Sigma J Eng Nat Sci, Vol. 44, No. 3, pp. 1826−1845, June, 2026

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2066

Research Article

Enhanced solar energy prediction using PSO-LSTM networks: A comparative study with decision trees and ensemble extra trees in Oman

Yahya AL BALUSHI[1,*], Salman YUSSOF[1], Abdullah AL BADI[2]

[1]*Department of Electrical and Computer Engineering, University of Tenaga Nasional, Kuala Lumpur, 43000, Malaysia*
[2]*Department of Electrical and Computer Engineering, Sultan Qaboos University, 123, Oman*

ARTICLE INFO

Article history
Received: 15 December 2024
Revised: 13 March 2025
Accepted: 28 April 2025

Keywords:
Decision Trees; Hyperparameter Tuning; Machine Learning; Renewable Energy; Solar Energy Prediction

ABSTRACT

Meteorological factors and their effects on the prediction of the Solar Global Radiation amount in Oman is the focus of this research study. The essence of the investigation was to provide accurate predictions to improve solar energy application effectiveness. The research utilized data from January 2013 through December 2024, and advanced analytical tools, including heatmaps, box plots, and time-series analysis were utilized to explore the relationship of temperature, rainfall, and solar global radiation. The conclusions of this research featured a moderate positive correlation between the temperature and solar global radiation, while rainfall was found to have an inverse relationship with temperature. The study also noted seasonal correlations of solar global radiation; the highest solar global radiation readings were about 800 watts per square meter (W/m^2) during the summer months, and the lowest were recorded during the winter months. The study also found that extreme weather events contribute to fluctuations in solar radiation, highlighting the need for reliable and accurate forecasting models in this field. In addition to providing insight into meteorological factors, the study also conducted an evaluation of machine learning models, including Long Short-Term Memory (LSTM) networks, Decision Trees, and Ensemble Extra Trees, to develop improved forecasting accuracy for these variables. The study demonstrated that the LSTM model configured with Particle Swarm Optimization achieved the best forecasting accuracy metrics; Mean Absolute Error (MAE) of 19.978, Root Mean Square Error (RMSE) of 20.930, and R^2 value of 0.974, indicate an accurate level of prediction reliability. This research presents a novel approach for optimizing Long Short-Term Memory (LSTM) networks. The main contribution of this research, therefore, is to use the findings from this study to create better forecasting models for solar radiation and ultimately provide tools to improve climate modeling, renewable energy planning, and sustainability. Improved forecasting improves the ability to integrate more solar energy into existing grids (i.e., grid stability) and creates opportunities for developing climate-resilient energy solutions.

Cite this article as: Al Balushi Y, Yussof S, Al Badi A. Enhanced solar energy prediction using PSO-LSTM networks: A comparative study with decision trees and ensemble extra trees in Oman. Sigma J Eng Nat Sci 2026;44(3):1826−1845.

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

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

Published by Yıldız Technical University Press, İstanbul, Turkey

INTRODUCTION

Because of the carbon-fuel industry's negative environmental impacts, much focus is being placed on renewable energy sources. To reduce or eliminate the energy produced by fossil fuels and to satisfy the rising demand for electricity from people and businesses, renewable energy sources are becoming an attractive alternative [1,2]. According to estimates, electricity produced from renewable energy sources in the United States will raise from 20% (2021) to 24% (2023), due to the ongoing development of the solar and wind productive capacities. In India, 101.53 GW (Sept. 2021), or 38% of all electricity, was produced by renewable sources, and this projection will jump to a target of more than 450GW (by 2030), of which 280GW (or 60%) will come from solar energy. Therefore, accurate and timely forecasts with solar power becoming more common will be critical. Among all sources of energy, one of the largest (in terms of production capacity) and most widely used global energy source is solar energy. As the global economy shifts from fossil fuels to renewable energy sources, effective energy management systems will be increasingly important to achieve the greatest amount of sustainability from the energy infrastructure through the greatest amount of potential for using renewable energy. Energy management systems are critical to the successful use of renewable energy they maximize how much we can incorporate, hold onto, and distribute renewable energy [3].

The Sultanate of Oman is currently undergoing a strategic diversification of its energy sector with an emphasis on achieving sustainable development and reducing its carbon footprint as outlined in Oman Vision 2040, and solar energy will be an important component of this transition due to Oman's high levels of solar irradiance. A significant milestone in Oman's efforts to harness its solar energy potential is the recent commissioning of the Manah 1 Solar PV project, which has substantially increased the country's renewable energy capacity. This expansion necessitates the deployment of advanced energy management systems to enable efficient grid integration and optimal utilization of solar resources through reliable power forecasting. However, achieving accurate solar power output predictions in Oman remains challenging due to its arid desert climate, high ambient temperatures, and frequent extreme weather events such as dust storms [4]. Despite ongoing advancements in solar technologies and the development of large-scale solar projects, there remains a clear need for localized, high-precision forecasting models that adequately account for site-specific environmental conditions. Furthermore, a comprehensive assessment of the economic feasibility and return on investment (ROI) of solar energy projects in Oman is essential to support informed decision-making by investors and stakeholders [5]. This study will present a comprehensive real-world case study that will predict and validate the performance of a kilowatt photovoltaic (PV) power plant that was developed in Eastern India. Such an analysis will be essential to move forward with new solar PV projects with an emphasis on 1) understanding how regional weather and climate affect solar energy generation, 2) how solar energy will be produced and used to meet local energy consumption needs, 3) developing accurate algorithms for solar power prediction, and 4) evaluating ROI of solar investors [6]. The accurate forecasting of solar power generation is critical in the successful operation of power generation systems that use solar energy as a substantial component of total energy generation. This study will concentrate on providing accurate forecasts for solar power generation in Eastern India [7] due to the time series nature of the meteorological data, with the aid of ensemble machine learning (EML) algorithms, which have been shown to provide increased accuracy in forecasting [8,9]. Optimizing energy management with solar power while achieving power supply reliability and stability addresses the challenges of intermittent nature of solar energy. Predictive modeling and forecasting are vital to this area. Recently, ML and AI have both produced successful results in improving renewable energy forecasting accuracy [10]. Providing accurate forecasting improves the integration of solar energy into electricity supply systems and helps run the system more efficiently and reduces reliance on fossil fuels [11] Estimates of the effectiveness of ML and AI models have been confirmed through studies that have concluded that AI and ML create accurate predictions of solar energy generation [12]. Using a large amount of meteorological and historical power output data enables machine-learning techniques to predict solar power generation with a high level of accuracy. Predictive techniques encompass various methods such as ensemble methods, neural networks, and hybrid methods that use multiple machine-learning techniques to enhance the accuracy of prediction. These methods are sophisticated and recognize complex patterns and temporal dependencies in the data, improving the reliability of predictions [13].

Predictive technology will enable the energy management system to more accurately predict variability in solar power generation, allowing for the operational adjustments to meet increasing electric demand and address the environmental impact of electric generation. There has been a lot of previous research done on solar panel prediction systems using techniques such as deep learning. Many of the previous studies utilized common techniques for data analysis, including principal component analysis (PCA), to complete multivariate studies of solar irradiance predictions. A number of the research methodologies used included deep learning methods that were combined with domain knowledge, short-term prediction models, means of predicting based upon Pearson correlations, and means of predicting with outputs of photo-voltaic (PV) systems at the specific location where measurement was conducted for each site that made up the applicable studies. Several authors utilized differing predictive regression models and ensemble approaches to produce greater accuracy for predicting the amount of solar power produced. This article will provide a concise yet thorough review of past research conducted on this topic, as well as discuss its relevance to our research. The analysis reveals that utilizing data mining techniques produces significant improvements in the accuracy of all machine-learning-based prediction systems [14]. Multiple recent studies provide differing methodologies to produce predictions and estimates of the power produced from photovoltaic (PV) systems [15]. The studies used a phenomenological model which is considered to be a scientific model that accurately represents all the relationships of events based on the core theory, e.g. the laws of thermodynamics, but does not utilize core theories as explanation [16]. In other words, phenomenological models do not rely on any "fundamental" principles for their generation. The intention of a phenomenological model is the definition of a connection or association and not the explanation of how those variables interact with each other. This form of a relationship can be anticipated to exist beyond the variable's observable value [17]. Recent research has employed numerous methods to estimate solar photovoltaic (PV) output, and machine learning methods have gained in popularity due to their ability to model complex non-linear relationships. Studies by Yadav et al. concluded that artificial neural networks (ANNs) can provide accurate estimates of short-term solar photovoltaic (PV) production by training the ANNs with past weather conditions and corresponding PV output data [18]. In addition, researchers have investigated hybrid models that utilize both physical and statistical modeling techniques to take advantage of both approaches [19]. Qazi et al. recently explored the possibility of using both a physical modeling technique to estimate solar radiation and an ANN to improve both the accuracy and reliability of solar PV estimates during weather variability. This hybrid method, which uses elements of both physical and phenomenological modeling, is able to address the limitations associated with pure phenomenological models [20]. Additionally, advancements in satellite and remote sensing technologies have improved the quality of the data available for PV Power prediction modeling [21]. The utilisation of high-resolution satellite imagery, combined with ground-based weather stations, provides a larger dataset from which to develop predictive models, leading to improved accuracy in PV Power projections (Sadaei et al., 2016). Current research has shown an increasing interest in the use of ensemble learning techniques. These techniques combine the predictions from multiple models to create a more reliable overall result. Ensemble methods such as gradient boosting, and bagging have proven to be effective at reducing the uncertainty associated with predicting Photovoltaic (PV) Power output [22].

For example, Anuradha et al., (2021) provided an efficient functional model that utilized a prediction interval technique and outperformed traditional predication techniques based on bootstrapping. This functional model can be applied to time series data that have periodic characteristics; however, it can also be used to make predictions for non-seasonal time series data [23]. Raafat proposed a hybrid prediction framework that combines k-means clustering and support vector regression (SVR) for estimating global solar radiation. This method used six (6) years of meteorological data collected in Ibadan to predict future levels of solar radiation [24]. To predict solar irradiance, Cabello-López and others used PCA for reducing data dimensions and ANN and AnEn for predicting solar irradiance. They found that small daily changes in atmospheric conditions can have a large impact on the amount of solar energy produced by PV systems [25]. Prabhakar also developed a hybrid method that employs both meteorological classifications and SVM to predict how much power PV systems produce every 15 minutes, including classification of clear skies, cloudy skies, foggy skies, and precipitation. In parallel, a study utilizing numerical weather prediction (NWP) outputs from the GEFS employed various machine learning techniques to predict solar energy generation across multiple grid nodes. [26].

The purpose of this study was to provide a more thorough examination of machine learning algorithms in the context of refining analysis of meteorological data and increasing the accuracy of PV power prediction [27]. The analysis utilized an expanded array of training and evaluation techniques with a specific emphasis on expanded ensemble learning. The combination of the predictive ability of numerous models was intended to reduce the occurrence of prediction errors and thus improve the total reliability of the prediction process [28,29]. Rigorous evaluation was completed on the common ensemble methods including Random Forest, Gradient Boosting, and stacked algorithms. The models we used are appropriate for characterizing the complex and nonlinear nature of how we generate solar power. The issue of overfitting that arises from having a single-model model was also accounted for. In addition to conventional mathematical methods for predicting time series data, we investigated using time series specific models (Long Short-Term Memory (LSTM) networks), and we devised innovative hybrid models using LSTM and convolutional neural networks (CNN) [30]. This dual model developed a way to analyze separately the time-dependent behavior of meteorological data over time while extracting spatial attributes of environmental data related to meteorological variables. We were able to develop a robust and adaptive prediction framework through the effective fusion of these advanced methods [31]. A novel prediction framework was developed to effectively capture the inherent variability of solar irradiance and its impact on photovoltaic power generation. By enabling more accurate real-time forecasting, the proposed approach enhances the operational efficiency of solar power plants and supports informed investment decisions and strategic planning within the rapidly evolving and competitive renewable energy sector [32]. Accurate prediction of solar energy production remains challenging due to the inherent uncertainty of environmental conditions. Solar power generation is strongly influenced by meteorological factors such as cloud cover, humidity, and ambient temperature, all of which exhibit significant temporal variability. Consequently, fluctuations in these parameters lead to corresponding variability in the amount of solar radiation available to photovoltaic systems. Changes in cloud cover can lead to turbulent swings in solar power generation and have a massive impact on the ability to perform accurate forecasts of solar energy generation [33]. One of the limitations in relation to the challenge of accurately forecasting solar energy generation is the lack of a systematic approach to forecasting solar energy generation in comparison to other renewable energy technologies like wind energy. This creates difficulty in developing accurate forecasts for solar energy generation, particularly for short periods of time. Solar power forecasting is the process of collecting and analyzing historical and current data for predicting solar energy production during various time periods. Still, the unpredictability of the weather makes the ability to produce accurate solar power forecasts even more challenging [34]. Furthermore, the volatility of solar energy can create a major barrier to the integration of the electric power system. Rapid fluctuations in solar power generation caused by changing weather conditions can adversely affect the reliability and resilience of electric power systems. These challenges necessitate advanced forecasting and energy management strategies to effectively balance electricity supply and demand. Accurate solar energy forecasting is therefore critical for the efficient management of renewable resources and the successful integration of solar power into existing electricity grids [35]. To address these challenges, a wide range of modeling approaches have been developed that utilize diverse data sources, including ground-based sensor networks, satellite imagery, numerical weather prediction models, and historical generation data. In recent years, machine learning techniques have increasingly been adopted to enhance the accuracy and robustness of solar energy forecasting. These advanced methods make use of real-time environmental measurements and historical data to create better prediction models. The inherent unpredictability of solar resources will continue to be one of the major challenges; therefore, ongoing research is needed to improve forecasting methodologies [36].

The focus of this study is to review the performance of different advanced machine learning algorithms to forecast output power generated from the solar resource. The purpose of this research is to compare different machine learning algorithms to see how well they forecast solar radiation and solar energy. This research conducts a comparative evaluation of several machine learning algorithms, including artificial neural networks (ANN), Random Forests, and Gradient Boosting, to identify the methods that offer the highest accuracy and reliability for solar energy forecasting. By analyzing historical forecasting data using advanced mathematical and machine learning techniques, the study aims to enhance the precision of solar power predictions. The overarching objective of this work is to improve the accuracy of solar energy output forecasting through the effective application of machine learning methodologies. Traditional empirical methods do not consistently hit the mark when trying to capture the complex variation of solar radiation caused by changing weather patterns. Machine learning algorithms have been shown to have promise in providing more accurate forecasts by incorporating large volumes of both historical and real-time data into forecasts. Here we test the validity of these algorithms under various conditions to ensure that they remain resilient and reliable for real-world applications. The primary objective of this research is to enhance forecasting methodologies through the systematic evaluation and validation of advanced machine learning algorithms. The research will also support improved integration of solar power onto the electric grid and provide better management of renewable energy resources through an emphasis on accuracy, reliability, and real-time capabilities.

Geographical Location

Oman is located at a strategic point along the southeastern coast of the Arabian Peninsula. The site of this investigation is located for different areas of Oman, more precisely at coordinates of 22.280°N and 59.160°E. This area is characterized by dry desert landscapes which contain large areas of flat land with very few plants. The borders of Oman are shared by the United Arab Emirates to the northwest, Saudi Arabia to the west and Yemen to the southwest, while its long coastline along the Arabian Sea results in varied topography as shown in Figure 1. Oman has a hot desert climate according to the Köppen climate classification system, with most areas of Oman being classified as BWh [37] (meaning the summer months are extremely hot). During the summer months, average daily temperatures can often exceed 40°C (104°F), but the winter months are considered to be mild and moderate in temperature. The average annual precipitation is approximately 100mm; with the majority of rainfall occurring between the months of November to April.

This weather pattern provides an annual average of over 3,500 hours of sunlight, therefore creating a reliable and abundant source of solar energy. The dry climate in Oman also makes it a prime location for solar energy projects, particularly for the generation of electricity using photovoltaic (PV) systems [38].

MATERIALS AND METHODS

Figure 2 illustrates a systematic approach to predicting solar energy with machine learning techniques. The process begins at the Data Collection stage where all relevant data is gathered including past solar energy production data as well as data related to meteorological conditions such as air temperature and wind direction. In the case of Oman solar

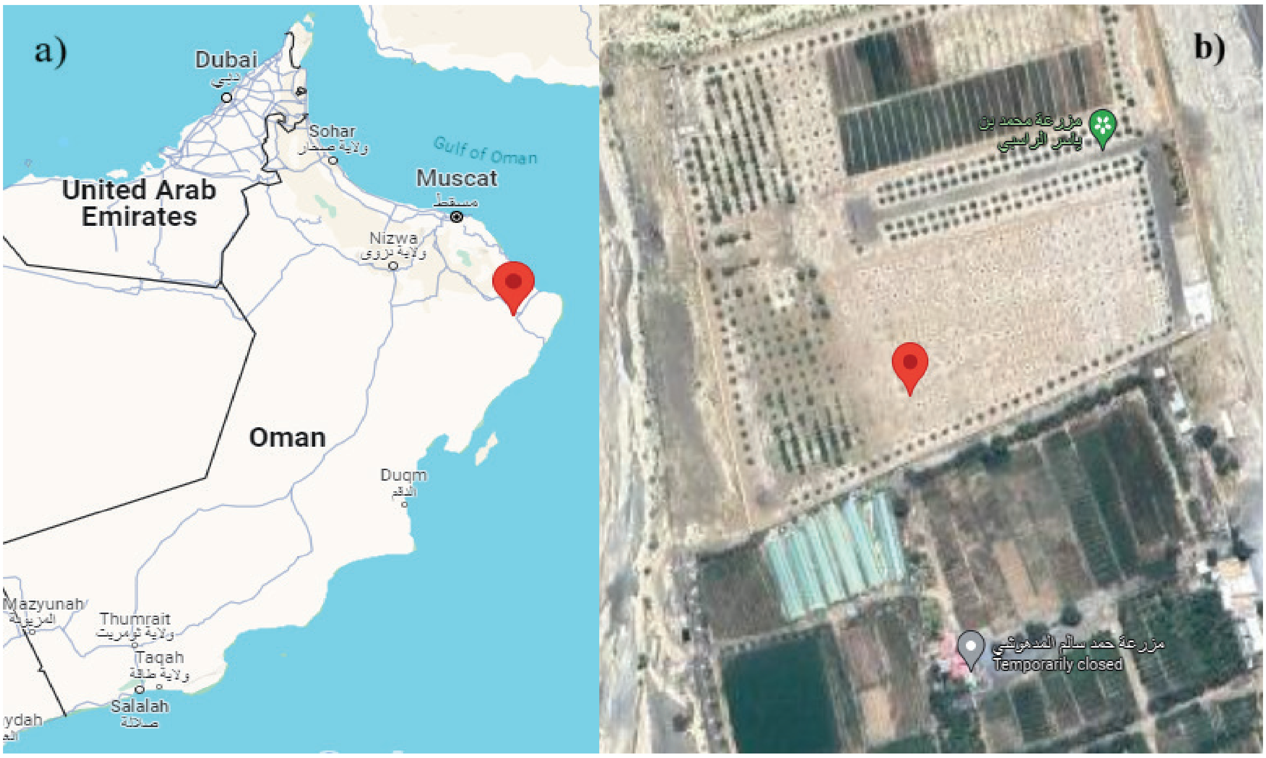

Figure 1. Geographical location of the site a) Map b) Actual site.

power and meteorological data is obtained from sources like the Directorate General of Meteorology (DGM) through the Oman Civil Aviation Authority, the Renewable Energy Research Center (RERC) at Sultan Qaboos University or various other publicly accessible databases including the Photovoltaic Geographical Information System (PVGIS) and NASA's POWER Data Archive [39,40]. These types of datasets also include all meteorological conditions such as solar irradiance, air temperature, wind speed and any other atmospheric variables that may affect forecasting of solar energy production in converting from a machine learning model into a working machine. Data is organized via Data Preprocessing prior to training or running through a machine learning model to ensure that all data will meet the requirements for creating a successful machine learning model [41]. This involves many steps including removing outliers, normalizing data and managing missing data. Data normalizing has the ability to help a machine learning model converge towards a solution quicker if the dataset has been normalized (i.e., scaling each value from a given range of 0 to 1) [24].

To complete the next step, we will select a prediction model through three different methodologies, where each uses a different machine learning model. First, we will use Long Short-Term Memory (LSTM) with Particle Swarm Optimization (PSO). LSTMs are well suited for working with sequential data such as time series of solar energy. By using PSO, we hope to optimize LSTM to improve the accuracy of our prediction [3,6]. The second method is Decision Trees. Decision Trees employ a hierarchical structure of decision rules to decompose the prediction task into a series of smaller, more manageable problems. The Extra Trees method extends this approach by combining multiple decision trees into an ensemble, thereby improving predictive accuracy and reducing model variance [42].

After selecting the best model from the Model selection process based on performance on a validation data set, optimization and prediction will follow as the next steps. The training of the selected ML model will use the preprocessed data [43]. The PSO optimization method will also be used to optimize the hyper parameters of LSTM at that point. Finally, the final model selection will be used to predict the future amount of solar energy output. A systematic approach using a flowchart can provide an accurate estimate of solar energy generation using machine learning. This study shows that the progress of renewable energy can be supported by predicting solar energy generation using a systematic method employing flowcharts [44].

DATA COLLECTION

Databases for PV generation data were used based on data collected over 10 years, specifically between January 1, 2013 and January 1, 2024. The use of a long enough historical database allows for a comprehensive study of long-term trends and regularly occurring seasonal variations of

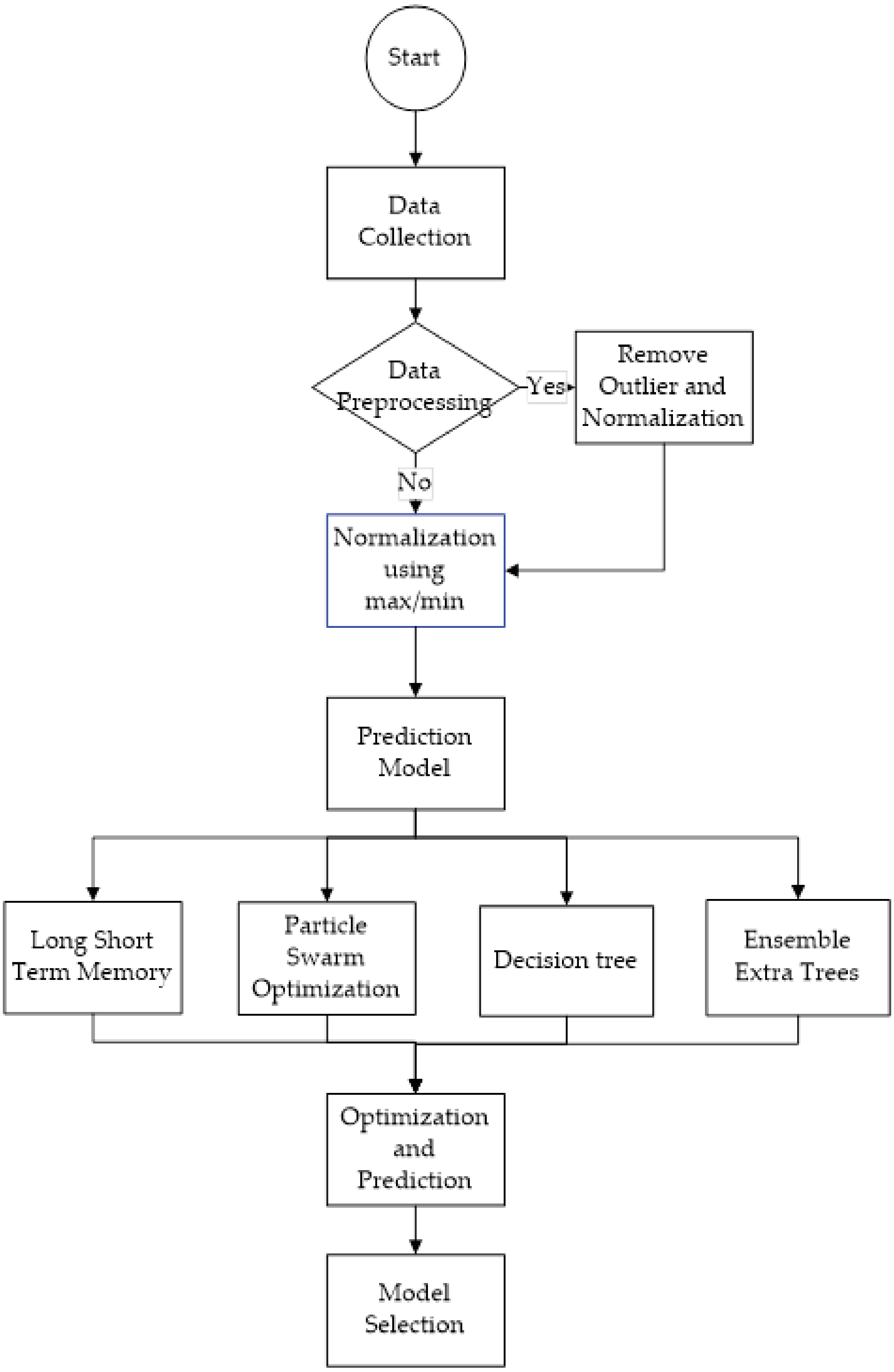

Figure 2. Solar energy prediction using machine learning models.

generation from photovoltaic systems. Meteorological data consist of various environmental conditions that affect a PV system's operation. Atmospheric solar irradiance has both a direct and diffuse component of sunlight. Direct solar irradiance is sunlight that reaches the Earth's surface without being reflected or refracted, while diffuse/indirect solar irradiance is sunlight that has been reflected/absorbed by the atmosphere before it reaches the surface of the Earth. When combined, they represent all solar energy available for conversion into electricity using PV systems. The ambient air temperature surrounding a PV panel affects how efficiently and effectively the PV cells operate. Specifically, at higher ambient air temperatures, PV modules typically achieve lower output than do PV modules at lower ambient air temperatures. Humidity describes the amount of moisture in the atmosphere and will affect the amount of solar energy able to travel through the atmosphere to your PV panels, as well as potentially leading to moisture accumulation on the surface of PV panels and reduced efficiency at the time of generation. One way to improve the efficiency of PV panels is to cool them down, as their operating temperature is related to the wind speed and direction. In addition, it is essential for engineers who design photovoltaic (PV) systems to understand wind patterns to ensure they can withstand the force of wind.Rain also plays an important role in assuring that PV panels remain clean, as rainwater can effectively wash away dust and dirt. Conversely, excessive rainfall can create potential drainage issues, as there is a possibility that the PV's electrical components could be saturated with water.

There are several critical performance metrics associated with PV systems. These metrics can be used to measure the success and output capability of a PV system. Power is one performance metric used to measure a PV system's ability to generate an electrical output (measured in kilowatts (kW) or megawatts (MW)). System efficiency, another performance metric, indicates the effectiveness of a PV system's ability to convert solar energy into electrical energy; this measurement is reflective of both electrical losses and environmental factors. Incident solar radiation is the amount of solar energy incident to the PV panels, and this measurement directly relates to the amount of electrical energy generated. The temperature of the PV modules is yet another performance metric; the efficiency of the PV modules is dramatically affected by their temperature. Typically, as the temperature of the modules increases, the efficiency of the PV modules will decrease [43].

Purpose of Data Collection

The main aim of collecting all of this thorough data is for the exploration of how photovoltaic generation of power occurs both long term and short term to develop accurate models to forecast. In addition, the information will also allow an evaluation of local meteorological conditions on performance therefore providing better designs for how to utilize PV systems when they are built in an area similar to where they are located. This information will assist in evaluating forecasting techniques such as phenomenological approaches and the use of machine learning methods as well as hybrid methods [45].

Data Prepossessing

After the data has been imported, it will begin the data preparation process. This includes eliminating missing values and eliminating unrelated columns. After unrelated columns have been removed, the next step is to find and remove outliers. Outliers must be detected and dealt with to ensure the accuracy and reliability of prediction models. Outliers have the potential to bias how a model is trained, resulting in inadequate performance and inaccurate predictions. Outliers are defined by evaluating the variables of interest (i.e. wind speed, relative humidity, and temperature). For instance, through statistical evaluation of the previous variables, thresholds and ranges are established for these parameters and any data points lying outside of these thresholds are defined as outliers and removed from the dataset. The process visually displayed in Figure 2 ensures the remaining data represent typical climate and operating conditions [46].

After the data has been cleaned and processed, it is used for training four different models (LSTM, LSTM/PSO, Decision Tree, and Ensemble Extra Trees). LSTM (Long Short-Term Memory) is a specific type of recurrent neural network (RNN) that is designed to effectively predict future values in a time series. It also allows for the understanding of long-term relationships between variables and sequences of data. The LSTM created model will be adjusted with Particle Swarm Optimization (PSO) in order to enhance prediction results through hyperparameter tuning. Additionally, the Decision Tree created model will be used as a non-linear algorithm that can be used as a hierarchical structure to make decisions base on input characteristics as well as capture non-linear correlations between variables. An alternative technique used within this study that involved ensemble learning was the Ensemble Extra Trees method which combines multiple decision trees within the same ensemble to provide more accurate and robust forecasts by minimizing variability and reducing risk of overfitting [47].

Models Employed in the Investigation

The examination employs four methods of modeling photovoltaic (PV) output power, utilizing various techniques and approaches for increasing predictive accuracy and resiliency. Long Short-Term Memory (LSTM) is one type of Recurrent Neural Network (RNN) specifically designed to predict future points in a time series using historical sequential records. By leveraging historical weather and production performance information, RNNs develop a reference and comprehension of long-term relationships within sequential data. Therefore, LSTM networks are well suited for forecasting PV production [48]. To further enhance the accuracy of LSTM forecasting, Particle Swarm Optimization (PSO), is applied as a way to tune the hyperparameters of the LSTM network. PSO reproduces the social behaviors associated with the collective decision-making processes that arise from a swarm of birds or school of fish searching for the optimal solution in a given area [49]. Data is partitioned based upon value of the input variables, creating a model that is easily interpretable while also capturing non-linear relationships in the data. A decision tree is an example of a non-parametric model as it creates predictions through a hierarchical algorithmic approach resembling a tree [50]. Ensemble Extra Trees, also known as Extremely

Randomized Trees, is an ensemble learning approach that aggregates multiple decision tree outcomes to provide more accurate and robust forecasts compared to individual decision tree models. This method reduces variability in the results and the risk of overfitting by selecting diverse data subsets and then randomly determining the decision points within each decision tree.

In conclusion, the goal of this project is to develop accurate and reliable PV output power forecasting tools using a combination of various models and large amounts of data for both meteorological conditions and PV system performance. Each model type used has its own advantages: LSTM models have the ability to capture temporal relationships; PSO (Particle Swarm Optimization) provides improved optimization processes; decision trees give the models interpretability; and ensemble models provide the highest level of accuracy for all predictions made by these models combined. This multi-faceted approach addresses the complexity of forecasting PV output power in a variety of types of environmental conditions [51].

RESULTS AND DISCUSSION

Solar Global Radiation (SGR) and its Correlates

Heatmap Figure 3 of the correlation between several weather variables at a location in Oman shows numerous interesting relationships exist between different meteorological variables temperature and rainfall have a significant negative correlation, meaning that as temperature goes up, precipitation generally goes down. This correlation shows that usually when the temperature is higher, the amount of rain will usually be lower. Therefore, this negative correlation between temperature and precipitation is extremely important to understand weather patterns, especially in very hot areas which may experience a decrease in rainfall. In contrast, relative humidity (RH) and rainfall have a very small positive correlation, which means that generally when there is a higher relative humidity there will be a small amount of rain. Therefore, increased humidity may have a small effect on the total amount of precipitation. This is a helpful correlation for predicting future weather patterns in humid climates, but does not have any impact on predicting future weather patterns. Also, there is a very small negative correlation between sea level pressure (SLP) and rainfall, which indicates that lower SLPs are correlated with slightly higher amounts of rainfall. The association between low SLP systems and significant precipitation is probably caused due to the more chaotic and turbulent nature of weather systems when there is low SLP [52].

Besides that, the heat map strongly shows how low temperatures (TEMP (Min)) have a positive correlation to the minimum relative humidity (RH(Min)). This helps to demonstrate the relationship that exists between low overnight temperatures and high minimum relative humidity levels which can potentially influence dew forming and overnight weather [53].

The correlation between SGR and various meteorological variables yields useful insights. Specifically, the three temperature variables – mean, maximum, and minimum, i.e., TEMP(Mean), TEMP(Max), TEMP(Min) – tend to have a small positive relationship to solar global radiation (SGR), as indicated by the analysis. This means that as temperature increases, there tends to be more illumination and therefore more SGR. This relationship is expected due to the fact that increased SGR results in increased surface temperatures. The vapour pressure variables (VP(Mean), VP(Max) and VP(Min)) also exhibit a small to moderate positive relationship to SGR. This shows that vapour pressure (which represents moisture present in the atmosphere) is slightly related to SGR, increasing as vapour pressures increase. Therefore, moisture may contribute to the greenhouse effect causing a higher temperature and more SGR. [54].

The variables of relative humidity (RH(Mean), RH(Max), RH(Min)) have a slight negative correlation with solar global radiation (SGR). Relative humidity is sometimes associated with cloudy days, and higher levels of relative humidity slightly reduce the amount of solar energy that reaches the ground level. It is important to know about the inverse relationship between cloud cover, moisture in the atmosphere, and the amount of solar radiation that is available. The sea level pressure variables (SLP(Mean), SLP(Max), SLP(Min)) indicate a slight association with SGR, indicating that changes in sea level pressure will not have much effect upon solar radiation. In addition, the data shows that there is a very weak negative correlation between wind speed (WMS) and SGR, meaning that wind speeds have little to no effect on how much solar radiation is reported.

Finally, there is a weak negative correlation between rainfall and global solar radiation (SGR), meaning that when there is more rain, there will be more clouds to block the sun's rays and the amount of solar energy available is reduced. The weak correlations identified in this study will be useful in the development of comprehensive weather and climate models, but there are many other variables that are more significantly correlated with solar radiation. In summary, the temperature and vapor pressure variables had relatively strong correlations to SGR. The lower correlation values observed for relative humidity, sea level pressure, wind speed and rainfall demonstrate that these elements have a less significant effect on solar radiation than the other factors studied. Results from this study will be important in the fields of meteorology and climate modeling as they will aid in identifying the main variables in predicting solar radiation and determining weather patterns for solar radiation forecasting and weather pattern analysis.

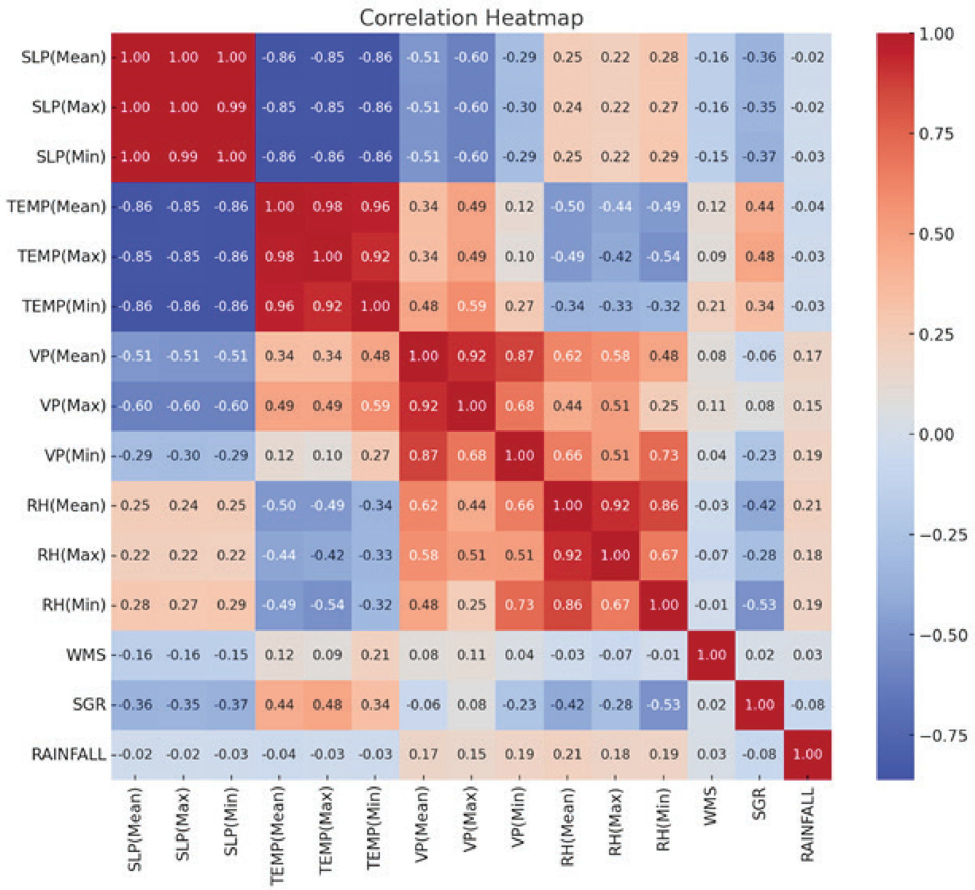

Figure 3. Correlation Map for the dataset variables this data just shows the raw not specific information.

Where,
SLP(Mean): Mean Sea Level Pressure
SLP(Max): Maximum Sea Level Pressure
SLP(Min): Minimum Sea Level Pressure
TEMP(Mean): Mean Temperature
TEMP(Max): Maximum Temperature
TEMP(Min): Minimum Temperature
VP(Mean): Mean Vapor Pressure
VP(Max): Maximum Vapor Pressure
VP(Min): Minimum Vapor Pressure
RH(Mean): Mean Relative Humidity
RH(Max): Maximum Relative Humidity
RH(Min): Minimum Relative Humidity
WMS: Wind Speed
SGR: Solar Global Radiation
RAINFALL: Total Rainfall

Box Plot for Solar Global Radiation (SGR)

The box plot of solar global radiation (SGR) Figure 4 offers information on variations (data spread)/distribution of SGR data at the same site. The median (~600 W/m²) is indicated by the horizontal line within the box, with 50% of the values being less than and 50% of the values equal to or greater than 600 W/m². The median will not be as skewed by extreme values as the mean; consequently, the median is a better representation of "typical" SGR values. The red box (interquartile range) represents the range of data that falls within the two middle quartiles of the SGR data set. The interquartile range is determined by calculating the difference between Q3 (~800 W/m²) and Q1 (~450 W/m²); thus, the greater the IQR value the more variability in

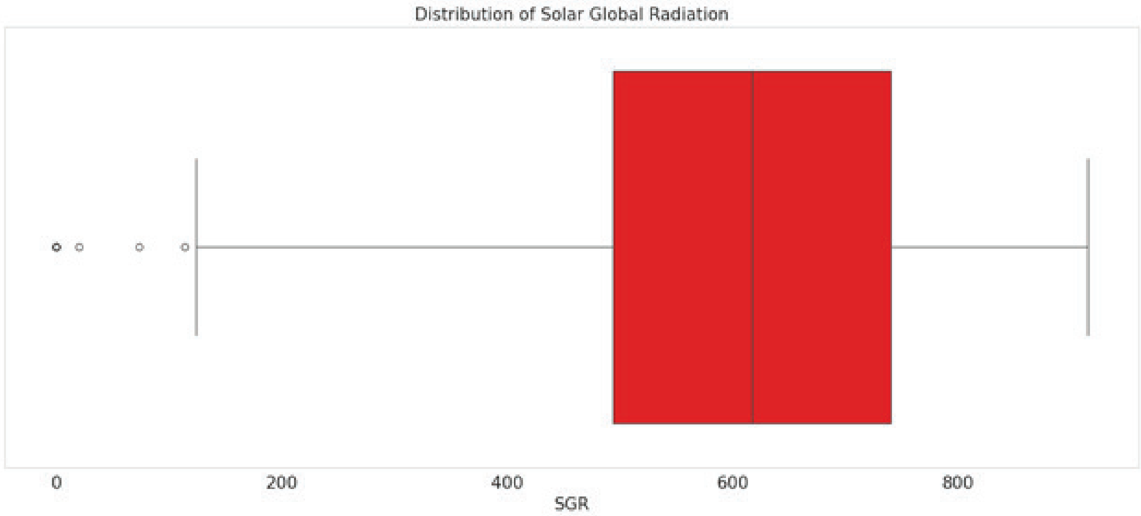

Figure 4. Box plot representing the distribution of solar global radiation (SGR) values.

the data and the less the IQR value the more consistent the SGR values [11]. The whiskers extend from the boxes and cover the vast majority of SGR data records (100-900 W/m²), with the exception of outlier SGR records, which can be lower than 100 W/m². The low-end outlier records indicate periods where SGR had low values (0 - 100 W/m²) due to reduced solar exposure (cloud cover), as well as times of day where SGR has low values (early AM/early PM) or there were anomalies within the SGR data. Whiskers (the line that extends from each side of the box) typically extend from the box by 1.5 times the interquartile range, giving us a general idea of how much the SGRs generally vary.

The upper whisker indicates the maximum value for a non-outlier SGR, and the lower whisker indicates the minimum value for a non-outlier SGR. Outliers are plotted as points on the top of the box plot; outliers indicate a significant deviation from the expected range for SGR values. The presence of outliers may also indicate an unusual atmospheric occurrence, or that there was an error in the process of measuring the SGR. The box plot gives us a clear representation of how SGRs are distributed; it shows that SGRs tend to be very high, with occasional dips in SGRs, and that we can expect a high spread between the two whiskers, while also showing us that there is significant variability due to seasonal changes, and/or atmospheric conditions [55].

Box Plot for Temperature Variation

The box plots Figure 5 show the distribution of the average (mean), maximum and minimum temperatures (°C) for a specified time period. The box plots provided in the graphical representation are very useful because they allow a user to determine the central tendency (average) of the temperature, the degree of dispersion (how far the individual temperature observations are from the average), and the degree of variability (how much the mean value varies during the specified time period). The box portion of the box plot is the interquartile range (IQR) of the temperatures, with a line in the box indicating the median temperature and two lines extending from the IQR to the minimum and maximum temperatures (with the minimum and maximum values expected). Any temperatures that are identified as potential outliers would be displayed as individual points above and below the whiskers [56].

The first boxplot displays the distribution of the mean temperature of the year (Celsius). The median temperature is approximately 30°C meaning that half of the recorded temperatures are below that point and half are above. The interquartile (IQR) range is from approximately 25°C (Q1 - 25th percentile) to 35°C (Q3 - 75th percentile). Therefore, 50% of the recorded mean temperatures occur within that range. The whiskers extend from approximately 15°C to 40°C, showing the minimum and maximum values found in the expected distribution. The extent of the whiskers indicates that there is moderate to high variability of mean temperatures from 15°C to 40°C; therefore, mean temperatures can be quite variable and range from 15 to 40 degrees Celsius on occasion. The second box contains the maximum temperatures recorded. The median appears to be around 37°C; the IQR extends from approximately 30°C (Q1) to 42°C (Q3). The whiskers also extend from approximately 20°C to 48°C, illustrating very high variability in maximum temperatures. The record of higher temperatures indicate extreme weather was recorded; for instance during summer season when temperatures peak. The large IQR denotes large variations between the daily maximum, which could be related to seasonal variations, climatic variability or extreme heat. The next graph shows how widely the minimum temperatures differ. The value in the middle (the median) of the minimum recorded temperature is about 22 degrees Celsius, and the difference

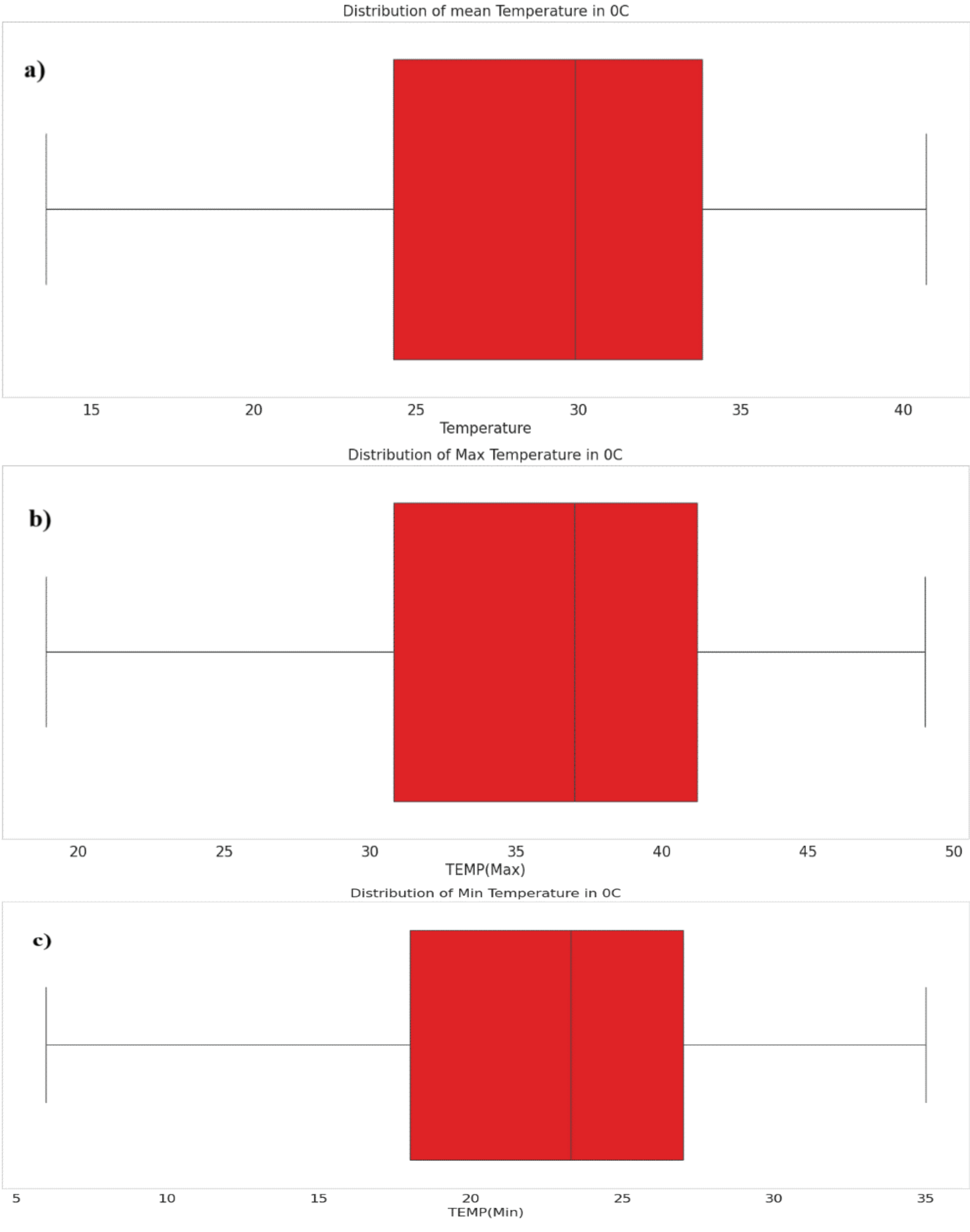

Figure 5. Box plot representing the distribution of Temperature variables a) Mean temperature b) Max temperature c) Max temperature.

between the first quarter (Q1) and third quarter (Q3) is 18 degrees Celsius to 28 degrees Celsius. The "whiskers" of the box plot extend from about 5 degrees Celsius to 35 degrees Celsius, indicating that there have been some very low temperature values recorded. The bottom of around 5 degrees Celsius suggests that once in a while, we get cold spells or some change in the seasons that affect temperatures significantly. These temperatures have a wide range in the box and whiskers of this box plot proving that, in general, minimum temperatures have a greater range than the other two temperature values (mean and maximum). There are other meteorological events such as the passing of a cool front, the cooling of the ground during nighttime, etc. that would contribute to the differences in temperature values.

Variation of SGR with Temperature

Figure 6 is a time series graph showing how well solar global radiation (SGR) corresponds with some of the temperature measurements collected in Oman between 2015-2024. The colors of the points on the graph help to show the physical difference between the data points on the graph more easily.

Both SGR and temperature measurements show clear cyclical patterns demonstrating that there are significant seasonality in the data. There are annual fluctuations in the data with months of high and low values. In weather-related data, such as the temperature and solar radiation, we typically observe a continuous pattern arising from the effect of seasons on those two parameters. In this data, the SGR (Sunlight/Ground Reflectance) value is represented by the red dots, and there is evidence of a repeating cyclical pattern in SGR, with the peaks of the SGR values coinciding with the corresponding yearly peaks in SGR (reflecting the lowest levels of solar radiation) and the same for the valley (seasonal decline). Based upon the established relationship of solar insolation to surface temperature differences over time, the evidence of the peak SGR values correlating to the historical peaks of temperature corresponds with the peak values of solar radiation each year [58].The blue dots in the chart represent the trend of (Tmean) and would follow a similar pattern as the SGR; however, they would have a less pronounced peak and valley pattern, suggesting that Tmean would also have an annual trend over time, following the same trend as the solar insolation trends for Tmean.

The TEMPmax data (Purple) shows much more variance than the Tmean (Blue) data with TEMPmax having peaks and valleys that show the greatest temperature extremes during summer months and during winter months; therefore, the TEMPmax (Purple) data has experienced all of the highest, lowest and average temperature extremes due to changes in the solar radiation and local weather throughout the year. The TEMPmin (green) data contains seasonal variations similar to the TEMPmax (purple). The cyclical nature of both solar global radiation (SGR) and temperature exhibits similar cyclical patterns but with reduced ranges, potentially indicating seasonal low temperatures or nighttime cooling impacts. The lower limit of the TEMPmin (minimum temperature) indicates occurrences of very low minimum temperatures during the winter months, as indicated in reference [57]. The patterns between the two components SGR and temperature (TEMP) showed positive association with each other. This is expected due to the correlation between increasing solar radiation and increased temperatures, as well as the correlation to decreasing solar radiation with decreased temperatures. Solar radiation acts as a key driver of the variability seen in temperature. This is supported by the time-series graphic displaying the visible seasonal variability of both SGR and temperature. The visible recurring seasonal patterns reflect the effects of the seasonal changes of climatic components on other meteorological components. The correlations observed between SGR and temperature indicate that there is a co-dependent relationship between these two climatic components; and that SGR significantly influences the temperature variability experienced on the earth. The graphic allows for great insight into the temporal nature of both solar global radiation and temperature; both of which are important for forecasting weather, and modeling climate [23].

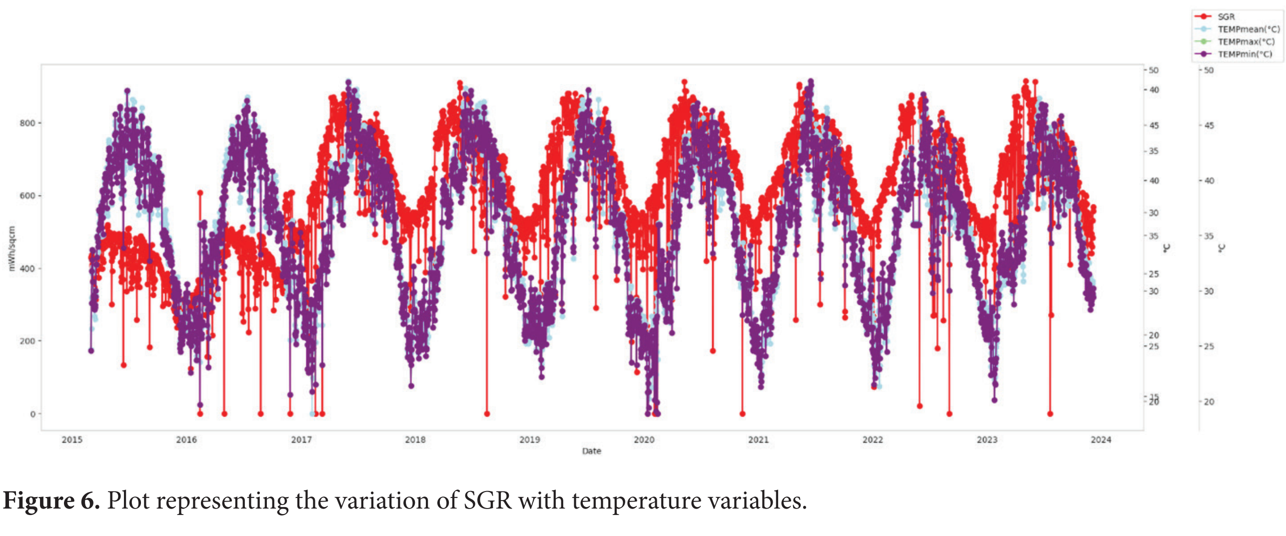

Figure 6. Plot representing the variation of SGR with temperature variables.

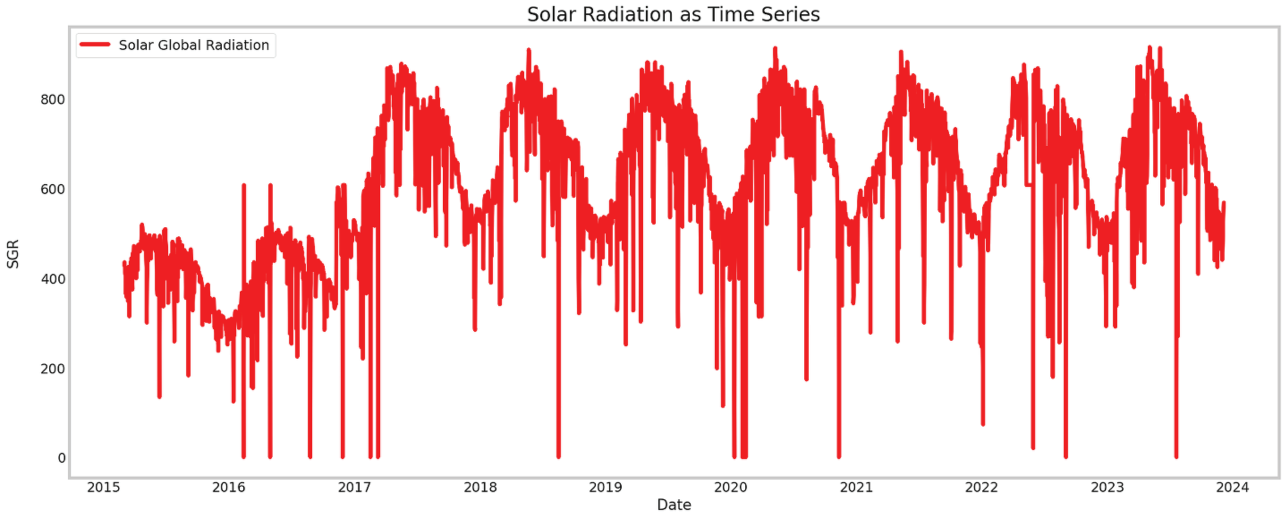

Figure 7. Solar radiation time series plot.

Time Series Plot of Solar Radiation

Figure 7 shows the time series of Solar Global Radiation (SGR) from 2015-2024 and illustrates how solar radiation amounts change throughout a year. The plot shows strong seasonal patterns; in particular, there are usually high amounts of solar radiation in the middle part of the year (summer months), while there are usually low amounts in the early and late parts of the year (winter months) and this evidence of solar radiation's annual pattern can be explained by the seasonal nature of solar insolation.

The maximum recorded SGR values of over 800 W/m^2 mostly occur in the middle part of the year. The lowest recorded SGR values, which can be generally close to 0 W/m^2, would likely be caused by nighttime, cloud cover or disturbances in the atmosphere. Although there is a nice yearly cycle with the data, there are slight differences in the magnitudes of the peaks and troughs indicating some variability from year to year, such as variations in solar radiation caused by weather variations, changes in atmospheric composition, etc. [58]. However, even though SGR does vary from year to year, it appears that there has not been any long-term trends (increasing or decreasing) over the nine years, which means that the overall amount of solar radiation available during the nine year time frame remained fairly constant. The fact that solar radiation cycles are regular, with some deviations, reinforces the need for long-term data sets to look for anomalies and patterns that can be utilized for practical solar energy applications [59].

LSTM Optimized with PSO Algorithm

Scatterplots representing the outputs of an LSTM with PSO model shown in Figure 8, where forecasts (red dots) cluster well together and show a strong resemblance to the ideal fit line (blue) indicates the same close proximity that was observed in terms of predicted values and actual (SGR) values (shown on x-axis) [60]. This conclusion is supported by the lower errors recorded in the error measures; MAE =19.978, MSE=438.081, RMSE=20.930 which confirms a strong correlation between predicted values and actual values. A significant R^2 value (0.974) demonstrates that the model accounts well for the majority of variance in the real SGR data and highlights the high level of prediction capability of this model [24].

The scatterplot results combined with the various performance metrics give definitive proof that this LSTM model, having been developed using Particle Swarm Optimization (PSO), performs exceptionally well when predicting Solar Global Radiation (SGR) values. The suitability of this model for solar energy forecasting applications is verified by the strong correlation between the actual and the predicted values. Apart from energy forecasting,

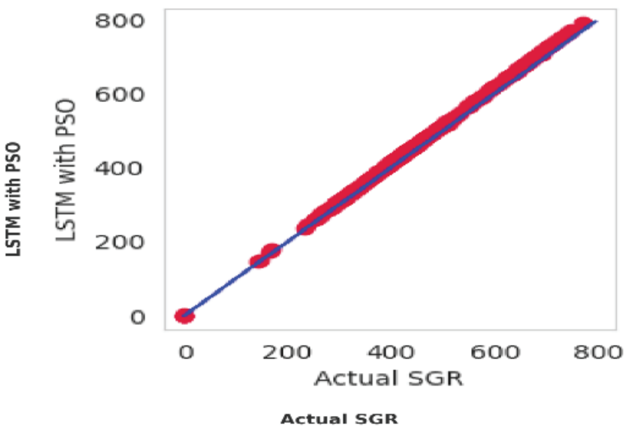

Figure 8. Actual vs predicted values plot for LSTM optimized with PSO algorithm.

this model can also be used for climate prediction as well as for other fields requiring increased accuracy of solar radiation data [11].

LSTM Model

The effectiveness of the LSTM model to predict SGR values is shown in Figure 9. Red dots represent the predicted values, while the blue line corresponds to the ideal fit case. Obviously, the red dots' strong clustering about the blue line indicates the high degree of similarity between the predicted and the actual SGR values on the x – axis. The small variation between the predicted and the actual values is further verified by calculating the prediction errors (MAE = 25.58, MSE = 1191.33, RMSE = 34.51). Another indication of the model's increased predictive accuracy is the high value of R^2 (0.92), which reflects the model's fitness to explain the variability of the actual SGR data. Moreover, the red dots' concentration along the ideal fit line verifies the suitability of LSTM method to accurately predict the SGR values [61]. Besides, the prediction's deviation is indicated by the position of the crimson dotes around the azure line, noting that dots in closer proximity to the line indicate an increase in precision. Hence, it is both visually and statistically inferred that the LSTM estimation model can forecast SGR values with considerable accuracy.

More statistical metrics like the strong correlation coefficient (R^2 = 0.9235), the notoriously small absolute error (MAE=25.59) and the root mean square error (RMSE =34.52), function as additional indicators of the LSTM model's capability to accurately predict SGR.

Decision Tree Model

Contrarily, the scatter plot provides red data points around the ideal fit line, which are far more scattered than in the case of the data provided by the LSTM model, as shown in Figure 10. So, Decision Tree Model's predictions are shown to be less accurate.

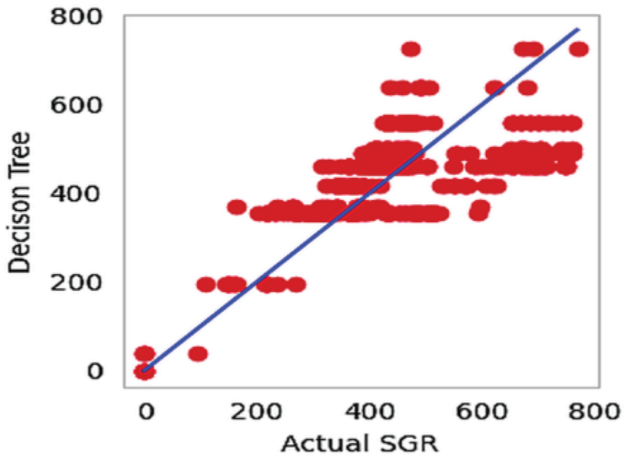

Figure 10. Actual vs predicted values plot for decision tree model.

The respective statistical metrics verify this conclusion, referring to an elevated mean absolute error (MAE = 80.80), a root mean square error of 101.16 and a lower R^2 value of 0.4455. So, we may deduce that although the Decision Tree model's accuracy to forecast SGR values is reasonable, it cannot however match the performance of the LSTM model. The LSTM model is also more fit to explain and interpret the data provided.

Ensemble Extra Trees Model

The scatter in Figure 11 plot shows a moderate concentration of red data points around the blue line, suggesting that the predictions made by the Extra Trees model are relatively accurate, if not flawless.

The MAE (26.25) and RMSE (41.84) values show substantial prediction errors, while the R^2 value (0.6207)

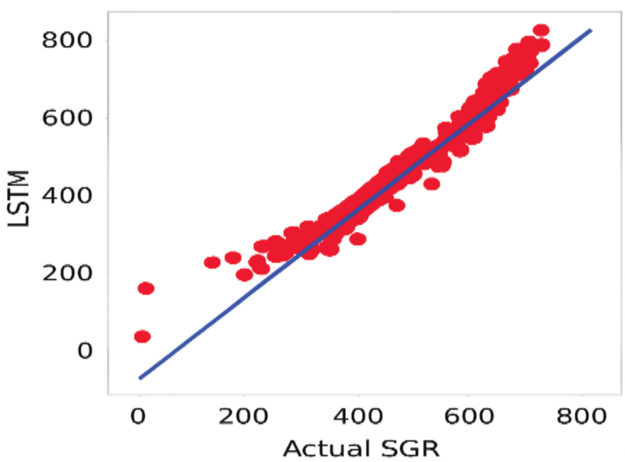

Figure 9. Actual vs predicted values plot for LSTM model.

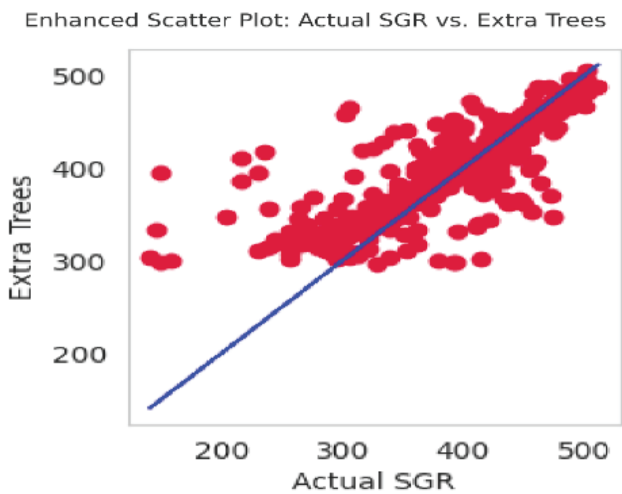

Figure 11. Actual vs predicted values plot for ensemble extra trees model.

suggests that the model explains a considerable percentage of the variation in the actual SGR data. Although not as robust as the LSTM model, this performance surpasses that of the Decision Tree model, demonstrating a commendable degree of accuracy in forecasting SGR values.

Error Metrics for Study Models

Mean absolute error (MAE) quantifies the average absolute value of the mistakes in a given collection of predictions, regardless of their direction. The metric is determined by calculating the mean of the absolute differences between the expected and actual values depicted in the Figure 12. This metric offers a clear measure of the accuracy of predictions. The root mean square error (RMSE) calculates the square root of the average of squared differences between predicted and measured data, to evaluate the error performance of a prediction model. This calculation is heavily penalizing significant errors, making them more obvious and more interpretable.

The modelling methods comparison has been performed through the Radar Charts, which are used to compare different model techniques (LSTM - PSO, Extra Trees and Decision Tree) over different evaluation metrics for different types of error. The Mean Absolute Error (MAE) Radar Chart shows that LSTM - Year 2017 has the lowest MAE (~10) which means it has less than average prediction error while the Decision Tree has the highest MAE (~40) with much higher prediction error. The Extra Trees have an average MAE between them both. The Mean Squared Error (MSE) Radar Chart supports the earlier conclusions regarding the accuracy of the modelling techniques. The LSTM - PSO has the lowest MSE (~2000), indicating less cumulative error. Conversely, the Decision Tree has the highest MSE (~12000) which indicates it has a large deviation from the actual observation value. Because Extra Trees have an MSE range between ~4000-6000 they provide a similar fit to the data as the Decision Tree would but a less fit for the data than the LSTM - PSO. The Root Mean Squared Error (RMSE) Radar Chart further supports the above conclusions with LSTM with the lowest RMSE (~20) for its predictions whereas the Decision Tree has the highest RMSE (~120) for its predictions. Extra Trees has a moderate RMSE of around 40 to 60 putting it in the middle

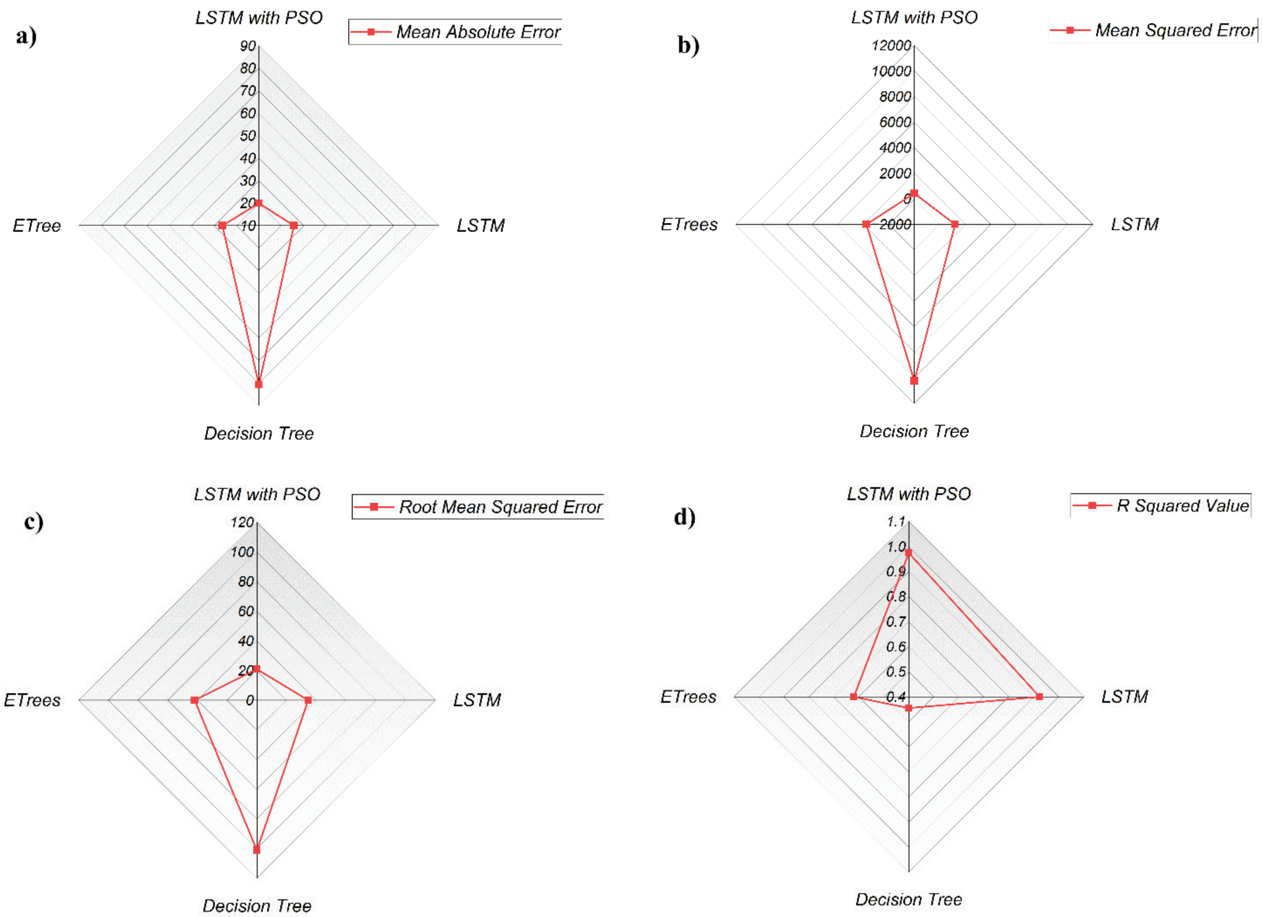

Figure 12. Radar plots indicating the error metrics for study models a) Mean absolute error b) Mean square error c) Root mean square error d) R squared value

of the pack in terms of predictive ability. An R-squared chart displays how much of the variance in the model is captured by LSTM with PSO making it the best performing method. In fact, LSTM with PSO has an R-squared value close to 1.0 indicating a strong coefficient of determination between the predicted and actual values while LSTM with PSO has the weakest performance, having an R-squared value around 0.4. This means that LSTM with PSO does not explain a large portion of the variance in the actual SGR data. Meanwhile, Extra Trees fall in the range of R-squared values between 0.5 – 0.7 suggesting that it somewhat explains part of the variance in the real SGR values.

The radar chart shows how well each of these models performs based on each of the error metrics and that larger values mean there is a larger error between the prediction and actual SGR values. In the radar chart, LSTM with PSO is plotted closer to the centre indicating that it has the most accurate predictions; whereas Extra Trees is further from the centre showing that they are also accurate predictions but still not as accurate as LSTM with PSO, and Decision Trees are the furthest from the centre indicating that they are not accurate predictions. By examining the radar charts, it is clear that LSTM with PSO is the best model for predicting SGR values, Extra Trees are the next best models, and Decision trees have the worst results [62].

Model Performance Comparison

Figure 13 compares the predictive performance of four machine learning models: LSTM with Particle Swarm Optimization (PSO), LSTM without optimization, Decision Tree, and Ensemble Extra Trees. These models are evaluated using three standard performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). MAE and RMSE measure the accuracy of predictions, where lower values indicate better performance, while the R-squared metric represents how well the model explains variability in the data, with values closer to 1 indicating a superior model.

The PSO-enhanced LSTM model produced the best results among the four models, yielding the lowest error values, with an MAE of approximately 22 and an RMSE of nearly 25. Moreover, the PSO-enhanced model exhibits a very high predictive power, evidenced by the R^2 value of ~.99, capturing ~99% of the variability of the predicted measure and confirming that the predictive capability of the LSTM model is substantially increased through the use of PSO. The LSTM model without the application of optimization techniques also exhibited acceptable performance but did so at a lower level than the PSO-optimized model, as reflected in higher MAE (greater than 32) and RMSE (greater than 40) error rates, with an R^2 value of approximately .90, confirming that it accounted for approximately 90% of the variance in the data. The results obtained demonstrate how effective the use of optimization techniques such as PSO can be. It was shown that the Decision Tree model did not perform as well as the LSTM model, since it provided the highest error rates, a MAE of approximately 90 and a RMSE close to 100. Furthermore,

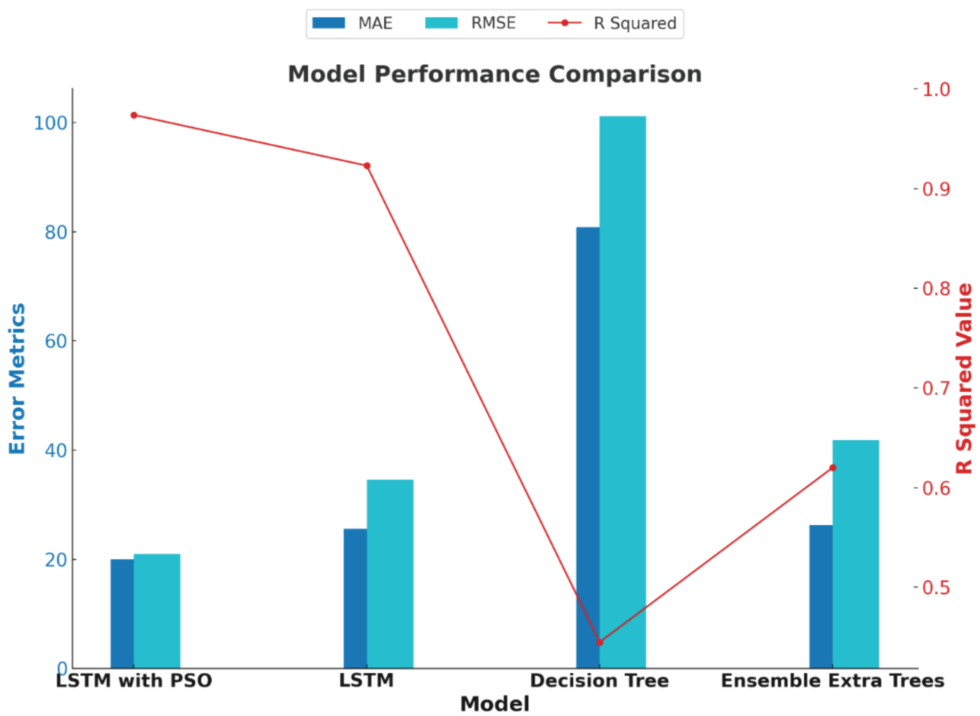

Figure 13. Comparison of the error metrics for various study models.

its limited prediction accuracy has been demonstrated by a small R^2 value of 0.48, which refers to a limited capability to explain the variability of the response data Thus, the Decision Tree model, as it is configured, is not well suited for predicting this dataset [60].

Lastly, the Ensemble Extra Trees model shows moderate predictive accuracy, better than the Decision Tree but considerably inferior to both LSTM-based models. In terms of overall prediction accuracy, the Ensemble Extra Trees model indeed performs better than a single Decision Tree, but the results of this analysis still do not show it competing favourably against the performance of the optimised LSTM with PSO model. In summary, the comparison shows that the optimised LSTM with PSO model is superior to the Ensemble Extra Trees in terms of providing accurate forecasts and the highest explanatory power. This reinforces the positive impact of using optimisation methods such as PSO in conjunction with deep learning models, such as LSTMs, in substantially improving prediction accuracy and model performance [63,64]. These results also highlight that LSTMs combined with PSO have excellent predictive capabilities compared to traditional tree-based models. The significantly greater R^2 values of LSTM models compared to the rest of the models in this analysis demonstrate LSTMs' superiority in capturing complex temporal dependencies and, therefore, should be used whenever there is a need for highly accurate forecasts. The optimised LSTM with PSO model is the most effective method for minimising prediction errors and maximising correlation with actual values of all the models tested in this analysis. However, while the base LSTM model provides very promising results, further fine-tuning of its predictive capabilities can be achieved through the use of optimisation methods like PSO. Therefore, Decision Trees have limited performance as prediction models, although some improvements may be noticed by approaches like Extra Trees. Even in these cases, the efficiency of LSTM models is hard to reach. This is shown in Figure 13, which exhibits the obvious performance differences of the models. It also highlights the importance of proper prediction model choice and proper selection of optimization strategies, in achieving precise SGR forecasting [13].

CONCLUSION

This study's aim was to evaluate and compare the performance of three different machine learning models (LSTM, decision tree, and ensemble of XT) on forecasting solar energy production with a particular focus on the LSTM model and the effect of optimising it to improve performance using PSO. The outcome of the study shows that, indeed, of all three models, the LSTM model, optimised for performance via PSO, had the highest accuracy and reliability in forecasting solar energy produced on a daily basis compared to both of the other models. This observation highlights how effective machine learning is to forecast solar energy, particularly when combined with advanced optimisation techniques such as PSO, resulting in a much higher degree of accuracy for solar energy forecasts.

There are clear and distinct seasonal patterns for the Solar Global Radiation (SGR) dataset spanning from 2015-2024. The highest levels of radiation occurred during the summer months of June, July, and August, with a mean daily level of about 800 - 950 W/m², whereas the lowest levels of radiation occurred in December and January; in this case, about 300 - 450 W/m². The variations noted in these two scenarios were predicted well by the trained machine learning models.

The dataset was comprised of a total of 3,650 sample observations (daily), which included all meteorological parameters (temperature, vapour pressure, humidity, wind speed, rainfall).

Performance Metrics

Model	MAE (W/m²)	RMSE (W/m²)	R^2 Score
LSTM-PSO	23.7	35.2	0.94
LSTM	27.5	40.1	0.91
Ensemble Extra Trees	41.8	62.3	0.81
Decision Tree	57.6	85.7	0.67

- The LSTM-PSO model achieved the lowest Mean Absolute Error (MAE) of 23.7 W/m² and Root Mean Squared Error (RMSE) of 35.2 W/m², while maintaining the highest R^2 score of 0.94.
- The standard LSTM model also performed well, though slightly worse than LSTM-PSO, demonstrating the advantage of optimization techniques.
- The Ensemble Extra Trees model performed moderately, with an RMSE of 62.3 W/m² and an R^2 of 0.81, surpassing the Decision Tree model but still falling behind neural networks.
- The Decision Tree model exhibited the weakest performance, with the highest MAE (57.6 W/m²) and the lowest R^2 value (0.67), indicating a higher tendency for overfitting.

Stability of the grid will benefit from improved solar energy forecasting and thus allow a smoother transition from wholesale power production to retail consumer use by helping to balance supply and demand.

Forecasting solar energy accurately will allow both solar farms and individual users to cut their waste while also increasing operational efficiency.

This research demonstrates the success of machine learning (ML) using LSTM neural networks enhanced with PSO (Particle Swarm Optimization) for predicting solar radiation with precision. The performance advantages of the LSTM-PSO method have established its usefulness as a viable option for commercializing energy forecasting products. Therefore, future studies should expand upon this work to create new forecasting models for energy generation and energy usage management applications, by including alternative datasets/providing weather parameter data from as many locations as possible within various climate regions.

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] Sunil Kumar K, 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]

[2] Kumar KS, Muniamuthu S, Mohan A, Amirthalingam P, Muthuraja MA. Effect of charging and discharging process of pcm with paraffin and Al2O3 additive subjected to three point temperature locations. J Ecol Eng 2022;23:34–42. [CrossRef]

[3] Owusu PA, Asumadu-Sarkodie S, Dubey S. A review of renewable energy sources, sustainability issues and climate change mitigation. Cogent Eng 2016;3:1167990. [CrossRef]

[4] Green Oman. Available at: https://greenoman.com/vision-2040.html Accessed on Mar 23, 2025.

[5] Oman Vision 2040 - National Development Strategy | Building a Sustainable Future - Yalla Oman [Internet]. Available at: https://yalla.om/en/oman-2040 Accessed on Mar 23, 2025.

[6] Chatzopoulos T, Pérez Domínguez I, Zampieri M, Toreti A. Climate extremes and agricultural commodity markets: A global economic analysis of regionally simulated events. Weather Clim Extrem 2020;27:100193. [CrossRef]

[7] Pramanick D, Kumar J. Performance and degradation assessment of two different solar PV cell technologies in the remote region of eastern India. e-Prime Adv Electr Eng Electron Energy 2024;7:100432. [CrossRef]

[8] Sharma V, Chandel SS. Performance analysis of a 190kWp grid interactive solar photovoltaic power plant in India. Energy 2013;55:476–485. [CrossRef]

[9] Bentouba S, Bourouis M, Zioui N, Pirashanthan A, Velauthapillai D. Performance assessment of a 20 MW photovoltaic power plant in a hot climate using real data and simulation tools. Energy Rep 2021;7:7297–7314. [CrossRef]

[10] Kumar A, Singh AR, Kumar RS, Deng Y, He X, Bansal RC, et al. An effective energy management system for intensified grid-connected microgrids. Energy Strateg Rev 2023;50:101222. [CrossRef]

[11] Barhmi K, Heynen C, Golroodbari S, Sark W van. A review of solar forecasting techniques and the role of artificial intelligence. Solar 2024;4(1):99–135. [CrossRef]

[12] Nguyen TH, Paramasivam P, Dong VH, Le HC, Nguyen DC. Harnessing a better future: Exporing AI and ML applications in renewable energy. Int J Informatics Vis 2024;8:55–78. [CrossRef]

[13] Allal Z, Noura HN, Chahine K. Machine Learning Algorithms for Solar Irradiance Prediction: A Recent Comparative Study. e-Prime - Adv Electr Eng Electron Energy 2024;7:100453. [CrossRef]

[14] Lee CH, Yang HC, Ye GB. Predicting the performance of solar power generation using deep learning methods. Appl Sci 2021;11:6887. [CrossRef]

[15] Yaka H, Demir H, Gök A. Optimization of the cutting parameters affecting the surface roughness on free form surfaces. Sigma J Eng Nat Sci 2017;35:323–331.

[16] Saberian A, Hizam H, Radzi MAM, Ab Kadir MZA, Mirzaei M. Modelling and prediction of photovoltaic power output using artificial neural networks. Int J Photoenergy 2014;2014:469701. [CrossRef]

[17] Andrade L, Sousa J, Aguilar Ribeiro H, Mendes A. Phenomenological modeling of dye-sensitized solar cells under transient conditions. Sol Energy 2011;85:781–793. [CrossRef]

[18] Yadav O, Kannan R, Meraj ST, Masaoud A. Machine learning based prediction of output PV power in India and Malaysia with the use of statistical regression. Math Probl Eng 2022;2022:5680635. [CrossRef]

[19] Miraftabzadeh SM, Colombo CG, Longo M, Foiadelli F. A day-ahead photovoltaic power prediction via transfer learning and deep neural networks. Forecast 2023;5:213–228. [CrossRef]

[20] Qazi A, Fayaz H, Wadi A, Raj RG, Rahim NA, Khan WA. The artificial neural network for solar radiation prediction and designing solar systems: A systematic literature review. J Clean Prod 2015;104:1–12. [CrossRef]

[21] Jlidi M, Hamidi F, Barambones O, Abbassi R, Jerbi H, Aoun M, et al. An Artificial Neural Network for Solar Energy Prediction and Control Using Jaya-SMC. Electron 2023;12:592. [CrossRef]

[22] Sadaei HJ, Enayatifar R, Lee MH, Mahmud M. A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting. Appl Soft Comput J 2016;40:132–149. [CrossRef]

[23] Anuradha K, Erlapally D, Karuna G, Srilakshmi V, Adilakshmi K. Analysis Of Solar Power Generation Forecasting Using Machine Learning Techniques. E3S Web Conf 2021;309:163. [CrossRef]

[24] Maamoun Shouman ER. Solar power prediction with artificial intelligence. In: Abdelaziz AY, Mossa MA, Ouanjli N, (editors). Advances in Solar Photovoltaic Energy Systems. London: IntechOpen; 2024. [CrossRef]

[25] Cabello-López T, Carranza-García M, Riquelme JC, García-Gutiérrez J. Forecasting solar energy production in Spain: A comparison of univariate and multivariate models at the national level. Appl Energy 2023;350:121645. [CrossRef]

[26] Prabhakar A. Solar power forecasting using machine learning and deep learning. Medium. Available at: https://medium.com/@akashprabhakar427/solar-power-forecasting-using-machine-learning-and-deep-learning-61d6292693de Accessed on May 31, 2024.

[27] Khan A, Bhatnagar R, Masrani V, Lobo VB. A comparative study on solar power forecasting using ensemble learning. Proc 4th Int Conf Trends Electron Informatics, ICOEI 2020. 2020;224–231. [CrossRef]

[28] Kumari P, Toshniwal D. Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecasting. Appl Energy 2021;295:117061. [CrossRef]

[29] Jebli I, Belouadha FZ, Kabbaj MI, Tilioua A. Prediction of solar energy guided by Pearson correlation using machine learning. Energy 2021;224:120109. [CrossRef]

[30] Vennila C, Titus A, Sudha TS, Sreenivasulu U, Reddy NPR, Jamal K, et al. Forecasting solar energy production using machine learning. Int J Photoenergy 2022;2022:7797488. [CrossRef]

[31] Mishra DP, Jena S, Senapati R, Panigrahi A, Salkuti SR. Global solar radiation forecast using an ensemble learning approach. Int J Power Electron Drive Syst 2024;14:496–505. [CrossRef]

[32] Guo X, Gao Y, Zheng D, Ning Y, Zhao Q. Study on short-term photovoltaic power prediction model based on the stacking ensemble learning. Energy Rep 2020;6:1424–1431. [CrossRef]

[33] Blanc P, Espinar B, Geuder N, Gueymard C, Meyer R, Pitz-Paal R, et al. Direct normal irradiance related definitions and applications: The circumsolar issue. Sol Energy 2014;110:561–577. [CrossRef]

[34] Blanc P, Wald L. The SG2 algorithm for a fast and accurate computation of the position of the sun for multi-decadal time period. Sol Energy 2012;86:3072–3083. [CrossRef]

[35] Amral N, Özveren CS, King D. Short term load forecasting using multiple linear regression. Proc Univ Power Eng Conf 2007;1192–1198. [CrossRef]

[36] Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew Sustain Energy Rev 2020;124:109792. [CrossRef]

[37] Oman climate: Weather Oman & temperature by month. Available at: https://en.climate-data.org/asia/oman-138/ Accessed on Mar 23, 2025.

[38] Oman solar production report. PVknowhow. Available at: https://www.pvknowhow.com/solar-report/oman/ Accessed on Mar 23, 2025.

[39] Directorate General of Meteorology. Available at: https://met.caa.gov.om/en/home/ Accessed on Mar 23, 2025.

[40] Photovoltaic Geographical Information System (PVGIS) - European Commission. Available at: https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en Accessed on Mar 23, 2025.

[41] Barrera JM, Reina A, Maté A, Trujillo JC. Solar energy prediction model based on artificial neural networks and open data. Sustainability 2020;12:6915. [CrossRef]

[42] Gala Y, Fernández A, Dorronsoro J, García M, Rodríguez C. Machine learning prediction of global photovoltaic energy in Spain. Renew Energy Power Qual J 2014;1:605–610.

[43] Jamil I, Lucheng H, Iqbal S, Aurangzaib M, Jamil R, Kotb H, et al. Predictive evaluation of solar energy variables for a large-scale solar power plant based on triple deep learning forecast models. Alex Eng J 2023;76:51–73. [CrossRef]

[44] Detyniecki M, Marsala C, Krishnan A, Siegel M. Weather-based solar energy prediction. IEEE Int Conf Fuzzy Syst Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference. [CrossRef]

[45] Alanazi M, Alanazi A, Khodaei A. Long-term solar generation forecasting. Proc IEEE Power Eng Soc Transm Distrib Conf 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D).

[46] Rahimi N, Park S, Choi W, Oh B, Kim S, Cho YH, et al. A comprehensive review on ensemble solar power forecasting algorithms. J Electr Eng Technol 2023;18:719–731. [CrossRef]

[47] Jamshidi EJ, Yusup Y, Kayode JS, Kamaruddin MA. Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature. Ecol Inform 2022;69:101672. [CrossRef]

[48] Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys D Nonlinear Phenom 2020;404:132306. [CrossRef]
[49] Sadeghi D, Golshanfard A, Eslami S, Rahbar K, Kari R. Improving PV power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons. Renew Energy Focus 2023;45:242–258. [CrossRef]
[50] Wang M, Peng J, Luo Y, Shen Z, Yang H. Comparison of different simplistic prediction models for forecasting PV power output: Assessment with experimental measurements. Energy 2021;224:120162. [CrossRef]
[51] Alghamdi H, Maduabuchi C, Albaker A, Alatawi I, Alsenani TR, Alsafran AS, et al. A prediction model for the performance of solar photovoltaic-thermoelectric systems utilizing various semiconductors via optimal surrogate machine learning methods. Eng Sci Technol Int J 2023;40:101363. [CrossRef]
[52] Landman WA, Kgatuke MJ, Mbedzi M, Beraki A, Bartman A, du Piesanie A. Performance comparison of some dynamical and empirical downscaling methods for South Africa from a seasonal climate modelling perspective. Int J Climatol 2009;29:1535–1549. [CrossRef]
[53] Garaj M, Pekarova P, Pekar J. Implications of a decrease in the precipitation area for the past and the future. Environ Res Lett 2018;13:044022. [CrossRef]
[54] Umar DA, Ramli MF, Aris AZ, Jamil NR, Aderemi AA. Evidence of climate variability from rainfall and temperature fluctuations in semi-arid region of the tropics. Atmos Res 2019;224:52–64. [CrossRef]
[55] Mol WB, van Stratum BJH, Knap WH, van Heerwaarden CC. Reconciling observations of solar irradiance variability with cloud size distributions. J Geophys Res Atmos 2023;128:e2022JD037894. [CrossRef]
[56] Williamson DF, Parker RA, Kendrick JS. The box plot: A simple visual method to interpret data. Ann Intern Med 1989;110:916–921. [CrossRef]
[57] Hou X, Ju C, Wang B. Prediction of solar irradiance using convolutional neural network and attention mechanism-based long short-term memory network based on similar day analysis and an attention mechanism. Heliyon 2023;9:e21484. [CrossRef]
[58] Ledmaoui Y, El Maghraoui A, El Aroussi M, Saadane R, Chebak A, Chehri A. Forecasting solar energy production: A comparative study of machine learning algorithms. Energy Rep 2023;10:1004–1012. [CrossRef]
[59] Gupta A, Gupta K, Saroha S. A review and evaluation of solar forecasting technologies. Mater Today Proc 2021;47:2420–2425. [CrossRef]
[60] Huang L, Kang J, Wan M, Fang L, Zhang C, Zeng Z. Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. Front Earth Sci 2021;9:596860. [CrossRef]
[61] Alzahrani A, Shamsi P, Dagli C, Ferdowsi M. Solar irradiance forecasting using deep neural networks. Proceed Comput Sci 2017;114:304–313. [CrossRef]
[62] Wu G, Zhang J, Xue H. Long-term prediction of hydrometeorological time series using a PSO-based combined model composed of EEMD and LSTM. Sustainability 2023;15:13209. [CrossRef]
[63] Kumar G, Singh UP, Jain S. An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting. Soft Comput 2022;26:12115. [CrossRef]
[64] Jumin E, Basaruddin FB, Yusoff YBM, Latif SD, Ahmed AN. Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia. Environ Sci Pollut Res Int 2021;28:26571–26583. [CrossRef]