 2020;150:178–188. [\[CrossRef\]](#)
- [13] Chowdhuri AR, Bhattacharya D, Sahu SK. Magnetic nanoscale metal–organic frameworks for potential targeted anticancer drug delivery, imaging and as an MRI contrast agent. *Dalton Trans* 2016;45:2963–2973. [\[CrossRef\]](#)
- [14] Cheng G, Li W, Ha L, Han X, Hao S, Wan Y, et al. Self-assembly of extracellular vesicle-like metal–organic framework nanoparticles for protection and intracellular delivery of biofunctional proteins. *J Am Chem Soc* 2018;140:7282–7291. [\[CrossRef\]](#)
- [15] Wuttke S, Zimpel A, Bein T, Braig S, Stoiber K, Vollmar A, et al. Validating metal–organic framework nanoparticles for their nanosafety in diverse biomedical applications. *Adv Healthc Mater* 2017;6:1600818–1600828. [\[CrossRef\]](#)
- [16] Zhou HC, Long JR, Yaghi OM. Introduction to metal–organic frameworks. *Chem Rev* 2012;112:673–674. [\[CrossRef\]](#)
- [17] Rojas S, Carmona FJ, Maldonado CR, Horcajada P, Hidalgo T, Serre C, et al. Nanoscaled zinc pyrazolate metal–organic frameworks as drug-delivery systems. *Inorg Chem* 2016;55:2650–2663. [\[CrossRef\]](#)
- [18] Chen Y, Li P, Modica JA, Drout RJ, Farha OK. Acid-resistant mesoporous metal–organic framework toward oral insulin delivery: protein encapsulation, protection, and release. *J Am Chem Soc* 2018;140:5678–5681. [\[CrossRef\]](#)
- [19] Horcajada P, Chalati T, Serre C, Gillet B, Sebrie C, Baati T, et al. Porous metal–organic-framework nanoscale carriers as a potential platform for drug delivery and imaging. *Nat Mater* 2010;9:172–178. [\[CrossRef\]](#)
- [20] Lawson HD, Walton SP, Chan C. Metal–organic frameworks for drug delivery: a design perspective. *ACS Appl Mater Interfaces* 2021;13:7004–7020. [\[CrossRef\]](#)

- [21] Wang L, Feng X, Ren L, Piao Q, Zhong J, Wang Y, et al. Flexible solid-state supercapacitor based on a metal–organic framework interwoven by electrochemically-deposited PANI. *J Am Chem Soc* 2015;137:4920–4923. [\[CrossRef\]](#)
- [22] Yu S, Xu K, Wang Z, Zhang Z, Zhang Z. Bibliometric and visualized analysis of metal–organic frameworks in biomedical application. *Front Bioeng Biotechnol* 2023;11:1190654. [\[CrossRef\]](#)
- [23] Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr* 2017;11:959–975. [\[CrossRef\]](#)
- [24] Van Eck N, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84:523–538. [\[CrossRef\]](#)
- [25] Bilir C, Güngör C, Kökalan Ö. Bibliometric overview of operations research/management science research in Turkey. *Sigma J Eng Nat Sci* 2019;37:797–811.
- [26] Sun Y, Zheng L, Yang Y, Qian X, Fu T, Li X, et al. Metal–organic framework nanocarriers for drug delivery in biomedical applications. *Nano Micro Lett* 2020;12:1–29. [\[CrossRef\]](#)
- [27] Cai W, Gao H, Chu C, Wang X, Wang J, Zhang P, et al. Engineering phototheranostic nanoscale metal–organic frameworks for multimodal imaging-guided cancer therapy. *ACS Appl Mater Interfaces* 2017;9:2040–2051. [\[CrossRef\]](#)
- [28] Simon-Yarza T, Mielcarek A, Couvreur P, Serre C. Nanoparticles of metal–organic frameworks: on the road to in vivo efficacy in biomedicine. *Adv Mater* 2018;30:1707365. [\[CrossRef\]](#)
- [29] Mao D, Hu F, Kenry, Ji S, Wu W, Ding D, et al. Metal–organic-framework-assisted in vivo bacterial metabolic labeling and precise antibacterial therapy. *Adv Mater* 2018;30:1706831. [\[CrossRef\]](#)
- [30] Sava Gallis DF, Butler KS, Agola JO, Pearce CJ, McBride AA. Antibacterial countermeasures via metal–organic framework-supported sustained therapeutic release. *ACS Appl Mater Interfaces* 2019;11:7782–7791. [\[CrossRef\]](#)
- [31] Gandara-Loe J, Ortuño-Lizarán I, Fernández-Sánchez L, Alió JL, Cuenca N, Vega-Estrada A, et al. Metal–organic frameworks as drug delivery platforms for ocular therapeutics. *ACS Appl Mater Interfaces* 2018;11:1924–1931. [\[CrossRef\]](#)
- [32] Zhou Z, Vázquez-González M, Willner I. Stimuli-responsive metal–organic framework nanoparticles for controlled drug delivery and medical applications. *Chem Soc Rev* 2021;50:4541–4562. [\[CrossRef\]](#)
- [33] Li H, Li L, Lin RB, Zhou W, Zhang Z, Xiang S, et al. Porous metal–organic frameworks for gas storage and separation: status and challenges. *Energy Chem* 2019;1:100006–100044. [\[CrossRef\]](#)
- [34] Getman RB, Bae YS, Wilmer CE, Snurr RQ. Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal–organic frameworks. *Chem Rev* 2012;112:703–723. [\[CrossRef\]](#)
- [35] Hu W, Wang S, Jiang C, Zheng M, Bai Z, Srivastava D, et al. Recent advances in sonodynamic therapy by MOFs-based platforms for biomedical applications. *Dyes Pigments* 2023;219:111596. [\[CrossRef\]](#)
- [36] Zhang S, Wang J, Kong Z, Sun X, He Z, Sun B, et al. Emerging photodynamic nanotherapeutics for inducing immunogenic cell death and potentiating cancer immunotherapy. *Biomaterials* 2022;282:121433. [\[CrossRef\]](#)
- [37] Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. *Nature* 2011;480:480–489. [\[CrossRef\]](#)
- [38] Rahman MM, Behl T, Islam MR, Alam MN, Islam MM, Albarrati A, et al. Emerging management approach for the adverse events of immunotherapy of cancer. *Molecules* 2022;27:3798–3824. [\[CrossRef\]](#)
- [39] Meng X, Gui B, Yuan D, Zeller M, Wang C. Mechanized azobenzene-functionalized zirconium metal–organic framework for on-command cargo release. *Sci Adv* 2016;2:e1600480. [\[CrossRef\]](#)
- [40] Ge X, Wong R, Anisa A, Ma S. Recent development of metal–organic framework nanocomposites for biomedical applications. *Biomaterials* 2022;281:121322. [\[CrossRef\]](#)
- [41] Rocca JD, Liu D, Lin W. Nanoscale metal–organic frameworks for biomedical imaging and drug delivery. *Acc Chem Res* 2011;44:957–968. [\[CrossRef\]](#)
- [42] Gao P, Chen Y, Pan W, Li N, Liu Z, Tang B. Antitumor agents based on metal–organic frameworks. *Angew Chem Int Ed Engl* 2021;60:16901–16914. [\[CrossRef\]](#)
- [43] Yang J, Yang YW. Metal–organic frameworks for biomedical applications. *Small* 2020;16:1906846. [\[CrossRef\]](#)



Review Article

A review on utilisation of nanofluid in solar energy harnessing

Sanjeev Kumar GUPTA^{1,*}, Vivek RATHORE²

¹Department of Mechanical Engineering, IET, GLA University Mathura, 281406 India

²Engineer, UNO Minda Limited Manesar Gurugram, 122004, India

ARTICLE INFO

Article history

Received: 13 June 2024

Revised: 21 August 2024

Accepted: 03 October 2024

Keywords:

Heat Transfer; Nanofluid; Solar Energy; Stability; Thermal Conductivity; Thermal Efficiency

ABSTRACT

The current global epidemic and population growth, which led to an increase in energy consumption, highlighted the need to maximize solar energy extraction. Because nanofluids have good intrinsic thermal properties, they are employed to improve the effectiveness of solar collectors. One of the primary research themes in solar energy consumption is to boost the effectiveness of the harvesting technologies. Thus, it is essential to examine the performance of solar energy harvesting devices based on nano fluids. Therefore, it is crucial to look at how well nanofluid-based solar energy production systems work. The article includes information on how various properties can be changed to efficiently harvest solar energy, as well as a summary of recent advancements in nanofluid-based solar energy extraction devices. From this review, it is observed that 85.4% enhancement in solar collector utilizing graphite as TES material, 38.4% enhancement in solar water heater using CuO nanofluid, 84.74% enhancement in thermal efficiency of PV/T system using NEPCM and 196% gain in exergy efficiency of solar still using Ag nanofluid.

Cite this article as: Gupta SK, Rathore V. A review on utilisation of nanofluid in solar energy harnessing. Sigma J Eng Nat Sci 2025;43(5):1852–1885.

INTRODUCTION

In the 21st century, finding sustainable and renewable energy sources has become a top priority. The world is facing two big problems: fossil fuel stocks are running out, and using them continues to harm the environment. Now is the time to switch to cleaner, more sustainable energy sources [1–2]. Solar energy is one of the most promising renewable energy sources since it is plentiful, reliable, and can meet a large part of the world's energy needs. The Sun is a huge fusion reactor that sends a huge amount of energy to Earth. This energy is more than enough to meet the world's energy

needs several times over [3]. Even while solar energy systems have a lot of potential, their current efficiency is still a barrier to their widespread use. Because of this inefficiency, a lot of research has been done on novel technologies and materials that can improve the capture and conversion of solar energy [4]. Using nanofluids to collect solar energy is one area of research that looks promising.

Solar energy is one of the best sources of renewable energy because it is almost always available and doesn't directly release greenhouse gases when it is used. Fossil fuels are limited and found in certain areas of the world, but solar

*Corresponding author.

*E-mail address: sanjeev.gupta@gla.ac.in

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



energy is available almost everywhere on Earth [5]. The International Energy Agency (IEA) says that the quantity of solar energy that hits the Earth's surface in only one hour is enough to power the world's economy for a whole year. This shocking number shows how solar energy might become a major source of energy over the world [6]. Also, using solar energy could help make energy more secure, cut down on the need to import fossil fuels, and lessen the negative effects on the environment that come with traditional energy production [7]. Solar power can control the climate change by controlling the emission of carbon dioxide and hence it is most powerful renewable energy source. This is the reason behind investment in solar energy technology. Engineers and scientist can get maximum from solar energy by enhancing its efficiency and effectiveness of solar energy system [8].

The world energy scene is changing a lot because of the growing need for energy, population increase, economic growth, and the need to cut carbon emissions right away. According to IEA, the consumption of global energy will enhance in the coming years. The electricity demand along is raised approximately 60% by 2040 [6]. Figure 1 shows the world's population, the amount of CO₂ in the air, the GDP per person, and the median temperature anomaly. This rising demand is a big problem because most of the world's energy still comes from fossil fuels, which are not only limited but also release a lot of greenhouse gases into the atmosphere [9]. Now the entire world is focused on solar energy to find a balance between energy security and environmental sustainability [10]. The world energy demand can be fulfilled by solar energy without and harmful effect on the environment. This energy can be harnessed by both medium directly using PV system and indirectly by solar

thermal system. Solar power is now more competitive with traditional energy sources since solar technology is improving quickly and prices are going down. But to keep up with the growing demand, solar energy systems need to become even more efficient and useful [11–12].

Nanofluids are made by mixing nanoparticles—very small particles that are around the size of nanometers—into a base fluid like water, oil, or ethylene glycol [13]. You may make these nanoparticles out of metals (like copper and gold), oxides (like aluminium oxide and titanium dioxide), and carbon-based materials (like graphene and carbon nanotube). Nanofluids are very appealing for use in solar energy systems because they have unique qualities, such as a high surface area-to-volume ratio, better thermal conductivity, and the ability to change their optical properties [14–15]. Nanofluids can greatly increase how well solar thermal systems absorb and transport solar energy [16]. Nanoparticles in the fluid make it better at conducting heat, which makes heat transmission more efficient. Also, some nanoparticles can be designed to absorb certain wavelengths of sunlight better, which means the device can capture more energy [17].

This better thermal performance could make solar collectors work better and lessen the cost of making solar thermal energy. Nanofluids can be utilised as a cooling medium in photovoltaic systems to get rid of the heat that PV cells make. Too much heat can make PV cells less effective and shorten their lifespan, therefore cooling them down is quite important. Due to better thermal characteristics of nanofluid, it can assist PV cell properly which make the entire system more effective and efficient [18]. Researchers and people in the business world are very interested in nanofluids

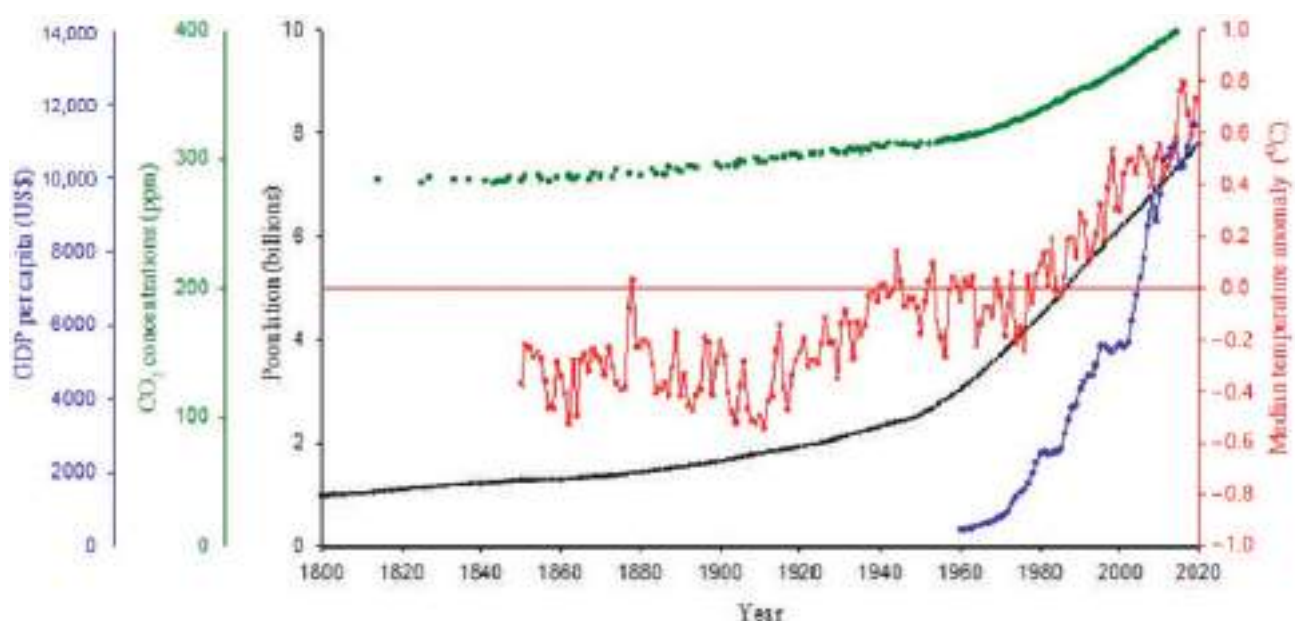


Figure 1. World population, CO₂ concentration, GDP per capita and Median Temperature Anomaly [From Holechek et al. [1], with permission from MDPI].

because they could make solar energy systems work better. Many studies have shown that nanofluids can be used in a variety of solar applications [19–21]. Researchers and scientist are still searching best combination of nanomaterial which can make the system more efficient. Utilization of nanofluid in solar energy harnessing devices can make the solar system more efficient and also fulfill the world energy demand without any harmful impact on the environment.

Already Known

- Nanofluids are colloidal suspension different nanomaterials in a base fluid. Nanofluid offers better thermophysical properties compared to traditional fluid.
- Nanofluid is most widely used in solar harnessing devices because of its better heat transfer ability.
- Energy conversion in solar system has become more effective due to absorption capacity of solar radiation of nanoparticles present in the nanofluid.
- Most of researches are performed on nanoparticles such as Al_2O_3 , TiO_2 , MWCNT, etc.
- Researchers are focused on efficiency, effectiveness and stability and cost of nanofluid utilization in solar energy harnessing devices.

Gap that Need to be Addressed

- Stability of nanofluid is the major concern when it is used in different environmental condition.
- The utilization of nanoparticles in solar energy devices is very hazardous for health and its long term utilization can also affect the environment.
- Only limited research are available on amount of nanoparticles used in the preparation of nanofluid which can cause problem of higher viscosity and clogging.
- Commercial utilization of nanofluid in solar energy devices to harness solar energy is limited.
- Utilization of nanofluid in solar system is not cost effective.
- Hybrid nanofluid is now used in lab and real life solar system in place of mono nanofluid because of its enhanced thermophysical properties.

Novelty

The novelty of this review article lies in the development of new nanofluid and its utilization in its solar energy

harnessing devices. In review article different types of solar system are studies with commercial utilization of nanofluid in the solar system. This review article also focuses on the different way to enhance its long term stability of nanofluid and wide application in solar energy devices.

Objectives and Scope of Study

The primary objective of this review article is to summarize the recent development and utilization of nanofluid in solar energy harnessing systems. In this review article, the development, progress, possible utilization of nanofluid in different solar system is discussed in details along with their commercial usage and limitations. Also this article helps to understand how different nanofluid used in different solar application to optimize system efficiency and effectiveness to use them as an alternate source of renewable energy. The objective covered is as follows:

- Analysis of different types of nanoparticles and base fluid along with their thermophysical properties and characteristics.
- Analysing the behaviour of different types of nanofluid in different solar applications to optimize the system efficiency and effectiveness.
- Exploration of nanofluids in PV/T systems, with a focus on how they can help photovoltaic cells cool down and produce more electrical and thermal energy.
- A summary of computer models, simulations, and experimental studies that look at how nanofluids act in solar energy systems, taking into account things like flow dynamics, heat transfer, and system efficiency.
- There will be a discussion of the real-world problems that come up when using nanofluids, such as stability, cost, scalability, and environmental impact. There will also be a list of possible research directions, technological advances, and new uses that could make nanofluids even better at collecting solar energy.

Table 1 shows how this study's methodology is different from that of previous studies. The authors believe that no other study has looked at these problems in such depth. Hamzat et al. [22] did a similar study; however this one includes the most recent developments, difficulties, and trends in the field since it was published.

Table 1. Description of Earlier Published Review Paper on Similar Topic after 2018

Ref.	Review Detailed
[23], 2018	This study only focused on solar powered membrane distillation for enhancement of energy using nanofluid.
[24], 2019	This is limited to renewable system such as hydrogen production, biofuel and geothermal for energy harnessing using nanofluid
[25], 2020	This review only focused to enhance solar collector efficiency using nanofluid
[26], 2020	This review focused on only harnessing of solar thermal energy using nanofluid
[27], 2021	This research only focused on harnessing solar thermal energy using hybrid nanofluid
[28], 2022	Exclusively solar energy harvesting systems for photovoltaic self-powered applications were reviewed.
[22], 2022	An overview of current developments in nanofluid-based solar energy harvesting systems is provided in this article.

NANOMATERIALS AND NANOFLUID

Nanomaterials are materials with structural components smaller than 100 nanometers, typically consisting of atoms, molecules, or particles at the nanoscale. At this size, materials often have physical, chemical, and biological properties that are very different from those of their bulk

equivalents. This is because the surface area is larger, quantum effects are stronger, and surface phenomena are more important [29-30]. Detailed classification of nanomaterials is shown in Fig. 2 and Fig. 3.

There are two main types of nanomaterials: organic and inorganic. Each nanomaterial has its own characteristics. The nanomaterials made by carbon molecules such

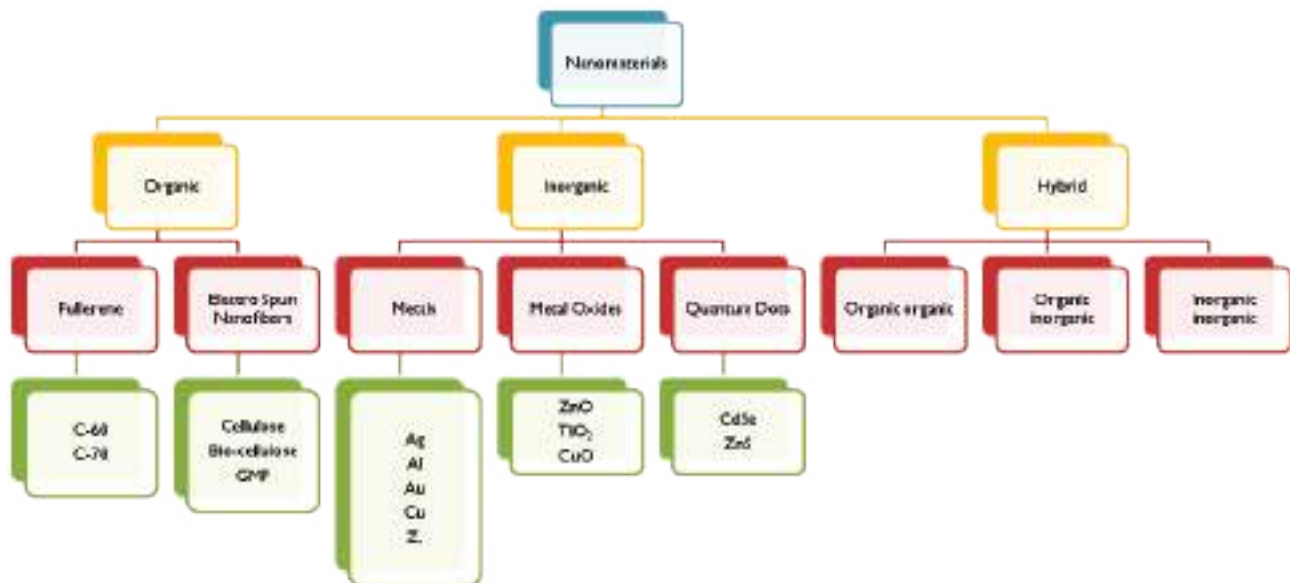


Figure 2. Various Types of Nanomaterials.

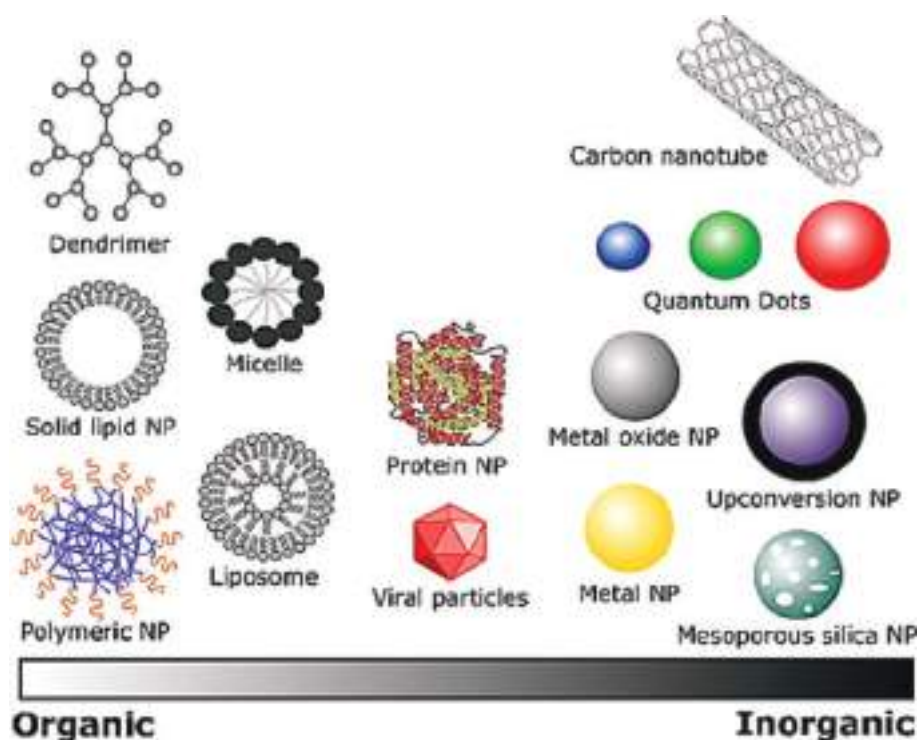


Figure 3. Organic and Inorganic nanoparticles compositions [From Bodunde et al. [31], with permission from Elsevier].

as polymer or organic compound are known as organic nanomaterials. Organic nanomaterials are most widely used in solar cells, optoelectronics, sensors, biomedicines like drug delivery and many more. However, inorganic nanomaterials do not contain any carbon. They are made by metals, metals oxide and ceramics. Such materials possess good strength, electrical conductivity and also stable at very high temperature. These characteristics are very useful for their utilization in electronic industry, solar system and also cleaning the environmental pollution. The materials which contain organic and inorganic material are known as hybrid nanomaterials which have the characteristics of both. So hybrid nanomaterials are versatile and useful in solar applications [31–32].

Nanofluid are the mixture of tiny nanomaterials (1-100 nm) in the base fluid like water, ethylene glycol etc having better heat transfer characteristics. Nanomaterials such as CNT, quantum dots, metal oxides etc. possess high heat transfer and thermophysical properties [33]. Nanofluids were first thought of in the mid-1990s as a new way to improve heat transfer efficiency. After the discovery of nanomaterials, they have got attention of researchers because it makes the system more effective and [34]. The various applications of nanofluid are depicted in Fig. 4.

Nanofluids have better thermal characteristics than regular fluids because the nanoparticles have a large surface area and excellent thermal conductivity, which makes energy transmission more efficient. Nanoparticles make the surface area accessible for heat exchange more effective, which speeds up the processes of heat transmission. Also, the Brownian motion of nanoparticles causes micro-convection in the fluid, which makes thermal conductivity even better [35]. Studies have shown that even a small amount of nanoparticles can make a big difference in how well something conducts heat. This makes nanofluids very useful for things like cooling systems in cars, electronic devices, and renewable energy systems like solar collectors [36].

In renewable energy system, nanofluid has got special attention since it makes renewable energy system more effective and reliable. In solar collector, nanofluid absorbs the solar radiation and converts it into thermal energy more prominently [37]. The direct conversion of energy makes solar collector more effective compared to other heat transfer fluids. Nanofluids are also used in nuclear reactors for cooling purpose to make reactor more efficient and work better. [38].

Nanofluids offer many advantages and utilizations in various solar energy systems along with possible problems of clogging and stability. Main concern about nanofluid is

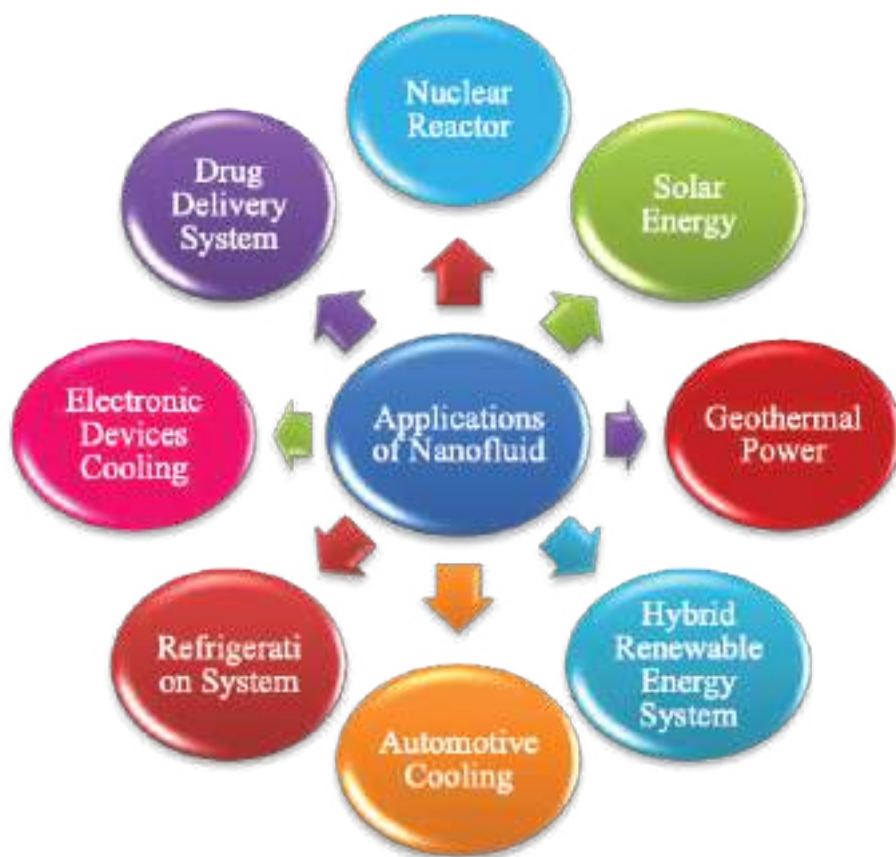


Figure 4. Applications of Nanofluid.

its stability over time since nanoparticles settled together after certain period of time and clogging created and ultimately lose their thermal characteristics [39]. Advance synthesis and characterization techniques are required to make nanofluid stable and its utilization in real life applications. The high cost of nanomaterials also create problem in mass production and its industrial utilization [40].

Researchers are mainly focused on Newtonian base fluid while experimentation on nanofluid since its viscosity does not affected with shear deformation. In case of non-newtonian fluids viscosity affected with shear deformation which cause problem clogging and stability of nanofluid. You can typically find these fluids in more specialized or industrial uses, including in biological fluids, slurries, or some types of lubricants. The behavior of nanofluids in non-Newtonian fluids can be more complicated to predict because the interactions between nanoparticles and the non-linear viscosity of the base fluid can result in complex flow characteristics.

Research on non-Newtonian nanofluids is less common, but it is expanding. This is because learning more about these interactions could lead to new uses for non-Newtonian fluids in areas where they are common. But the difficulty of modeling and testing non-Newtonian nanofluids

has probably caused researchers to focus on Newtonian fluids in the beginning phases of nanofluid research.

PROPERTIES OF NANOMATERIALS (NANOFLUID)

The size, shape, concentration, and material composition of the nanoparticles have a big effect on the properties of nanofluids. Figure 5 shows some of the different properties of nanofluid. One of the most interesting things about nanofluids is that they have better thermal conductivity than regular fluids, which makes them very good for transferring heat [41]. Nanofluids also have better viscosity and heat capacity, which can be changed by changing the concentration of nanoparticles. Another important attribute is their stability; well-dispersed nanoparticles stop aggregation and settling, which keeps the thermal performance steady [42]. Nanofluids can also have different optical, magnetic, and electrical properties depending on the nanoparticles utilized. This makes them useful for many different industrial uses, such as solar energy harnessing and cooling systems. Different properties of nanofluids used by different researchers are summarized in Table 2.

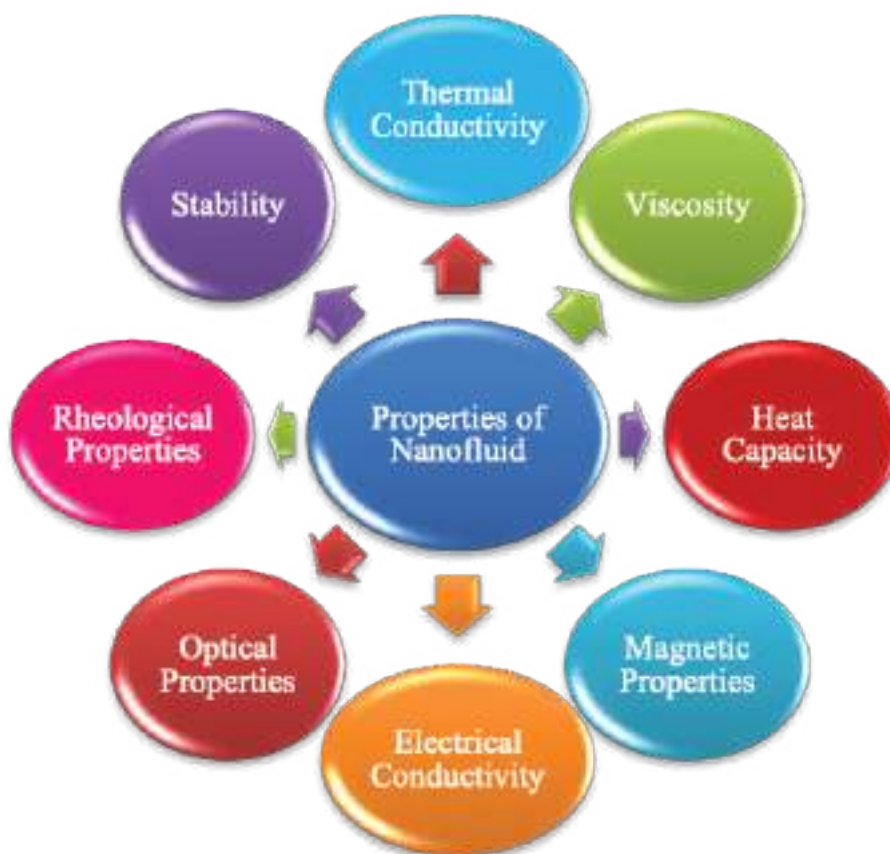


Figure 5. Different Properties of Nanofluid.

Table 2. Earlier Researchers Investigation on Nanofluid Properties

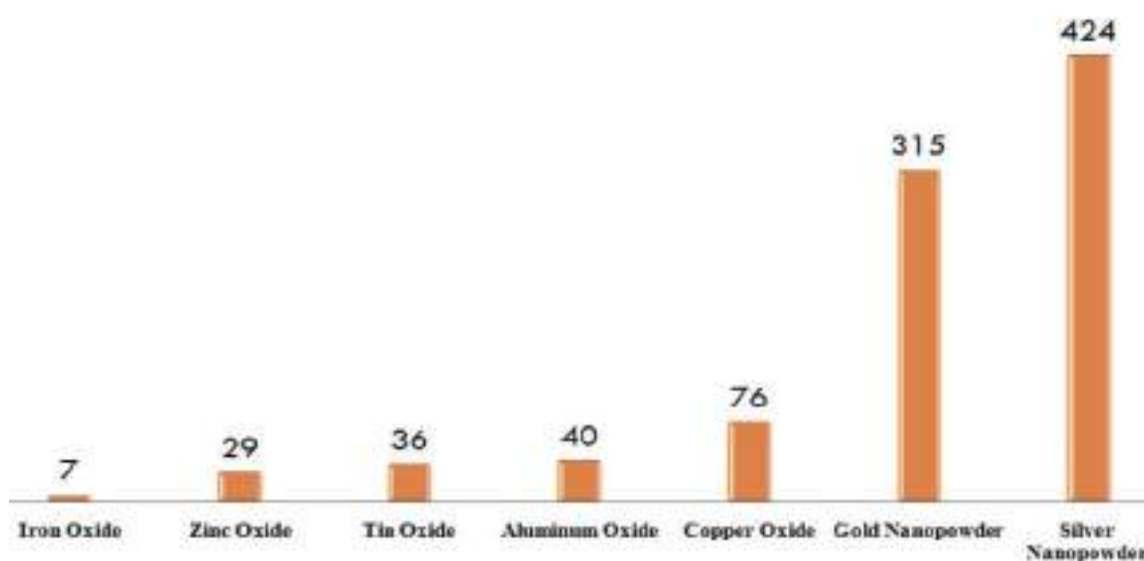
Ref. No.	Nanofluid	Properties of Nanofluid	Remark
[46]	Al ₂ O ₃ /H ₂ O	Thermal conductivity	Maximum boost of 10.1% at the 0.15 weight percent nanoparticle suspension
[47]	Ti ₃ C ₂ Tx/water	Thermal conductivity	30.6% increase TC due to Brownian movement
[51]	Al ₂ O ₃ /H ₂ O	Viscosity	2.3 times rise in viscosity was observed at 5% volume concentration
[52]	Ti ₃ SiC ₂ /propylene glycol	Viscosity	A 0.2 weight percent inclusion of MXene resulted in a viscosity drop
[55]	Ti ₃ C ₂ -soybean oil	Specific heat capacity	The addition of MXene caused a noticeable rise in the heat capacity.
[56]	CuO/ethylene glycol	Specific heat capacity	Maximum 6% increase in SHC
[61]	Cu-Zn/ vegetable oil, paraffin oil, and SAE oil	Stability	All of the samples displayed increased stability for over 72 hours.
[62]	Ti ₃ C ₂ -H ₂ O	Stability	Prolonged storage periods were associated with increased aggregation in the stability of CTAB specimens

Thermal Conductivity

Nanofluids, which are fluids that contain nanoparticles, have a very important property called thermal conductivity that helps solar energy systems work better. Nanofluids have better heat conduction than regular fluids because the nanoparticles have a larger surface area and better heat transfer capabilities [43]. This better thermal conductivity makes solar collectors better at absorbing and transferring heat, which makes the process of turning solar energy into thermal energy more efficient. Researchers want to improve the thermal performance of nanofluids by finding the best kind, concentration, and dispersion of nanoparticles. This makes them good candidates for improved solar

energy systems [44,45]. The thermal conductivity of different nanomaterials is shown in Fig. 6.

The impact of pH value fluctuations on the thermal conductivity (TC) of Al₂O₃/H₂O nanofluid was investigated by Zhu et al. [46]. They found that varying SDBS surfactant concentrations in nano-suspensions and pH values had a significant impact on the degree of stability and improvements in TC of Al₂O₃/H₂O nanofluids. By adding optimal SDBS dispersing agent, the TC increased. To increase the TC, it was suggested to combine the pH and chemical surfactant treatments. Ti₃C₂Tx/water TC was studied by Mao et al. [47] at various loading fractions ranging from 0 to 0.5 wt%. It was discovered that at 0.5 weight percent, there was the greatest improvement in TC, measuring 0.8431 W/m.K.

**Figure 6.** Thermal conductivity of different nanomaterials (W/mK).

The big basal plane of mXene and its Brownian movement in water both contributed to a 30.6% increase.

Viscosity

Nanofluids are fluids that have nanoparticles in them. Their viscosity is an important aspect in deciding if they can be used to collect solar energy. Adding nanoparticles to the fluid can greatly improve its heat conductivity, but it can also change the fluid's viscosity. Higher viscosity can mean that more power is needed to pump the fluid, which could cancel out the advantages in thermal efficiency [48]. To employ nanofluids effectively in solar energy systems, it is important to find the right balance between better thermal conductivity and viscosity that is easy to handle. To get the viscosity just right, researchers are looking at the size, shape, concentration, and composition of nanoparticles. This will make sure that heat is transferred efficiently and that the system doesn't lose too much energy. Figure 7 show that several factors changed the viscosity of the nanofluid.

Nanofluids usually get thicker as there are more nanoparticles in them because the particles interact with each other and with the base fluid more. Smaller nanoparticles usually make the fluid thicker because they have a larger surface area compared to their volume, which makes them interact with the fluid more. Irregularly shaped or long nanoparticles can also make liquids thicker than spherical nanoparticles because they impede flow more

[49]. With increase of temperature, viscosity of nanofluid decreases, but this change totally depends upon the type of nanoparticle. At higher temperatures, the thermal motion of particles may overcome the interparticle forces, reducing viscosity [50].

In a laboratory investigation, Chandrasekhar et al. [51] investigated the viscosity of $\text{Al}_2\text{O}_3/\text{H}_2\text{O}$ nanofluids. The outcome showed that as the volume fraction of nanoparticles rises, so does the viscosity of nanofluids. At 5% volume fraction, a 2.3-fold increase in viscosity was noted in comparison to the base fluid in this study. Mahesh et al. [52] calculated a reduction in viscosity for Ti_3SiC_2 /propylene glycol. A 0.2 weight percent inclusion of MXene resulted in a viscosity drop. Propylene glycol and MXenes demonstrated good surface associations at 0.25 weight percent, as evidenced by a small rise in viscosity. There could be several causes for the MXenes viscosity reduction. One explanation is that there is little interaction between particles at tiny amounts due to charges' tendency to pull away from one another.

Heat Capacity

The heat capacity of nanofluids, which refers to their ability to store thermal energy, is a critical parameter for solar energy harnessing applications. Nanofluids usually have better thermal conductivity; however their heat capacity might change depending on the type and amount



Figure 7. Factors affecting viscosity of nanofluid.

of nanoparticles utilized. A large heat capacity is ideal for solar energy systems because it lets the fluid soak up and store more heat from the solar collector before sending it to a storage unit or directly to where it will be used [53]. However, adding nanoparticles can sometimes lower the fluid's overall heat capacity. A solar thermal system performs much better if its thermal conductivity and specific heat capacity of nanofluid is enhanced [54].

Rubbi et al. [55] examined the SHC of Ti_3C_2 -soybean oil at various MXene adding fraction. The heat capacity of nanofluid is enhanced due to inclusion of MXene for temperature range between 25°C to 80°C . Zhou et al. [56] found that their CuO /ethylene glycol nanofluid has an ultimate increase of roughly 6% of its SHC.

Stability of Nanofluid

For nanofluids to work well for collecting solar energy, they need to be stable. The term stability refers to proper distribution of nanoparticles in the base fluid without clogging over time. Solar thermal system works more effectively and efficiently when nanofluid is stable over period of time [57]. Various methods such as magnetic steering, homogenizer, ultrasonication, surfactant addition and many more are used to enhance the stability of nanofluid. Stable nanofluids keep solar energy systems working consistently, requiring less maintenance, and make them last longer and work better overall. As a result, one of the most important areas of research is how to make nanofluids very stable so that they can be used reliably and effectively for solar energy over a long period of time [58]. Various ways to enhance the stability of nanofluid is shown in Fig. 8.

In magnetic stirring, a magnetic stirrer is used to move the nanoparticles regularly and also properly mix them into the base fluid while in ultrasonication a sound wave is used to break the agglomerated particle and distribute them properly. This is possible due to high frequency of sound wave. This method can make the fluid more stable for a short time by making the nanoparticles more evenly spread out [59]. Changing the pH of the nanofluid can make the

surface charge of the nanoparticles better, which makes them repel each other more strongly. Stability can be greatly enhanced by getting the pH just right, which is usually where the zeta potential is at its highest. To help nanoparticles spread out better in the base fluid, surfactants and dispersants are added to nanofluids. They work by reducing the surface tension and preventing particle agglomeration through electrostatic or steric stabilization [60].

In order to investigate TC and other rheological parameters, Kumar et al. [61] conducted investigations utilizing Cu/Zn 0.1–0.5% and several base fluids, including vegetable oil, paraffin oil, and SAE oil. For almost 72 hours, all samples showed improved stability. Abdelrazik et al. [62] employed zeta potential and optical examination to assess the stability of $\text{Ti}_3\text{C}_2\text{-H}_2\text{O}$ nanofluids at multiple concentrations containing CTAB and SDBS surfactants. Prolonged storage periods were associated with increased aggregation in the stability of CTAB specimens. The lower concentration CTAB specimens began to precipitate after a single day of preparation, and by day five, each specimen had settled down. But whereas some SDBS specimens remained stable for just a few hours, others did stay stable for a few days with minimal precipitation.

UTILIZATION OF NANOFLUID IN DIFFERENT SOLAR ENERGY HARNESSING SYSTEM

Nanofluids are engineered colloidal suspensions of nanoparticles in a base fluid. They have shown a lot of promise in making solar energy systems work better. Due to enhanced thermal conductivity and heat transfer rate, solar collector performs much better by utilization of nanofluid. In PV/T system, nanofluid is used as a coolant to keep solar panel cool and get maximum out of it. In concentrated solar power (CSP), solar radiation is absorbed by nanofluid and this radiation is stored as thermal energy for further uses. The optical and thermal properties of nanofluid make it more special to enhance the solar system performance.

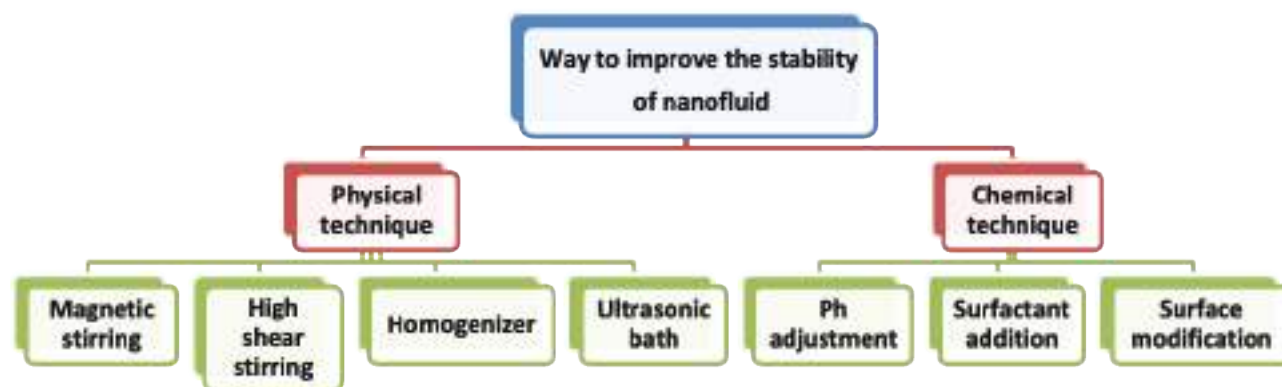


Figure 8. Way to improve the stability of nanofluid.

Solar Concentrator or Collector

A device which collects solar radiation and converts it into heat energy for further utilization is known as solar collector [63]. Figure 9 shows a simple flat plate solar collector. A solar collector's main parts are an absorber plate, a clear cover, insulation, and a system for moving fluid around. The materials such as aluminum, copper are used to make the absorber plate due to its high thermal conductivity. This plate is covered so that heat is not escaped from the plate. Insulation on the back and sides of the collector helps keep the heat it collects [64].

A solar collector works by collecting sunlight and turning it into heat. The absorber plate takes in energy from the sun and turns it into heat. This heat is then transferred to a working fluid (such as water or air) that flows through or across the absorber plate. The heated fluid is either used immediately or stored for later use. The efficiency of the collector depends on the quality of materials used, the design of the collector, and the intensity of the sunlight. Solar collectors are commonly used in residential and commercial settings for water heating, space heating, and even

for powering certain types of solar energy systems [65]. The detailed classification of different types of solar collectors is shown in Fig. 10 and the summary of recent research on solar collector for solar energy harnessing is provided in table 3. The mathematical formula used for the calculation of different parameters is mentioned below:

It is possible to assess the flowing fluid's usable heat transfer via

$$Q_u = \dot{m} c_p (T_{out} - T_{in}) \quad (1)$$

The solar collector thermal efficiency is calculated as

$$\eta_{th} = \frac{Q_u}{Q_s} = \frac{\dot{m} c_p (T_{out} - T_{in})}{I A_c} \quad (2)$$

The solar collector exergy efficiency is calculated as

$$\eta_{ex} = 1 - \frac{T_a S_{gen}^0}{\left(1 - \frac{T_a}{T_s}\right) Q_{solar}} \quad (3)$$

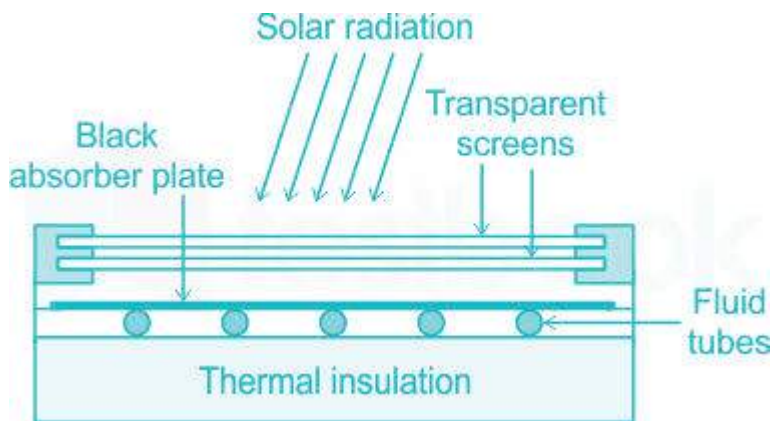


Figure 9. A Simple Flat Plate Solar Collector.

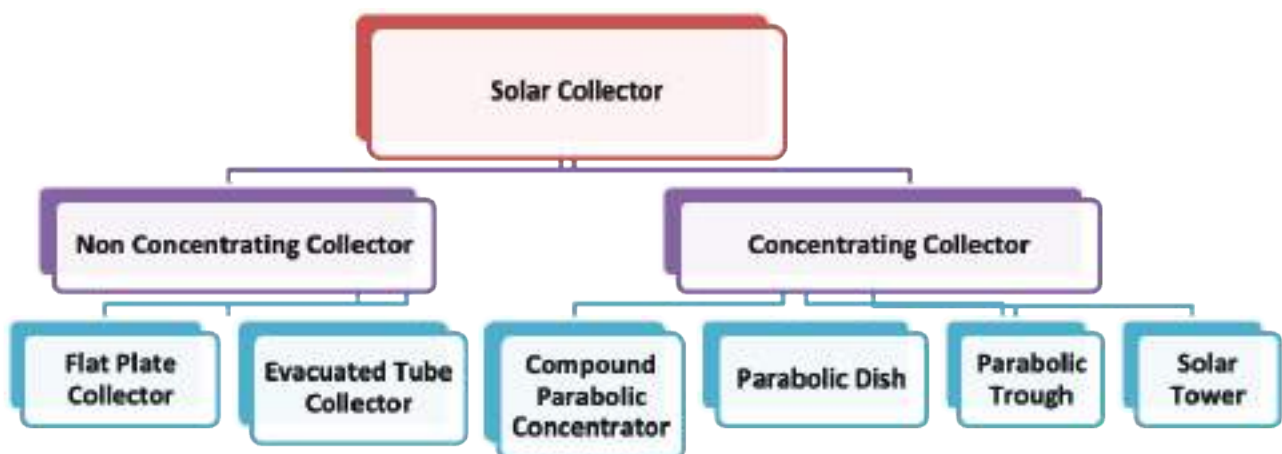


Figure 10. Classification of Solar Collector.

Table 3. Recent Researches Performed on Solar Collector using Nanofluid

Ref. No.	Type of Collector	Nanofluid	Remark
[66], 2016	FPSC	MgO/Water	A 9.34% increase in thermal efficiency is established for a 0.75% component volume fraction.
[67], 2017	FPSC	MWCNT/water, Graphene/water, CuO-H ₂ O, Al ₂ O ₃ -H ₂ O, TiO ₂ -H ₂ O, SiO ₂ -H ₂ O	Maximum improvement in energy efficiency is observed in MWCNT/water that is 29.3%.
[68], 2020	FPSC	Al ₂ O ₃ -H ₂ O, CuO-H ₂ O	The HTC can rise to 78.25% when zigzag pipes and CuO/water nanofluid are used.
[69], 2021	FPSC	Al ₂ O ₃ -H ₂ O and TiO ₂ -H ₂ O	Maximum 30% increment is observed using TiO ₂ -H ₂ O nanofluid
[70], 2022	FPSC	Al ₂ O ₃ , MWCNT, MWCNT/ Al ₂ O ₃	For 1.5 l/m, MWCNT/ Al ₂ O ₃ provide an efficiency increase of 26%.
[71], 2023	FPSC	Al ₂ O ₃ -H ₂ O, CuO-H ₂ O, Si ₂ O ₂ -H ₂ O	Efficiency of 2% Al ₂ O ₃ -H ₂ O is 29.75% higher than 1%.
[72], 2024	FPSC	SiC/DW	At a SiC volume percentage of 0.1, the FPSC's energy efficiency climbed to 37.4%.
[73], 2020	PTC	MWCNT/H ₂ O	49.52% enhancement in thermal efficiency at 0.3% volume concentration
[74], 2020	PTC	Syltherm 800/Copper	When the cermet layer is applied to the bare tubes, the highest improvement is 7.16%.
[75], 2021	PTC	Ag-ZnO/ Syltherm 800, Ag-MgO/ Syltherm 800, Ag- TiO ₂ / Syltherm 800	Maximum 31% enhancement in thermal efficiency using Ag-MgO/Syltherm 800 at 4% volume fraction
[76], 2022	PTC	Al ₂ O ₃ , CuO	Maximum thermal efficiency is 14.79% using CuO nanofluid at 0.0224 kg/s mass flow rate
[77], 2022	PTC	Al ₂ O ₃ , Cu, SWCNT	Maximum energy efficiency 79.9% for SWCNT
[78], 2023	PTC	CuO/H ₂ O	Highest efficiency increase of 69.07% at mass flow rate of 140 l/h and volume concentration of 0.1%.
[79], 2023	PTC	TiO ₂ and Graphite as TES material	85.4% maximum enhancement in energy efficiency using Graphite as TES material
[80], 2023	PTC	ZnO/EG- H ₂ O	100% increase in Nusselt number at volume fraction 4%
[81], 2023	PTC	Al ₂ O ₃ -Cu/ H ₂ O	A 32.8% decrease in the total amount of entropy generated
[82], 2024	FPSC, PTC	Carbon nanotube/water and carbon nanotube/ EG	80.6% increase in thermal efficiency of PTC using carbon nanotube/ EG nanofluid
[83], 2020	ETSC	MWCNT, SiO ₂ , TiO ₂ and CuO	Maximum thermal efficiency enhancement is 14.0.9% using CuO nanofluid.
[84], 2020	PT-ETSC	Al ₂ O ₃	Maximum gain in thermal efficiency was 69.5% 0.035 kg/s mass flow rate and 0.3% volume fraction.
[85], 2022	ETSC	hBN/H ₂ O	The ETSC's optimum energy efficiency is increased by 84%.
[86], 2023	ETSC	MgO-MWCNT/H ₂ O	MWCNT nanoparticles can boost the collector's optical effectiveness by up to 78.1%.
[87], 2023	ETSC	Al ₂ O ₃ , CuO, Fe ₃ O ₄ , MWCNT	Highest energy and exergy efficiency for 0.01% of MWCNT
[88], 2023	ETSC	MWCNT, Al ₂ O ₃ , MWCNT/ Al ₂ O ₃	Maximum energy efficiency obtained is 60% for MWCNT/ Al ₂ O ₃ hybrid nanofluid

Verma et al. [66] tested a FPSC using MgO/water nanofluid with particles measuring 40 nm and an aggregate volume fraction that ranged from 0.25 to 1.5% at various mass flow rates. Thermal efficiency gain of 9.34% is established by experimental evidence for 0.75% component volume fraction at 1.5 lpm flow rate. The variation of efficiency with mass flow rate at varying volume fraction is shown in Fig. 11. For the same composition and flow rate, there was

a 32.23% increase in energetic efficiency. Verma et al. [67] again concentrated on a broad range of nanofluids to assess the FPSC effectiveness in terms of several parameters. The results of the experiments show that the energy efficiency of MWCNT/water nanofluid is improved by 29.32% at a particle volume fraction of 0.75% and a mass flow rate of 0.025 kg/s. MWCNT/water nanofluids also have minimal entropy creation, which can be a disadvantage. The greatest

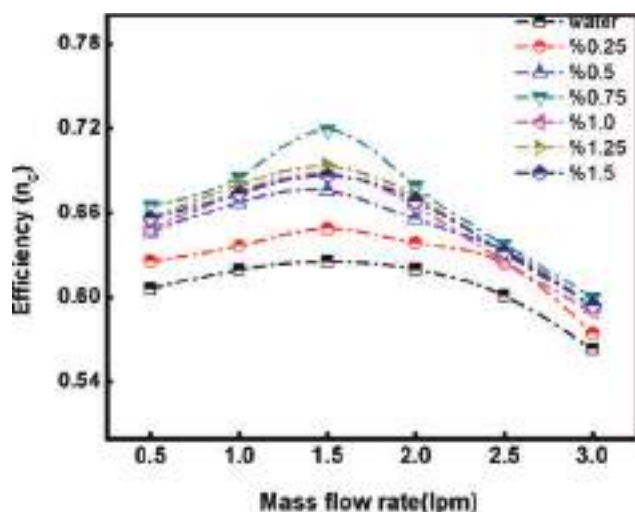


Figure 11. Variation of thermal efficiency of FPSC at varying mass flow rate and volume fraction [From Verma et al. [66], with permission from Elsevier]

measured decrease in production of entropy in MWCNT/water, at 65.55%, has been noticed under identical thermo-physical conditions.

Saffarian et al. [68] utilized two different nanofluid $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ and $\text{CuO-H}_2\text{O}$ having volume fraction varying from 1% to 4% to enhance the heat transfer of FPSC using different shape of flow path. The coefficient of heat transfer and Nusselt number can be greatly increased by employing spiral and zigzag pipes, according to the observations. Furthermore, it is noted that the zigzag pipes exhibit the most pressure decrease. When nanofluid is used in place of water, the coefficient of heat transfer rises in every situation. Gad et al. [69] conducted research, In order to investigate the impact of adding nanoscale $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ and $\text{TiO}_2\text{-H}_2\text{O}$ as an operating fluid on solar collector effectiveness. To improve spreading and decrease amalgamation, 25 nm-sized nanoparticles were employed. A surfactant was incorporated to the produced nanofluids to increase their stability, and their concentration was set at 2% by weight. Regarding base fluid, the largest increases in collector efficiency for $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ and $\text{TiO}_2\text{-H}_2\text{O}$ were approximately 22 and 30%, correspondingly. It was suggested that $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ and $\text{TiO}_2\text{-H}_2\text{O}$ nanofluid be used to improve the efficiency of solar collectors. Elshazly et al. [70] utilized three different nanofluids Al_2O_3 , MWCNT and, MWCNT/ Al_2O_3 to analyze the FPSC experimentally. Al_2O_3 substitution for 50% of MWCNT nanofluid results in a smaller efficiency gain and fewer hazards to the environment. According to the findings, employing hybrid MWCNT/ Al_2O_3 offers a 26% boost in efficiency for 1.5 L/m. Desisa [71] utilized $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ nanofluid for heat transfer investigation of FPSC experimentally as well as numerically at different mass flow rate and mean diameter of nanomaterial varying from 15 to 50 nm. Comparing CuO and Si_2O nanoparticles with

comparable nanoparticles, they exhibit varying degrees of enhancement. Compared to 1% $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ and H_2O , the thermal efficiency of 2% $\text{Al}_2\text{O}_3\text{-H}_2\text{O}$ is 5.3% and 29.75% higher, respectively. Ajeena et al. [72] examined the effects of using SiC/DW nanofluids on a FPSC's efficiency in the climate of Gödöllő, Hungary. The results showed that the inclusion of SiC nanoparticles increased the fluid's TC by as much as 30.3%. A 35.5% increase in thermal efficiency was seen in the FPSC with the addition of 0.1% SiC. At a SiC volume percentage of 0.1, the FPSC's energy efficiency climbed to 37.4%.

Hachicha et al. [73] conducted mathematical and experimental research to examine the impact of MWCNT/ H_2O nanofluid with varying concentration (0.05% to 0.3%) for moderate to medium-level temperature PTCs. 49.52% enhancement in thermal efficiency is achieved. At 0.3% concentration, 21% enhancement in Nusselt number is achieved. A thorough analysis of the application of Syltherm 800/Copper nanofluids in a PTC is conducted by Bellos et al. [74]. The findings show that situations with larger heat losses exhibit the greatest improvements, indicating that the majority of the bare tube's effectiveness is improved by the incorporation of nanofluids. When the cermet layer is applied to the bare tubes, the highest improvement is 7.16%.

Ekiciler et al. [75] looked into the 3D heat transmission and flow properties of the hybrid nanofluids at different volume fraction (1% to 4%) listed in table 3 under turbulence conditions in a PTC. During investigation, Reynolds number was varied from 10000 to 80000. It is observed that the maximum enhancement in convective heat transfer is 31% by utilizing Ag-MgO/Syltherm 800 nanofluid at 4% volume fraction while Ag-ZnO/Syltherm 800 nanofluid enhance the friction factor approximately by 15%. Farooq et al. [76] used CFD simulation to analyze the PTC with the application of two nanofluids namely Al_2O_3 and CuO at 0.01% concentration for environmental sustainability. Maximum thermal efficiency is 14.79% using CuO nanofluid at 0.0224 kg/s mass flow rate. Copper has consistent temperature dispersion throughout the entire span of the tube used for absorption because of its superior thermal conductivity.

Vahedi et al. [77] investigated the exergoeconomic and thermal performance of PTC using different nanofluid. The energy efficiency movement ran counter to the energy the demise tendency. Additionally, the maximum energy efficiency that achieved was approximately 74.4% for oil, aluminum oxide, and copper, and 79.9% for SWCNT. Ram et al. [78] used CuO/ H_2O nanofluid to enhance the performance of PTC with three different volume concentration at two mass flow rate. The increase in efficiency with varying volume concentration and mass flow rate is shown in Fig. 12 and observed that maximum efficiency gain 69.07% at 140 l/h mass flow rate at volume concentration of 0.1%.

Using a variety of techniques, involving the application of nanofluid such as TiO_2 and TES materials, Hamada et al. [79] attempted to improve the energy efficiency of the

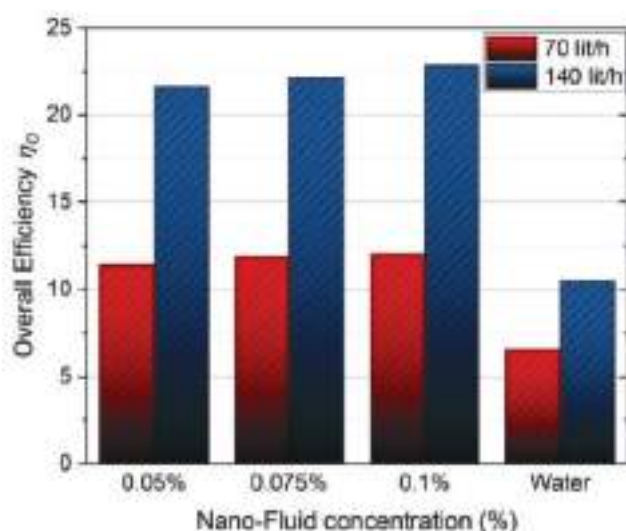


Figure 12. Variation of overall efficiency of PTC at varying mass flow rate and volume fraction [From Ram et al. [78], with permission from Elsevier].

conventional PTC. The result showed that 57.7% enhancement in thermal efficiency using TiO_2 nanofluid and 85.4% using graphite as TES material. The transport and flow properties of $\text{ZnO/EG-H}_2\text{O}$ nanofluid flowing in a PTC were studied both computationally and experimentally by Ekiciler et al. [80] at various flow rates ranging from 20 l/h to 80 l/h and volume fraction varying from 1% to 4%. The results show that the $\text{ZnO/EG-H}_2\text{O}$ nanofluid greatly increases the collector's performance. The greatest useable energy was produced using 4% $\text{ZnO/EG-H}_2\text{O}$ nanofluid at a flow rate of 80 l/h. As opposed to $\text{EG-H}_2\text{O}$, the data indicate that using $\text{ZnO/H}_2\text{O}$ causes a higher temperature

disparity. Esmaili et al. [81] used $\text{Al}_2\text{O}_3\text{-Cu/H}_2\text{O}$ hybrid nanofluid on PTC to investigate its performance. Heat transmission rises by 35.7% when CHGR are utilized as discs in the absorber tube as opposed to the absence of turbulators. In this instance, there is a 32.8% reduction in the overall creation of entropy. Abu-Zeid et al. [82] investigated the performance of FPSC and PTC using carbon nanotube/water and carbon nanotube/EG nanofluid. By employing EG nanofluid, FPSC and PTSC equipment demonstrated a notable increase in mean thermal efficiency of 64.1% and 80.6%, correspondingly.

Yurddaş [83] used CFD modeling to optimize the ETSC's thermal effectiveness after analyzing it with various nanofluids with varying concentration (0.5% to 5%) listed in Table 3. Maximum thermal efficiency enhancement is 14.0.9% using CuO nanofluid. Sasikumar et al. [84] experimentally examined the effectiveness of PT-ETSC with Al_2O_3 nanofluid at different concentrations (0.1% to 0.3%). Maximum gain in thermal efficiency was 69.5% 0.035 kg/s mass flow rate and 0.3% volume fraction. Kumar and Tiwari [85] use $\text{hBN/H}_2\text{O}$ nanofluid with different volume fractions (0.25 to 2%) to analyze the thermal effectiveness of an ETSC. Higher thermal conductivity has been proven using $\text{hBN/H}_2\text{O}$ nanofluid. The ETSC's optimum energy efficiency is increased by 84%. Henein et al. [86] used $\text{MgO-MWCNT/H}_2\text{O}$ hybrid nanofluid for 3E analysis of ETSC at different mass flow rate and weight ratio experimentally. According to the findings, increasing the weight ratio of MWCNT nanoparticles can boost the collector's optical effectiveness by up to 78.1%. There has been a boost in mean thermal energy absorption from 240 W to 495 W which is shown in Fig. 13. The use of the hybrid nanofluid results in a 36% decrease in area of ETSC and a 56% increase in the fluid's intake–exhaust variation in temperature.

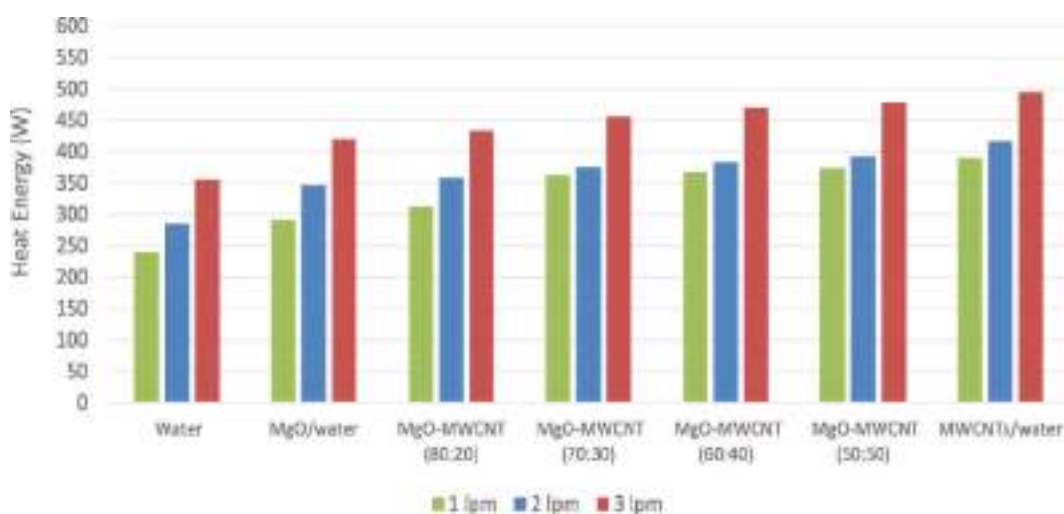


Figure 13. Heat energy gain in different nanofluid at different mass flow rate [From Henein et al. [86], with permission from Elsevier].

By employing several nanofluids listed in table 3 as operational fluids, Tong et al. [87] sought to assess the thermal endurance of an ETSC in a STS. The ETSC demonstrated the maximum energy consumption and energy effectiveness at 33.1% greater compared to the amount of water when the 0.01% MWCNT nanofluid was utilized. In terms of efficiency of energy, the highest value was 3.17 under ideal conditions. Elshazly et al. [88] utilized three different nanofluids as mentioned in table 3 to evaluate the thermal performance of ETSC at different volume fraction varied from 0.5% to 0.005% at three different mass flow rates. The results show that compared to employing Al_2O_3 , which has previously been reported, employing hybrid MWCNT/ Al_2O_3 50:50% results in an efficiency improvement of almost 20%. Ultimately, it was discovered that, under identical testing circumstances, the ETSC's energy and energy effectiveness could be increased to 73.5% and 51%, respectively, by using 0.5% MWCNT/ H_2O nanofluid at mass flow rate of 3.5 l/m.

Solar Water Heater

A solar water heater is an eco-friendly and energy-efficient system designed to harness solar energy for heating water. It typically consists of solar collectors, usually mounted on rooftops that absorb sunlight and convert it into heat. A solar water heating system is shown in Fig. 14 [89]. This heat is then sent to water, either directly or through a heat exchanger. The water is then kept in an insulated tank for later use. Solar water heaters work best in sunny places and can greatly minimize the need for traditional energy sources, which lowers utility bills and carbon emissions. These systems are great for homes and businesses because they provide a long-lasting means to provide hot water while also helping to protect the environment.

The way a solar water heater works is by turning sunlight into thermal energy, which is then utilized to heat water. Solar collectors, which are normally on rooftops, are what the system uses to capture sunlight. These collectors usually have an absorber plate that gets hot when the sun shines on it. Then, the heat is sent to a fluid, which can be water or a heat-transfer liquid that flows through pipes that are attached to the absorber plate. The heated fluid goes to a storage tank, where the heat is given to the water so it can be used. In passive systems, the fluid moves about on its own thanks to thermosiphon activity. In active systems, pumps move the fluid around. The hot water that is saved is then available for use in the home. A backup heater makes sure that there is always enough hot water, even when solar energy is low. Table 4 summarizes current research on solar water heating systems for solar energy harvesting using nanofluid.

Modi et al. [90] examined the thermal endurance of SWHS utilizing chitosan as an organic and Al_2O_3 as an inorganic nanofluid. The Chitosan produced the highest possible heat elimination factor (0.627) and collector efficiency factor (0.727). Chitosan exhibited a 5.96% and 33.6% better efficiency when compared with Al_2O_3 and H_2O , respectively. The variation of efficiency with time for three different cases is shown in Fig. 15.

Darbari and Rashidi [91] looked into the thermal effectiveness of the flat plate thermosyphon SWH employing the various nanofluids listed in table 4. The inclusion of Cu nanoparticles, subsequent to CuO, across the different nanoparticles produces the most increase in useable energy and effectiveness. Cu and CuO nanoparticles added to the base fluid at a volume percentage of 5% increase the solar collector's usable energy by 6% and 3%, correspondingly. Mandal et al. [92] performed parametric evaluation of SWH using paraffin wax as PCM and CuO-PCM. The highest

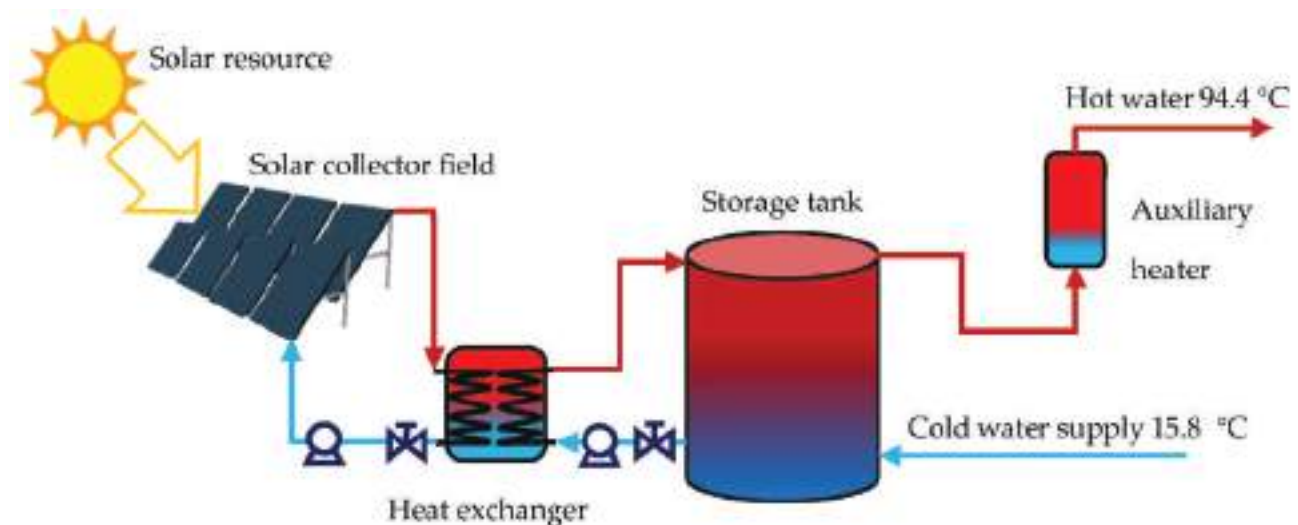
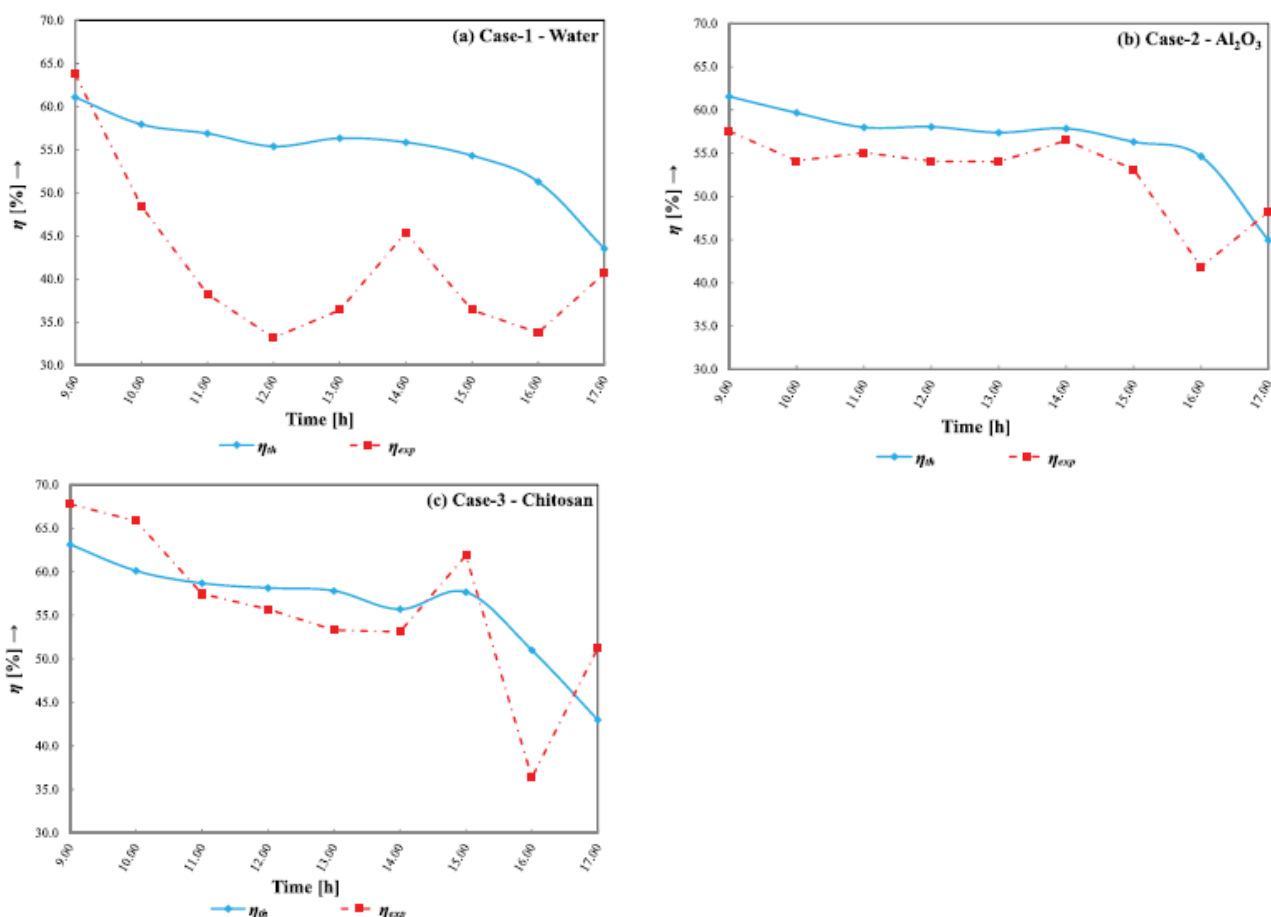


Figure 14. Solar water heating system [From López-Pérez et al. [89], with permission from MDPI].

Table 4. Current research on solar water heating systems for solar energy harvesting using nanofluid

Ref. No.	Type	Nanofluid	Remark
[90], 2020	FPC-SWH	Chitosan, Al_2O_3	It is possible to use chitosan and reach an experimental efficiency of 55.86%.
[91], 2021	FPC-SWH	Al_2O_3 , TiO_2 , CuO, Cu	Maximum 6% enhancement in thermal efficiency using CuO nanofluid
[92], 2021	FPC-SWH	CuO, CuO-PCM	Maximum heat transfer rate 5.7 kW is observed using CuO-PCM
[93], 2022	Vacuumed tube SWH	Al_2O_3 /Oil, TiO_2 /Oil	Al_2O_3 /oil nanofluid has 31% thermal efficiency as compared to TiO_2 /oil nanofluid
[94], 2023	PTC - SWH	DW/ZnO	Thermal efficiency is enhanced using DW/ZnO nanofluid
[95], 2023	ETSWH	TiO_2 /engine oil	When only porous material is used, the system's energy efficiency increases by 19.6%.
[96], 2023	ETSWH	TiO_2 /engine oil	24.4% thermal efficiency at 0.4% of mass using inserts
[97], 2023	Polymer SWH	Al_2O_3	27.7% increment in thermal efficiency of SWH
[98], 2024	FPC-SWH	TiO_2 - H_2O	Maximum 16.48% increment in SWH efficiency
[99], 2023	SWH	ZrO_2 - H_2O	Maximum 20.68% increment in SWH efficiency
[100], 2024	DTPTSWH	TiO_2 , Al_2O_3 , CuO, SiO_2	CuO nanofluid had the highest efficiency, at 38.4%.

**Figure 15.** Variation of efficiency of SWH with time with different cases [From Modi et al. [90], with permission from Elsevier].

possible rate of heat transfer for CuO-PCM is found to be 5.7 KW, provided that a gap in air of 40 mm is maintained among the plate that absorbs heat and the outer layer of glazing.

Shafiee et al. [93] used oil based nanofluid in vacuumed tube SWH to evaluated its thermal performance using three different nanofluid mentioned in table 4 at fixed volume

fraction of 1wt%. Al_2O_3 /oil nanofluid has 31% thermal efficiency as compared to TiO_2 /oil nanofluid. When the outcomes were assessed against the standard exergy efficiency calculation, a 6.1% discrepancy was found amongst them. Arun [94] proposed a PTC-equipped SWHS-ZO-NF. A boost in the volume fraction of the nanofluid lowers the exergy loss of the entire system. The thermal effectiveness and energy assessment are improved by the practical outcome of the proposed SWHS-ZO-NF. Firoozzadeh et al. [95] used TiO_2 /engine oil nanofluid to thermodynamically analyze the evacuated tube solar water heater (ETSWH) at two different mass concentrations. Consequently, the temperature rises by 6.4 °C when mass % 0.4 % TiO_2 in engine oil is merged into a porous substance. The rise in temperature results in an additional 41% increase in thermal effectiveness. Furthermore, there are increases in thermal effectiveness of 5.4% and 19% when a porous medium is used, and when 0.4 mass% TiO_2 /oil is applied concurrently to a porous medium. The study conducted by Lotfi et al. [96] centered on examining ways to improve the energy and performance of ETSWH installed in the warm weather of Dezful, Iran. Using TiO_2 nanoparticles added in two volume fractions, the investigation particularly assesses the effects of engine oil used as a circulatory fluid inside copper U-tubes placed into evacuated tubes. As a consequence, adding metallic insertion raises the water's temperature by 1.5 °C. Furthermore, a variations of 5.6 °C was noted in the evening hours among the 0.4% volume fraction TiO_2 + insertion and the base case, leading to increased thermal performance and exergy efficiency by 24.52% and 24.4%, correspondingly.

Nishit and Bekal [97] used Al_2O_3 nanofluid to enhance the performance of polymer SWH experimentally. The results of the research showed that compared to polymer SWH using water to serve as the absorption fluid, the inclusion of Al_2O_3 in polymer SWH increased thermal efficiency approximately 27.7%. At two separate flow rate of mass (0.008 kg/s and 0.012 kg/s), as well as mass fractions of 0.4% and 0.2%, Nayak et al. [98] both theoretically and experimentally analyzed an FPC-SWH employing $\text{TiO}_2 - \text{H}_2\text{O}$ nanofluids. The findings show that the energy effectiveness of SWH rises by 3.58% for H_2O when the flow rate is increased. Additionally, for TiO_2 nanofluids with 0.4% and 0.2% mass fraction, correspondingly, the SWH efficiency is increased by 11.64% and 16.48%. Rashmi et al. [99] used $\text{ZrO}_2/\text{H}_2\text{O}$ nanofluid to analyze the performance of SWH at constant mass flow rate. With a mass flow rate of 0.0185 kg/sec, nanofluid was shown to boost the efficiency of SWH by 20.68%. Arun et al. [100] experimentally and also used CFD technique to analyzed DTPTSWH using different nanofluid whose size varied from 10 nm to 15 nm at different volume fraction varied from 0.1% to 0.5%. Best efficiency is obtained for CuO nanofluid which is 38.4%.

Solar Photovoltaic Thermal (PV/T) System

A technology which converts solar energy into both electrical and thermal energy is known as Photovoltaic Thermal (PV/T) system. The energy efficiency of the system is enhanced by PV/T system since it directly transforms solar energy into electrical energy and also captures the remaining energy in the form of heat energy [101]. It is possible due to combination of PV panels and solar collector.

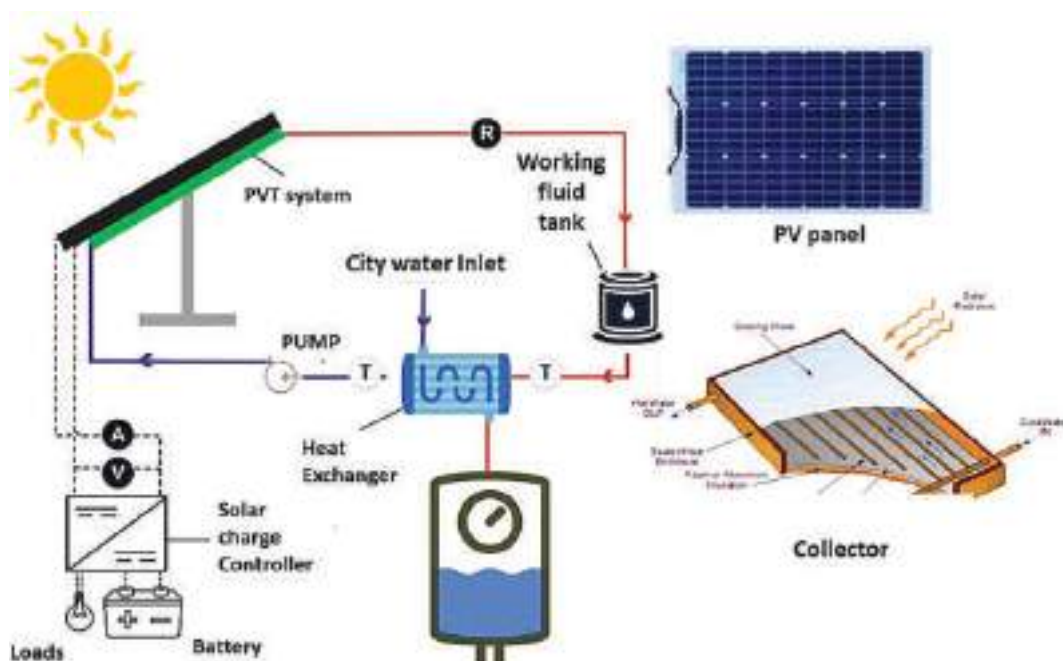


Figure 16. Block diagram of solar photovoltaic thermal system [From Mostakim and Hasanuzzaman [103], with permission from Elsevier].

The energy loss is reduced because of its dual combination and get maximum from solar radiation. Thermal energy is utilized in many appliances such as water heater, air heater etc. The PV/T system is very usefully in industrial application where both thermal and electrical energy is required. PV/T system enhances the energy production per unit area and also it is sustainable resource of renewable energy [102]. A typical block diagram of solar photovoltaic thermal system is shown in Fig. 16 [103].

A Solar Photovoltaic Thermal (PV/T) system works by putting together photovoltaic (PV) panels and a thermal collector. The PV panels make energy and the thermal collector collects heat from the sun. The photovoltaic effect turns sunshine into power, and the panels also soak

up extra heat, which can make them less efficient. To stop this from happening, the system uses a cooling medium, like water or air that flows via a heat exchanger that is built into the PV panels. This process takes out the additional heat, which can then be used to heat things like water or air. The PV/T system does this by making the PV panels work better electrically and giving off usable thermal energy. This maximizes the amount of energy that can be generated from the same solar resource. The detailed classification of solar PV/T system is shown in Fig. 17 and the recent research performed on solar PV/T system using nanofluid is summarized in table 5.

Salari et al. [104] used three different nanofluids with varying mass fraction mentioned in table 5 to evaluate PV/T

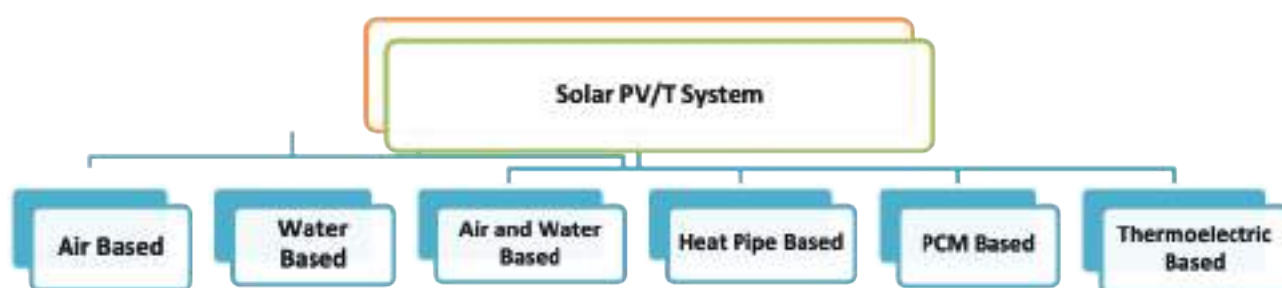


Figure 17. Classification of solar PV/T system.

Table 5. Recent research performed on solar PV/T system using nanofluid

Ref. No.	Nanofluid	Volume fraction	Remark
[104], 2020	MgO, MWCNT, MgO-MWCNT	3% to 6%	Maximum thermal efficiency 61.07% for MWCNT at 6% volume fraction
[105], 2021	CuO-MgO-TiO ₂	0.01%	58.38% thermal efficiency of PV/T is obtained
[106], 2022	TiO ₂	0.2%	PVS cells currently have an efficiency that is 2.11% more than PV-ground cells
[107], 2020	Ti ₃ C ₂ /silicone oil	0.05% to 0.1%	At 0.1 weight percent, the maximum TC augmentation is shown to be 64%
[108], 2024	Non-toxic GNP/water	0.025% and 0.1%	Maximum thermal efficiency is 55.22%, electrical efficiency is 14.05%.
[109], 2023	Sic and Nano-PCM	0.3% and 0.6%	84.74% thermal efficiency and 9.61% photovoltaic efficiency
[110], 2023	TiO ₂ and TiO ₂ -Fe ₂ O ₃	0.2% and 0.3%	47.2% thermal efficiency at hybrid nanofluid at 0.3% volume fraction
[111], 2024	MgO/H ₂ O and Cu ₂ SnS ₃	-	5.7% greater efficiency is achieved overall with the addition of a reflection.
[112], 2023	CuO/H ₂ O and CNT/H ₂ O	0.5% to 2%	Maximum 29.82% increase in thermal efficiency for BVPTM using CNT/H ₂ O nanofluid
[113], 2023	PCM and micro fin tube	1% paraffin and 0.6% SiC for nanofluid	Maximum 77.5% enhancement in thermal efficiency using micro fin tube nanofluid with 0.6% SiC
[115], 2023	ZnO/ H ₂ O	0.25%	Highest thermal efficiency of 33.4%
[116], 2023	ND-Co ₃ O ₄ / H ₂ O	0.05% to 0.15%	Maximum 15.44% enhancement in thermal efficiency
[117], 2023	Ag, CNT, and CNT/Ag	-	9.9% increment in thermal efficiency using hybrid CNT/Ag nanofluid

system numerically as well as experimentally. MgO nanofluid has the least total energy efficiency (55.24%), while MWCNT has the best overall efficiency of energy (61.07%). The system's overall heat capacity is increased when the layer of PCM thickness is increased from 0.5 cm to 1.5 cm, lowering the outside and outflow temperatures. Adun et al. [105] used CuO-MgO-TiO₂ nanofluid in solar PV/T system to evaluate its performance numerically and experimentally. An enhanced electrical efficiency of 13.54% was demonstrated by the outcome. At an ideal fraction of volume of 0.01, additional results included thermal efficiency of 58.38%. The investigation also finds that the introduction of tripartite nanofluid increased the PV/T technique's overall efficiency to an optimum of 11.14%. Arifin et al. [106] used TiO₂ nanofluid PV/T system to evaluate its performance. The fluid's ability to facilitate the exchange of heat provides an explanation for this. PVS cells currently have an efficiency that is 2.11% more than PV-ground cells. The mean PV temperature produced by TiO₂-based PV/T systems is 58.5 °C, which results in a PV efficiency of 13.04%. Aslfattahi et al. [107] used Mxene (Ti₃C₂)/silicone oil nanofluid of varying concentration to enhance the performance of concentrated solar PV/T system. When Ti₃C₂/silicone oil nanofluid is concentrated to 0.1 weight percent, the maximum TC augmentation is shown to be 64% at 150°C when contrasted with plain silicone oil. The concentration of 0.1 weight percent Ti₃C₂/silicone oil yields the maximum electrical efficiency of concentrated PV/T across all concentration ratios. At a larger concentration ratio, electrical efficiency is, nevertheless, modest.

Sheikhholeslami et al. [108] used non-toxic GNP/water nanofluid to enhance the performance solar PV/T system at two distinct weight fractions in the optimal scenario, there is an improvement in velocity at the inlet, leading to a 5.8% boost in the entire system's electrical efficiency. The best scenario results in an optimal exergy efficiency of 15.32%, an electrical efficiency of 14.05%, and a thermal efficiency of 55.22% when employing GNP nanoplatelets at Reynolds number = 1611 and weight fraction of 0.1%. Al-Aasam et al. [109] assessed the thermal endurance of a PVT collector using Nano-PCM and a twisting absorber tube using Sic nanofluid at two different volume fractions. Thermal efficiency and photovoltaic efficiency for the T-PVT-PCM-SiC-0.6% arrangement were 84.74% and 9.61%, correspondingly. Alktranee et al. [110] use mono TiO₂ nanofluid and an innovative hybrid TiO₂-Fe₂O₃ nanofluid to study the energy efficiency and exergy efficiency of a PV/T technology at different volume proportions. Thermal efficiency is observed 47.2% using hybrid nanofluid at 0.3% of volume fraction. The various efficiency of solar PV/T system is shown in Fig. 18.

Sheikhholeslami [111] concentrated on developing a novel PV/T module design that made advantage of a cylindrical reflector. Effectiveness is improved by combining a nanofluid optical filter made of a combination of MgO/H₂O with a TEG sheet made of Cu₂SnS₃ (CTS) inside PV/T. The addition of a filter tube, fins, and a cylindrical reflector improves the effectiveness by 36.3%. An estimated 5.7% and 4.7% aggregate effectiveness gain is achieved when using and not using a nanofluid filtration when a reflector is added to the coated design. Ahmadinejad and

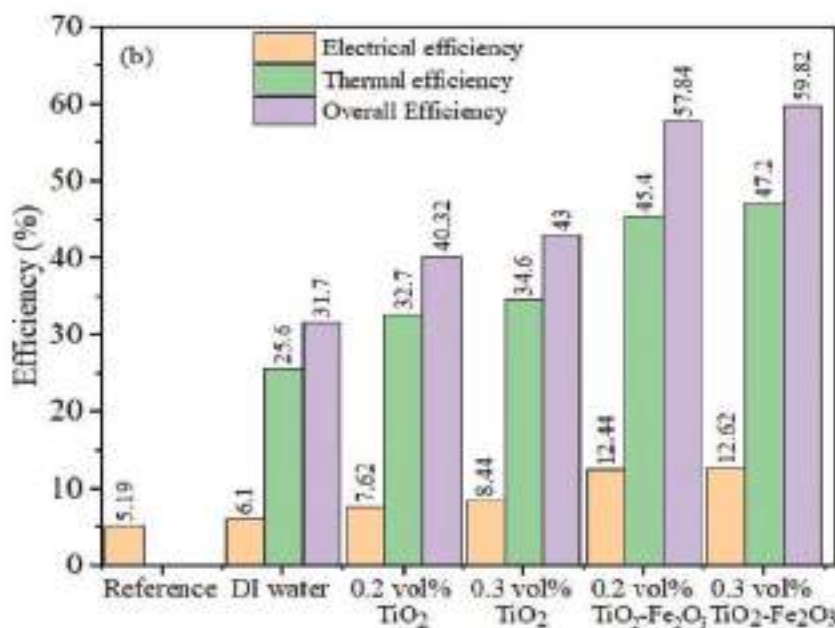


Figure 18. Various efficiency of solar PVT system for different nanofluid at varying volume fraction [From Alktranee et al. [110], with permission from Elsevier].

Moosaviwere [112] used three fluids H_2O , CuO/H_2O , and CNT/H_2O nanofluid to study the PV/T system's performance. When compared to the BPVTM using water, the total energy effectiveness enhances by 0.44% and 2.38% using CuO/H_2O and CNT/H_2O , correspondingly. Bassam et al. [113] experimentally analyzed PVT system using nano enhanced PCM and micro fins tube as nanofluid. When micro fin nanofluid was employed, the system's maximum thermal efficiency was 77.5%.

Gu et al. [114] modeled and analyzed an innovative bifacial PVT technology using multi-physics. Excellent insulating material and excellent transmissivity combine to provide an outstanding energy efficiency rating of 85.74%. Poor-quality production of heat accounts for the 12.26% total exergy efficiency. It is expected that when the transmission coefficients of the glass and nanofluid both approached 1, the power generated by the bPV/T technology will be greater compared to that of simple bPV. Shen et al. [115] looked at the nanofluid's impact on the PV/T through experimentation. The experimentation was performed using ZnO nanofluid at different mass flow rate varied from 0.008 kg/s to 0.012 kg/s and concentration 0.25%. Considering a hydrogen generation rate of 17.4 milliliters per minute, the highest thermal efficiency of 33.4% was recorded at 0.012 kg/s across the various mass flow rates. Khalili and Sheikholeslami [116] have presented a unique approach for chilling the PV unit employing $ND-Co_3O_4$ hybrid nanofluid at different mass concentrations ranging from 0.05% to 0.15%. With the mass fraction of 0.05%, the most effective system achieved electrical and thermal efficiencies of around 84% and 15.44%. Xia et al. [117] evaluated the solar spectrum properties of PV/T systems utilizing Ag, CNT, and CNT/Ag nanofluids at varying concentrations. The CNT/Ag nanofluid-based system's electrical and thermal efficiency rose by 15% and 9.9%, correspondingly.

Solar Stills (Evaporator)

A solar still, also known as a solar evaporator, is a device that harnesses solar energy to purify water through the process of evaporation and condensation. It typically consists of a transparent cover and a basin that holds the contaminated water. A simple solar still is shown in Fig. 19. The cover lets sunshine through, which warms the water in the basin and makes it evaporate. The water vapor then turns back into water on the cooler surface of the cover, which slopes into a collection channel where the clean water is stored. Solar stills are great for places with a lot of sun and not a lot of clean water because they are cheap and long-lasting ways to desalinate and clean water [118-119].

A solar still works by copying the natural water cycle. It uses solar energy to clean water by evaporating and then condensing it. The process starts when sunlight goes through the still's clear cover and warms up the dirty or salty water in a basin below. When the water gets hot, it evaporates, leaving behind dirt, salts, and other things that shouldn't be there. The water vapor rises and then collects

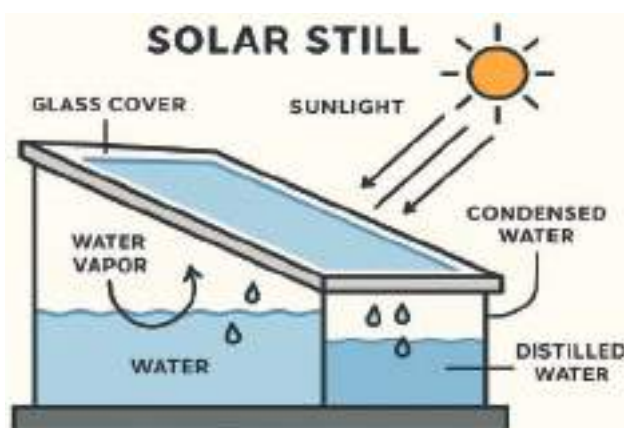


Figure 19. Schematic diagram of solar still.

on the colder, sloped inside of the lid. The cleansed water that has condensed runs down the sloped cover and is collected in a separate channel or reservoir [120]. This simple but effective approach makes it possible to make clean drinking water using only sunshine. This is especially useful in dry areas or places that are hard to get to that don't have easy access to fresh water. The detailed classification of solar still is shown in Fig. 20 and the summary of recent research on solar still using nanofluid for solar energy harnessing is summarized in table 6.

A solar still works by copying the natural water cycle. It uses solar energy to clean water by evaporating and then condensing it. When the water gets hot, it evaporates, leaving behind dirt, salts, and other things that shouldn't be there. The water vapor rises and then cools down on the inside of the cover, which is usually sloping. The cleansed water that has condensed runs down the sloped cover and is collected in a different channel or reservoir [121]. This process is very simple and also effective to use drinking water in sunshine only. This is especially useful in dry locations or places that are hard to get to and don't have easy access to fresh water [122]. The detailed classification of solar still is shown in Fig. 18 and the summary of recent research on solar still using nanofluid for solar energy harnessing is summarized in table 6.

Kumar et al. [123] conduct a laboratory study into the effectiveness of a solar still having single slope utilizing hybrid nanofluids and an outside parabolic reflector. The findings showed that using hybrid NFs and an outside parabolic reflector boosts the rate of production. Without utilizing nanofluid, the mean total distillate production was 1005 mL/day; however, productivity climbed to 1432.5 mL/day after employing nanofluids, indicating a 29.84 % additional rise in performance for the whole summer period. The effectiveness of a SSPSS using ordinary H_2O and Al_2O_3/H_2O nanofluid was compared by Azam and Akhtar [124]. According to the findings, adding Al_2O_3/H_2O nanofluid to the solar still increases the water distillation yield by



Figure 20. Classification of Solar Still (Evaporator).

Table 6. Summary of recent research performed on solar still system using nanofluid

Ref. No.	Nanofluid	Type	Remark
[123], 2024	hybrid	Single slope	29.84 % additional rise in performance for the whole summer period
[124], 2024	$\text{Al}_2\text{O}_3/\text{H}_2\text{O}$	Single slope	Solar still increases the water distillation yield by 24.11%
[125], 2024	ZnO , SiO_2 and $\text{ZnO} - \text{SiO}_2$	Single slope	Maximum efficiency is achieved using ZnO nanofluid which is 29.5%
[126], 2023	TiO_2 /Jackfruit peel nanofluids	Double-effect solar distiller	JPTN-equipped DESS has 41% energy efficiency.
[127], 2020	Ag	SODIS	Maximum enhancement in exergy efficiency is achieved 196% using nanofluid in Tehran
[128], 2021	$\text{Al}_2\text{O}_3 - \text{CuO}$	Conventional solar still	The average energy efficiency increases with hybrid nano by 13.4% in the winter and 12.6% in the summer.
[129], 2021	$\text{Al}_2\text{O}_3 - \text{SiO}_2/\text{H}_2\text{O}$	Conventional solar still	Hybrid nanofluid resulted in 37.76% thermal efficiency.
[130], 2022	$\text{Al}_2\text{O}_3 - \text{TiO}_2$	Double slope	The solar desalination process improved the environmental characteristics by 12.45%.
[131], 2023	$\text{Fe}_3\text{O}_4 + \text{NbO}_2 + \text{paraffin}$	Stepped solar still	The rate of production is increased by 98% when hybrid NPCM is used in a magnetic environment.
[132], 2023	$\text{CuO} + \text{Al}_2\text{O}_3 + \text{ZnO}$	SBGIDSS	Maximum energy efficiency is enhanced by 78.6%.
[133], 2023	Al_2O_3 , MgO , GO	Single slope	38.52% enhancement in production of water distillation using 1% wt of MgO
[134], 2024	$\text{ZnO}/\text{H}_2\text{O}$ and $\text{ZnO} - \text{WO}_3/\text{H}_2\text{O}$	Active solar still	The highest thermal efficiency was 39.9% for mono ZnO nanofluid.

24.11%, with a concentration of 0.15% being the most effective. Additionally, the investigation notes that evaporation heat transmission is more strongly facilitated by Al_2O_3 nanofluids than convection transfer of heat. Sahu and Tiwari [125] utilized three different nanofluids ZnO , SiO_2 and $\text{ZnO} - \text{SiO}_2$ to evaluate the performance of solar still having single slope at different depth of water. Maximum efficiency is achieved using ZnO nanofluid which is 29.5%. By using nanofluids, you may increase the rate of heat distribution, reduce the amount of time you need to warm up the water prior using it, and keep the difference

in temperature within the container higher for maximal condensing rate. Kumar et al. [126] used different JPTN at different concentrations ranging from 5% to 30% alongside silver color balls to illustrate the DESD's effectiveness. DESS with JPTN has an energy efficiency and exergy efficiency of 41% and 5.63%, respectively. The variation of thermal efficiency with time at different concentration of JPTN is shown in Fig. 21. From Fig. 21, it was observed that the highest thermal efficiency is obtained at 20% JPTN at water depth of 0.8 cm.

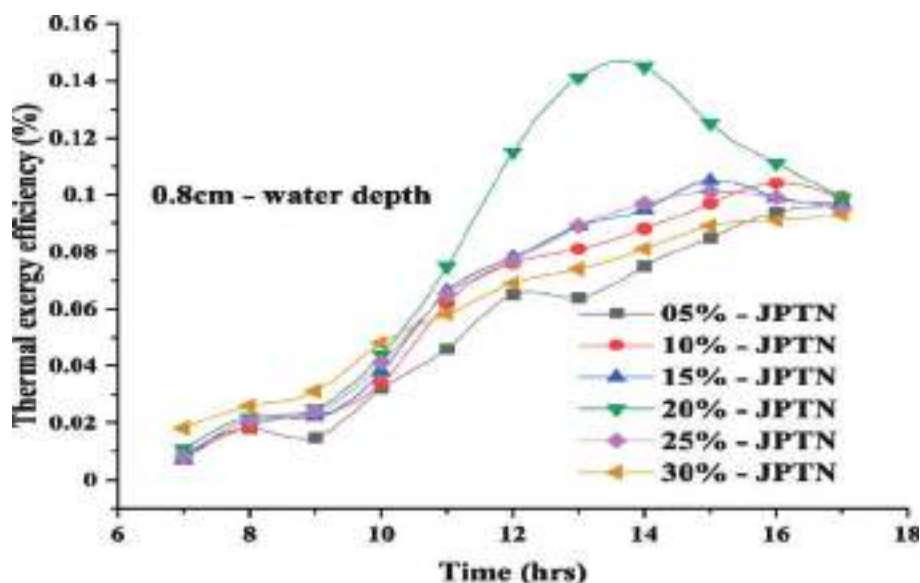


Figure 21. Variation of thermal efficiency with time at different concentration of JPTN [From Kumar et al. [126], with permission from Elsevier].

Parsa et al. [127] used Ag nanofluid to evaluate the performance of solar still at 0.04 wt%. Fig. 22 shows the SEM picture of the Ag nanofluid used in this work. This allows for interface consistency and uniformity to be confirmed, indicating a suitable distribution of nanomaterials in the base fluid. Maximum enhancement in exergy efficiency is achieved 196% using nanofluid in Tehran. From an energetic perspective, solar stills situated at high altitudes achieve the highest exergy efficiency

El-Gazar et al. [128] demonstrated the thermal characteristics of a standard solar still and illustrate the impact of utilizing a hybrid Al_2O_3 - CuO nanofluid with fixed concentration of 0.025% on the purification system. According to the findings, the use of hybrid nanofluid increases still daily production by 27.2% and 21.7% in comparison to still with no nanoparticles, reaching $5.5239 \text{ kg/m}^2\text{-day}$ in the hottest months and $3.1079 \text{ kg/m}^2\text{-day}$ in the colder months.

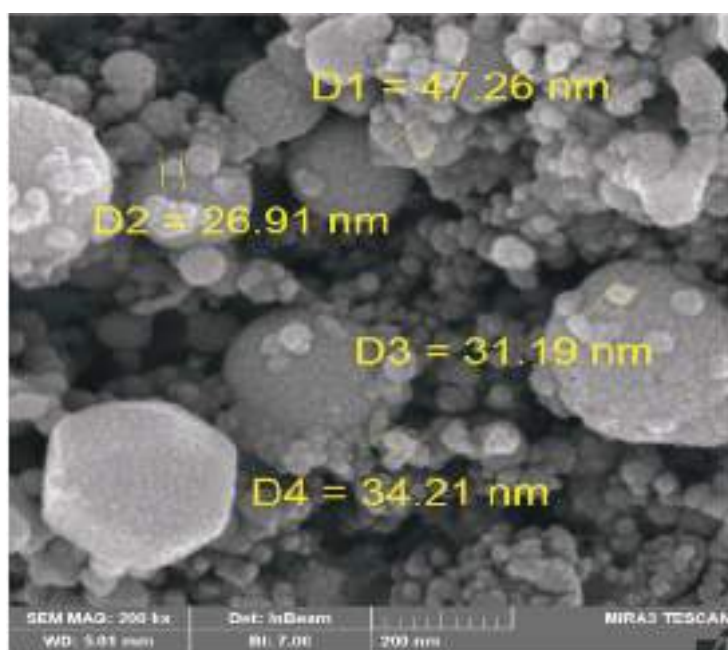


Figure 22. SEM image of Ag nanofluid at 0.4 wt% [From Parsa et al. [127], with permission from Elsevier].

Using Al_2O_3 -CuO nanofluid increases the average energy efficiency by 13.4% in the colder months and by 22.5% in the heat of the summer. Rabbi and Sahin [129] used Al_2O_3 - $\text{SiO}_2/\text{H}_2\text{O}$ hybrid nanofluid to enhance the performance of solar still. The utilization of hybrid nanofluids yields the greatest results. When hot water was used in the heat exchanger, the effectiveness of the solar still was enhanced. Hybrid nanofluid resulted in 37.76% thermal efficiency. Shoeibi et al. [130] used Al_2O_3 - TiO_2 hybrid nanofluid at fixed concentration of 0.4% to evaluate the performance of glass cooled double slope solar still using CFD analysis. According to the computational findings, the application of hybrid nanofluid optical cooling in sun desalination boosted the clean water productivity and energy efficiency by 11.09% and 28.21%, correspondingly, as compared to conventional solar desalination without hybrid nanofluid.

Toosi et al. [131] established and built an innovative stepped solar still incorporating hybrid NPCM underneath steady magnetism. The PCM's thermal conductivity enhanced with the addition of hybrid nanomaterials. Utilizing hybrid NPCM in a magnetic environment increases the stepped solar still's capacity for manufacturing by 98%. A steady magnetic attraction reduced the hybrid NPCM's thermal resistance. Three hours after nightfall, the distillation procedure is carried out utilizing NPCM. Naveenkumar et al. [132] used three different nanofluid mentioned in table 6 having volume concentration of 0.1% to improve the performance SBGIDSS using a vacuum fan operated by solar system. Maximum energy efficiency is enhanced by 78.6% by using vacuum fan and outside condenser. The increase in the rate at which water evaporation occurs in the solar still reservoir results in a 64.29% increase in production rate. Kajal et al. [133] performed comparative analysis using different nanofluid mentioned in table 6 to investigate the performance of solar still. When 1.0% wt MgO nanofluid is used, the highest increase in the amount produced of water from distillation above that obtained with water is around 48.91%. Almohammadi et al. [134] utilized hybrid nanofluid mentioned in table 6 to investigate the performance and productivity of solar still at a volume fraction of 0.035%. Maximum thermal efficiency was 39.9% for ZnO nanofluid. ZnO/ H_2O nanofluid is produced 3.14 kg of water per day at 0.035% volume fraction.

APPLICATION OF UTILIZATION OF NANOFLUID IN SOLAR ENERGY HARNESSING

The ultimate objective of this research is to produce maximum energy from solar energy system using nanofluid with highest efficiency and effectiveness. Nanofluid is the mixture of nanoparticles in the base fluid. Their thermal, optimal and chemical properties of much better than basic heat transfer fluid due to which they absorb solar energy much effectively. The performance of solar thermal system is enhanced due to its enhanced thermophysical and optical properties. These fluids are also very useful in PV/T system for cooling of the system

which can enhance the system performance and also boost the electrical efficiency. This review article mainly focused on utilization of nanofluid to harness solar energy to make solar system much better and sustainable solution of energy crisis.

Enhanced Thermal Conductivity

Enhanced thermal conductivity of nanofluid makes solar system more effective and prominent solution for renewable energy. The appliances like solar collector, heat exchanger perform much better with enhanced thermal conductivity.

Improved Solar Collector Efficiency

In solar thermal collector, sun light is absorbed and it is converted into thermal energy. Enhanced optical properties of nanofluid maximize the amount of absorbed solar radiation in solar collectors. These are the reasons behind the enhanced performance of all type of solar collectors.

Enhanced Photovoltaic Thermal System

PV/T system used nanofluid to cool the system properly and maximize the system performance. Nanofluids assist keep solar panels at the right operating temperatures by quickly getting rid of heat. This makes them work better and last longer.

Advanced Solar Thermal Energy Storage

Nanofluids can also be used in systems that store thermal energy. In these systems, they can make the materials that store the energy have a higher specific heat capacity and thermal conductivity. This change makes it easier to store and get back thermal energy, which makes solar power plants work better, especially when there isn't any sun.

Potential in Solar Desalination

Nanofluids make solar desalination more efficient by enhancing the transfer of heat between the evaporation and condensation stages. This saves money and energy, especially in places where there isn't much freshwater available.

ENVIRONMENTAL AND ECONOMIC IMPACT OF NANOFLUID IN SOLAR ENERGY HARNESSING

Nanofluids improve the collection of solar energy, which is good for the environment and the economy, but they also come with some problems. Due to enhanced thermal efficiency of solar collector, nanofluid enhanced the conversion of energy, minimising the dominance on fossils fuel and also diminishing the emission of green house gases [135]. The ability to conserve water in CSP makes them more environments friendly while production of nanoparticle for enhancement of energy production can be harmful for environment if it is not handled properly. By improving energy efficiency and lowering operating costs, nanofluids make it possible for solar systems to be more compact and resource-efficient, which can lower the original capital investment [136]. The high production cost and problem

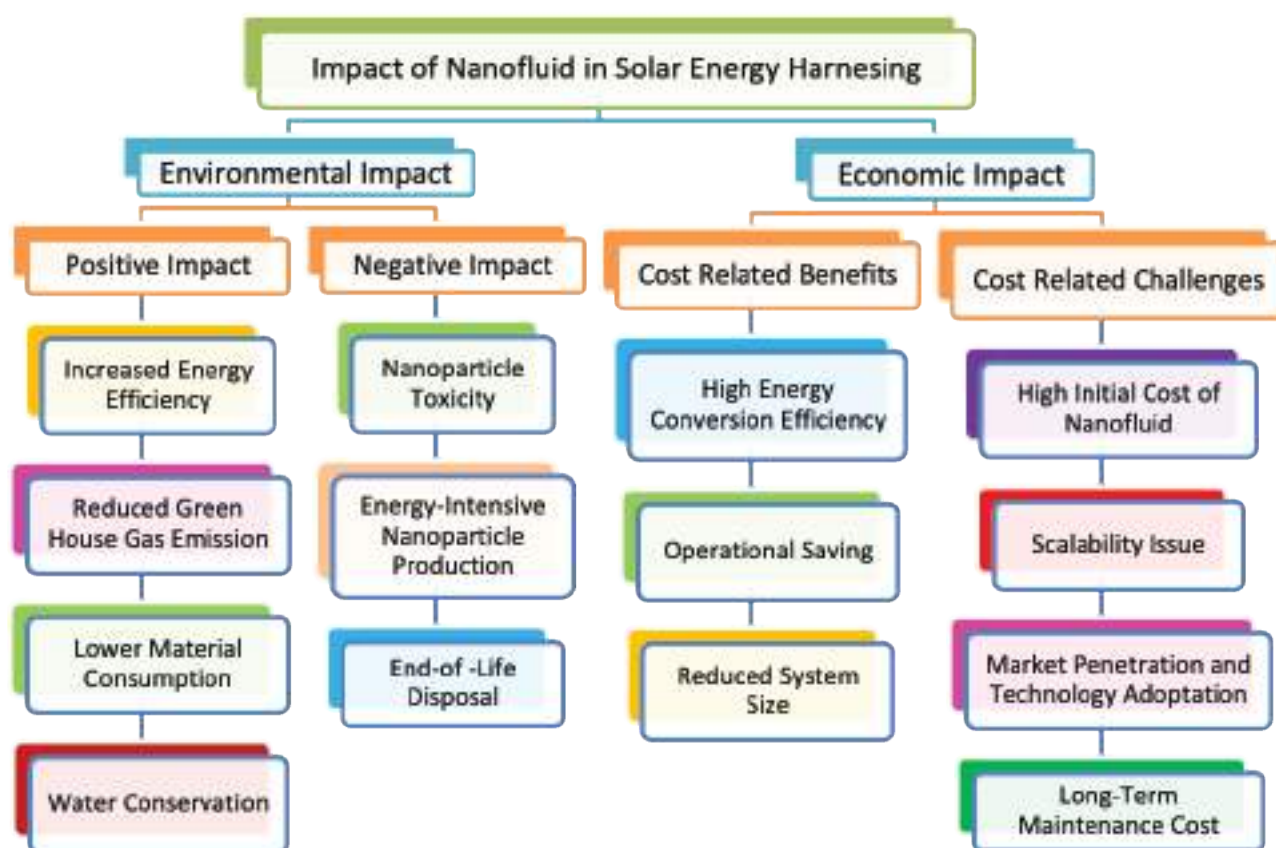


Figure 23.Environmental and Economic Impact of Nanofluid in Solar Energy Harnessing.

of scaling of nanoparticle create problem in its wide utilization. Nanofluids are unstable over period of time therefore its long term maintenance is also very expensive. New kind of research is required to make nanofluid utilization viable in academic and industry for long term utilization. Fig. 23 depicted the environmental and economic impact of nanofluid utilization in solar energy harnessing systems.

Environmental Impact

Solar thermal system performance is enhanced due to better thermal conducting capacity of nanofluid. Therefore the performance of solar system is enhanced and dependency on fossils fuel is minimized [137]. Nanofluids can help reduce greenhouse gas (GHG) emissions over time by making solar energy more efficient. Improved solar thermal technology could help the world move away from coal, natural gas, and oil, which would lower carbon footprints. Because nanofluids transport heat better, solar collectors might be able to do the same job with fewer or smaller components. This would mean that fewer building materials, including glass, metal, and plastic, would be needed. This cuts down on resource exploitation and lessens the overall impact on the environment [138]. Nanofluids can help make concentrated solar power (CSP) systems more efficient at transferring heat, which means they need less water to cool down. This is a big deal in places where water is scarce [139].

The effects of Al_2O_3 -MWCNT/ H_2O hybrid nanofluids on how well energy is used and how well the environment works Nanofluids offer better thermal qualities, like increased thermal conductivity, which can make solar thermal systems work better. This means that solar energy can be captured and used more effectively, which could mean less dependency on fossil fuels [137]. Nanofluids can help reduce greenhouse gas (GHG) emissions over time by making solar energy more efficient. Improved solar thermal technology could help the world move away from coal, natural gas, and oil, which would lower carbon footprints. Because nanofluids transport heat better, solar collectors might be able to do the same job with fewer or smaller components. This would mean that fewer building materials, including glass, metal, and plastic, would be needed. This lessens the overall environmental impact and the amount of resources that need to be extracted [138]. Nanofluids can help improve the thermal efficiency of concentrated solar power (CSP) systems, which means they use less water for cooling. This is a big deal in places where water is hard to come by [139].

How Al_2O_3 -MWCNT/ H_2O hybrid nanofluids affect energy efficiency and the environment was examined by Mahhadian et al. [140] in a DAPTC at different volume fractions. By employing hybrid nanofluid, the thermal performance of DAPTC has been improved by 197.1%, 69.2%, and 6.1% in comparison to water, Al_2O_3 , and MWCNT

nanofluids of 0.4% concentration. In order to examine the financial and environmental effects of ZnO/ H₂O, Al₂O₃/ H₂O, and TiO₂/ H₂O nanofluids at 0.2% volume fraction on the PV/T system, Abadeh et al. [141] conducted experimental research on these materials. In comparison to a PV system, it has been demonstrated that the nanofluid-based PV/T system reduced environmental emissions by 17%. A thorough investigation of the 4E was conducted by Said et al. [142] employing Ti₃C₂ nanoparticles in a PTC at different volume fractions (0.05% to 0.1%). Using Ti₃C₂ nanofluid, it was shown that 2.25 to 2.3 tons of CO₂ were mitigated annually.

Economic Impact

Nanofluids can make solar thermal systems (such flat-plate or parabolic trough collectors) work much better, which could cut the levelized cost of energy (LCOE). This makes investing in solar energy more cost-competitive with other types of electricity [143]. Nanofluids can lower the expenses of running solar power plants by making heat transfer more efficient. For example, they might let systems work with less fluid or at lower temperatures, which would save energy and make equipment last longer [144]. More efficient heat transfer means that you may utilize smaller solar collectors to make the same amount of energy. This can lower the upfront costs of land, infrastructure, and materials, which makes solar projects cheaper [145]. It's evident that nanofluids work better, but making nanoparticles and mixing them into fluids is still rather expensive. This could make people less likely to use it, especially in marketplaces where pricing is important. It is very difficult to produce nanofluid on mass level for industrial utilization since it requires huge amount of energy, nanoparticles are rare and also very expensive [146]. Nanofluid offers many advantages but still not utilized in mass level because the process of making nanofluid is very difficult, testing and characterization of nanomaterials are not easily available. A lot of funds are required to make nanofluid research available at commercial level [147-148]. The summary of economic impact of nanofluid in solar energy harnessing is presented in table 7.

Lari and Sahin [145] looked into the economics of a silver nanofluid with volume fraction of 0.5% in PV/T system. The recommended approach was shown to have reduced home electricity prices by 82%, and the initial expenditure paid for itself in just over two years. Lari and Sahin [149] conducted additional research on the impact of retrofitted on Ag/H₂O nanofluid in PV/T systems using PCM-thermal batteries. The predicted price repayment time frame for the recovery of investments was 5.5 years, resulting in a 58.71% decrease in energy produced costs when weighed against the cost of household power. Additionally, the release of 10,195.91 tons of CO₂ into the environment per year must be stopped. Stalin et al. [150] focused on 3E research of a FPSC using CeO₂/H₂O nanofluid with varying concentration of (0.01% to 0.1%). At 0.05% concentration, the collector efficiency is enhanced by 28.07% and energy augmentation of 5.8% was attained. The use of CeO₂/H₂O in FPSC resulted in a longer payback period of 2.12 years, and reduced CO₂ emissions to the environment by 175 kg. The 3E and financial advantages of a redesigned ETSWH with CuO nanomaterials having concentration of 0.5% were investigated by Saxena and Gaur [146]. In comparison to the standard design, the ETSWH required a larger initial expenditure, a longer payback period, and a longer time. Additional power is saved, maintenance expenses are decreased, and the transmission of heat is improved with the updated ETSWH. The impacts of GAMWCNT/H₂O nanofluid at different concentrations (0.025 wt% to 0.1 wt %) on the economic and environmental implications of FPSC were studied by Amar et al. [151]. According to experimental results, at a weight concentration of 0.1%, the maximum enhancement in efficiency of energy is 30.88%. It emerged that using GAMWCNT nanofluid instead of H₂O as base fluid reduced the dimensions of FPSC and saved 321.72 MJ of intrinsic energy. When GAMWCNT nanofluids were used at 0.1 weight percent in FPSC, the time to pay back became 1.897 years, that's 6.228% less than when H₂O was used as the heat transmission fluid.

Table 7. Summary of Economic Impact of nanofluid in Solar Energy Harnessing

Ref. No.	Nanofluid	System	Remark
[86], 2023	MgO-MWCNT/H ₂ O	ETSC	There is a drop of 36.2% and 34.6%, correspondingly, in the Carbon Repayment Period and Conventional Payback Time.
[145], 2017	Ag/H ₂ O	PV/T	82% reduction in house electricity prices.
[146], 2020	CuO	ETSWH	Due to enhanced ETSWH, energy is conserved, reduces the maintenance cost and heat transmission is enhanced
[149], 2018	Ag/H ₂ O	PV/T	The predicted price repayment time frame for the recovery of investments was 5.5 years.
[150], 2020	CeO ₂ /H ₂ O	FPSC	Resulted in a longer payback period of 2.12 years, and reduced CO ₂ emissions to the environment by 175 kg.
[151], 2023	GAMWCNT	FPSC	When GAMWCNT nanofluids were used at 0.1 weight percent in FPSC, the time to pay back became 1.897 years, that's 6.228% less than when H ₂ O was used

KEY STRATEGIES FOR OPTIMIZATION OF NANOFLUID PROPERTIES IN SOLAR ENERGY HARNESSING

Optimizing nanofluid properties for solar energy harnessing involves adjusting several parameters to enhance the overall efficiency of solar thermal systems. Nanofluids, which are engineered by dispersing nanoparticles into conventional base fluids like water, ethylene glycol, or oil, improve thermal conductivity, heat transfer rates, and optical absorption, which are critical for solar energy applications like solar collectors and photovoltaic thermal systems. The key strategies for optimizing nanofluid properties for solar energy harnessing are represented in Fig. 24.

Choice of Nanoparticles

Use nanoparticles with high thermal conductivity such as copper (Cu), silver (Ag), carbon nanotubes (CNT), and aluminum oxide (Al_2O_3) to enhance heat transfer. Select nanoparticles with excellent solar absorption characteristics like graphene, carbon black, or titanium dioxide (TiO_2). This makes it easier for solar radiation to be absorbed and cuts down on reflection losses. To get more surface area and less chance of particles sticking together, change the size and shape of the particles (usually between 10 and 100 nm).

Smaller particles have a larger surface-to-volume ratio, which makes them more stable and helps heat transfer. Spherical nanoparticles are usually more stable, but rod- or plate-like forms may make optical characteristics better.

Nanoparticles Concentration

Too many nanoparticles can make solar collectors less efficient by making them thicker and harder to pump. Usually, concentrations of 0.1% to 5% by volume strike a good balance between improving heat conductivity and keeping viscosity levels reasonable. Make sure that the particles are properly dispersed so that they don't clump together. This can lower the thermal performance and dirty up solar thermal equipment.

Base Fluid Selection

The thermal conductivity, specific heat, and density of the base fluid are very important. Water is the most frequent base fluid since it has great thermophysical qualities. However, for some conditions, such high-temperature settings, mixes like ethylene glycol-water or oil-based fluids can also be used. The base fluid should stay stable over a wide variety of temperatures so that it works the same way in different amounts of solar energy.



Figure 24. Key Strategies for Optimization of Nanofluid Properties in Solar Energy Harnessing.

Surfactants and Stabilizers

The clogging problem of nanofluid is improved by adding surfactant because it make nanoparticle to spread out throughout the base fluid. Addition of surfactant makes nanofluid stable over certain period of time and hence improves its characteristics and ultimately the performance of solar system is enhanced.

pH Optimization

The stability of nanofluid is measured by optimizing the Ph of nanofluid. A tinny amount of electroststic repulsion between nanoparticles makes them more stable over certain time period due to which heat transfer of nanofluid is enhanced.

Viscosity Optimization

Viscosity plays a major role in enhancement of thermal conductivity of nanofluid. Proper volume concentration of nanofluid enhances its thermal conductivity and improves the heat transfer rate and ultimately improves the solar system

Thermal Stability and Operating Temperature Range

Heat transfer in solar thermal system is enhanced due to their enhanced thermal conductivity of nanofluid. Nanofluids express excellent thermal conductivity at high temperature operations ranging from 50°C to 400°C. This temperature range depends on the various factors such as

volume fraction, type of base fluid, type of solar system and many more.

Hybrid Nanofluids for Synergy

When you mix different kinds of nanoparticles, like metal with oxide or carbon-based with metallic, you can get synergistic effects. For example, adding copper to alumina or graphene can make both thermal conductivity and optical absorption better.

Advanced Characterization Techniques

Advanced characterization techniques are very useful to know the stability condition of nanofluid. Many characterization techniques are available now these days such as SEM, TEM, dynamic light scattering (DLS) and many more to make nanofluid applicable for wide range of industrial applications.

LONG-TERM STABILITY AND SCALABILITY OF NANOFLUID

Enhanced properties of nanofluid make it stable for long term and scalable. Proper dispersion of nanoparticles within base fluid and advanced stabilization technique improves the performance of nanofluid. Over time, instability can reduce heat transfer efficiency and increase system maintenance costs [152-153]. Scalability, on the other hand, addresses the challenges of producing large volumes of nanofluids cost-effectively while maintaining consistent

Table 8. Summary of Characterization Techniques to Enhance the Stability of Nanofluid in Solar Energy Harnessing

Ref. No.	Nanomaterials	System	Particle Size	Surfactant	Characterization Technique
[66], 2016	MgO/Water	FPSC	40 nm	Cetyl Trimethyl Ammonium Bromide	TEM
[67], 2017	Al ₂ O ₃ , TiO ₂ , SiO ₂ , CuO, Graphene, MWCNT	FPSC	45 nm, 44 nm	Triton 100X	TEM
			10 nm, 42 nm		
			20 nm, 7 nm		
[69], 2021	Al ₂ O ₃ , TiO ₂	FPSC	25 nm	Triton-X100	SEM, TEM, XRD
[70], 2022	Al ₂ O ₃ , MWCNT	FPSC	30 nm, 10-40 nm	-	SEM, TEM, TGA
[72], 2024	SiC	FPSC	45-65 nm	-	XRD, TEM, FESEM
[73], 2020	MWCNT	PTC	8-15 nm	Arabic gum	Zeta Potential
[78], 2023	CuO	PTC	10 nm	-	-
[82], 2024	MWCNT	FPSC, PTC	50 nm	SDBS	Zeta Sizer
[84], 2020	Al ₂ O ₃	PT-ETSC	40 nm	-	-
[88], 2023	MWCNT, Al ₂ O ₃	ETSC	10-40 nm, <30 nm	-	TEM, XRD
[90], 2020	Chitosan, Al ₂ O ₃	FPC-SWH	15 nm	Polyvinyl pyrrolidone (PVP-K30)	-
[92], 2021	CuO-PCM	FPC-SWH	-	-	FESEM
[110], 2023	TiO ₂ and TiO ₂ -Fe ₂ O ₃	PV/T	-	-	TEM, XRD
[126], 2023	TiO ₂	Double-effect solar distiller	10 nm	-	-

quality. As production scales up, ensuring that the properties of the nanofluid—such as thermal conductivity, viscosity, and optical absorption—remain uniform becomes more difficult. Solutions include improving nanoparticle synthesis techniques and enhancing dispersion methods that are adaptable for industrial-scale operations [154-155]. Together, long-term stability and scalability are essential for making nanofluids a viable option in commercial solar energy systems and other high-performance heat transfer applications. The various characterization techniques to enhance the stability of nanofluid in solar energy harnessing are mentioned in table 8.

CONCLUSION

Using nanofluids to collect solar energy is a big step forward in the field of renewable energy. Solar systems such as solar collector, PV/T, solar still perform much better due to enhanced thermal characteristics of nanofluid. It is one of the best choices for solar energy conversion since it enhanced the heat transfer rate and ultimately enhanced the optical performance of the system. The major concern behind the utilization of nanofluid is its stability over long period of time, cost of production and environmental sustainability. If further research in this field improves these problems associates with nanofluid then definitely nanofluid will be a major player in the field of solar system to improve its performance and also this kind of research will be very helpful for future renewable system and diminish the dependency on fossils fuels.

The thermal management system of engine is also improved by utilization of nanofluid due to its enhanced properties. This result in better engine performance and emission of fuel is lowered. But nanofluids have a higher viscosity, especially when there are more nanoparticles in them, which mean they may need more energy to move around, which could cancel out any efficiency advantages. So, while nanofluids in engines could improve performance, their commercial viability depends on finding the right mix between better engine performance, cost-effectiveness, and long-term reliability.

Every researchers and scientists have believed on the utilization of nanofluid in solar system to harness its energy with enhanced efficiency and effectiveness of the system. The present research, nevertheless, took a comprehensive approach, covering a wide range of solar energy collecting equipment not before addressed in details. This investigation yields numerous investigations areas, gaps, various consequences. Furthermore, a detailed discussion on the recent development and nanofluid utilization in solar system discussed with its theoretical, experimental and also computational approach. The important findings from this review article are mentioned below:

- Maximum 85.4% gain in thermal efficiency of solar collector using graphite as TES material.

- Maximum 38.4% enhancement in efficiency of solar water using CuO/H₂O nanofluid.
- Maximum 84.74% enhancement in thermal efficiency of PV/T system SiC-H₂O with NEPCM.
- 196% enhancement in exergy efficiency of solar evaporator using Ag/H₂O nanofluid
- The utilization of multi walled carbon nanotube in all solar energy harnessing system improves the system performance. In case of solar desalination system, it improves the productivity of the system also.
- From recent research, it is observed that the utilization of hybrid nanofluid compared to mono nanofluid improves the performance of solar energy harnessing.
- Nanofluids speed up the evaporation process, which means that the solar still can purify water faster and make more distilled water in a shorter amount of time than ordinary water.
- Nanofluids can make solar stills work better even when the sun isn't shining very brightly, like on cloudy days or in the early morning and late afternoon. This makes sure that the still works well all day long.

FUTURE RESEARCH DIRECTION

Although, a lot of research has been performed on solar energy harnessing using nanofluid but still various works that need to be addressed for better performance which are mentioned below:

- More research is needed to make nanofluids more stable over time. This includes looking into problems like nanoparticles sticking together and settling to make sure they work the same way over lengthy periods of time.
- To search new technique for production nanofluid at low cost to make it more viable at commercial and industrial level utilization.
- New kind of research is also required to produce hybrid nanofluid to get maximum from solar system with highest effectiveness and efficiency.
- Characterization technique of nanomaterials must be improved to make it stable over long period of time and also environment friendly.
- Combining nanofluid with hybrid solar system is the new challenge for researcher and scientist. Future research must focus on it for enhancement of the system.
- Nanofluid preparation technique must be improved in future to make it environment friendly with low cost production to make it viable for commercial utilization.

CHALLENGES AND LIMITATIONS OF USING NANOFLUID IN SOLAR ENERGY SYSTEM

A thorough and organized review of high-quality past research has found many ways to improve the efficacy of solar energy collecting equipment. Even if the use of

nanofluids has been successful, there are still certain issues that should be taken very seriously since they restrict the use of solar energy harnessing systems that use nanofluids. The key challenges and limitations for performing research on the utilization of nanofluids in solar energy harnessing are mentioned below:

- Ensuring the stability of nanoparticles within the fluid is a major challenge, as nanoparticles tend to agglomerate or settle over time, leading to reduced effectiveness and inconsistent thermal properties.
- Nanofluids are hard to make and mix because high-quality nanoparticles are expensive and it's hard to get them to distribute evenly. This makes it hard to do a lot of research and use them in a lot of different ways.
- The behavior of nanofluids in heat transmission is intricate and not yet completely understood. Because of this complexity, it's hard to anticipate how well they'll do, which requires advanced research and a lot of money.
- Limited data exists on the environmental and health impacts of producing, using, and disposing of nanofluids, raising concerns about their safety and sustainability, which complicates securing research funding and regulatory approval.
- No standardized methods exist for preparing, testing, or evaluating nanofluids, complicating data comparison across studies and hindering the development of universal guidelines.
- Translating lab scale research findings to practical solar energy systems is challenging. For instance, nanofluid properties must be tailored to suit diverse environments and system configurations.
- There is a huge difference between optical and thermal properties of nanofluid due to concentration of base fluid, nanoparticle type etc. This difference requires very tough optimization technique and also requires much better characterization and testing.
- It is hard to make sure that nanofluids will last a long time when they are exposed to high temperatures, UV radiation, and other environmental conditions for a long time. This is because degradation can make solar energy systems less effective over time.
- The system's operation and maintenance become more complex due to the need for precise control and supervision of nanofluid characteristics.
- Experimental results can be very different from each other because of changes in how nanofluids are made, how systems are designed, and how they are used. This makes it hard to know how to use them in real life.

NOMENCLATURE

A	Solar collector area (m ²)
c _p	Specific heat capacity (kJ/kgK)
I	Solar radiation flux (W/m ²)
m ^o	Mass flow rate (kg/s)
η _{th}	Thermal efficiency

Q _u	Usable heat transfer (kW)
Q _s	Solar irradiation availability (kW)
Q _{solar}	Collector input solar energy (kW)
S ^o _{gen}	System entropy generation (kJ/K)
T _{out}	Outlet temperature (K)
T _{in}	Inlet temperature (K)
T _a	Ambient temperature (K)
T _a	Sun temperature (K)

Abbreviations

3E	Energy, Economy and Environment
4E	Energy, Exergy, Economy and Environment
BVPTM	Baffled PVT module
Bpv	Bifacial photovoltaic
bPV/T	Bifacial photovoltaic-thermal
CTAB	Cetyltrimethylammonium bromide
CHGR	Conical helical gear rings
DSC	Differential Scanning Calorimetry
DESD	Double-effect solar distiller
DTPTSWH	Dimple tube parabolic trough solar water heater
EG	Ethylene glycol
ETSC	Evacuated tube solar collector
ETSWH	Evacuated tube solar water heater
FPSC	Flat plate solar collector
FPC-SWH	Flat plate collector- Solar water heater
FESEM	Field emission scanning electron microscopy
GAMWCNT	Gallic acid-treated multiwall carbon nanotubes
GNP	Graphene nanoplatelets
GO	Graphene Oxide
HTC	Heat transfer coefficient
JPTN	Jackfruit peel TiO ₂ nanofluids
MWCNT	Multi walled carbon nanotube
Nano-PCM	Nanoparticle-enhanced phase change material
NPCM	Nano phase change material
PCM	Phase change material
PTC	Parabolic trough collector
PV/T	Photovoltaic thermal system
PVS	Photovoltaic solar
PV	Photovoltaic
PT-ETSC	Parallel type evacuated tube solar collector
SODIS	Solar water desalination
SBGIDSS	Single-basin glasswool-insulated double-slope solar still
SSPSS	Single-slope passive solar still
SDBS	Sodium dodecyl benzene sulfonate
SWCNT	Single walled carbon nanotube
SWHS-ZO-NF	Solar water heating system using zinc oxide-based nanofluid
SHC	Specific heat capacity
STS	Solar thermal system
SWH	Solar water heater
SEM	Scanning electron microscopy

TEM	Transmission electron microscopes
TC	Thermal conductivity
TES	Thermal energy storage
TEG	Thermoelectric generator
TGA	Thermal gravimetric analysis
XRD	X-ray diffraction

AUTHORSHIP CONTRIBUTION

All the 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 manuscript.

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] Holechek JL, Geli HM, Sawalhah MN, Valdez R. A global assessment: can renewable energy replace fossil fuels by 2050? *Sustainability* 2022;14:4792. [\[CrossRef\]](#)
- [2] Kumar S, Rawat MK, Gupta S. An evaluation of current status of renewable energy sources in India. *Int J Innov Technol Explor Eng* 2019;8:1234–1239. [\[CrossRef\]](#)
- [3] Kannan N, Vakeesan D. Solar energy for future world: a review. *Renew Sustain Energy Rev* 2016;62:1092–1105. [\[CrossRef\]](#)
- [4] Yavuz C, Botsalı AR. A weather-based forecasting system for the solar power plants in the Konya region. *Sigma J Eng Nat Sci* 2022;40:144–149. [\[CrossRef\]](#)
- [5] Qadem M, Filik ÜB. Solar radiation forecasting by using deep neural networks in Eskişehir. *Sigma J Eng Nat Sci* 2021;39:159–169. [\[CrossRef\]](#)
- [6] International Energy Agency. World energy outlook 2023. <https://www.iea.org/reports/world-energy-outlook-2023>
- [7] Kabir E, Kumar P, Kumar S, Adelodun AA, Kim KH. Solar energy: potential and future prospects. *Renew Sustain Energy Rev* 2018;82:894–900. [\[CrossRef\]](#)
- [8] Kumar N, Gupta SK. Progress and application of phase change material in solar thermal energy: an overview. *Mater Today Proc* 2021;44:271–281. [\[CrossRef\]](#)
- [9] Kumar S, Gupta SK, Rawat M. Resources and utilization of geothermal energy in India: an eco-friendly approach towards sustainability. *Mater Today Proc* 2020;26:1660–1665. [\[CrossRef\]](#)
- [10] Gupta SK, Pradhan S. A review of recent advances and the role of nanofluid in solar photovoltaic thermal (PV/T) system. *Mater Today Proc* 2021;44:782–791. [\[CrossRef\]](#)
- [11] Gupta SK, Rawat MK, Kukreja N. Analysis of heat transfer enhancement of electronic chip using CFD. *Int J Sci Technol Res* 2019;8:1017–1020.
- [12] Kumar N, Gupta SK, Sharma VK. Application of phase change material for thermal energy storage: an overview of recent advances. *Mater Today Proc* 2021;44:368–375. [\[CrossRef\]](#)
- [13] Mahian O, Kolsi L, Amani M, Estellé P, Ahmadi G, Kleinstreuer C, et al. Recent advances in modeling and simulation of nanofluid flows – Part I: fundamentals and theory. *Phys Rep* 2019;790:1–48. [\[CrossRef\]](#)
- [14] Li Y, Tung S, Schneider E, Xi S. A review on development of nanofluid preparation and characterization. *Powder Technol* 2009;196:89–101. [\[CrossRef\]](#)
- [15] 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. [\[CrossRef\]](#)
- [16] Gupta SK, Gupta S, Gupta T, Raghav A, Singh A. A review on recent advances and applications of nanofluids in plate heat exchanger. *Mater Today Proc* 2021;44:229–241. [\[CrossRef\]](#)
- [17] Cuce E, Cuce PM, Guclu T, Besir AB. On the use of nanofluids in solar energy applications. *J Therm Sci* 2020;29:513–534. [\[CrossRef\]](#)
- [18] Akaje TW, Olajuwon BI, Musiliu Tayo RAJI. Computational analysis of the heat and mass transfer in a Casson nanofluid with a variable inclined magnetic field. *Sigma* 2023;41:512–523. [\[CrossRef\]](#)
- [19] Ahmadi MH, Ghazvini M, Sadeghzadeh M, Nazari MA, Ghalandari M. Utilization of hybrid nanofluids in solar energy applications: a review. *Nano Struct Nano Objects* 2019;20:100386. [\[CrossRef\]](#)
- [20] 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;37:340–360.
- [21] Gupta SK, Verma H, Yadav N. A review on recent development of nanofluid utilization in shell & tube heat exchanger for saving of energy. *Mater Today Proc* 2022;54:579–589. [\[CrossRef\]](#)
- [22] Hamzat AK, Omisanya MI, Sahin AZ, Oyetunji OR, Olaitan NR. Application of nanofluid in solar energy harvesting devices: a comprehensive review. *Energy Convers Manag* 2022;266:115790. [\[CrossRef\]](#)

- [23] Zhang Y, Liu L, Li K, Hou D, Wang J. Enhancement of energy utilization using nanofluid in solar powered membrane distillation. *Chemosphere* 2018;212:554–562. [\[CrossRef\]](#)
- [24] Ahmadi MH, Ghazvini M, Alhuyi Nazari M, Ahmadi MA, Pourfayaz F, Lorenzini G, et al. Renewable energy harvesting with the application of nanotechnology: a review. *Int J Energy Res* 2019;43:1387–1410. [\[CrossRef\]](#)
- [25] Sahin AZ, Uddin MA, Yilbas BS, Al-Sharafi A. Performance enhancement of solar energy systems using nanofluids: an updated review. *Renew Energy* 2020;145:1126–1148. [\[CrossRef\]](#)
- [26] Mausam K, Kumar S, Ghosh SK, Tiwari AK, Sehgal M. Solicitation of nanoparticles/fluids in solar thermal energy harvesting: a review. *Mater Today Proc* 2020;26:2289–2295. [\[CrossRef\]](#)
- [27] Hu G, Ning X, Hussain M, Sajjad U, Sultan M, Ali HM, et al. Potential evaluation of hybrid nanofluids for solar thermal energy harvesting: a review of recent advances. *Sustain Energy Technol Assess* 2021;48:101651. [\[CrossRef\]](#)
- [28] Hao D, Qi L, Tairab AM, Ahmed A, Azam A, Luo D, Pan Y, et al. Solar energy harvesting technologies for PV self-powered applications: a comprehensive review. *Renew Energy* 2022;188:678–697. [\[CrossRef\]](#)
- [29] Rao CNR, Cheetham AK. Science and technology of nanomaterials: current status and future prospects. *J Mater Chem* 2001;11:2887–2894. [\[CrossRef\]](#)
- [30] Rizwan M, Shoukat A, Ayub A, Razzaq B, Tahir MB. Types and classification of nanomaterials. In: *Nanomaterials: synthesis, characterization, hazards and safety*. 2021. p. 31–54. [\[CrossRef\]](#)
- [31] Bodunde OP, Ikumapayi OM, Akinlabi ET, Oladapo BI, Adeoye AOM, Fatoba SO, et al. A futuristic insight into a “nano-doctor”: a clinical review on medical diagnosis and devices using nanotechnology. *Mater Today Proc* 2021;44:1144–1153. [\[CrossRef\]](#)
- [32] Kukreja N, Gupta SK, Rawat M. Performance analysis of phase change material using energy storage device. *Mater Today Proc* 2020;26:913–917. [\[CrossRef\]](#)
- [33] Rawat MK, Kukreja N, Gupta SK. Effect of reinforcing micro sized aluminium oxide particles on mechanical properties of polymer based composite. *Mater Today Proc* 2020;26:1306–1309. [\[CrossRef\]](#)
- [34] Lee JH, Hwang KS, Jang SP, Lee BH, Kim JH, Choi SUS, et al. Effective viscosities and thermal conductivities of aqueous nanofluids containing low volume concentrations of Al_2O_3 nanoparticles. *Int J Heat Mass Transf* 2008;51:2651–2666. [\[CrossRef\]](#)
- [35] Gupta SK, Dixit S. Progress and application of nanofluids in solar collectors: an overview of recent advances. *Mater Today Proc* 2021;44:250–259. [\[CrossRef\]](#)
- [36] Dmitriev AS, Klimenko AV. Prospects for the use of two-dimensional nanomaterials in energy technologies. *Therm Eng* 2023;70:551–572. [\[CrossRef\]](#)
- [37] Younes H, Mao M, Murshed SS, Lou D, Hong H, Peterson GP. Nanofluids: key parameters to enhance thermal conductivity and its applications. *Appl Therm Eng* 2022;207:118202. [\[CrossRef\]](#)
- [38] Bolarin G, Yusuf A, Adekunle ST, Ayiesimi YM, Jiya M. Analysis of a boundary layer flow of a nanofluid over an inclined plane via ADM. *Sigma J Eng Nat Sci* 2019;37:475–488.
- [39] Arshad A, Jabbar M, Yan Y, Reay D. A review on graphene based nanofluids: preparation, characterization and applications. *J Mol Liq* 2019;279:444–484. [\[CrossRef\]](#)
- [40] Gupta SK, Gupta S. The role of nanofluids in solar thermal energy: a review of recent advances. *Mater Today Proc* 2021;44:401–412. [\[CrossRef\]](#)
- [41] Gupta M, Singh V, Kumar R, Said Z. A review on thermophysical properties of nanofluids and heat transfer applications. *Renew Sustain Energy Rev* 2017;74:638–670. [\[CrossRef\]](#)
- [42] Gupta SK, Sharma A. A brief review of nanofluids utilization in heat transfer devices for energy saving. *Mater Today Proc* 2023. doi:10.1016/j.matpr.2023.03.364 [Epub ahead of print] [\[CrossRef\]](#)
- [43] Khanafer K, Vafai K. A review on the applications of nanofluids in solar energy field. *Renew Energy* 2018;123:398–406. [\[CrossRef\]](#)
- [44] Mahian O, Kianifar A, Kalogirou SA, Pop I, Wongwises S. A review of the applications of nanofluids in solar energy. *Int J Heat Mass Transf* 2013;57:582–594. [\[CrossRef\]](#)
- [45] Elango T, Kannan A, Murugavel KK. Performance study on single basin single slope solar still with different water nanofluids. *Desalination* 2015;360:45–51. [\[CrossRef\]](#)
- [46] Zhu D, Li X, Wang N, Wang X, Gao J, Li H. Dispersion behavior and thermal conductivity characteristics of Al_2O_3 - H_2O nanofluids. *Curr Appl Phys* 2009;9:131–139. [\[CrossRef\]](#)
- [47] Mao M, Lou D, Wang D, Younes H, Hong H, Chen H, Peterson GP. $\text{Ti}_3\text{C}_2\text{Tx}$ MXene nanofluids with enhanced thermal conductivity. *Chem Thermodyn Therm Anal* 2022;8:100077. [\[CrossRef\]](#)
- [48] Gupta SK, Gupta S, Singh R. A comprehensive review of energy saving in shell & tube heat exchanger by utilization of nanofluids. *Mater Today Proc* 2022;50:1818–1826. [\[CrossRef\]](#)
- [49] Koca HD, Doganay S, Turgut A, Tavman IH, Saidur R, Mahbubul IM. Effect of particle size on the viscosity of nanofluids: a review. *Renew Sustain Energy Rev* 2018;82:1664–1674. [\[CrossRef\]](#)
- [50] Mishra PC, Mukherjee S, Nayak SK, Panda A. A brief review on viscosity of nanofluids. *Int Nano Lett* 2014;4:109–120. [\[CrossRef\]](#)
- [51] Chandrasekar M, Suresh S, Bose AC. Experimental investigations and theoretical determination of thermal conductivity and viscosity of Al_2O_3 /water nanofluid. *Exp Therm Fluid Sci* 2010;34:210–216. [\[CrossRef\]](#)

- [52] Mahesh KV, Linsha V, Peer Mohamed A, Ananthakumar S. Processing of 2D-MAXene nanostructures and design of high thermal conducting, rheo-controlled MAXene nanofluids as a potential nanocoolant. *Chem Eng J* 2016;297:158–169. [\[CrossRef\]](#)
- [53] Adun H, Wole-Osho I, Okonkwo EC, Kavaz D, Dagbasi M. A critical review of specific heat capacity of hybrid nanofluids for thermal energy applications. *J Mol Liq* 2021;340:116890. [\[CrossRef\]](#)
- [54] Gupta SK, Saxena A. A progressive review of hybrid nanofluid utilization in solar parabolic trough collector. *Mater Today Proc* 2023 doi: 10.1016/j.matpr.2023.06.204. [\[CrossRef\]](#)
- [55] Rubbi F, Habib K, Saidur R, Aslfattahi N, Yahya SM, Das L. Performance optimization of a hybrid PV/T solar system using soybean oil/MXene nanofluids as a new class of heat transfer fluids. *Sol Energy* 2020;208:124–138. [\[CrossRef\]](#)
- [56] Zhou LP, Wang BX, Peng XF, Du XZ, Yang YP. On the specific heat capacity of CuO nanofluid. *Adv Mech Eng* 2010;2:172085. [\[CrossRef\]](#)
- [57] Chakraborty S, Panigrahi PK. Stability of nanofluid: a review. *Appl Therm Eng* 2020;174:115259. [\[CrossRef\]](#)
- [58] Wang J, Yang X, Klemeš JJ, Tian K, Ma T, Sunden B. A review on nanofluid stability: preparation and application. *Renew Sustain Energy Rev* 2023;188:113854. [\[CrossRef\]](#)
- [59] Dey D, Kumar P, Samantaray S. A review of nanofluid preparation, stability, and thermo-physical properties. *Heat Transf Asian Res* 2017;46:1413–1442. [\[CrossRef\]](#)
- [60] Nikolov A, Wu P, Wasan D. Structure and stability of nanofluid films wetting solids: an overview. *Adv Colloid Interface Sci* 2019;264:1–10. [\[CrossRef\]](#)
- [61] Kumar MS, Vasu V, Gopal AV. Thermal conductivity and rheological studies for Cu–Zn hybrid nanofluids with various base fluids. *J Taiwan Inst Chem Eng* 2016;66:321–327. [\[CrossRef\]](#)
- [62] Abdelrazik AS, Tan KH, Aslfattahi N, Arifutzzaman A, Saidur R, Al-Sulaiman FA. Optical, stability and energy performance of water-based MXene nanofluids in hybrid PV/thermal solar systems. *Sol Energy* 2020;204:32–47. [\[CrossRef\]](#)
- [63] Tian Y, Zhao CY. A review of solar collectors and thermal energy storage in solar thermal applications. *Appl Energy* 2013;104:538–553. [\[CrossRef\]](#)
- [64] Ghritleahre HK, Prasad RK. Application of ANN technique to predict the performance of solar collector systems: a review. *Renew Sustain Energy Rev* 2018;84:75–88. [\[CrossRef\]](#)
- [65] Gupta SK. A short and updated review of nanofluids utilization in solar parabolic trough collector. *Mater Today Proc* 2023. [\[CrossRef\]](#)
- [66] Verma SK, Tiwari AK, Chauhan DS. Performance augmentation in flat plate solar collector using MgO/water nanofluid. *Energy Convers Manag* 2016;124:607–617. [\[CrossRef\]](#)
- [67] Verma SK, Tiwari AK, Chauhan DS. Experimental evaluation of flat plate solar collector using nanofluids. *Energy Convers Manag* 2017;134:103–115. [\[CrossRef\]](#)
- [68] Saffarian MR, Moravej M, Doranehgard MH. Heat transfer enhancement in a flat plate solar collector with different flow path shapes using nanofluid. *Renew Energy* 2020;146:2316–2329. [\[CrossRef\]](#)
- [69] Gad MS, Said M, Hassan AY. Effect of different nanofluids on performance analysis of flat plate solar collector. *J Dispers Sci Technol* 2021;42:1867–1878. [\[CrossRef\]](#)
- [70] Elshazly E, Abdel-Rehim AA, El-Mahallawi I. 4E study of experimental thermal performance enhancement of flat plate solar collectors using MWCNT, Al₂O₃, and hybrid MWCNT/Al₂O₃ nanofluids. *Results Eng* 2022;16:100723. [\[CrossRef\]](#)
- [71] Desisa TR. Experimental and numerical investigation of heat transfer characteristics in solar flat plate collector using nanofluids. *Int J Thermofluids* 2023;18:100325. [\[CrossRef\]](#)
- [72] Ajeena AM, Farkas I, Víg P. Energy and exergy assessment of a flat plate solar thermal collector by examining silicon carbide nanofluid: an experimental study for sustainable energy. *Appl Therm Eng* 2024;236:121844. [\[CrossRef\]](#)
- [73] Hachicha AA, Said Z, Rahman SMA, Al-Sarairah E. On the thermal and thermodynamic analysis of parabolic trough collector technology using industrial-grade MWCNT based nanofluid. *Renew Energy* 2020;161:1303–1317. [\[CrossRef\]](#)
- [74] Bellos E, Tzivanidis C, Said Z. A systematic parametric thermal analysis of nanofluid-based parabolic trough solar collectors. *Sustain Energy Technol Assess* 2020;39:100714. [\[CrossRef\]](#)
- [75] Ekiciler R, Arslan K, Turgut O, Kurşun B. Effect of hybrid nanofluid on heat transfer performance of parabolic trough solar collector receiver. *J Therm Anal Calorim* 2021;143:1637–1654. [\[CrossRef\]](#)
- [76] Farooq M, Farhan M, Ahmad G, Tahir ZR, Usman M, Sultan M, Hanif MS, et al. Thermal performance enhancement of nanofluids based parabolic trough solar collector (NPTSC) for sustainable environment. *Alex Eng J* 2022;61:8943–8953. [\[CrossRef\]](#)
- [77] Vahedi B, Golab E, Sadr AN, Vafai K. Thermal, thermodynamic and exergoeconomic investigation of a parabolic trough collector utilizing nanofluids. *Appl Therm Eng* 2022;206:118117. [\[CrossRef\]](#)
- [78] Ram S, Ganesan H, Saini V, Kumar A. Performance assessment of a parabolic trough solar collector using nanofluid and water based on direct absorption. *Renew Energy* 2023;214:11–22. [\[CrossRef\]](#)
- [79] Hamada MA, Khalil H, Abou Al-Sood MM, Sharshir SW. An experimental investigation of nanofluid, nanocoating, and energy storage materials on the performance of parabolic trough collector. *Appl Therm Eng* 2023;219:119450. [\[CrossRef\]](#)

- [80] Ekiciler R, Arslan K, Turgut O. Application of nanofluid flow in entropy generation and thermal performance analysis of parabolic trough solar collector: experimental and numerical study. *J Therm Anal Calorim* 2023;148:7299–7318. [\[CrossRef\]](#)
- [81] Esmaili Z, Akbarzadeh S, Rashidi S, Valipour MS. Effects of hybrid nanofluids and turbulator on efficiency improvement of parabolic trough solar collectors. *Eng Anal Bound Elem* 2023;148:114–125. [\[CrossRef\]](#)
- [82] Abu-Zeid MAR, Elhenawy Y, Bassyouni M, Majazi T, Toderas M, Al-Qabandi OA, et al. Performance enhancement of flat-plate and parabolic trough solar collector using nanofluid for water heating application. *Results Eng* 2024;21:101673. [\[CrossRef\]](#)
- [83] Yurddaş A. Optimization and thermal performance of evacuated tube solar collector with various nanofluids. *Int J Heat Mass Transf* 2020;152:119496. [\[CrossRef\]](#)
- [84] Sasikumar SB, Santhanam H, Noor MM, Devasenan M, Ali HM. Experimental investigation of parallel type-evacuated tube solar collector using nanofluids. *Energy Sources Part A Recover Util Environ Eff* 2020;1–13. [\[CrossRef\]](#)
- [85] Kumar S, Tiwari AK. Performance evaluation of evacuated tube solar collector using boron nitride nanofluid. *Sustain Energy Technol Assess* 2022;53:102466. [\[CrossRef\]](#)
- [86] Henein SM, Abdel-Rehim AA, El-Nagar K. Energy, economic and environmental analysis of an evacuated tube solar collector using hybrid nanofluid. *Appl Therm Eng* 2023;219:119671. [\[CrossRef\]](#)
- [87] Tong Y, Wang R, Wang S, Wang H, Huang Li Shao C, et al. Comparison and evaluation of energetic and exergetic performance of an evacuated tube solar collector using various nanofluid. *Process Saf Environ Prot* 2023;174:585–594. [\[CrossRef\]](#)
- [88] Elshazly E, El-Mahallawi I. Thermal performance enhancement of evacuated tube solar collector using MWCNT, Al₂O₃, and hybrid MWCNT/Al₂O₃ nanofluids. *Int J Thermofluids* 2023;17:100260. [\[CrossRef\]](#)
- [89] López-Pérez LA, Torres-Díaz T, Pérez Grajales SG, Flores Prieto JJ, Romero DJ, Hernández Pérez JA, et al. Solar water heating system with absorption heat transformer for annual continuous water heating. *Processes* 2024;12:1650. [\[CrossRef\]](#)
- [90] Modi K, Shukla D, Bhargav B, Devaganiya J, Deshle R, Dhodi J, et al. Efficacy of organic and inorganic nanofluid on thermal performance of solar water heating system. *Cleaner Eng Technol* 2020;1:100020. [\[CrossRef\]](#)
- [91] Darbari B, Rashidi S. Thermal efficiency of flat plate thermosyphon solar water heater with nanofluids. *J Taiwan Inst Chem Eng* 2021;128:276–287. [\[CrossRef\]](#)
- [92] Mandal SK, Singh PK, Kumar S, Mishra SK. Parametric investigation of CuO-doped charged nanofluid in solar water heater. *Int J Environ Sci Technol* 2021;18:2855–2864. [\[CrossRef\]](#)
- [93] Shafiee M, Farbeh A, Firoozzadeh M. Experimental study on using oil-based nanofluids in a vacuumed tube solar water heater: an exergy analysis. *Int J Ambient Energy* 2022;43:7387–7399. [\[CrossRef\]](#)
- [94] Arun M. Experimental investigation on energy and exergy analysis of solar water heating system using zinc oxide-based nanofluid. *Arab J Sci Eng* 2023;48:3977–3988. [\[CrossRef\]](#)
- [95] Firoozzadeh M, Shafiee M. Thermodynamic analysis on using titanium oxide/oil nanofluid integrated with porous medium in an evacuated tube solar water heater. *J Therm Anal Calorim* 2023;148:8309–8322. [\[CrossRef\]](#)
- [96] Lotfi M, Firoozzadeh M, Ali-Sinaei M. Simultaneous use of TiO₂/oil nanofluid and metallic-insert as enhancement of an evacuated tube solar water heater. *J Therm Anal Calorim* 2023;148:9633–9647. [\[CrossRef\]](#)
- [97] Nishit J, Bekal S. Experimental investigation on polymer solar water heater using Al₂O₃ nanofluid for performance improvement. *Mater Today Proc* 2023;92:249–257. [\[CrossRef\]](#)
- [98] Nayak S, Hassan MA, Paswan M. Investigation on thermo-fluid behavior of flat-plate solar water heater employing titanium dioxide–water nanofluid: an experimental and computational approach. *J Mech Sci Technol* 2024;38:1507–1516. [\[CrossRef\]](#)
- [99] Rashmi M, Sagar PS, Devi NL, et al. Analyzing performance of a solar water heater using nano-ZrO₂/water-based nanofluid. *Mater Today Proc* 2023. doi: 10.1016/j.matpr.2023.09.012 [Epub ahead of print] [\[CrossRef\]](#)
- [100] Arun M, Barik D, Sharma P, Gürel AE, Ağbulut Ü, Medhi BJ, et al. Experimental and CFD analysis of dimple tube parabolic trough solar water heater with various nanofluids. *Appl Nanosci* 2024;14:291–337. [\[CrossRef\]](#)
- [101] Chow TT. A review on photovoltaic/thermal hybrid solar technology. *Renew Energy* 2018;4:88–119. [\[CrossRef\]](#)
- [102] Al-Waeli AH, Kazem HA, Chaichan MT, Sopian K. A review of photovoltaic thermal systems: achievements and applications. *Int J Energy Res* 2021;45:1269–1308. [\[CrossRef\]](#)
- [103] Mostakim K, Hasanuzzaman M. Global prospects, challenges and progress of photovoltaic thermal system. *Sustain Energy Technol Assess* 2022;53:102426. [\[CrossRef\]](#)
- [104] Salari A, Kazemian A, Ma T, Hakkaki-Fard A, Peng J. Nanofluid based photovoltaic thermal systems integrated with phase change materials: numerical simulation and thermodynamic analysis. *Energy Convers Manag* 2020;205:112384. [\[CrossRef\]](#)
- [105] Adun H, Adedeji M, Dagbasi M, Bamisile O, Senol M, Kumar R. A numerical and exergy analysis of the effect of ternary nanofluid on performance of photovoltaic thermal collector. *J Therm Anal Calorim* 2021;145:1413–1429. [\[CrossRef\]](#)

- [106] Arifin Z, Prasetyo SD, Tjahjana DDDP, Rachmanto RA, Prabowo AR, Alfaiz NF, et al. The application of TiO₂ nanofluids in photovoltaic thermal collector systems. *Energy Rep* 2022;8:1371–1380. [\[CrossRef\]](#)
- [107] Aslfattahi N, Samylingam, L, Abdelrazik AS, Arifutzzaman A, Saidur R. MXene based new class of silicone oil nanofluids for the performance improvement of concentrated photovoltaic thermal collector. *Sol Energy Mater Sol Cells* 2020;211:110526. [\[CrossRef\]](#)
- [108] Sheikholeslami M, Khalili Z. Solar photovoltaic-thermal system with novel design of tube containing eco-friendly nanofluid. *Renew Energy* 2024;222:119862. [\[CrossRef\]](#)
- [109] Al-Aasam AB, Ibrahim A, Sopian K, Abdulsahib B, Dayer M. Nanofluid-based photovoltaic thermal solar collector with nanoparticle-enhanced phase change material (nano-PCM) and twisted absorber tubes. *Case Stud Therm Eng* 2023;49:103299. [\[CrossRef\]](#)
- [110] Alktranee M, Shehab MA, Németh Z, Bencs P, Hernadi K. Thermodynamic analysis of mono and hybrid nanofluid effect on the photovoltaic-thermal system performance: a comparative study. *Heliyon* 2023;9:e22535. [\[CrossRef\]](#)
- [111] Sheikholeslami M, Khalili Z, Scardi P, Ataollahi N. Environmental and energy assessment of photovoltaic-thermal system combined with a reflector supported by nanofluid filter and a sustainable thermoelectric generator. *J Clean Prod* 2024;438:140659. [\[CrossRef\]](#)
- [112] Ahmadinejad M, Moosavi R. Energy and exergy evaluation of a baffled-nanofluid-based photovoltaic thermal system (PVT). *Int J Heat Mass Transf* 2023;203:123775. [\[CrossRef\]](#)
- [113] Bassam AM, Sopian K, Ibrahim A, Fauzan MF, Al-Aasam AB, Abusaibaa GY. Experimental analysis for the photovoltaic thermal collector (PVT) with nano PCM and micro-fins tube nanofluid. *Case Stud Therm Eng* 2023;41:102579. [\[CrossRef\]](#)
- [114] Gu W, Wang X, Lu H, Mutailipu M. Energy and exergy analyses of a bifacial photovoltaic/thermal system with nanofluids. *Sol Energy* 2023;262:111875. [\[CrossRef\]](#)
- [115] Shen T, Xie H, Gavurová B, Sangeetha M, Karthikeyan C, Praveenkumar TR, et al. Experimental analysis of photovoltaic thermal system assisted with nanofluids for efficient electrical performance and hydrogen production through electrolysis. *Int J Hydrogen Energy* 2023;48:21029–21037. [\[CrossRef\]](#)
- [116] Khalili Z, Sheikholeslami M. Analyzing the effect of confined jet impingement on efficiency of photovoltaic thermal solar unit equipped with thermoelectric generator in existence of hybrid nanofluid. *J Clean Prod* 2023;406:137063. [\[CrossRef\]](#)
- [117] Xia X, Cao X, Li N, Yu B, Liu H. Study on a spectral splitting photovoltaic/thermal system based on CNT/Ag mixed nanofluids. *Energy* 2023;271:127093. [\[CrossRef\]](#)
- [118] Mohiuddin SA, Kaviti AK, Rao TS, Sikarwar VS. Historic review and recent progress in internal design modification in solar stills. *Environ Sci Pollut Res* 2022;29:38825–38878. [\[CrossRef\]](#)
- [119] Chekifi T, Boukraa M. Solar still productivity improvement using nanofluids: a comprehensive review. *Int J Ambient Energy* 2023;44:1396–1416. [\[CrossRef\]](#)
- [120] Kaviti AK, Yadav A, Shukla A. Inclined solar still designs: a review. *Renew Sustain Energy Rev* 2016;54:429–451. [\[CrossRef\]](#)
- [121] Hussein AK, Rashid FL, Rasul MK, Basem A, Younis O, Homod RZ, et al. A review of the application of hybrid nanofluids in solar still energy systems and guidelines for future prospects. *Sol Energy* 2024;272:112485. [\[CrossRef\]](#)
- [122] Gunay T, Gumus C, Sahin AZ. The impact of using nanofluid on the performance of solar stills: a comprehensive review. *Process Saf Environ Prot* 2024;189:1464–1516. [\[CrossRef\]](#)
- [123] Kumar M, Patel SK, Mishra V, Singh D, Giri BS, Singh D. Performance analysis of modified solar still with parabolic reflector and nanofluids. *J Taiwan Inst Chem Eng* 2024;105651. [\[CrossRef\]](#)
- [124] Azam F, Akhtar N. Optimizing the performance of solar still with Al₂O₃ nanofluid: an experimental study. *Int J Energy Water Resour* 2024;9:941–954. [\[CrossRef\]](#)
- [125] Sahu R, Tiwari AC. Performance enhancement of single slope solar still using nanofluids at different water depth. *Desalination Water Treat* 2024;317:100046. [\[CrossRef\]](#)
- [126] Kumar R, Chanda J, Elsheikh AH, Ongar B, Khidolda Y, Praveen Kumar S, et al. Performance improvement of single and double effect solar stills with silver balls/nanofluids for bioactivation: an experimental analysis. *Sol Energy* 2023;259:452–463. [\[CrossRef\]](#)
- [127] Parsa SM, Rahbar A, Koleini MH, Javadi YD, Afrand M, Rostami S, et al. First approach on nanofluid-based solar still in high altitude for water desalination and solar water disinfection (SODIS). *Desalination* 2020;491:114592. [\[CrossRef\]](#)
- [128] El-Gazar EF, Zahra WK, Hassan H, Rabia SI. Fractional modeling for enhancing the thermal performance of conventional solar still using hybrid nanofluid: energy and exergy analysis. *Desalination* 2021;503:114847. [\[CrossRef\]](#)
- [129] Rabbi HMF, Sahin AZ. Performance improvement of solar still by using hybrid nanofluids. *J Therm Anal Calorim* 2021;143:1345–1360. [\[CrossRef\]](#)
- [130] Shoeibi S, Kargarsharifabad H, Rahbar N, Ahmadi G, Safaei MR. Performance evaluation of a solar still using hybrid nanofluid glass cooling – CFD simulation and environmental analysis. *Sustain Energy Technol Assess* 2022;49:101728. [\[CrossRef\]](#)
- [131] Toosi SSA, Goshayeshi HR, Zahmatkesh I, Nejati V. Experimental assessment of new designed stepped

- solar still with Fe_3O_4 +graphene oxide+paraffin as nanofluid under constant magnetic field. *J Energy Storage* 2023;62:106795. [\[CrossRef\]](#)
- [132] Naveenkumar R, Shanmugam S, Veerappan A. Performance and exergy analysis of solar-operated vacuum fan and external condenser integrated double-slope solar still using various nanofluids. *Environ Sci Pollut Res* 2023;30:12883–12902. [\[CrossRef\]](#)
- [133] Kajal G, Malik P, Garg H, Lamba R. Thermophysical properties analysis of Al_2O_3 , MgO and GO nanofluids with water for solar still. *Mater Today Proc* 2023. doi: 10.1016/j.matpr.2023.06.383 [Epub ahead of print] [\[CrossRef\]](#)
- [134] Almohammadi BA, Alharthi MA, Alshareef RS, Sharafeldin MA, Refaey HA, Abd El-Ghany HA. Effects of hybrid ZnO-WO_3 /water nanofluid on the performance of active solar still equipped with a heat exchanger. *J Therm Anal Calorim* 2024;149:8631–8645. [\[CrossRef\]](#)
- [135] Kasaeian A, Eshghi AT, Sameti M. A review on the applications of nanofluids in solar energy systems. *Renew Sustain Energy Rev* 2015;43:584–598. [\[CrossRef\]](#)
- [136] Rasheed T, Hussain T, Anwar MT, et al. Hybrid nanofluids as renewable and sustainable colloidal suspensions for potential photovoltaic/thermal and solar energy applications. *Front Chem* 2021;9:737033. [\[CrossRef\]](#)
- [137] Elsheikh AH, Sharshir SW, Mostafa ME, Essa FA, Ali MKA. Applications of nanofluids in solar energy: a review of recent advances. *Renew Sustain Energy Rev* 2018;82:3483–3502. [\[CrossRef\]](#)
- [138] Mahian O, Bellos E, Markides CN, Taylor RA, Alagumalai A, Yang L, et al. Recent advances in using nanofluids in renewable energy systems and the environmental implications of their uptake. *Nano Energy* 2021;86:106069. [\[CrossRef\]](#)
- [139] Alghamdi H, Maduabuchi C, Okoli K, Alanazi M, Fagehi H, Alghassab M, et al. Harnessing solar power: innovations in nanofluid-cooled segmented thermoelectric generators for exergy, economic, environmental, and thermo-mechanical excellence. *Alex Eng J* 2024;106:147–163. [\[CrossRef\]](#)
- [140] Mashhadian A, Heyhat MM, Mahian O. Improving environmental performance of a direct absorption parabolic trough collector by using hybrid nanofluids. *Energy Convers Manag* 2021;244:114450. [\[CrossRef\]](#)
- [141] Abadeh A, Rejeb O, Sardarabadi M, Menezo C, Passandideh-Fard M, Jemni A. Economic and environmental analysis of using metal-oxides/water nanofluid in photovoltaic thermal systems (PVTs). *Energy* 2018;159:1234–1243. [\[CrossRef\]](#)
- [142] Said Z, Ghodbane M, Boumeddane B, Tiwari AK, Sundar LS, Li C, et al. Energy, exergy, economic and environmental (4E) analysis of a parabolic trough solar collector using MXene based silicone oil nanofluids. *Sol Energy Mater Sol Cells* 2022;239:111633. [\[CrossRef\]](#)
- [143] Hotaa SK, Mata-Torres C, Cardemil JM, Diaz G. Techno-economic assessment of carbon-based nanofluid dispersions in solar stills for rural coastal locations in the Northern and Southern hemispheres. *Desalin Water Treat* 2022;245:72–84. [\[CrossRef\]](#)
- [144] Salari A, Taheri A, Farzanehnia A, Passandideh-Fard M, Sardarabadi M. An updated review of the performance of nanofluid-based photovoltaic thermal systems from energy, exergy, economic, and environmental (4E) approaches. *J Clean Prod* 2021;282:124318. [\[CrossRef\]](#)
- [145] Lari MO, Sahin AZ. Design, performance and economic analysis of a nanofluid-based photovoltaic/thermal system for residential applications. *Energy Convers Manag* 2017;149:467–484. [\[CrossRef\]](#)
- [146] Saxena G, Gaur MK. Energy, exergy and economic analysis of evacuated tube solar water heating system integrated with heat exchanger. *Mater Today Proc* 2020;28:2452–2462. [\[CrossRef\]](#)
- [147] Sohani A, Shahveridian MH, Sayyaadi H, Samiezadeh S, Doranehgard MH, Nizetic S, et al. Selecting the best nanofluid type for a photovoltaic thermal (PV/T) system based on reliability, efficiency, energy, economic, and environmental criteria. *J Taiwan Inst Chem Eng* 2021;124:351–358. [\[CrossRef\]](#)
- [148] Hussain MI, Kim J. Conventional fluid- and nanofluid-based photovoltaic thermal (PV/T) systems: a techno-economic and environmental analysis. *Int J Green Energy* 2018;15:596–604. [\[CrossRef\]](#)
- [149] Lari MO, Sahin AZ. Effect of retrofitting a silver/water nanofluid-based photovoltaic/thermal (PV/T) system with a PCM-thermal battery for residential applications. *Renew Energy* 2018;122:98–107. [\[CrossRef\]](#)
- [150] Michael Joseph Stalin P, Arjunan TV, Matheswaran MM, Dolli H, Sadanandam N. Energy, economic and environmental investigation of a flat plate solar collector with CeO_2 /water nanofluid. *J Therm Anal Calorim* 2020;139:3219–3233. [\[CrossRef\]](#)
- [151] Amar M, Akram N, Chaudhary GQ, Kazi SN, Soudagar MEM, Mubarak NM, et al. Energy, exergy and economic (3E) analysis of flat-plate solar collector using novel environmental friendly nanofluid. *Sci Rep* 2023;13:411. [\[CrossRef\]](#)
- [152] Chiney A, Ganvir V, Rai B, Pradip. Stable nanofluids for convective heat transfer applications. *J Heat Transf* 2014;136:021704. [\[CrossRef\]](#)
- [153] Mukherjee S, Mishra PC, Chaudhuri P. Stability of heat transfer nanofluids – a review. *ChemBioEng Rev* 2018;5:312–333. [\[CrossRef\]](#)
- [154] Sharaf OZ, Taylor RA, Abu-Nada E. On the colloidal and chemical stability of solar nanofluids: from nanoscale interactions to recent advances. *Phys Rep* 2020;867:1–84. [\[CrossRef\]](#)
- [155] Khan F, Karimi MN, Khan O. Exploring the scalability and commercial viability of biosynthesized nanoparticles for cooling panels with the help of artificial intelligence and solar energy systems. *Green Technol Sustain* 2023;1:100036. [\[CrossRef\]](#)



Review Article

Assessing the role of circular economy principles in reducing waste by sustainable manufacturing practices: A review

Mustafa M. MANSOUR^{1,*}

¹Department of Mechanical Engineering, College of Engineering, University of Thi-Qar, Thi-Qar, 64001, Iraq

ARTICLE INFO

Article history

Received: 11 August 2024

Revised: 03 October 2024

Accepted: 25 October 2024

Keywords:

Circular Economy; Recycling;

Sustainable Manufacturing;

Waste

ABSTRACT

The goal of this research is to assess the potential of circular economy principles in reducing waste through the identification of sustainable manufacturing practices. By focusing on the symbiotic net, the re-manufacturing layer, and waste prevention through recycling, the following aims are addressed: to identify sustainable waste prevention practices within manufacturing companies and classify sustainable practices related to the pre-manufacturing layer, symbiotic net, re-manufacturing layer, and waste prevention through recycling; to explore and assess the potential of established sustainable practices to foster Circularity by using Wasted controller, Redirector, Re-producer, and Recycling processes; and to analyze the impact of identified sustainable practices on waste reduction. To achieve these aims, exploratory multiple case studies are employed, focusing on both manufacturing companies and non-manufacturing companies perceived as sustainable. Qualitative interviews with key informants from each company are utilized, followed by a qualitative content analysis. By identifying a set of sustainable practices within the manufacturing industry, companies can better understand the potential of established strategies to be applied and integrated into the company's business model. This understanding can help bridge the gap between the Circular Economy and the Manufacturing industry and clarify the often ambiguous concept of the Circular Economy. The generation of waste is very much correlated to industrial activities and has a significant environmental impact. It is a key target in many countries to focus on the manufacturing sector to achieve more reasonable waste management and to open up markets for recycled goods. An effective and efficient approach to this end is to integrate the principles of sustainable manufacturing based on a circular economic standpoint. Therefore, the objective of this manuscript is to review the existing work carried out in this regard. We have reviewed many previous papers related to implementing different strategies of a circular economy in manufacturing. Our findings show that there is no systematic review available in the literature that explores the implementation of the principles of circular economy in the manufacturing system.

Cite this article as: Mansour MM. Assessing the role of circular economy principles in reducing waste by sustainable manufacturing practices: A review. Sigma J Eng Nat Sci 2025;43(5):1886–1896.

*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



INTRODUCTION

Sustainable manufacturing practices, driven by concerns over climate change, resource depletion, and ecosystem degradation, are gaining prominence in manufacturing sectors [1]. The rapid growth of the world's population, combined with increased consumption patterns, is leading to environmental degradation, threatening the planet's well-being and longevity. Industrial waste, notably from the manufacturing sector, is a significant contributor to climate change and resource depletion. Currently, 92% of materials used in consumer products aren't recycled or reused, resulting in the production of 92 billion tons of waste and threatening resource and natural capital depletion [2]. This crisis has brought attention to sustainable manufacturing practices, with the Circular Economy (CE) as an emerging alternative to prevent resource depletion, pollution, and environmental crises.

The circular economy proposes designing resource-efficient products with closed-loop life cycles achievable through a deep, up-front evaluation of production and manufacturing processes. Although products designed for Circular Economy principles have advantages over conventionally manufactured products, the associated costs and complexity currently preclude their wide adoption in manufacturing industries. Organizations face difficulties in assessing the influence of CE principles on product life cycles holistically and precisely evaluating gains and constraints at different life cycle stages. This paper aims to fill this gap by quantitatively investigating the influence of “circularity” on a product's life cycle in terms of waste generation allowed in manufacturing

systems and sustainably on relocating the CE concept to existing manufacturing systems. Factors that influence Sustainable Manufacturing Systems unable to adopt CE principles are also considered.

Background and Rationale

Manufacturing still represents around 16 percent of the world's economic activity, though its growth has recently slowed due to many factors, including COVID-19 [3]. Some global manufacturers have been especially impacted, such as global land and sea transportation firms, managing every day the movement of over ten million transport containers throughout the world. Another example is global automotive manufacturers, which, facing a semiconductor supply scarcity, are likely to produce 11 million fewer vehicles in 2021 than in 2020. Industry 4.0 technologies, for example, big data and deep learning, can facilitate the development of new intelligent industrial systems capable of maintaining and developing manufacturing competitiveness under unexpected environmental scenarios as showed figure 1. The introduction of new intelligent technologies to manufacturing firms, simulating the information processing capabilities of the human brain, can foster manufacturing capabilities dependent on correlated data processing at exponential speeds and volumes.

Waste manufacturing can happen inadvertently, as manufacturing systems run under a combination of global uncertainties and several coupling variables, generating non-intended event chains, called emergent behavior. Recent studies show how these factors can create parallel scenarios ranged from a slow and sub-optimized system to a chaotic and failed system. A solution to avoid unexpected

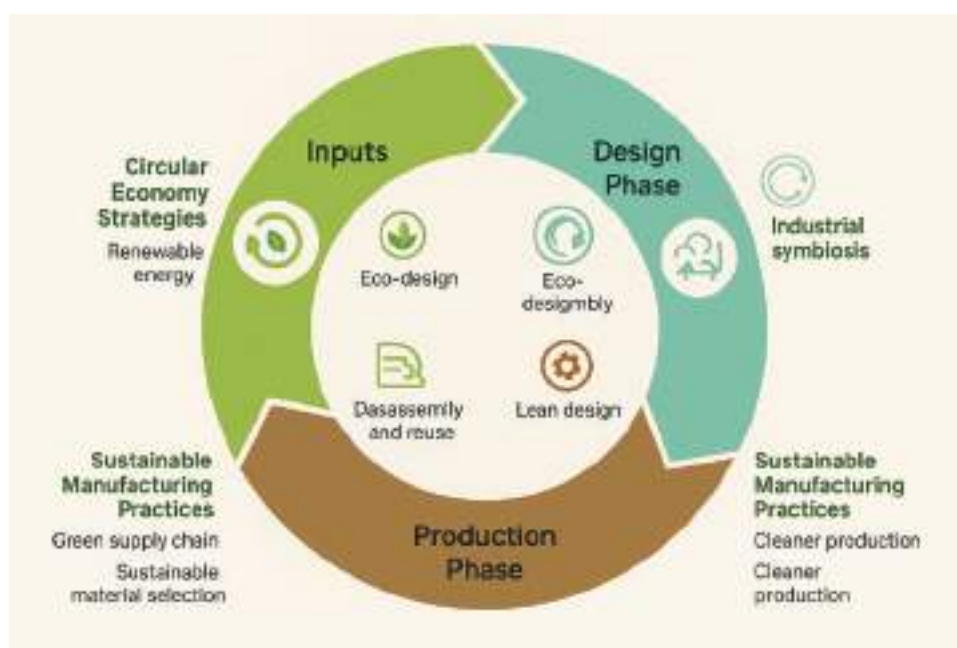


Figure 1. Framework and circular strategy loops. Adapted and redrawn from [4].

waste generation is to monitor those systems, develop explanatory models of knowledge of the factor correlations behind those emergent behaviors, and use that knowledge to pre-empt those unexpected events through decision-making systems. In this regard, methods of artificial intelligence used to infer knowledge models from data—can be useful, especially neuro-fuzzy systems, which mix knowledge of the process with models inferred from data, creating systems interpretable by humans that can be relevant to detect hazards [4].

Understanding Sustainable Manufacturing

The paradigm of sustainable manufacturing has been derived from the well-known phrase ‘sustainable development’ in order to overcome the environmental challenges created by all nations as a result of rapid growth in industrialization and mass production [1]. Sustainable manufacturing has already been recognized as a general concept worldwide, with the aim of framing the definition, need, and different principles governing sustainable manufacturing in manufacturing industries. Sustainable manufacturing practices recognized under policies frame a sustainable future by remaining competitive in industrial markets. Sustainable manufacturing is essential to eliminate waste discharged from factories, which have created a dangerous situation in the present world by polluting water and air, and threatening the existence and health of plants, animals, and the human race.

The principles on which the paradigm of sustainable manufacturing is based are significantly utilized by industries to mitigate the harmful impacts of the manufacturing sector on the environment. An analysis of different models of sustainable manufacturing and their focus assists in understanding the importance of sustainability among industries. Continuous climate change has created a demand for reducing greenhouse gas emissions due to industrial production and hence, industries have empirical needs to limit their carbon footprint on the environment. Proponents have been increasingly advocating that sustainable manufacturing should go hand-in-hand with circular economy principles in design, manufacture, use, and reuse [5].

Definition and Concepts

Sustainable manufacturing refers to the creation of manufactured products through economically-sound processes that minimize negative environmental impacts while conserving energy and natural resources. The sustainability of a product depends on its whole life cycle, including the extraction of raw materials, processing, distribution, use, and disposal. Manufacturing is a key industry segment that contributes significantly to economic growth by providing jobs, wealth, and quality of life. However, the manufacturing sector’s share of GDP in many developed countries has declined due to globalization, with industries moving to developing nations where

resources and labor are more abundant and less expensive. Rapid industrialization has caused environmental deterioration around the world, leading to the emergence of environmental regulations and compliance issues. The introduction of the ISO 14000 series of environmental management standards has been an effective way to address environmental issues and improve competitiveness in countries where such regulations are a concern. In developing countries, significant progress has been made in establishing environmental management systems based on ISO 14000, as well as compliance with regulatory requirements. However, there is a need to improve environmental management and economic viability for the manufacturing sector to strengthen competitiveness in the global economy [1]. Sustainable manufacturing and processing technologies should be developed to minimize the negative impacts of manufacturing on the environment and improve overall sustainability performance. Sustainable technologies are those that use renewable materials and energy sources with non-toxic chemicals and processes, minimizing waste generation, while maintaining the economic viability of the business. The move from local to global competition in the manufacturing sector has forced companies to rethink operations. Increasing global competition due to the globalization of markets and innovations in transportation and information technologies has intensified competition [6].

IMPORTANCE IN THE CURRENT INDUSTRIAL LANDSCAPE

At present, “sustainability” is considered essential in every discipline, including manufacturing and its technologies [1]. Sustainability drives a company’s outlook toward profit and future. Sustainable development is defined in many ways, but a clear and widely accepted definition is: Development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains two key concepts: the concept of “needs,” in particular, the essential needs of the world’s poor, to which overriding priority should be given, and the idea of limitations imposed by the state of technology and social organization on the environment’s ability to meet present and future needs.

In manufacturing and its technologies, sustainable manufacturing is the idea of manufacturing that meets the needs without compromising the need of future generations [7]. It can be defined clearly as “the creation of manufactured products that use processes that are non-polluting, socially responsible, and economically viable.” It can also be interpreted as designing, producing, using, and recycling manufactured products in such a way that there is the least environmental impact, the least social impact, and the most profitability to the manufacturer. Sustainable manufacturing is the most trending topic in today’s world.

Circular Economy Principles

The principles of the circular economy (CE) are outlined to highlight its significance in sustainable manufacturing. The elaborated understanding of CE principles is important in the context of this research. It provides foundational knowledge regarding key principles underlying CE initiatives in the context of manufacturing. The extensive findings of the literature review are described as subthemes or principles of the CE. Broadly, CE principles drive circular initiatives through proficient product design, sustainable sourcing of materials and services, and end-of-life product recovery [8]. Resource efficiency and waste reduction through recycling and recovery (regenerative approaches) is an alternative or supplementary approach.

Brief descriptions of the principles of the CE as showed table 1. Product design for longevity and recovery highlights the importance of product design in driving circular initiatives in manufacturing. This may include shorter life cycle product designs, designs that enable disassembly and innovation in remanufacturing components. Sustainable sourcing of materials and services is another principle under CE that closely mirrors a green initiative. Using ethically sourced materials, developing local supply chains and developing synergies between industries to exchange by-product materials are allowed. Recovery of end-of-life products discusses approaches to establish backward supply chains for the recovery of products after the consumer end of product life. Product take-back policies, establishing by-product exchanges assisted by third parties and

innovative recycling solutions to product diversion from landfills fall under this principle.

OVERVIEW AND KEY PRINCIPLES

The industrial production pattern followed by most countries worldwide can be classified as linear, which violates certain ecological principles. Therefore, to reduce waste and strive for sustainable growth, circular economy principles have been proposed for industrial production, as commonly done in nature. The basic principles of a circular economy are that after product uses all resources, including material and energy, should either be returned to the same production system for the same use (reuse) or separated and returned to a different production system for different use (recycling). Generally, resource use should be minimized and resources should be kept in production as long as possible [9].

In addition to the circularity principles, four ease-of-implementation principles have been proposed. These easy-to-implement principles are a minimum requirement for compliance with the circularity principles. Under the ease-of-implementation principles, no resource-comprehensive actions or advanced technology innovations are required. The number of inputs to a product design must not increase and other changes to the core components must also not occur. Anything that could decrease product durability must also be avoided. There must be no other major technical redesigns or

Table 1. Circular economy integration in solid waste management Data extracted and tabulated from [9]

Stage	Focus area	Circular economy strategies (loops)	SWM practices
Waste prevention	Source reduction, eco-design	- Product redesign - Lightweighting - Avoid single-use	- Eco-packaging - Green procurement
Collection & sorting	Waste segregation & logistics	- Separate at source - Smart bins - Reverse logistics	- Recycling centers - Digital monitoring
Recycling & recovery	Material and resource recovery	- Mechanical & chemical recycling - Upcycling - Material cascading	- MRFs (Material Recovery Facilities) - Waste-to-resource plants
Treatment	Biological/thermal treatment	- Composting (organic loop) - Anaerobic digestion - Waste-to-energy	- Biogas plants - Controlled incineration
Reuse & remanufacture	Extending product/component lifecycle	- Refurbishment - Repurposing - Remanufacturing	- Second-hand markets - Industrial reuse networks
Landfilling (last resort)	Residual waste disposal	- Only for non-recoverable materials	- Sanitary landfills with methane capture
Enablers	Cross-cutting mechanisms	- Policies (EPR, bans) - Digitalization - New business models	- PPP (public–private partnerships) - Awareness programs

replacements of basic components (e.g., structure). To assess the compliance with these principles, a circularity index has been proposed as a quantification instrument. The circularity index quantifies the degree to which the studied implementation conforms to the circularity principles. The result is presented as a dimensionless quotient between 0 and 1, where 1 indicates complete compliance and 0 indicates no compliance [10].

Applications in Manufacturing

The results of the interviews and workshop show that specifications for energy, water, chemicals, and waste are already being implemented and further development will continue along these lines. Bespoke machine designs for resource efficiency are still rare. The subsequent quantification of the resource consumption of relevant textile processes by the participants indicates the need for additional actions to reduce resource consumption [8].

Implementation issues for plant efficiency and device performance would be very helpful, as would shareable reliable information on efficiency for common devices. Interesting pilot programs would be the development of device performance measures based on clear specifications and the establishment of a benchmarking platform for industrial devices on resource consumption. The existing European Company Awards for the ecodesign of the production process and product could be updated to include material recovery and recycling criteria [2].

WASTE REDUCTION STRATEGIES IN MANUFACTURING

Lean Manufacturing Techniques

Lean manufacturing, or lean production, denotes a manufacturing strategy to reduce wastes. Originated from the mass manufacturing practices of the Toyota production system (TPS), it derives its name from the ideal of succinct resource usage and a fully value-adding process. Lean typically describes improvements in five aspects: product and service features, process ownership, staff involvement, and drawing on partnership arrangements and forecasting system [11].

Lean considers any activity that does not add value, or waste, to a product is non-lean. Thus, waste reduction is considered a strategy for improvement. Waste is not only defined as excess or scrap but also in terms of time, workflow, arrangement, employees, etc. Similarly, value is considered in an enlarged sense, related to the employee satisfaction together with the customer. The aim of lean, therefore, involves multiple facets of efficiency and optimization related to the resource usage in a manufacturing process [12].

Recycling and Upcycling Practices

Recycling and upcycling practices have gained increasing attention for fresh approaches to waste reduction and

circularity in manufacture. These techniques represent added-value strategies compatible with the circular economy approach, which involves limiting the waste generation and resource consumption from manufacture processes [13]. Sustainable waste management implies minimizing waste before it is generated. Waste reduction techniques can be classified into several groups, such as product or input redesign, manufacturing process redesign, management and control mechanisms, recycling and upcycling practices, and product or process reuse [2].

Recycling is the transformation of waste into a new product, often altering its physical or chemical properties to make it reusable. Recycling manufacturing often results in a change in the material structure, which prevents reusing the product in the same manufacturing process due to the loss of its initial shape, performance, or qualities. Upcycling is a recycling practice gaining increased attention, referring to the transformation of waste into a new product of equal or higher value. This approach is often applied in premature technologies when the removal of contaminants or the technical limitations of the mechanical recycling process prevents the return of the end-of-life products back to the same manufacturing process and quality.

Case Studies and Best Practices

The research findings have been validated using published case studies and exemplary practices to describe how manufacturers have successfully adopted circular economy principles at different levels. It describes practical insights and approaches that can be replicated by others. Circular economy principles are a relatively new development in manufacturing and, therefore, it would be appropriate to use published examples to demonstrate the successful implementation of these principles.

These case studies have been reported in international journals with an established reputation for scholarly rigor and integrity, as well as through presentations at conferences organized by established organizations. The cases have been chosen to be representative and show a diversity of client industries, methods of implementation, and locations. Their presentation reflects ongoing progression concerning early movers and introduces the reader to a wider array of circular economy principles within manufacturing contexts [10].

Successful Implementations of Circular Economy in Manufacturing

As circular economy (CE) moves to the forefront of international agendas, potential avenues for its adoption are being investigated with haste. This research concentrates on significant instances of CE's implementing in a manufacturing context. Manufacturing has been chosen as the main focus of the study since it delivers a considerable opportunity for resource consumption reduction, justifying the need for a paradigm shift of this nature. Accordingly, large firms worldwide already practicing some form of CE

and its principles are sought out through a combination of a detailed literature review and examination of sought industry examples [10].

1. The Pratt & Whitney [11] case shows how a product-based solution in combination with end-of-pipe treatment effectively reduces the volume of waste. The company estimates that the solution saves it as much as \$2 million per year in disposal costs and presents a lower volume of waste to increasingly environmentally conscious customers.
2. The case study on mini-mills highlights that it is good business to invest in manufacturing processes that reduce environmental impact. There was no compulsion to clean up bottom and fly ash, depositing it typically in tailing ponds [12]. The process was standard industry practice. Faced with problems in obtaining new water resources, Chapleau used the locally available river and drew its supplies upstream, giving rise to significant differences from inlet to outlet for river and mill discharges. This acute sensitivity to effluent, in turn, gave rise to the new system for removing redundant materials to reduce contamination.
3. Cummins Engine Co. set a 5-year target to reduce all air pollutants by 40%. By 1998, it had managed 63%. This move has not only made business sense, but it has also been profitable in looking to the long-term \$45 million in environmental and safety costs per annum. In the engine-manufacturing business, the company rolls out three new products per day [13]. Six months before production of a new engine model begins, a cross-functional team starts work. Cummins is the best engine maker and has won awards mainly for waste-reduction work. Such a market focus makes it exceedingly sensitive to environmental costs.
4. Interface pioneered the radical concept of selling carpet as a service or lease and only guaranteeing that it was sufficiently clean and free from defects [14]. The payback times were four and six months, respectively. The company now sells its eco-efficiency carpet to 8% of the market. It is the market leader and has been so for many years. Overall, the company claims that it has made \$124 million of efficiency profit.
5. Nike [15] found that by installing a wholly integrated waste treatment system, it was able to safeguard both the product and the environment and, if necessary, guarantee to withdraw waste. This solution recovers and recycles 98% of the water used in production, eliminating waste from the production process. Here, production waste is reused with all of the recycling economic, environmental, and operational benefits. This system is compact enough to require only ten workers to operate it. Furthermore, with this design, the company stores waste materials that are reused throughout the production process.

Manufacturing places enormous demands on a limited supply of global resources, creating significant

environmental burdens and CO₂ emissions. CE is defined as a regenerative economic model wherein resource input, waste, emissions, and energy leakage are minimized by closing resource loops. CE aims to erect a strong manufacturing base while stopping net material use expansion. Implementation examples with supporting factors from CE principles are additionally highlighted to distinguish best practices from CE functions. This addresses how the CE could be tailored specifically for the manufacturing context, in turn prompting the uptake of CE through significant illustrations and managed expectations from its adoption. Exploring CE's principles in a manufacturing context is found to be timely, relevant, and important [16]. Of the 212 companies adopting in some form a CE and its principles, all but two operate at a manufacturing level. CE principles include closing material loops, keeping resources in use, and promoting systems thinking and cooperation (up to 12 illustrations thereof).

Challenges and Barriers to Implementing Circular Economy in Manufacturing

Although significant advances have been made in addressing waste management in manufacturing, many difficulties and challenges remain. Therefore, this study investigates the CE principles that reduce waste generation as well as their implementation challenges in manufacturing. In general, circular economy principles can reduce waste generation in manufacturing by accommodating several approaches, such as sustainable supply chain management, cleaner production, industrial ecology, and life cycle assessment [17]. However, in practice, implementing manufacturing initiatives, other than technology initiatives, is mostly neglected.

The principles of CE and green manufacturing are very important, however difficult to implement and very challenging in developing countries. It is because from the nation perspective, even though there are policies that promote CE or green manufacturing that a lot of small and medium enterprises (SMEs) cannot comply with these policies. The principles of the CE have concerned with the reduction, reuse and recycling (3R) of waste [18]. Various efforts such as development of strategies and methodologies, improvement of machines and equipment, and recycling innovations have been taken, however a significant reduction in current waste generation has not been achieved.

Technological Limitations

Focus is placed on technological constraints which hinder a seamless adoption of circular economy practices. Industry stakeholders may encounter technology challenges, such as lack of advance technology and equipment, inadequate technical capabilities as constraints to CE implementations. Scientific artifacts are presented as a novel model to streamline boundary conditions of CE adoption to these technological challenges and barriers

[18]. In this model, CE visions can only be attained generally by realizing technological transitions, value chain reconfigurations, and market expansion efforts when available scientific artifacts conform to boundary conditions specified by these realizations. Systematic searches are needed to identify and onboard scientific artifacts conform to boundary conditions whereby a certain technology challenge or barrier is circumvented. It is also conceptually examined how and how much scientific artifacts used within CE implementations should be processed to account for boundary conditions (i.e. upper and/or lower limits thereof). Realizing relational implementation procedures formats have to be investigated in future research to assess CE visions attainment likelihoods and derivable boundary conditions. Extensions to broader longer time horizon frames CE visions that might change with a retrofitting procedure based on learning effect, better understanding of devices or knowledge about them would be potential future researches. Since CE intends using energy and materials at manufacturing upmost beneficially, providing scientific artifacts enhancing energy and materials better utilization with a CE concept such as Design of Active Manufacturing Systems (as having methodologies from macro strategic ones to enhance energy and material usage) could yield interesting research results [19].

This section describes the approach of the literature review adopted in this paper, which began with identifying articles related to “circular economy” and “manufacturing,” followed by refining the set of documents to include those contained in databases and sorting them on a time basis up to January 2022. The aim of this comprehensive review is to ensure that academic and practical evidence can be seen from more than one perspective, with the view that the sources chosen provide alternatives for many different research and development profiles. The electronic databases used in this literature review include various academic sources. By not excluding or including specific databases, the search strategy encompasses a wide range to overcompensate for missing information across just one or two search engines, thus ensuring more inclusivity in terms of results. The multi-database strategy seeks to improve the exploration process within research and avoid any possibility of underlying errors or misinformation. Moreover, it is to be noted that, at this point, this review is based neither on a meta-analysis of a specific topic nor on a decisive aggregation of themes according to quantitative criteria. Furthermore, it is pivotal to emphasize and draw attention to the fact that both databases and engines are indispensable tools, since no specific academic engine can be considered a standalone site for comprehensive research. The exhaustive search herein includes searching for the keywords, with the engine sorted by “relevance.”

The approach was extracted from the literature of “study-level design,” and its work state stemmed from the five definitions developed for the “epistemic context.” In summary, epistemic studies focus on research communities, work,

and products. Also, this review has focused on addressing the frequently cited scholarly literature published in top-tier journals, as well as commonly referenced trade press/media. Excluded were surveys, secondary literature, and purely theoretical research outputs. The approach outlined above supports criteria and research questions to facilitate the review process and collate information dealing with factors driving implementation agendas and activities related to the circular economy and opportunities for manufacturing. The emphasis is driven by the potential to use tried and tested technologies, processes, tools, models, and practices that are adaptable across supported activities, given the alignments.

Economic Considerations

The exploration of the Circular Economy (CE) as an emerging alternative business model is a direct response to the multitude of socio-economic and ecological sustainability concerns facing today’s society and the planet. Considered to be part of the trend away from the traditional linear “take-make-dispose” business model, CE principles provide a more holistic conceptual framework and way of thinking concerning sustainability and sustainable resource management [20]. However observable in society over the previous couple of decades, the increased attention and prominence afforded to CE principles has predominantly emerged over the past 5-10 years, leading to the term being variously coined as a new movement, trend, paradigm shift, the next industrial revolution, and so on.

There is a distinct feeling of excitement, hype, utopian vision, and optimism about embracing CE principles. Nevertheless, there are on-going and fundamental economic considerations, factors, and drawback / hurdles standing in the way of the broader acceptance and adoption of CE principles within the relevant decision making / policy influence circle. Hence, this analysis will explore some economic factors and considerations, standing as challenges to the broader acceptance and adoption of CE principles. It is appreciated that whilst an amplification of CE principles may result in net economic benefits, improved economic growth, reduced socio-economic inequalities and some other economic benefits, there are nevertheless economic concerns, factors, and considerations that do stand in the way of wider societal acceptance and adoption.

Policy and Regulatory Frameworks

Several EU Member States have adopted comprehensive policy frameworks to support the transition to a Circular Economy (CE), with a focus on eco-innovation and the support of CE business models in India by specific financing instruments. The European Commission regards the CE policy initiative as a priority, motivated by economic considerations and resource dependency, as well as the aspiration to re-establish Europe as a leader in sustainable loss-prevention technologies [20].

The EU's CE policy initiatives comprise a package of proposals related to the waste framework directive, packaging waste directive, land-filling directive, and the integrated policy state-of-play on the CE. The European Strategy for Plastics in a Circular Economy was updated in January 2018, followed by the Circular Economy Action Plan. Considerable efforts have been made to formulate complete regulatory frameworks within broader CE initiatives in a number of EU Member States. The policy mix concerning the CE is, by most definitions, multi-faceted, comprising, among others, regulatory standards to Smith to foster eco-innovation, with a set of measures directed at the whole production-distribution-consumption chain. Given the complexity and multi-facetedness of the CE concept, it is essential to focus on major systemic components, and these components should be included in the CE policy frameworks. Firstly, recent waste management targets concerning separate collection, recycling, and recovery of wastes are barely sufficient standards on their own [21].

Government Initiatives and Incentives

In recent years, a variety of initiatives and incentives have been adopted by governments at national and regional levels to promote the more sustainable manufacturing and consumption of products. Regulatory agencies and trade promotion organizations have provided guidelines on sustainable manufacturing practices, and broad incentives have been provided in the form of tax credits and more targeted initiatives to foster new technologies for more sustainable manufacturing [21]. Regarding the latter, improved interactions of policymakers with different stakeholders have fostered competitive advantages in Europe for higher sustainability standards in the manufacturing sector. The role of government is seen as key for the transition towards a circular economy, with an active involvement in the establishment of required infrastructures for reused or recycled products, strict legislation for resource consumption reduction with transparent indicators, and facilitation towards research and development on new technologies [22].

In European countries, current government initiatives and incentives for sustainable manufacturing were reviewed. The majority of countries have established broad guidelines on sustainable manufacturing and consumption. A more product-centered approach was adopted by regarding sector-specific guidelines on eco-design and eco-labeling. In conjunction with these guidelines, broad incentives are provided in the form of tax credits and voluntary agreements with industries. Currently, many countries are implementing new sector-specific taxes for certain goods and materials, environmental disclosure and reporting, and environmental standards for products. Broad business models have been proposed for increased sustainable practices in manufacturing. Some initiatives propose the establishment of a directory of available technologies (processes, materials, equipment, etc.) with well-documented evaluations of their life cycle environmental impacts, costs,

and effects on work environments. By making such information widely available and easy to understand, companies can compare different options for improved sustainability.

Measuring and Assessing the Impact of Circular Economy in Manufacturing

Assessing and Evaluating the Impact of Circular Economy in the Manufacturing Sector is one of the important tasks to consider Industrial Revolution 4.0, covering Industry trends, Green Manufacturing and Sustainable Manufacturing. It is a dilemma to evaluate this opportunity adequately by finding Key Performance Indicators and Methods of Assessment to evaluate this CE application properly and give a handy tool for further research. As the transition to Sustainable, Green and CE manufacturing seems unavoidable, current and future manufacturing companies must meet new challenges and opportunities in order to stay competitive and on a safe side [23]. There is a problem that current CE definitions in manufacturing are vague. A business model based on shifting to CE would require comprehensible and unambiguous definitions of how this principle should change current manufacturing strategies by covering environmental aspects and flexibility in term of disposable resources [24]. If CE practices in manufacturing are properly defined, new Key Performance Indicators would be needed, as a reflect of CE principals, but also a reflect of sectors within the manufacturing itself. In this manner, it is hard to talk on universally adopted KPIs for the CE manufacturing approach without having clearly defined CE principals themselves.

Assessment and evaluation of Sustainable Manufacturing has not received attention from researchers at this time. Most of the methodologies are mainly focused on very vague and subjective a Sustainable index, which in reality has no base in strict and numeric terms. Likewise, Option 3 could serve as a good framework for various discussions and works; but, practical calculations are impossible as this Option is vague and has not been developed to a level of applicability, where it could provide numeric coverage and performance assessment of sustainability to almost any system. This dilemma simply calls for further research [25].

Future Trends and Innovations in Sustainable Manufacturing

A new generation of savvy, sustainable products and production systems is emerging as environmental, economic, and social pressures shape the future of industrial practice. The focus on sustainability will gain momentum as eco-efficiency and greener products and processes become vital facets of the competitive strategies of industry leaders [1]. Market demands for sustainable products are connected to manufacturers being more environmentally friendly in their production processes. Market pressures exist on organization producers to demonstrate that their processes and practices are more sustainable—using less harmful materials, minimizing energy use, and extricating

the creation of waste. Investigations into applying particular resource-efficient technologies to process systems are expected. Greater engagement by organizations in the economic, social, and environmental sustainability of their supply chain is anticipated.

Radical changes in manufacturing products, processes, and systems are anticipated. The evolution of new sustainable or greener products and production processes will likely maintain ultimate product performance and function while enhancing sustainability [25]. Many manufacturers have adopted sustainable practices in product design, sourcing, production, distribution, and marketing, and environmental compliance programs. Advances in sustainable manufacturing processes based on renewable energy, particularly solar, hydropower, and bioenergy, will trace a 30-year innovation and investment trajectory. New materials and composites for manufacturing will emerge, adapted to rationalized and flexible production technologies, with improved performance and lower environmental impacts.

Emerging Technologies

This section delves into the latest technological innovations and advancements that hold promise for further enhancing sustainable manufacturing practices. It sheds light on future trends and developments.

Emerging Technologies in the Manufacturing Plastic Value Chain. A total of 15 emerging technologies related to the transition towards circularity in the plastic materials value chain were identified [27]. The largest group addresses the early stages of the value chain, feedstock and polymer producers, while the second largest group addresses the end-of-life stages of waste management and recycling. Very few specifically address the mid stages of plastic value chains. The literature highlights the complementarities between the technologies:

- (1) Enabling a shift in the production-consumption system (distributed economies),
- (2) Facilitating the development of high-added value products from biological materials (bio-based systems), and
- (3) Reducing the need for raw materials to produce high-quality recycled plastic (chemical recycling).

These technologies exhibit great potential for increased circularity in the manufacturing stages of the plastic value chain. Favourable forces behind this group include a promise of efficiency and productivity increase, the generation of positive marketing messages towards consumers, enabling materials' re-utilisation, a potential seizing of mixed waste sources, and enabling transparency and collaboration among actors [27]. The concept of distributed economies refers to shifting the economic paradigm into more local sourcing, manufacturing, consumption, and recycling systems. The combination of emerging technologies that enable this concept are 3D printing, IoT, blockchain, AI, and cloud computing as well as small-scale chemical transformation processes [28].

Machine Learning and Artificial Intelligence in Circular Economy: A Bibliometric Analysis and Systematic Literature Review. Their study found that large global enterprises are not concerned about their packaging waste; however, the problem can be solved by using the right technology and providing the appropriate motivations to the involved actors. Vlachokostas et al. proposed a generic model using a decision support system that can concurrently examine environmental, economic, and social criteria to generate three different units of alternative biowaste treatment to enhance the production of bioenergy and bioproducts. Bio energy is clean and renewable which completely satisfies the idea of CE. Their experiments showed notable results [29].

CONCLUSION

Sustainable manufacturing practices are essential for industries looking to gain competitive advantage while contributing positively to the environment. This study evaluated these practices and their effect on waste reduction in manufacturing industries based on the principles of Circular Economy. To meet its objectives, the study conducted qualitative interviews with manufacturing industry professionals. The findings indicate that sustainable manufacturing practices are effective in waste reduction. However, the extent of reduction varies based on the application of different Circular Economy strategies. Further, the study presents an evaluation model that provides reference to industry professionals looking to adopt these practices. The model details a sustainable manufacturing practices framework that categorizes the practices based on effective Circular Economy strategies into resource efficiency, recovery, remanufacturing, green technologies, and recycling. The study assesses the level of adoption and implementation of these practices to provide an estimate for waste reduction potential. Further, the study details waste reduction estimation calculations based on the assessment of sustainable manufacturing practices. To assist with calculations, the study presents a cost equation that allows industry stakeholders to evaluate monetary gains from adopting these practices.

Based on the findings, the study highlights two recommendations that can be beneficial for industry stakeholders pursuing sustainable manufacturing efforts. First, an immediate priority for companies is to focus on improving resource efficiency practices such as resource recovery, reuse of natural resources, and the development of energy-efficient technologies. Engaging in these practices will catalyze the adoption of other practices while reducing waste by preserving natural resources [30]. Second, there is an urgent need for academic researchers to shift focus toward investigating sustainable manufacturing practices centered on Circular Economy principles. Research in this area will create awareness in the manufacturing industry about the significance of adopting sustainable manufacturing and Circular Economy practices for waste reduction.

Further, it can guide industry professionals in appropriately adopting practices for maximum waste reduction potential.

Summary of Findings

Sustainable manufacturing practices are developing rapidly worldwide, and major companies are undergoing industrial transitions to more sustainable processes and resources. Waste is a major global problem, and concerning manufacturing processes, wastes have been produced since the beginning of the industrial revolution. The need for sustainable manufacturing practices has arisen, and such practices are under a generalized acceptance stage globally. A circular economy of sustainable manufacturing practices has been envisioned to close waste loops. Such circular economy principles enhance the manufacturer's efficiency by utilizing the major waste sources as resources and closer to the zero-waste scenario. This research endeavors to clarify the link between sustainable manufacturing practices and circular economy principles while providing insights into indicators. It has been demonstrated that there are noticeable links between sustainable manufacturing practices and circular economy principles. There are numerous circular economy indicators concerning manufacturing processes; however, not all questions of the desired practice are inferred in these indicators. A complementary and non-overlapping indicator list has been proposed to assess the role of circular economy principles in these desired sustainability circumstances. Additionally, it has been validated whether these indicators are reasonable concerning the manufacturing processes. In light of the increasing industrial awareness, more stringent petitions, and industrial competition, sustainable manufacturing and circular economy assessment will be hourly anticipated in the completing decade. It is hoped that the indicator list can engender awareness and inspire more disclosures and works on sustainability in academia and industry.

Recommendations for Industry Stakeholders

The interest in implementing circular economy principles in manufacturing processes has been growing, as evidenced by the rising number of publications discussing the topic and the increasing popularity of the phrase. Manufacturers interested in capitalising on the advantages offered by circularity may find it daunting to know where or how to start embracing these principles in practice. They may also be unsure what difficulties and challenges may arise along the way. To this end, a set of recommendations was formulated for industry stakeholders, taking the academic knowledge found in scientific literature and sector-specific documentation on circular manufacturing, sustainability, and industry 4.0, and connecting it to the practical requirements to make small first steps or larger transformational efforts to drive this change towards circularity [31].

To ensure that the changes undertaken by pilot projects can be replicated in other projects across industrial contexts, this research effort must be backed up with

full stakeholder engagement and provide detailed and hands-on direction for circularity implementation. In other words, the recommendations should be undertaken as a set of guiding principles that can serve as building blocks in different industrial settings. Hence, the recommendations aim at showcasing and collectively explaining approaches and tools that can readily be used as the foundation for circularity implementation. It must be clearly stated that, in general, these recommendations should not only be considered techno-centric solutions, but rather considers the strategy, stakeholders, technological aspects, product characteristics, and closures as integrated and interdependent aspects that must be addressed simultaneously. Circularity cannot be achieved unless a thorough understanding of the system as a whole has been gained, which can be accomplished through these sets of investigations and chain of events as a first step in a change process for circularity [32].

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] Chowdhury N, Yasmin J. A study on sustainable manufacturing practices in RMG industries of Bangladesh. MPRA Paper No. 87104. London: City University, 2018.
- [2] Noman AA, Akter UH, Pranto TH, Haque AKMB. Machine learning and artificial intelligence in circular economy: a bibliometric analysis and systematic literature review. ArXiv abs/2205.01042, 2022. DOI:10.33166/AETiC.2022.02.002 [CrossRef]
- [3] Nascimento DLM, Alencastro V, Quelhas OLG, Caiado RGG, et al. Exploring industry 4.0 technologies to enable circular economy practices in a manufacturing context: a business model proposal. J Manuf Technol Manag 2019;30:607–627. [CrossRef]

- [4] Almulhim AI. Toward a greener future: applying circular economy principles to Saudi Arabia's food sector for environmental sustainability. *Sustainability* 2024;16:786. [\[CrossRef\]](#)
- [5] Salman HS, Mansour MM, Lafta AM, Shkariah AJ. Modification design and process of pipeline to reduce erosion rate and deposited. *Int J Comput Methods Exp Meas* 2024;12:165–173. [\[CrossRef\]](#)
- [6] Mansour MM, Uгла AA. Employing genetic algorithm to optimize manufacturing cells design. *Acad J Manuf Eng* 2024;22:3.
- [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] Mansour MM, Erabee IK, Lafta AM. Comprehensive analysis of water-based emulsion drilling fluids in Gharrar oil field in southern Iraq: properties, specifications, and practical applications. *Int J Comput Methods Exp Meas* 2024;12:297–307. [\[CrossRef\]](#)
- [9] Patil RA, Ghisellini P, Ramakrishna S. Towards sustainable business strategies for a circular economy: environmental, social and governance (ESG) performance and evaluation. In: Springer eBooks. 2020:527–554. [\[CrossRef\]](#)
- [10] Lafta AM, Mansour MM. Employing artificial neural networks to forecast gas consumption by a power plant. *Acad J Manuf Eng* 2025;23:1.
- [11] 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. [\[CrossRef\]](#)
- [12] Mansour MM. Recent advancements in refractory materials and digital innovations in precision manufacturing. *Sigma J Eng Nat Sci* 2025.
- [13] Zaman A. Zero-waste: A New Sustainability Paradigm for Addressing Global Waste Problem. New York: Springer Nature; 2022. [\[CrossRef\]](#)
- [14] Garcés-Ayerbe C, Rivera-Torres P, Suárez-Perales I, Leyva-de la Hiz DI. Is it possible to change from a linear to a circular economy? An overview of opportunities and barriers for European small and medium-sized enterprise companies. *Int J Environ Res Public Health* 2019;16:851. [\[CrossRef\]](#)
- [15] Iacovidou E, Hahladakis JN, Purnell P. A systems thinking approach to understanding the challenges of achieving the circular economy. *Environ Sci Pollut Res Int* 2021;28:24785–24806. [\[CrossRef\]](#)
- [16] Aristi Capetillo A, Bauer F, Chaminade C. Emerging technologies supporting the transition to a circular economy in the plastic materials value chain. *Circ Econ Sustain* 2022;1–30. [\[CrossRef\]](#)
- [17] Milios L. Advancing to a Circular Economy: three essential ingredients for a comprehensive policy mix. *Sustain Sci* 2018;13:861–878. [\[CrossRef\]](#)
- [18] Moraga G, Huysveld S, Mathieux F, Blengini GA, Alaerts L, Van Acker K, et al. Circular economy indicators: What do they measure? *Resour Conserv Recycl* 2019;146:452–461. [\[CrossRef\]](#)
- [19] Beghetto V, Cayzer S, Griffiths P. Design of indicators for measuring product performance in the circular economy. *Int J Sustain Eng* 2017;10:289–298. [\[CrossRef\]](#)
- [20] Pinna C, Demartini M, Tonelli F, Terzi S. How soft drink supply chains drive sustainability: key performance indicators (KPIs) identification. Preprint 2018. <https://re.public.polimi.it/retrieve/e0c31c0c-a1ba-4599-e053-1705fe0aef77/PROCIR-D-17-00728.pdf>
- [21] Schröder JJ, Smit AL, Cordell D, Rosemarin A. Improved phosphorus use efficiency in agriculture: a key requirement for its sustainable use. *Chemosphere* 2011;84:822–831. [\[CrossRef\]](#)
- [22] Scottish Government. Making things last – a circular economy strategy for Scotland. 2016.
- [23] SITRA. Leading the cycle – Finnish road map to a circular economy 2016–2025. 2016.
- [24] SITRA. McKinsey. The opportunities of a circular economy for Finland. 2015.
- [25] Sokka L, Lehtoranta S, Nissinen A, Melanen M. Analyzing the environmental benefits of industrial symbiosis: life cycle assessment applied to a Finnish forest industry complex. *J Ind Ecol* 2011;15:137–155. [\[CrossRef\]](#)
- [26] Stegeman H. The potential of the circular economy – from circular materials cycles to a circular macroeconomy with scenarios for the Netherlands. Rabobank Econ Res Dep 2015.
- [27] Su B, Heshmati A, Geng Y, Yu X. A review of the circular economy in China: moving from rhetoric to implementation. *J Clean Prod* 2013;42:215–227. [\[CrossRef\]](#)
- [28] Prasad MAN, Kotagi BP, Chandrashekar K. Enhancing road safety through blackspots mitigative measures: a review. *Sigma J Eng Nat Sci* 2024;42:1670–1682. [\[CrossRef\]](#)
- [29] Aladağ H, Güven İ, Ballı O. Contribution of artificial intelligence (AI) to construction project management processes: state of the art with scoping review method. *Sigma J Eng Nat Sci* 2024;42:1654–1669. [\[CrossRef\]](#)
- [30] Neeraj SN, Vijay BG, Al-Mansour A. Study on pedestrian crossing decision at the roundabouts. *Sigma J Eng Nat Sci* 2024;42:1646–1653. [\[CrossRef\]](#)
- [31] Gözüoğlu A, Özgönel O, Gezegin C. Modeling and implementation of demand-side energy management system. *Sigma J Eng Nat Sci* 2024;42:1628–1645. [\[CrossRef\]](#)
- [32] Özden OB, Gökçe B. Determination of stress magnification factor in welded joint by experimental and numerical analysis using ultra-high-strength steel. *Sigma J Eng Nat Sci* 2024;42:1621–1627. [\[CrossRef\]](#)