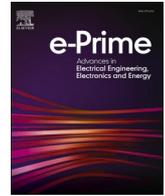




Contents lists available at ScienceDirect

# e-Prime - Advances in Electrical Engineering, Electronics and Energy

journal homepage: [www.elsevier.com/locate/prime](http://www.elsevier.com/locate/prime)

## Hybrid machine learning model combining of CNN-LSTM-RF for time series forecasting of Solar Power Generation

Mobarak Abumohsen<sup>a</sup>, Amani Yousef Owda<sup>a,\*</sup>, Majdi Owda<sup>c</sup>, Ahmad Abumihsan<sup>b</sup>

<sup>a</sup> Arab American University, Faculty of Graduate Studies, Department of Natural, Engineering and Technology Sciences, Ramallah, West Bank, Palestine

<sup>b</sup> Faculty of Data Science, Arab American University, Ramallah P600, Palestine

<sup>c</sup> Arab American University, Faculty of Data Science, UNESCO Chair in Data Science for Sustainable Development, Ramallah, West Bank, Palestine

### ARTICLE INFO

#### Keywords:

Solar power generation  
Hybrid machine learning  
Deep learning models  
Classical machine learning  
Power systems  
Solar energy

### ABSTRACT

Forecasting solar power generation (SPG) is vital for the development and planning of power systems, offering significant benefits in terms of technical performance and financial efficiency. It enhances system reliability, safety, stability and it reduces the operational costs. This paper's primary goal is to develop models that can precisely forecast solar power generation by analyzing real first-hand dataset of solar power. The value of these forecasting models lies in their ability to anticipate future solar power generation, thus optimizing resource use and minimizing expenses. To achieve this, the study utilizes various classical machine learning, deep learning, and hybrid machine learning techniques, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Recurrent Neural Networks (RNN), Random Forest (RF), Support Vector Regression (SVR), Bi-directional LSTM (Bi-LSTM), and Convolutional Neural Network (CNN). Among these, the hybrid model combining CNN-LSTM-RF demonstrated superior accuracy with R-squared of 92 %, a Root Mean Square Error (RMSE) of 0.07 kW, and a Mean Absolute Error (MAE) of 0.05 kW. This indicates that the hybrid machine learning model combining of CNN-LSTM-RF is effective in forecasting solar power generation.

### 1. Introduction

Renewable energy sources, continuously replenished by natural processes, play a pivotal role in our quest for sustainable energy solutions [1]. Solar energy, derived directly from the sun, is one of the most prominent and widely harnessed kinds of renewable energy, thanks to its abundant availability and potential for clean electricity generation [2-6]. The advantages of solar energy are manifold; it is an inexhaustible resource, reduces greenhouse gas emissions, and can lead to significant economic savings over time. As scientists and researchers worldwide strive to solve the challenges posed by climate variation and depleting fossil fuels, solar energy has emerged as a leading resolution due to its sustainability and environmental benefits.

The movement towards utilizing machine learning (ML) to forecast solar power generation (SPG) represents a considerable advancement in renewable energy technologies on a global scale. Given the unique geographical and climatic conditions across various regions worldwide, applying ML for solar energy prediction is both a challenge and an opportunity. It highlights the potential for leveraging advanced computational methods to improve the efficiency and reliability of solar power

systems everywhere. This approach not only addresses the pressing need for sustainable energy solutions to meet increasing global energy demands but also paves the method for a more adaptive and intelligent energy grid. The use of datasets from diverse locales, including places like Palestine, enriches the global understanding and applicability of these ML models, ensuring that the research and solutions developed are comprehensive and universally relevant.

The application of ML into solar power forecasting represents a crucial global initiative, as regions worldwide harness the power of abundant sunlight while advancing their renewable energy frameworks. The ML, lying at the confluence of historical data analysis and future pattern prediction, provides a disruptive strategy to increasing the accuracy of solar power estimates. However, the unpredictability of solar energy, influenced by meteorological factors such as solar irradiance, temperature, wind speed, and atmospheric pressure, presents a substantial barrier to accurate prediction and usability. By applying ML algorithms to extensive datasets, researchers aim to improve forecasting precision, a critical step for the efficient operation of solar energy facilities and the overall stability of global power grids. This precision facilitates the smoother combination of solar power into diverse energy

\* Corresponding author.

E-mail address: [amani.owda@aaup.edu](mailto:amani.owda@aaup.edu) (A.Y. Owda).

<https://doi.org/10.1016/j.prime.2024.100636>

Received 25 March 2024; Received in revised form 7 May 2024; Accepted 7 June 2024

Available online 22 June 2024

2772-6711/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

portfolios, ensuring that it can be a cornerstone of sustainable energy solutions worldwide [7]. The exploration of ML for solar energy forecasting transcends mere technical advancement; it represents a global stride toward enhancing energy security and reducing environmental footprints. This approach aligns with the broader objectives of sustainable growth and contributes significantly to the worldwide effort against climate change. As this field continues to evolve, the adoption of ML algorithms for more accurate solar power forecasting is poised to revolutionize energy landscapes around the world. This journey not only supports the economic and environmental goals of nations but also positions them at the forefront of adopting cutting-edge technologies in the renewable energy sector. At the heart of the global march towards a greener future lies the innovative application of machine learning (ML) to harness solar energy.

This work represents more than a leap in technological innovation; it embodies a beacon of hope for regions worldwide striving to overcome their energy challenges with ingenuity and resilience. The application of ML into solar power forecasting extends beyond mere energy yield calculations—it's about shaping a future where renewable resources are seamlessly integrated into the daily lives of communities everywhere. This endeavor demonstrates a deep understanding of how to counterbalance the unpredictability of natural elements with the quest for reliable energy sources. By leveraging the predictive power of ML, nations are charting a course for global communities to employ technology in navigating the complexities of renewable energy, advancing confidently towards energy independence and environmental stewardship. This study focuses on the hybrid model (CNN-LSTM-RF) to forecast SPG. The CNN-LSTM-RF hybrid model combines the strengths of convolutional neural networks (CNNs) for spatial feature extraction, Long Short-Term Memory (LSTM) networks for capturing temporal dependencies, and Random Forest (RF) for ensemble learning and robustness. CNNs extract spatial patterns from weather data, LSTMs capture temporal dynamics in solar energy production, and RF combines their outputs for more accurate forecasts. This integration enhances forecasting accuracy by utilizing the complementary nature of these algorithms, leading to more reliable predictions of SPG. The primary objectives of this research are as follows:

- Optimization of energy production: by accurately predicting solar power output, energy producers can optimize their operations to ensure maximum efficiency.
- Grid stability and reliability: accurate forecasting allows for better grid management, ensuring stability and reliability.
- Develop and implement advanced machine learning models to increase the reliability of solar energy predictions.
- Enhance the ability of utility companies to manage and predict solar power supply.
- Construct and refine a ML model specifically designed to forecast SPG with high accuracy and minimal error rates.

This section provides an overview of SPG forecasts, and the other sections of the research are structured as follows: [Section 2](#) explains the research reviews and past works, and [Section 3](#) presents the approach for creating classical ML, hybrid machine learning, and deep learning models to forecast SPG. [Section 4](#) presents the experimental findings and compares them to past research. [Section 5](#) presents a conclusion and future study.

## 2. Literature review

The field of solar power forecasting has experienced significant growth recently, driven by the global incorporation of renewable energy sources into power grids. Given the pivotal role of solar energy in transitioning to sustainable and clean energy, precise forecasting is crucial for efficient grid management, optimal energy utilization, and seamless integration into existing power infrastructures [8,9]. This

literature review comprehensively examines research studies that employ ML and deep learning (DL) techniques for solar power forecasting [10]. Categorizing studies based on the techniques used enables the identification of trends, challenges, and future approaches in the field. This review serves as an invaluable resource for academics, practitioners, and regulators in the renewable energy sector, facilitating informed decision-making and improvements in solar energy forecasting techniques.

### 2.1. Overview

Solar power predictions may be divided into different types, with the persistence (or smart persistence) model being one of the most basic yet essential methods. Machine learning models use past solar energy data to predict future energy production for a brief period, typically 2–3 h ahead. It acts as a standard for comparing the effectiveness of various forecasting techniques [11–20]. Essentially, this involves adapting numerical weather prediction (NWP) to specific times and locations and then using these predictions to estimate solar power output. The forecasts can be generated through various approaches, including physical models, statistical methods, or ML techniques. The persistence model, in particular, is foundational in understanding how solar power might vary in the short term based on recent historical data. Beyond this, models specifically designed for forecasting the power generation of photovoltaic (PV) plants have undergone significant advancements [21–23]. However, despite improvements, forecasting accuracy is challenged by the unpredictability of factors like cloud cover, which affects solar radiation (power per unit area generated from the sun) more unpredictably than other environmental variables like temperature [24]. Due to the impracticality of scrutinizing all interconnected meteorological forecasts, numerous alternative approaches have been devised. Some researchers have considered weather predictions from meteorological websites [25], while others have explored nonlinear modeling methods, such as artificial neural networks (ANN), to streamline the solar forecast model. Two frequently utilized network architectures for predicting various aspects of solar radiation, including global solar radiation, solar radiation on tilted surfaces, daily solar radiation, and short-term solar radiation, are radial basis function (RBF) networks and multilayer perceptron (MLP) networks.

### 2.2. Deep learning approaches in solar power forecasting

This subsection focuses on research studies employing deep learning (DL) methodologies for accurate solar power forecasting, utilizing advanced techniques such as Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and attention mechanisms. Researchers in [26] built a model for renewable energy power forecasting the energy gradated of 21 solar energy plants based on various DL and artificial neural network algorithms, including Deep Belief Networks, Auto Encoder, LSTM, and real historical data from Kassel, Germany. The result shows the Auto-LSTM emerges as the superior Deep Neural Network (DNN) model, demonstrating outstanding performance where an average Root Mean Squared Error (RMSE) of 0.0713, an average Mean Absolute Error (MAE) of 0.0366, and an average absolute deviation of 0.2765. [27] Develops a model aimed at forecasting short-term power generation from PV power plants. An assessment of the Long Short-Term Memory (LSTM) network's performance in this context was carried out and benchmarked against the performance of a Multi-layer Perceptron (MLP) network. The evaluation utilized key metrics such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Coefficient of Determination ( $R^2$ ), leveraging real data from Canada for the analysis. The findings revealed that the LSTM network consistently surpassed the MLP network in forecasting accuracy across all metrics. Specifically, the LSTM network achieved a MAE of 676.34, a MAPE of 0.275, an RMSE of 883.5, and an  $R^2$  of 0.745.

In [28], researchers employ DL techniques for solar energy forecasting, specifically utilizing Recurrent Neural Networks (RNN), LSTM, and Gated Recurrent Units (GRU). Real data from Errachidia province spanning from 2016 to 2018 was used, the results obtained show that RNN and LSTM slightly exceed GRU, particularly in their ability to maintain long-term dependencies in time series data, with the RNN MAE of 1.83, the MSE of 8.53, and the RMSE of 2.92.

In [29], a novel stacked ensemble algorithm named Dual-Stage Ensemble with XGBoost (DSE-XGB) is introduced, leveraging two DL techniques— Artificial Neural Networks (ANN) and LSTM as base models for solar energy forecasting. DSE-XGB is designed to provide a superior balance of consistency and stability across various case studies, irrespective of variations. The output shows that the introduced DSE-XGB method [29] outperforms other models, showcasing an improvement in the R-squared value by 10 % to 12 %. In [30], a model is built using DL techniques such as LSTM, GRU, Auto-LSTM, and Auto-GRU to predict SPG. The introduced model is validated on two real-time series datasets from Shagaya in Kuwait and Cocoa in the USA. The results indicate that LSTM exhibits lower error and higher accuracy, with RMSE and MAE values of 0.0739 and 0.0176, respectively. Furthermore, [31] develops a model using ML techniques, including multilayer feed-forward neural network (MLFFNN), radial basis function neural network (RBFNN), Support Vector Regression (SVR), fuzzy inference system (FIS), and adaptive neuro-fuzzy inference system (ANFIS) to forecast SPG. The real dataset from Iran is used, and the results demonstrate that SVR and MLFFNN models have the maximum effectiveness in predicting solar irradiance, with  $R^2$  of 0.9999 and 0.9795, respectively.

In [32], the authors developed a model to forecast stable power generation, comparing two ML techniques: DL and SVR. The study utilized real-world data from Spain. The findings indicated that DL outperformed SVR, with a MAPE of 7.9 % for DL and 8.52 % for SVR. This establishes DL as the more effective approach for solar energy forecasting in this context. In [33], researchers constructed a DL model for forecasting solar power production, combining LSTM with a data-driven technique known as Adaptive Neuro-Fuzzy Inference System (ANFIS). The comparative analysis showed that the LSTM model yielded superior outputs, with RMSE, MAE, and  $R^2$  values of 60.66 kW, 30.47 kW, and 0.9777, respectively. Additionally, [34] recommended various models for solar radiation prediction, including DL models such as ANN, CNN, and RNN, as well as traditional approaches like polynomial regression, SVR, and Random Forest (RF). Applied across four locations in Nigeria over 12 years, the study demonstrated that DL models exhibit superior prediction accuracy compared to ML models, with an  $R^2$  of 0.9546 and RMSE/MAE of 82.22 / 36.52 W/m<sup>2</sup>. Furthermore, [35] introduced a model designed to predict future output power values of solar cells in Palestine using Multilayer Feed-Forward with Backpropagation Neural Networks (MFFNNBP). The results from the proposed model achieved an average RMSE of 0.0349 for one year.

### 2.3. Machine learning techniques for solar power forecasting

In this subsection, various studies are discussed that employ different machine learning methods for solar power forecasting, highlighting the efficacy of algorithms such as Support Vector Regression (SVR), Random Forest (RF), Decision Trees (DT), and K-Nearest Neighbors (KNN). In [36], the researcher utilized ML algorithms, specifically RF and ANN, surpassing Linear Regression (LR) and SVR to forecast short-term solar energy based on data from Errachidia, a semi-desert climate province in Morocco spanning from 2016 to 2018. RF exhibited robust performance for both real-time and short-term predictions, achieving MAE=0.000264 [36]. Furthermore, in [37], various ML algorithms, including SP, SVM, and RF, the DL and SVR techniques were utilized for solar power forecasting at the Buruthakanda solar park. A comparison of the efficiency of these ML algorithms with the Smart Persistence (SP) method indicated that the ML models outperformed the SP model in

terms of accuracy and effectiveness, specifically, the SVM model demonstrated the lowest MAE at 0.01286. In [38], ML models, including NN, LR, SVM, and ANN, were employed to build a solar power forecasting model utilizing public data from the National Solar Radiation Database (NSRD). The results from this study [38] proved that the ANN model achieved the lowest error and was more accurate than other ML models, with an RMSE of 0.998. Moreover, in [39], a model for solar energy forecasting in Morocco was built using six ML algorithms: SVR, ANN, DT, RF, Generalized Additive Models (GAM), and Extreme Gradient Boosting (XGBOOST). The study, based on daily data from a Solar Power Plant in Benguerir city, Morocco, demonstrated that ANN achieved the RMSE of  $2.6e-08$  and high accuracy of 0.99.

In [40], the study applies ML methods to model PV power production for a system located in Medellin, Colombia. Four forecasting models were constructed using KNN, LR, ANN, and SVM. The findings indicate that all four methods yielded satisfactory forecast of PV energy generation. Nevertheless, in terms of RMSE and MAE, the ANN forecasting model emerged as the most accurate and effective among the examined approaches. Specifically, KNN achieved RMSE and MAE of 92.857 and 8.8279, LR achieved 94.5 and 8.96, and SVM achieved 93.644 and 9.96, while ANN achieved 86.446 and 8.409, respectively. Moreover, in [41], the research investigates various time-series methods, encompassing both DL and ML algorithms, to forecast PV power producing for prompt detection of equipment and panel defects. The study utilized PV power generation data from Ansan City, South Korea. Among the tested models, which included Holt-Winters, Multivariate Linear Regression, ARIMA, SARIMA, ARIMAX, SARIMAX, and LSTM, the LSTM model exhibited exceptional accuracy, boasting an outstanding  $R^2$  of 0.943 and an average mean Mean Absolute Percentage Error (mMAPE) of 5.79 [41].

In [42], a comprehensive comparison of different ML techniques and time series models is conducted across five different sites in Sweden. The analysis reveals that utilizing time series models becomes intricate due to the non-stationary nature of energy time series. On the other hand, the implementation of ML techniques proves to be more straightforward. The results [42] indicate that ANN and Gradient Boosting Regression Trees (GBR) outperform other models on average across all sites. The RMSE for ANN is 0.113, GBRT is 0.112, and KNN is 0.14. Furthermore, in [43], a solar energy forecasting model is built based on Support Vector Machines (SVM), and GBR using public data from Kaggle for the State of Oklahoma. Feature selection methods, including Linear Correlation, the ReliefF algorithm, and a novel approach based on local information analysis, were evaluated. The findings indicate that non-linear methods achieve lower errors compared to linear methods.

In [44], the authors aim to address the challenge of integrating renewables into the grid, given their intermittent and uncontrollable power generation. The study builds a ML model to automatically create site-specific forecasting models for solar energy generation using National Weather Service (NWS) weather forecasts. ML methods, including SVM and LR, were employed. The results indicate that SVM is more accurate and has fewer errors compared to LR. Additionally, in [45], researchers constructed a model using ML models such as SVM, Gaussian Process Regression (GPR), LR, and DT for predicting SPG based on real data from Malaysia. The results prove that DT achieves lower RMSE of 5.83 and  $R^2$  of 95.91.

Researchers in [46], a study built a model using ML models, including SVM, RF, and LR, to predict SPG based on real data from India. The results indicate that RF achieves lower RMSE of 27.32. Furthermore, in [47], authors constructed a model using ML models such as SVM and the Hidden Markov Model to forecast SPG, utilizing real data from Australia. The results reveal that SVM achieves lower error and higher accuracy than the Hidden Markov Model, with an accuracy of 94 % for SVM and 61.8 % for the Hidden Markov Model in sunny conditions.

In [48], researchers introduce an enhanced Radial Basis Function Neural Network (RBFNN) model using the standard RBF implemented in MATLAB (newrb). This enhancement relies on intelligent algorithms

such as K-means clustering, K-nearest neighbor, and singular value decomposition to enhance the centers, radii, and weights of the RBFNN. The proposed enhanced model is applied to forecast solar cell energy production in Palestine, utilizing data from already installed solar panels in Jericho city. The experimental results in [48] illustrate that the enhanced model outperforms traditional RBFNN and Multilayer Perceptron Neural Network methods, exhibiting a higher level of precision with a low mean square error, particularly when employing a relatively few neurons on the hidden layer, where it achieved the lowest RMSE of 0.004994 when the number of neurons used is 7.

This subsection provides a detailed exploration of studies employing various ML techniques for solar energy forecasting, showcasing the diversity of approaches and their effectiveness across different regions and datasets.

#### 2.4. Hybrid approaches and comparative studies

This subsection encompasses studies that introduce hybrid approaches combining clustering techniques, ensemble algorithms, and comparative investigations into the effectiveness of different forecasting models. In [49], authors built a model using DL techniques such as LSTM, GRU, RNN, MLP, WPD-LSTM, and hybrid ESN-CNN to forecast SPG, utilizing real data from Australia Solar Centre. The results show that the hybrid ESN-CNN model achieves lower RMSE of 0.1432. Furthermore, in [50], authors constructed a model using DL techniques, CNN, and ANN to predict SPG, utilizing real data from Stanford University Neural Network for Solar Electricity Trend. The authors created a hybrid model by first inserting the regression input and hybridizing it with sky image based on CNN as input. Subsequently, they inputted all data to train and test in the ANN model, and the results demonstrated that this procedure yields high accuracy and lower error.

In [51], the study built a model using DL and hybrid methods like CNN, Multi-headed-CNN, CNN-LSTM, ARMA, and MLR to forecast solar energy generation, utilizing a public dataset. The results show that the CNN-LSTM model achieves lower RMSE of 0.065213. Moreover, in [52], the research introduces a hybrid DL methodology that integrates clustering methods, CNN, LSTM, and an attention mechanism with a wireless sensor network to address the challenges in PV energy generation forecasting based on real historical data from Shaoxing, China. The study first employs correlation analysis and self-organizing mapping to identify the most relevant factors in historical data. Subsequently, a hybrid deep learning model is created by combining a CNN, LSTM, and attention mechanism for the forecasting task. Finally, the model for training is chosen using the month of testing results, the results obtained from purposed model RMSE of 2.04. In [53], researchers propose a monthly prediction model for photovoltaic (PV) power, intending to estimate potential PV solar power production at a new location using actual data from South Korea. They construct an RNN model featuring LSTM to identify temporal patterns within the time series data. The model is then tested to assess its forecasting accuracy for PV utilities not included in the training phase. The outcomes indicate that the model proposed in [53] attains an RMSE of 7.416 % and an MAPE of 10.805 % when applied to the testing dataset.

In [54], researchers introduce an innovative hybrid ML strategy designed for predicting SPG. This approach leverages a hybrid Ensemble Average (EA) technique, capitalizing on the strengths of various ML methods, including a non-linear autoregressive neural network (NAR-NN), a non-linear autoregressive neural network with exogenous signal (NARX-NN), an SVR with RBF kernel, and an Extreme Learning Machine (ELM). All these components are integrated into a unified model. The performance of this introduced model is evaluated using a real-world dataset from a 1 MW solar park located in Gujarat, India, the model EA achieved the lowest MAE of 0.1295. Furthermore, in [55], scientists introduce a hybrid model that integrates a CNN and LSTM for the forecasting of stable energy generation. In this model, the CNN detects meteorological conditions, while the LSTM learns energy

production patterns based on them, using data gathered in Busan, Korea. Both quantitative and qualitative assessments were used to determine the model's success. The results from the proposed model in [55] reveal a MAPE of 4.58 on sunny days and 7.06 on cloudy days during the quantitative evaluation.

This subsection provides insights into hybrid approaches that combine different methods for enhanced solar power forecasting and studies that conduct comprehensive comparative investigations

### 3. Methodology

The methodology commences with data collection and preparation as its initial phase. Following this, the second step involves the exploration of the collected data. Moving forward, the third step focuses on preprocessing the data specifically for ML applications. Subsequently, the fourth step entails the utilization of various ML algorithms such as LSTM, RNN, RF, SVR, CNN, and GRU for forecasting SPG. To assess and compare the efficacy of these algorithms, diverse performance metrics are employed. The objective is to identify the most effective approach. Ultimately, the optimization and tuning process is applied to determine the best model for electric load forecasting, summarizing the methodology as depicted in Fig. 1.

#### 3.1. Data collection

The data used in this paper were obtained from Tubas Electricity Company - Palestine. All loads are stored through the SCADA program in a database. This data contains 5045 records, dataset utilized in our study encompasses a period of approximately 14 months, spanning from the date Jun 3, 2022, to the date of July 31, 2023. The most important features found in this data are (date and hour, temperature, active power generation, and others). To examine the significance and influence of the pressure factor, wind speed, and humidity on solar energy generation, the data obtained from the Tubas Electricity Company lacked information on these variables. Consequently, data for these factors was sourced from NASA (nasa.gov) [56]. The coordinates of the solar station's location were used to extract relevant data from NASA's records. The dataset contains 5045 rows and 14 columns as shown in Table 1, which represents the five records of the data.

#### 3.2. Exploratory data analysis (EDA)

In this section, the extraction and examination of all exploratory data related to SPG have been undertaken. Exploratory Data Analysis (EDA) involves the scrutiny of the various characteristics, correlations, and hidden patterns inherent in SPG data. Diverse methods, including autocorrelation, box plots, and line plots, have been utilized for the analysis and exploration of data. The capability to visually inspect and explore the interconnections among different variables, revealing previously unnoticed patterns, is identified as a pivotal aspect of EDA, which is crucial for the development of time series forecasting models. For the formal modeling and validation of forecasting, the initiation of an EDA is deemed imperative [57].

##### 3.2.1. Correlation

The statistical method known as correlation measures the linear relationship between two or more variables. Correlation allows the prediction of one variable based on another. The rationale for employing correlation in feature selection is based on the idea that meaningful variables exhibit a robust correlation with the outcome. There must exist a correlation between the variable and the endpoint, while no correlation should be present among the variables themselves. To assess the adequacy of connection for constructing a regression model and forecasting short-term load, a heat map is employed to visualize the correlation ratios among features in the dataset. Calculates the Pearson correlation coefficient ( $r$ ) between components through a specific Eq.

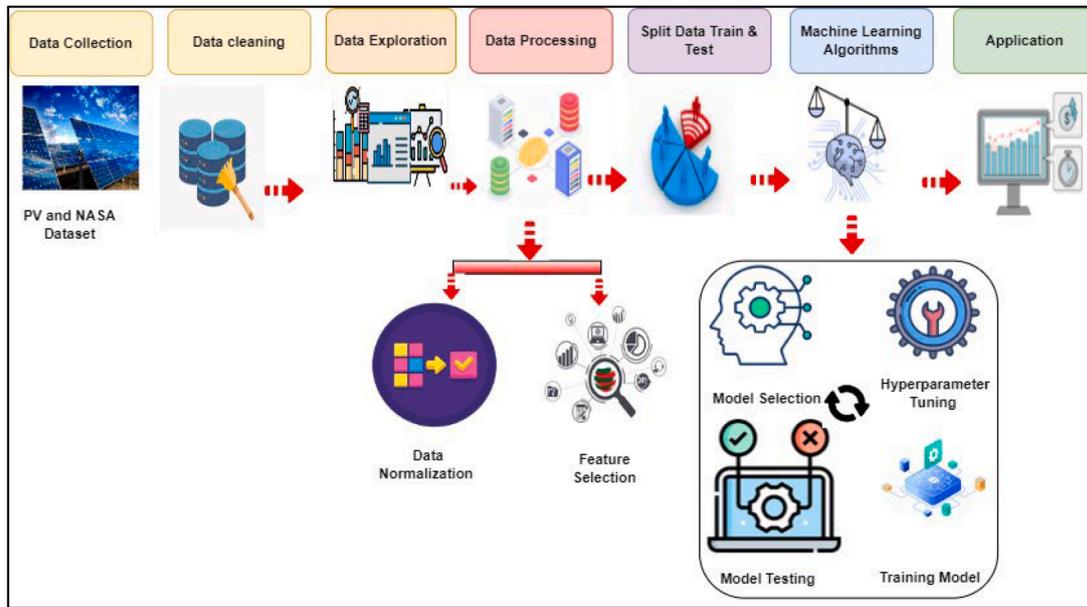


Fig. 1. Workflow for SPG forecasting models.

(1) [58].

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

Where  $\bar{x}$  represents the of the x-variable in a sample,  $r$  Pearson correlation coefficient,  $x_i$  values of the x-variable in a sample,  $\bar{y}$  represents the values of the y-variable, and  $y_i$  represents the y-variable in a sample.

Fig. 2 illustrates the correlations among various elements. Positive relationships are observed between certain elements, such as the week and month, as well as the week and the number of days in the year. Conversely, negative relationships are evident, such as those between the year and month, and the year and day of the year. From Fig. 2 it can be seen that the correlation between solar radiation and SPG is notably strong at 0.78, followed by the correlation with temperature at 0.6.

### 3.2.2. SPG behavior analysis

To identify the optimal days and hours for SPG, as well as the corresponding days of the month with the highest generation, Figs. 3 through 4 present a detailed analysis. Figs. 3 and 4 showcase key statistical metrics, including the minimum, first quartile (Q1), median, third quartile (Q3), and maximum values for each category. This comprehensive analysis not only aids in pinpointing specific times, months, and hours with the highest SPG but also serves as a valuable tool for electricity companies. By leveraging this information, companies can optimize their utilization of solar power, implementing best practices for enhanced efficiency and sustainability.

Fig. 3 illustrates the SPG spanning from June 2022 to July 2023. Notably, the highest power generation occurs in July and August, aligning with elevated temperatures and solar radiation during these months. In contrast, power generation is relatively lower in November, December, January, and February, corresponding to reduced temperatures and solar radiation. This finding is valuable for effective data management and forecasting SPG, particularly during the summer. The significance of summer forecasting lies in its ability to facilitate proactive technical measures, mitigating potential network issues stemming from increased consumption. Additionally, such forecasts play a crucial role in sourcing electricity from alternative and sustainable sources, such as solar stations.

Fig. 4 shows that solar power production peaks at noon, specifically

between 12 p.m. and 2 p.m., where the average production is about 800 kW. During this peak, the highest production can approach 900 kW while the lowest hovers around 650 kW. Essentially, this indicates that more than half of the time, solar power output is within this range in these hours. However, production significantly decreases before sunrise and after sunset. For example, in the early morning hours between 5 a.m. and 8 a.m., the average power generated is only about 200 kW, with the highest production at nearly 300 kW and the lowest at approximately 100 kW, showcasing much lower energy generation during these times. A similar trend is observed from 4 p.m. to 7 p.m. additionally, SPG tends to be more unpredictable in the afternoon, with a broader output range. The variation in power output, or the interquartile range (IQR), representing the middle 50 % of data, is wider in the afternoon than in the morning. For instance, the IQR from 1 p.m. to 4 p.m. is about 200 kW, compared to only about 100 kW from 8 a.m. to 11 a.m., highlighting the increased variability of solar power production in the afternoon hours.

### 3.2.3. Time series analysis for SPG

The description suggested in Figs. 5-11 presents an analysis of SPG across various time scales (daily, weekly, monthly, and hourly) with an emphasis on identifying the underlying nature of the data, whether it exhibits a seasonal, recurring, or random pattern. Additionally, these figures incorporate statistical measures to illustrate the distribution characteristics of specific dataset figures. This approach provides a holistic perspective on the temporal aspects of SPG, aiding in recognizing patterns, trends, and variations within the data. The incorporation of statistical measures further enhances the interpretability of the figures, allowing for a more comprehensive understanding of the dataset's distributional properties. Such analyses are crucial for making informed decisions, forecasting, and optimizing resource utilization in the context of solar energy generation.

In Fig. 5, the x-axis represents the hour, and the y-axis represents the generated solar power (kW). It is noted from the figure the generated power begins at 6 a.m. while the generated power starts to finish at almost 7 p.m. Moreover, the generated solar power reaching its highest point around midday, and then gradually decreases in the afternoon. Furthermore, the generated solar power does not related to the weekday, unlike the electricity consumption which is related to the weekday where the Friday has the lowest consumption electricity.

Fig. 6 shows the monthly average SPG specifically during the

**Table 1**  
The sample records for dataset samples.

Data	ActivePower (kW)	solar radiation (kW/m <sup>2</sup> )	Temperature (°C)	Humidity (g/kg)	WindSpeed (m/s)	Pressure (kPa)	Month	Year	Week	WeekDay	Daynum	Dayofyear	Hour	Day
2022-06-11 05:00:00	0	133.26	21.82	7.02	1.38	99.21	6	2022	23	6	11	162	5	Saturday
2022-06-11 05:30:00	20.77	201	25.34	6.04	1.14	99.25	6	2022	23	6	11	162	5	Saturday
2022-06-11 06:00:00	46.75696316	341.41	25.84	6.04	1.14	99.25	6	2022	23	6	11	162	6	Saturday
2022-06-11 06:30:00	120	360	25.84	6.04	1.14	99.25	6	2022	23	6	11	162	6	Saturday
2022-06-11 07:00:00	200	480	25.84	6.04	1.14	99.25	6	2022	23	6	11	162	7	Saturday
2022-06-11 07:30:00	288	640	32	4.1	0.73	99.24	6	2022	23	6	11	162	7	Saturday
2022-06-11 08:00:00	310.6138585	733.56	33.62	4.15	0.73	99.23	6	2022	23	6	11	162	8	Saturday
2022-06-11 08:30:00	450	760	34.9	4.15	0.73	99.23	6	2022	23	6	11	162	8	Saturday
2022-06-11 09:00:00	563.0857142	879.65	36.67	4.15	1.2	99.16	6	2022	23	6	11	162	9	Saturday

summer period. Notably, SPG reaches its zenith in the summer months, particularly in September and August. Subsequently, there is a gradual decline in SPG during the fall and winter seasons, with a notable decrease in December, January, and February. Observing the data, it becomes apparent that average daily and weekly SPG correlates with the annual season, particularly influenced by the strength of the sun's rays. Additionally, the distribution of the data on both a weekly and nearly daily basis exhibits proximity, suggesting a consistent pattern over shorter time interval. This analysis provides valuable insights into the seasonal variations and temporal dependencies associated with SPG.

Fig. 7 shows the variation of average temperature (in degrees Celsius) and SPG (in kilowatts) over time in Tubas. The x-axis displays periods from June 2022 to August 2023. The y-axis on the left side represents the average temperature, while the y-axis on the right side represents SPG. SPG appears to be highest in the summer months (June to August) and lowest in the winter months (December to February), which coincides with the pattern for average temperature. The average temperature peaks in July at around 40 °Celsius, while SPG peaks in July at around 300 kW. Conversely, the average temperature is lowest in January at around 10 °Celsius, and SPG is also lowest in January at around 50 kW. Solar panel efficiency can actually decrease with higher temperatures. There could be other factors influencing SPG, such as variations in daylight hours throughout the year.

Fig. 8 demonstrates a direct and positive link between solar radiation and solar power output, indicating that an increase in solar radiation leads to a rise in SPG. This relationship is logical, considering solar panels are designed to convert sunlight into electricity. For instance, in June 2023, with solar radiation at approximately 250 Wh/m<sup>2</sup>, the solar power system was able to produce around 200 kW of electricity. Essentially, this translates to the system generating 200 kwatt-hours of electricity for every 250 watt-hours of solar radiation received per square meter. Moreover, Fig. 8 highlights noticeable seasonal trends in both solar radiation and solar power production. The highest levels of solar radiation and power output are recorded during the summer months (June–August), while the lowest are observed in the winter months (December–February). This seasonal variation is attributed to the sun's higher position in the sky during the summer, allowing solar panels to capture more sunlight. Additionally, the data suggests variability in solar radiation and power generation from one year to the next. For example, the measurements for July 2023 were higher than those of July 2022, which could be influenced by several factors, including weather pattern changes or advancements in solar panel efficiency.

Fig. 9 shows that the average humidity levels in Tubas, Palestine, surpass the average SPG. Specifically, the graph indicates an average humidity of around 150 gs per cubic meter, while the SPG graph illustrates an average output of approximately 300 kW. This observation can be attributed to the adverse impact of high humidity on the efficiency of solar panels. Solar panels operate most effectively in dry conditions, and elevated humidity levels can diminish their electricity generation capacity. Notably, the SPG in Tubas peaks during the summer months when temperatures are at their highest. This observation suggests that, in the context of Tubas, temperature emerges as a more influential factor than humidity in determining SPG efficiency.

As depicted in Fig. 10 there exists a negative correlation between atmospheric pressure and SPG. This implies that an increase in atmospheric pressure is associated with a decrease in the amount of solar power generated. The explanation lies in the blocking of sunlight by clouds, which are more likely to form under high-pressure conditions. Notably, when the pressure reaches its peak at approximately 1000 Pa, SPG is at its lowest at around 100 kW. Conversely, when pressure is at its lowest, around 990 Pa, SPG peaks at approximately 300 kW. The robust negative correlation underscores the significance of atmospheric pressure as a major determinant of SPG. This insight could be utilized to enhance SPG efficiency by strategically installing solar panels in areas characterized by lower atmospheric pressure.

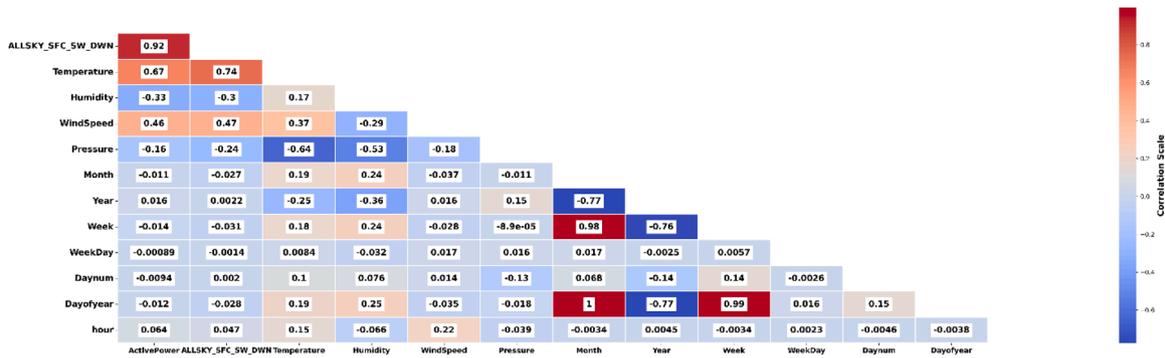


Fig. 2. Correlation matrix for the features in the dataset.

Comparison of Solar Power Generation by Month (2022 vs. 2023)

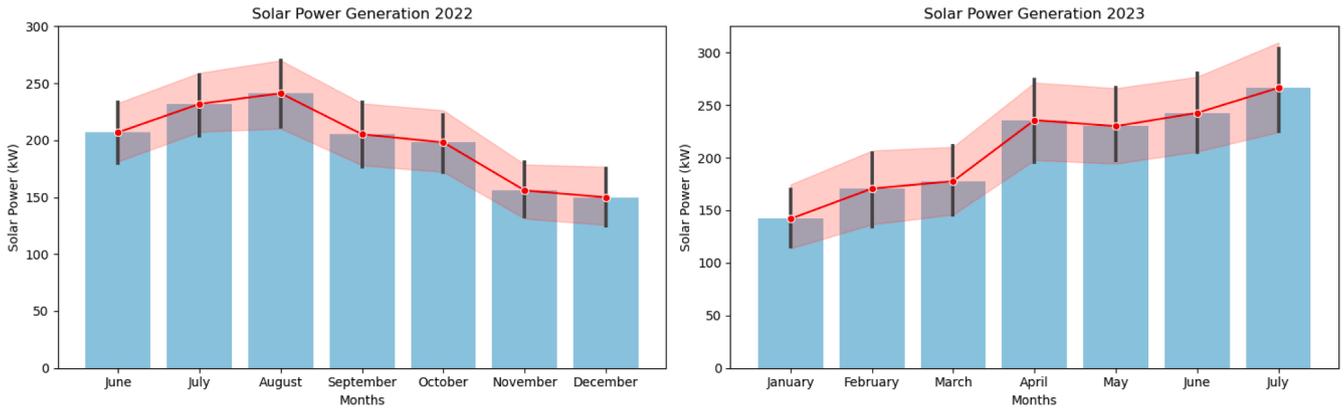


Fig. 3. SPG based on months from 2022 to 2023.

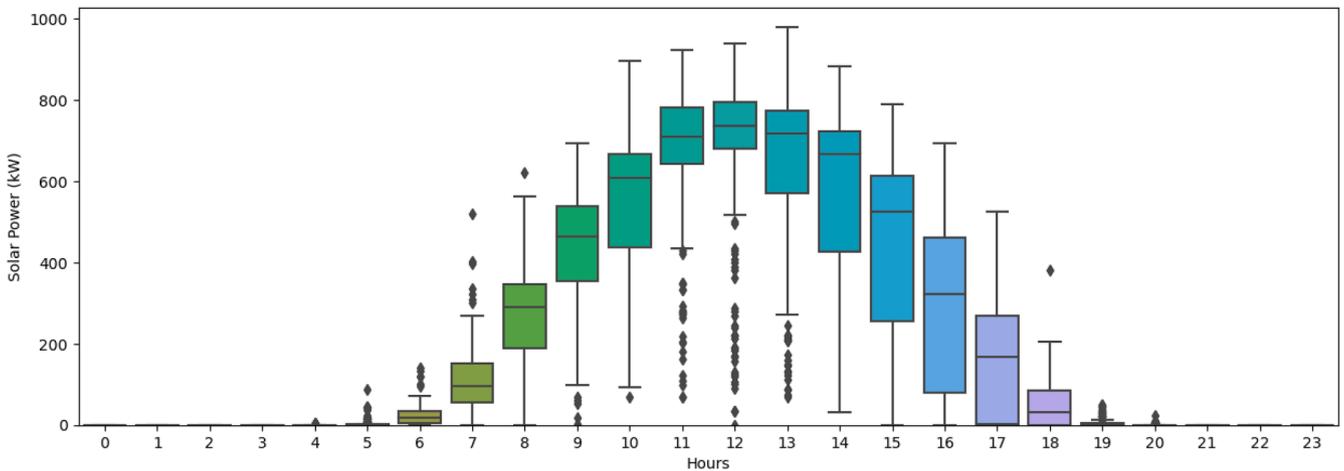


Fig. 4. Box plot SPG based on the hours of days.

Fig. 11 illustrates a compelling positive correlation between wind speed and SPG, indicating that an increase in wind speed corresponds to a heightened generation of solar power. This relationship is attributed to the enhanced efficiency of wind turbines in generating electricity at higher wind speeds. Moreover, wind turbines contribute to the cooling of solar panels, thereby further improving their efficiency. The robust positive correlation observed in the graph underscores the significance of wind speed as a major influencing factor in determining SPG. This insight can be leveraged to enhance the efficiency of SPG by strategically placing solar panels in areas characterized by elevated wind speeds.

This subsection encompassed an exploratory data analysis, providing

insights into data patterns and guiding the decision-making process for leveraging ML to extract knowledge from the dataset. Through comprehensive data visualization, the analysis yielded a key conclusion: solar energy generation is markedly influenced by solar radiation, where elevated solar radiation strongly correlates with increased production. Solar radiation emerges as the pivotal factor impacting solar cell generation. Moreover, the study illuminated the significance of additional environmental variables specifically, atmospheric pressure, wind speed, humidity, and temperature in shaping the overall performance and generation of solar energy. This holistic comprehension underscores the intricate interplay of factors shaping solar energy generation,

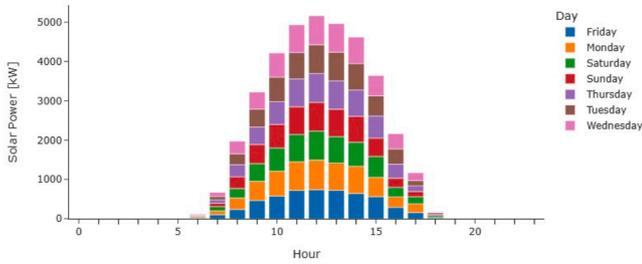


Fig. 5. Hourly SPG per Weekday.

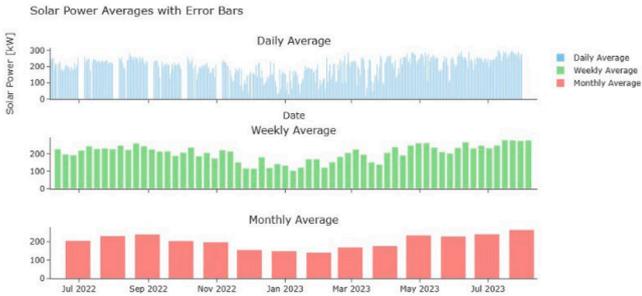


Fig. 6. Tubas station for SPG.

emphasizing the necessity of considering a spectrum of environmental variables for optimizing solar power systems. Building upon this exploration, the subsequent section will delve into the detailed presentation and discussion of the methodology employed for forecasting SPG and the accompanying dataset.

### 3.3. Forecasting methodology

This section outlines the proposed methodology for predicting SPG using ML, DL, and hybrid ML algorithms applied to a real dataset. The process initiates with the creation of the dataset and the description of the preprocessing steps, including normalization and feature selection. Subsequently, models are constructed and trained, employing ML algorithms for forecasting SPG based on historical data. The selected ML models encompass LSTM, RNN, RF, Bi-LSTM, SVR, CNN, and GRU.

Following model development, hyperparameter tuning is conducted for the ML models. This involves adjusting parameters such as the

optimizer, activation function, learning rate, number of epochs, batch size, and number of hidden layers. Concurrently, various performance metrics are employed to gauge the accuracy of each model. The final step involves the selection of the optimal forecasting model based on the comprehensive evaluation conducted throughout the process.

#### 3.3.1. Data preprocessing

The preprocessing of data plays a pivotal role in preparing the machine-learning model to effectively utilize an optimal data structure. In the absence of preprocessed data, the performance of machine-learning models may be suboptimal, leading to inaccuracies and undesirable outcomes. Depending on the inherent characteristics of the raw data, various sub-steps of preprocessing may be applied [59]. This paper specifically incorporates data normalization and the elimination of highly correlated features. Additionally, features exhibiting minimal correlation with the target feature are removed, outliers are addressed, and an evaluation is conducted to identify and handle null values in the original dataset.

**3.3.1.1. Data normalization.** Data normalization serves as a preprocessing technique designed to prevent certain features from disproportionately influencing others. The objective of data normalization is to ensure that features sharing the same scale contribute equally to the model. Various methods of data normalization exist, such as standardization and max-min normalization [60]. In the current study, all feature ranges were standardized to fall within the [0–1] range. The max-min normalization technique was applied for a linear transformation on the data, with the determination of the max-min normalization method governed by Eq. (2) [61].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Where  $x'$  represents the normalized value,  $x$  denotes the original feature value,  $x_{min}$  is the minimum value of the feature, and  $x_{max}$  signifies the maximum value of the feature.

**3.3.1.2. Feature selection.** Feature selection is a methodology focused on identifying and choosing features in the data that significantly contribute to the target variable. This technique is pivotal for achieving robust identification rates, particularly in challenging scenarios [62]. The correlation between features, a statistical measure depicting how one variable varies concerning another, was computed. In instances where the dataset contained highly correlated features, it led to

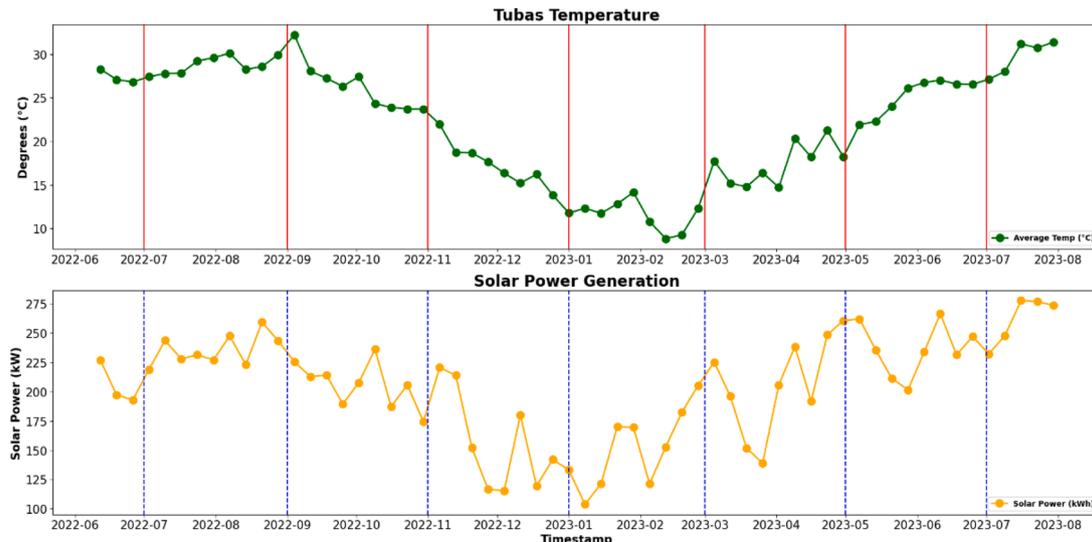


Fig. 7. The relation between temperature in degrees and the solar power in Kilowatts.

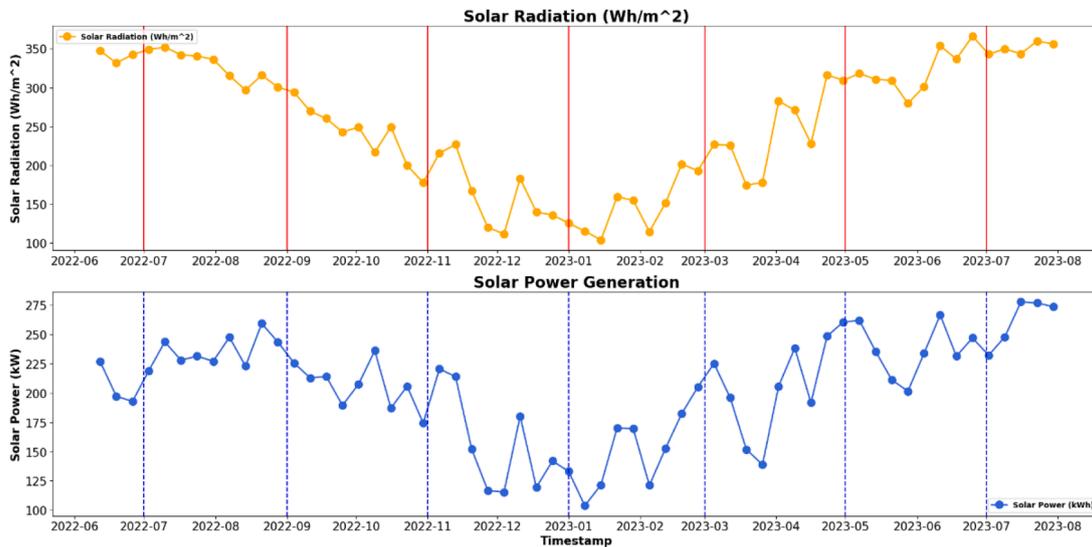


Fig. 8. The relation between solar radiation and the solar power in Kilowatts.

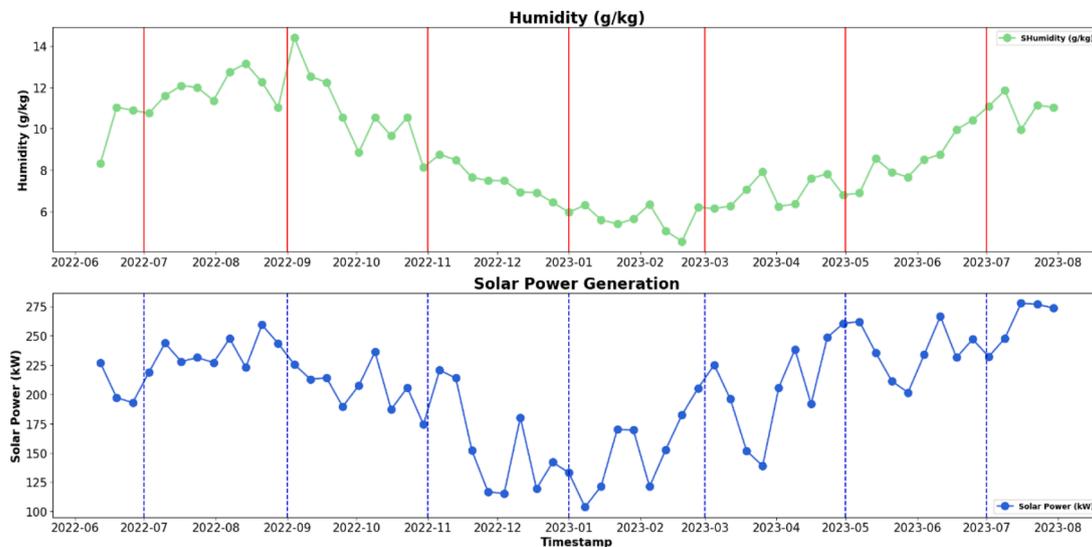


Fig. 9. The relation between humidity and the solar power in Kilowatts.

increased variance and unreliability [63]. To overcome this, the correlation coefficient was used as a filter for feature selection. A threshold was used to exclude characteristics with correlations greater than 90 %, thereby segregating highly correlated features from those with low correlation with the target variable.

3.3.2. Machine learning algorithms

This section offers a concise overview of the selected ML algorithms. The choice of these algorithms is grounded in their widespread use and demonstrated performance in previous studies. Despite their common objective, various ML and DL algorithms exhibit distinct mathematical models, strengths, and limitations. The DL approach delves into the intricate relationships among elements during the learning process, facilitating the prediction of dependent variable values based on independent variables. In this study, LSTM, RNN, Bi-LSTM, CNN, RF, SVR, and GRU were employed to forecast SPG.

3.3.2.3. Long short-term memory network (LSTM). A Long Short-Term Memory network (LSTM) is a specialized form of temporal cyclic neural network [64] engineered to tackle the long-term dependency

problem intrinsic to conventional Recurrent Neural Networks (RNNs). Unlike standard RNNs, LSTM networks feature memory units in lieu of hidden-layer neurons. These memory units are structured with input gates, forgetting gates, and output gates, allowing the network to selectively retain or discard information at each time step. This design effectively addresses the difficulty of preserving pertinent data over prolonged durations [65].

The LSTM recurrent network has gained widespread recognition for its capacity to capture temporal correlations, proving particularly effective across diverse domains such as language translation and speech recognition. In the realm of electrical load forecasting, the LSTM network is customized to discern load patterns from incoming SPG profiles, storing these states in memory for subsequent predictive tasks [66]. Fig. 12 offers a visual depiction of the LSTM cell block's architecture.

3.3.2.4. Recurrent neural network(rnn). Recurrent Neural Networks (RNNs) were initially devised for analyzing time-series data and have demonstrated efficacy across various domains such as voice recognition, machine translation, and picture captioning [68,69]. RNNs function by

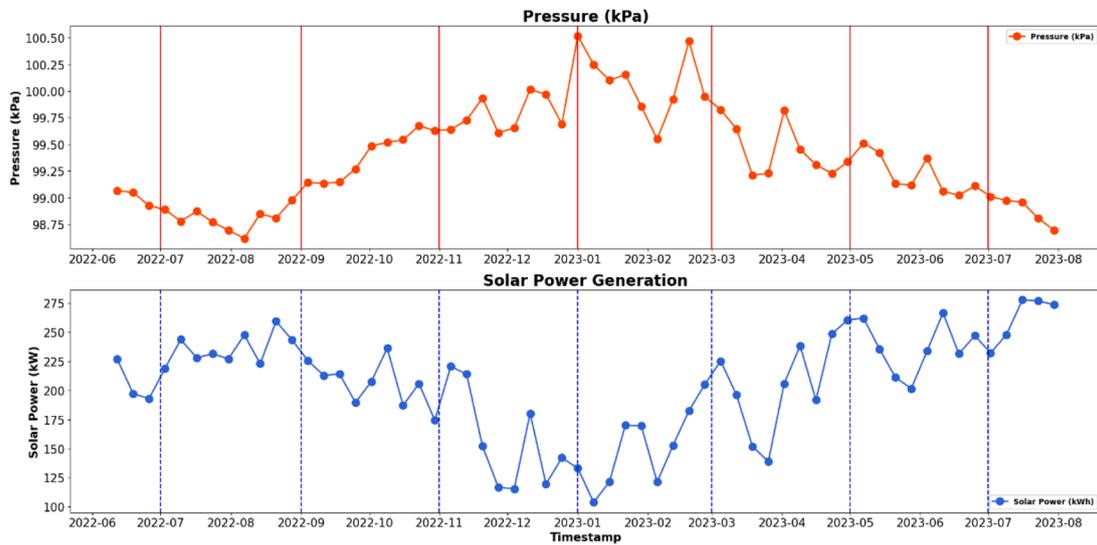


Fig. 10. The relation between pressure and the solar power in Kilowatts.

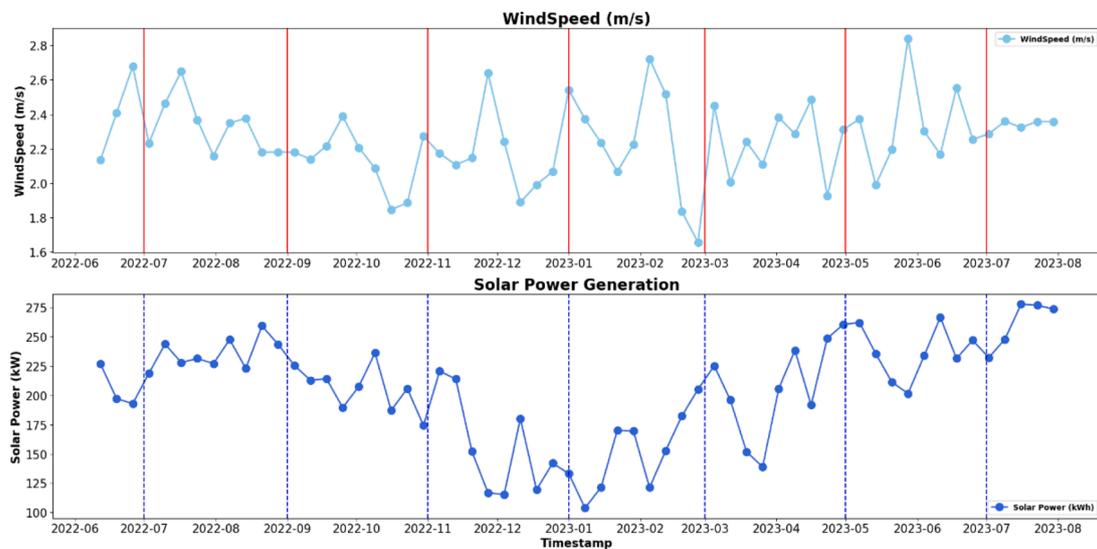


Fig. 11. The relation between wind speed and solar power in Kilowatts.

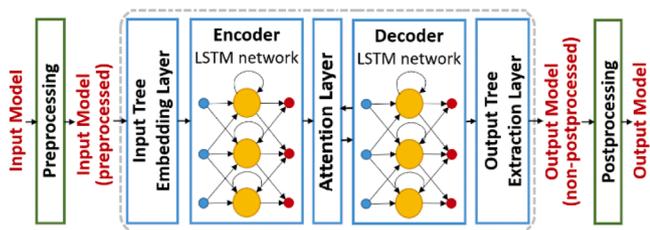


Fig. 12. A long short-term memory block diagram structure [67].

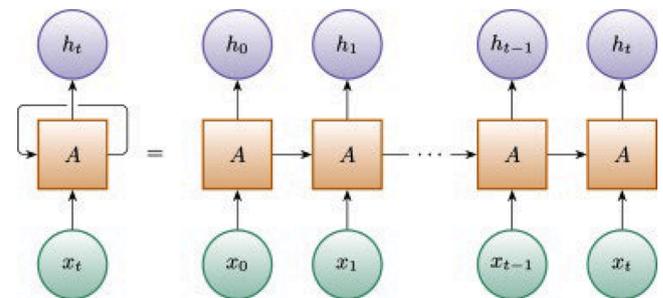


Fig. 13. RNN block diagram structure [70].

sequentially processing incoming sequence or time-series data through individual vectors at each step, thereby retaining information from preceding time steps in a concealed manner. Fig. 13 provides a visual representation of the construction of an RNN cell block.

3.3.2.5. Gated recurrent units (GRUs). Gated Recurrent Units (GRUs), introduced in 2014 as a gating mechanism within recurrent neural networks, exhibit similarities to LSTM networks while boasting fewer parameters by excluding an output gate. GRUs have demonstrated

superior performance across tasks such as polyphonic music modeling, speech signal modeling, and natural language processing compared to LSTMs. Notably effective when dealing with smaller and less frequent datasets, GRUs represent an enhancement over the hidden layer of classical RNNs, as illustrated in Fig. 14 [71-76].

3.3.2.6. *Bi-directional long short-term memory (Bi-LSTM)*. Bi-directional Long Short-Term Memory (Bi-LSTM), introduced by Schuster and Paliwal [78], enables a network to leverage both past and future input data sequences. The input data undergoes processing through two interconnected layers [79]. In Bi-LSTM, each element's sequence prediction or tagging is determined by considering the context of elements both in the past and future. This is achieved by running two LSTMs concurrently—one from left to right and the other from right to left. The combined output of these two LSTMs yields the forecast of the target signal, termed the composite output. Fig. 15 shows the structure of a bidirectional LSTM (bi-LSTM) algorithm.

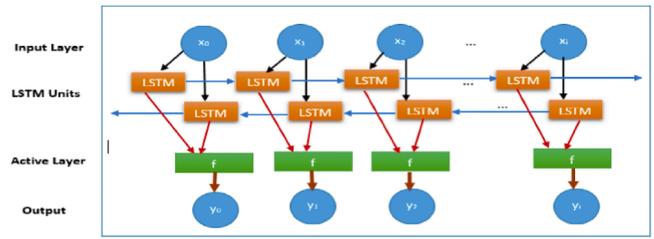


Fig. 15. Structure of a bidirectional LSTM (bi-LSTM) algorithm [80].

3.3.2.7. *Convolutional neural network (CNN)*. In DL, Convolutional Neural Networks (CNNs) are widely employed for image classification, drawing inspiration from the human visual system as proposed by [81] and [82]. CNNs stand as state-of-the-art approaches for pattern recognition, object detection, and various image-related applications. Fig. 16 illustrates a basic schematic representation of a simple CNN, comprising five distinct layers: an input layer, a convolution layer, a pooling layer, a fully connected layer, and an output layer [83].

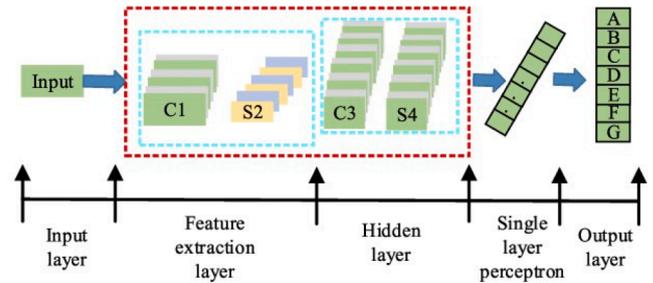


Fig. 16. Structure of convolutional neural network (CNN) algorithm [84].

3.3.2.8. *Support vector regression (SVR)*. Support Vector Regression (SVR) is a well-established technique utilized in engineering regression problems, renowned for its advantages such as rapid learning, exceptional generalization capabilities, and resilience to noise [85]. SVR utilizes a kernel function to map data from the sample space to a higher-dimensional characteristic space. By discerning intricate relationships between observed and response variables, this regression model adeptly transforms nonlinear problems into linear ones. The structure of SVR is illustrated in Fig. 17.

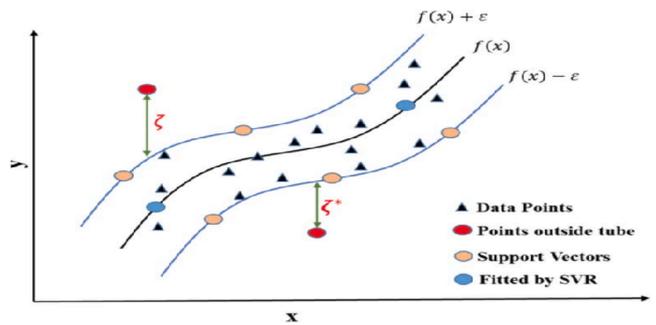


Fig. 17. Structure of Support Vector Regression (SVR) [86].

3.3.2.9. *Random forest (RF)*. The random forest model, introduced by [87], is a member of the decision tree family, adhering to the "divide and conquer" principle. It employs multiple random trees to make predictions, selecting different rows and columns from the training dataset through a bootstrapping process as illustrated in Fig. 18. This approach mitigates correlations between trees and reduces variance [88]. During runtime, each tree generates a prediction, and the model aggregates the mean of these individual predictions to produce an overall prediction result.

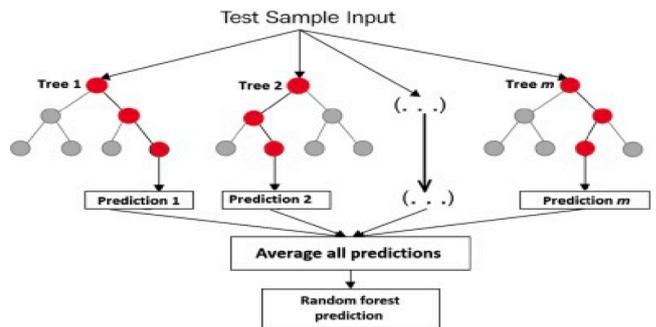


Fig. 18. Structure of Random Forest (RF) [89].

3.3.2.10. *Hybrid model [CNN-LSTM-RF]*. The hybrid model architecture comprises three main components: Convolutional Neural Network (CNN), LSTM, and RF. Here's a more in-depth explanation of each component and how they are integrated:

1. Convolutional Neural Network (CNN):

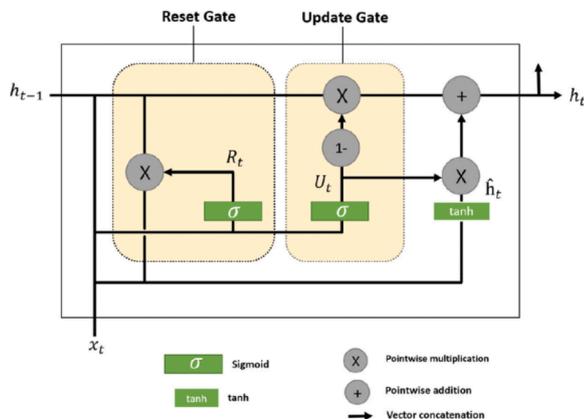


Fig. 14. GRU block diagram structure [77].

- The CNN component is responsible for extracting spatial features from the input data, which in our case, are the time series data representing SPG.
- CNN layers consist of convolutional and pooling layers that learn hierarchical representations of the input data, capturing patterns and relationships between neighboring data points.
- 2. Long Short-Term Memory (LSTM):
- The LSTM component is a type of RNN designed to model sequential data and capture temporal dependencies over time.

- LSTM cells contain a memory mechanism that allows them to retain information over long sequences, making them suitable for time series forecasting tasks.

- LSTM layers in our model process the output of the CNN layers and learn temporal patterns in the SPG data.

3. Random Forest (RF):

- The RF component serves as a decision-making ensemble model that combines the predictions of multiple decision trees.

- RF is known for its robustness and ability to handle complex, nonlinear relationships in the data.

- In our hybrid model, RF integrates the features extracted by both the CNN and LSTM components to make the final predictions of SPG.

The integration of these components follows a sequential process: the CNN layers extract spatial features from the input data, which are then fed into the LSTM layers to capture temporal dependencies. The output of the LSTM layers, along with additional features derived from the CNN and LSTM layers, is used as input to the RF model for final prediction.

### 3.3.3. Hyperparameters tuning for machine learning models

In this section, the hyperparameters utilized in the research are delved into to optimize the performance of the applied models for predicting SPG. The process of fine-tuning these parameters, known as hyperparameter tuning, is crucial for achieving the best results. The key parameters considered in this study include:

- Best optimizer
- Activation function
- Learning rate
- Number of epochs
- Batch size
- Number of hidden layers
- Dropout
- Number of estimators in RF

By systematically adjusting these hyperparameters, in aim to improve the predictive accuracy and overall efficiency of the ML models in forecasting SPG.

### 3.3.4. Metrics selection

Various metrics exist to statistically measure the performance of data regression [90]. This paper emphasizes specific metrics for evaluating deep learning models, including root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R-squared). Eqs. (3) to (5) illustrate these metrics.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_t - y_t^p)^2}{n}} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (4)$$

$$coefficient\ of\ determination\ (R^2) = 1 - \frac{SS_{regression}}{SS_{total}} \quad (5)$$

Where:

$SS_{regression}$ —The regression sum of squares (explained sum of squares).

$SS_{total}$ —The sum of all squares.

## 4. Results and discussion

After deploying the ML algorithms for forecasting solar power, the metrics were used to evaluate the models' efficiency with different ML algorithms. This section will present the outcomes of the forecasting process alongside the corresponding performance metrics, including  $R^2$ ,

RMSE, and MAE, for each ML methods. The dataset was split into a training set, accounting for 80 % of the data, and a test set, representing the remaining 20 %. The implementation of the models was conducted using the Python programming language. A Jupyter Notebook served as the experimentation platform. The experiments were carried out on a machine equipped with a dedicated NVidia 1080Ti GPU boasting 11 GB of memory.

### 4.1. Classical machine learning algorithms

In this subsection, ML models are employed, specifically RF and SVR, to forecast SPG and what the best results are. These models demonstrated promising performance, achieving a testing result as shown in Fig. 19-27. This indicates that the models could capture patterns and trends within the training data effectively. The optimal number of estimators for the Random Forest model was determined to be 100, further enhancing its predictive capabilities.

Fig. 19 illustrates the actual and expected outcomes of forecast SPG with the RF model. The test outcome values obtained from the test sample, and the ideal number of estimators is 100. The RF model has the greatest value,  $R^2 = 0.89$ . It was determined that the RF model had the best accuracy and the lowest error rate. This shows that the RF model may be marginally more accurate in anticipating real SPG numbers.

Fig. 20 depicts the real and expected outcomes of forecast SPG with the SVR model. The test result numbers obtained from the test sample. The SVR model produced the greatest value at the test dataset, with  $R^2 = 0.83$ . It was found that the SVR model had the best accuracy and the smallest error rate, with the mistake occurring in the peak value of solar power output.

Table 2 shows the forecasting test results for a classical ML model. The RF model appears to be a good choice for forecasting in this case. This is because it has a lower RMSE (0.09) and MAE (0.05) compared to the SVR model (RMSE: 0.13, MAE: 0.10). Lower RMSE and MAE indicate better accuracy in predicting the actual values and absolute errors, respectively. Both models have high  $R^2$  values (0.89 for RF and 0.83 for SVR). This means that both models explain a large proportion of the variance in the SPG variable, but the RF model performs slightly better in this aspect.

Overall, the RF model seems to be the more accurate and efficient choice for forecasting.

### 4.2. Deep learning algorithms

Upon the implementation of various neural network architectures, including Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), Recurrent Neural Network (RNN), and Gated Recurrent Unit (GRU), the performance evaluation of these models ensues. This segment presents the forecasted outcomes generated by the DL algorithms, accompanied by performance metrics such as R-squared, root mean square error (RMSE), and mean absolute error (MAE). Each algorithmic variant is subjected to a distinct analysis employing these statistical measures to gauge its predictive accuracy and overall efficiency.

This section delineates the methodology employed to attain optimal forecasting outcomes in each model, elucidated through considerations of critical parameters. The optimization process is explained concerning: (1) the optimal number of hidden layers, (2) batch size, (3) learning rate, (4) optimizer type, and (5) activation function. Initially, the four models (LSTM, Bi-LSTM, RNN, and GRU) exhibited superior performance when subjected to an 80 % training rate and a corresponding 20 % test rate. Multiple optimizers with varying learning rates were applied to each model to discern the most favorable combination. The optimal configuration was determined to be a learning rate of 0.01 coupled with the Adam optimizer, yielding the most favorable results. Furthermore, the optimal number of hidden layers was identified as two, with an epoch size of 50 and a batch size of 32, collectively producing the best

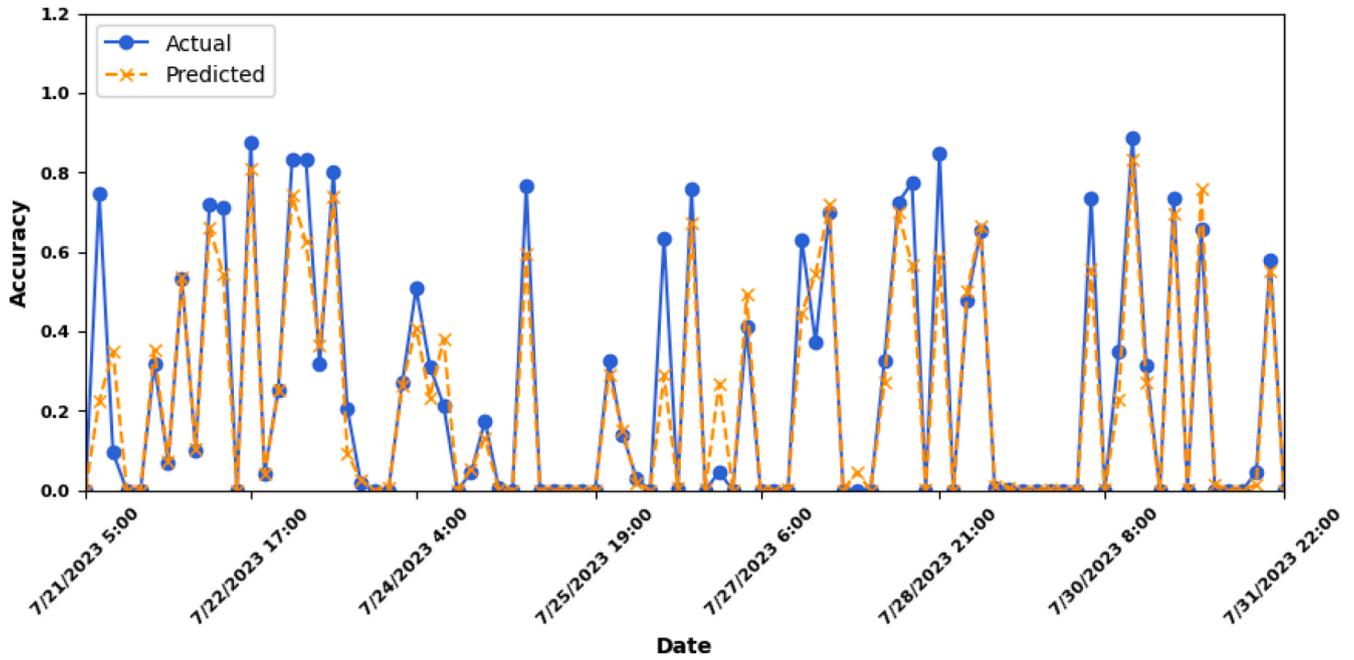


Fig. 19. Random Forest test results for forecasting SPG.

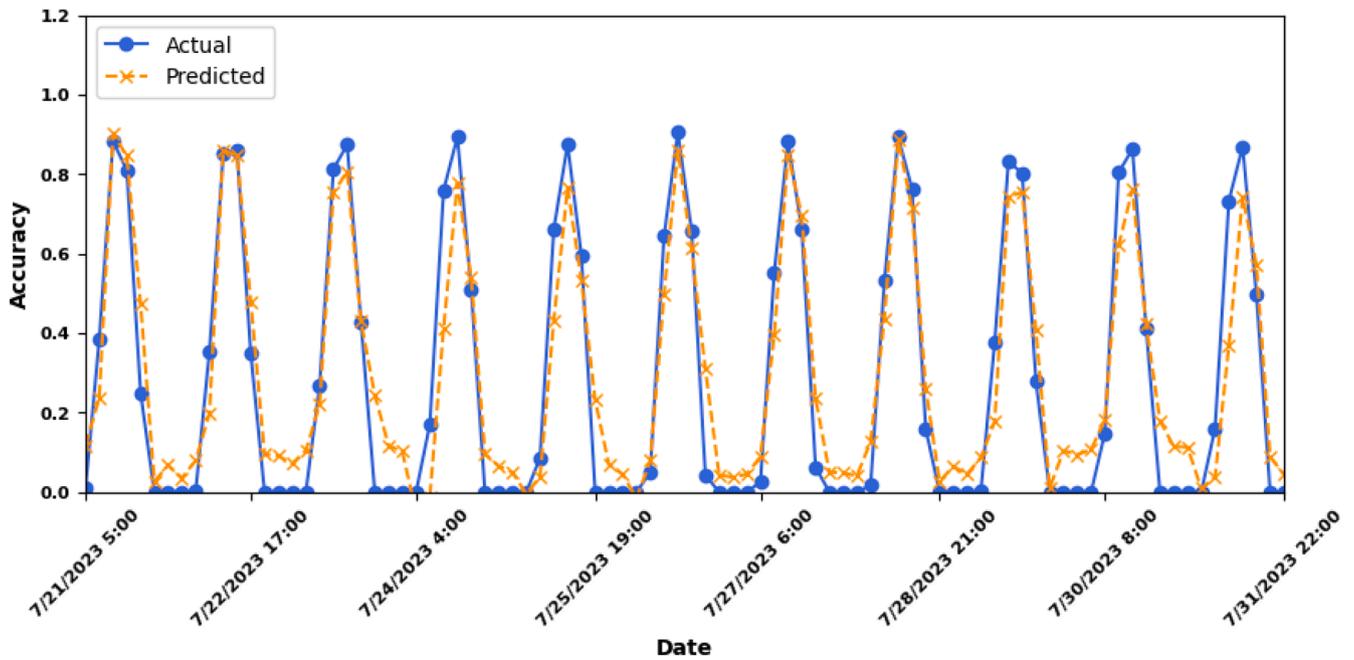


Fig. 20. SVR test results for forecasting SPG.

Table 2  
Forecasting test result for classical ML model.

	RF	SVR
RMSE (kW)	0.09	0.13
MAE (kW)	0.05	0.10
R <sup>2</sup>	0.89	0.83
Time training (second)	6.52	0.44
Memory training (MB)	614.76	783.33

forecasting outcomes.

Fig. 21 displays the real and expected outcomes of forecasting solar energy generation with the LSTM model. The test outcome values were

obtained from the test, and the Adam optimizer had two hidden layers. The LSTM model had the greatest value ( $R^2 = 0.898$ ). It was determined that the LSTM model had the best accuracy and the smallest error rate. This shows that the LSTM model may be marginally more accurate in anticipating real SPG levels.

Fig. 22 depicts both the real and expected outcomes of forecasting SPG with the Bi-LSTM model. Where the test outcome values were obtained from the test sample, and the Adam optimizer had two hidden layers. The Bi-LSTM model has the greatest value ( $R^2 = 0.9027$ ). It was determined that the Bi-LSTM model had the best accuracy and a small error rate. This shows that the Bi-LSTM model may be more accurate in projecting real SPG levels than the LSTM.

Fig. 23 represents the real and anticipated outcomes of forecast SPG

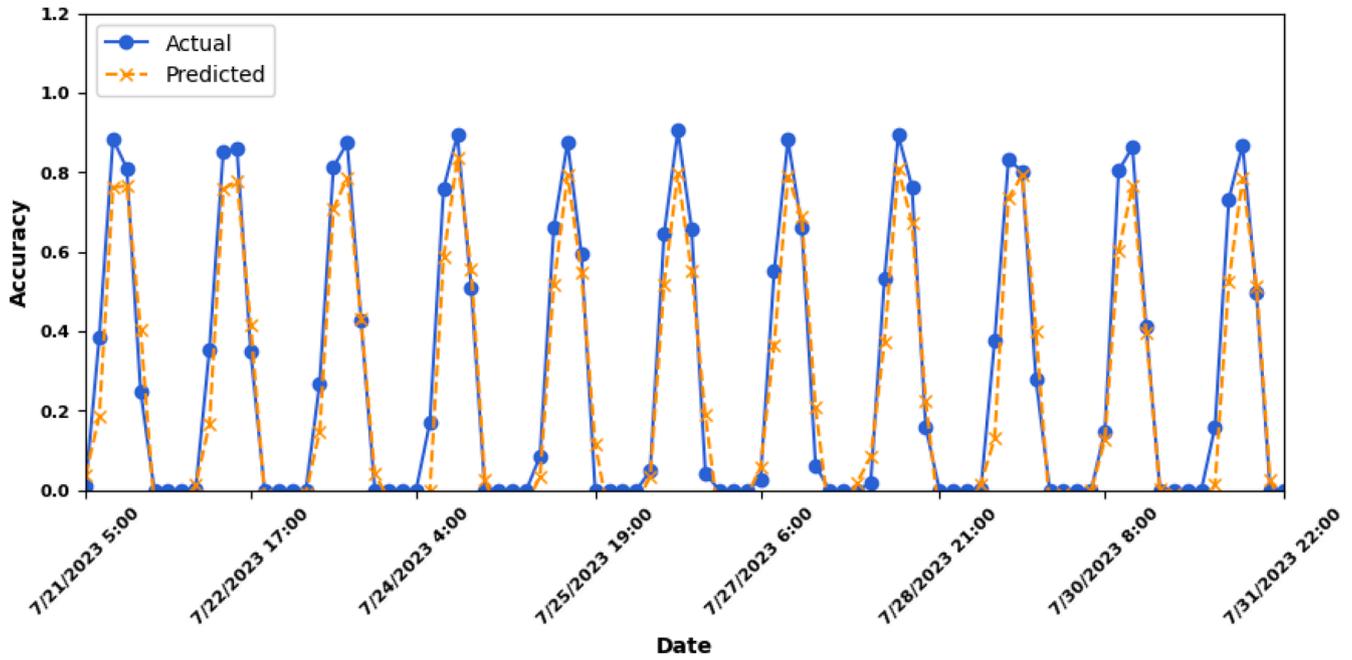


Fig. 21. LSTM test results for forecasting SPG.

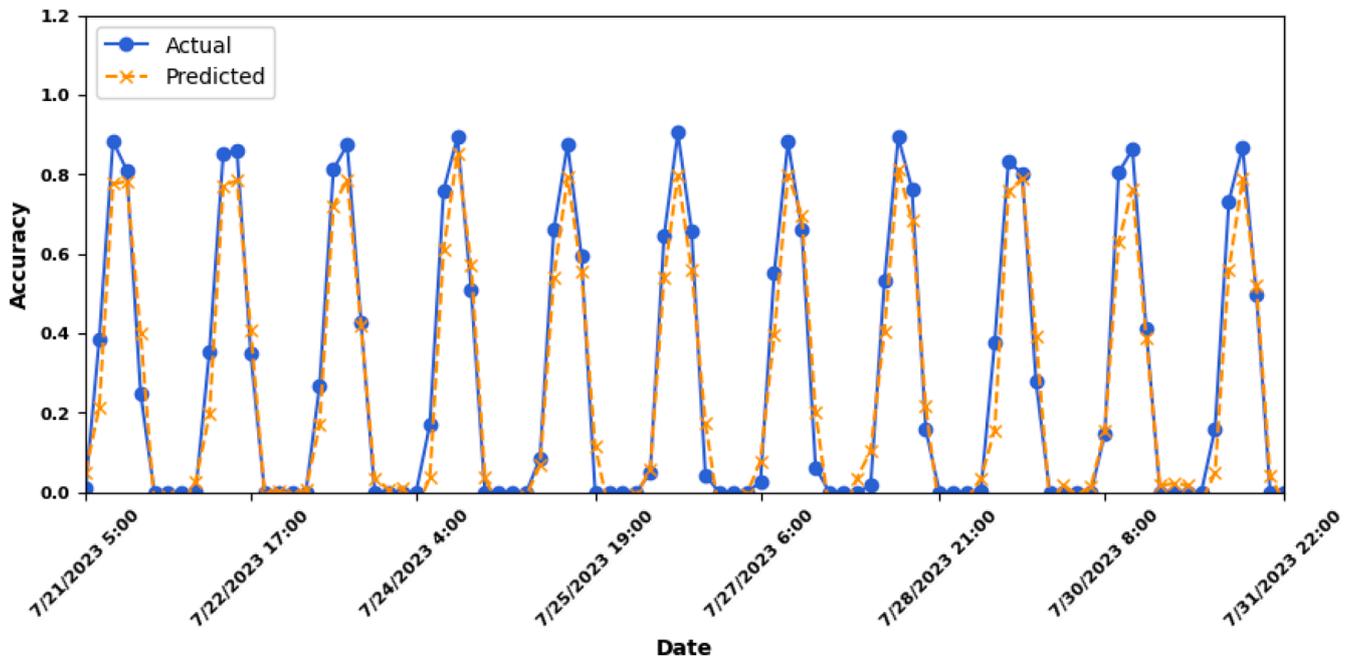


Fig. 22. Bi-LSTM test results for forecasting SPG.

with the RNN model. Where the test outcome values were obtained from the test, and the Adam optimizer had two hidden layers. The RNN model had the greatest value ( $R^2 = 0.8957$ ). It was determined that the RNN model had the best accuracy and the least error rate. This shows that the RNN model may be slightly more accurate in forecasting real SPG levels, with results comparable to those of the LSTM and Bi-LSTM models.

Fig. 24 illustrates the real and expected outcomes of forecasting SPG with the GRU model. Where the test result values were obtained from the test, and the Adam optimizer had two hidden layers. The GRU model had the greatest result ( $R^2 = 0.894$ ). It was determined that the GRU model had the maximum accuracy and the least error rate. This implies that the GRU model is somewhat better at forecasting real SPG levels, with results comparable to the RNN and LSTM models.

Table 3 shows the results of a forecasting test for four different DL models: LSTM, Bi-LSTM, RNN, and GRU. The metrics used to evaluate the models are RMSE, MAE, and R-square. All four models have similar RMSE and MAE values, which are all around 0.10 and 0.06, respectively. This suggests that all four models have similar levels of accuracy in terms of predicting the absolute error. The Bi-LSTM model has the highest  $R^2$  value, at 0.9027. This indicates that the Bi-LSTM model has the best fit for the data among the four models.

Overall, the Bi-LSTM model appears to be the best-performing model based on the R-square metric. The Bi-LSTM model outperforms the other three models in terms of  $R^2$ , which is a measure of how well the model fits the data. The Bi-LSTM model has an  $R^2$  value of 0.9027, while the other models have  $R^2$  values that are all around 0.89. This suggests that

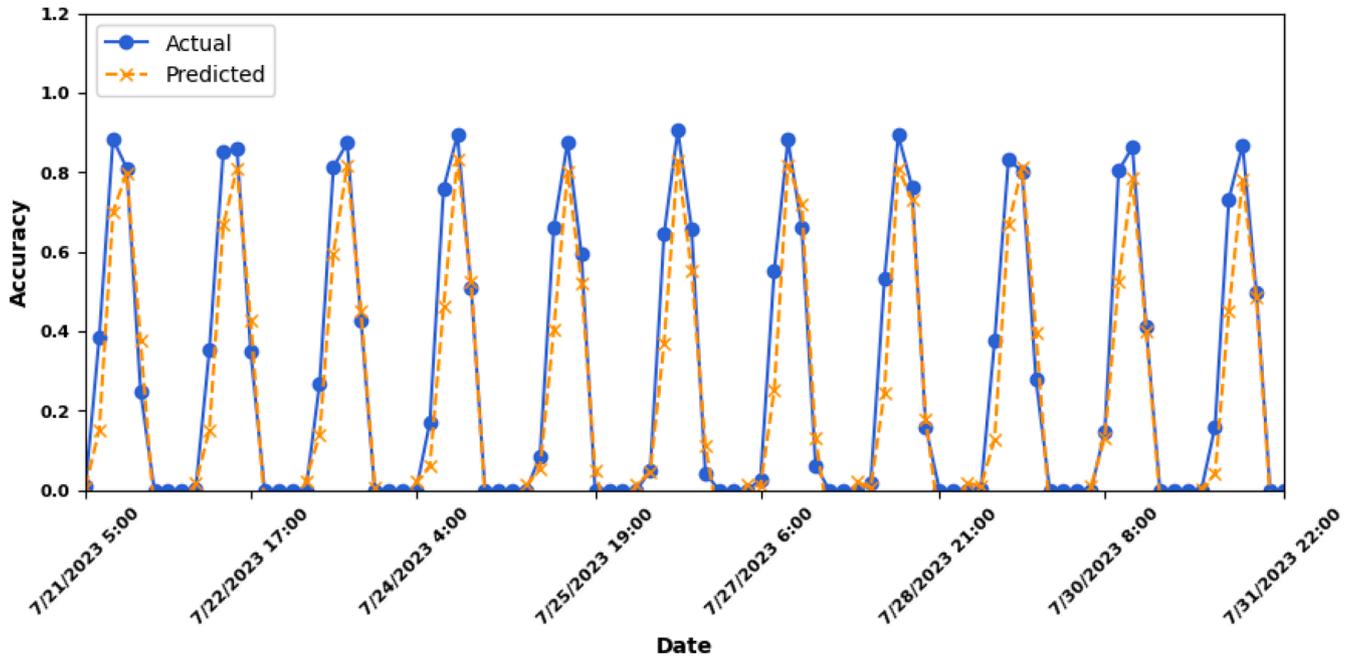


Fig. 23. RNN test results for forecasting SPG.

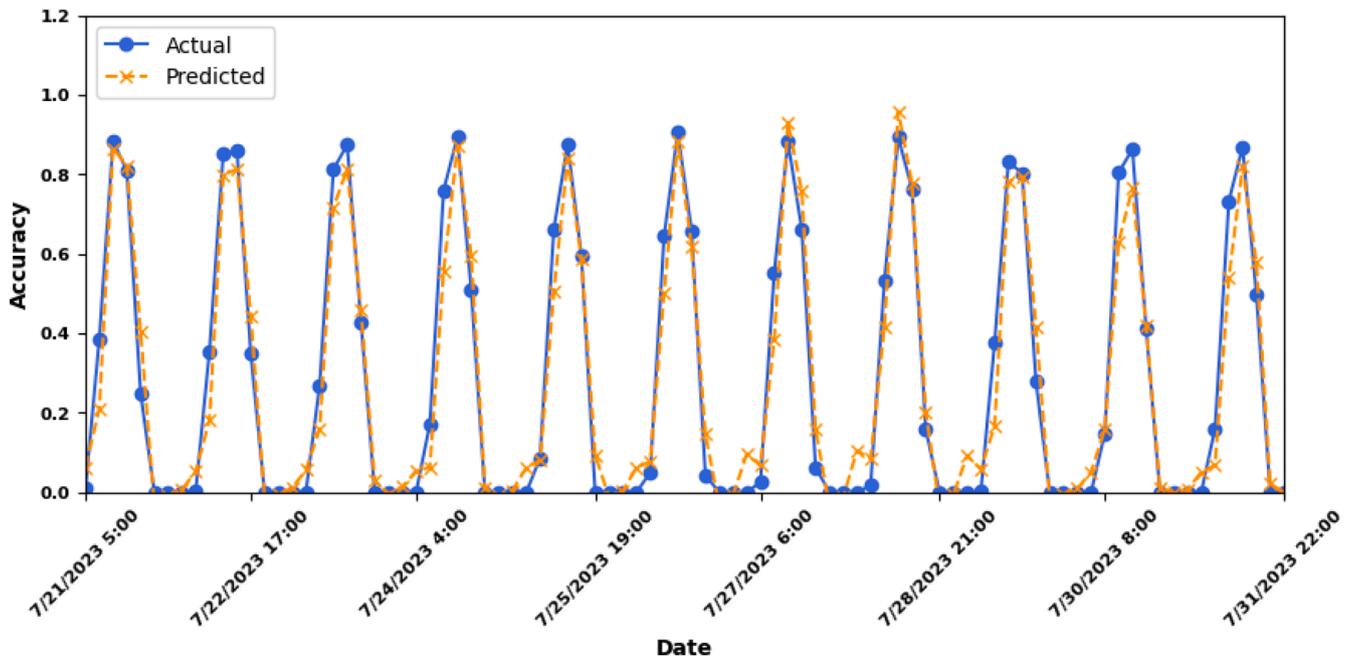


Fig. 24. GRU test results for forecasting SPG.

the Bi-LSTM model can capture the underlying relationships in the data more accurately than the other models.

### 4.3. Hybrid machine learning algorithms

In this section, delving into the exploration and analysis of hybrid machine learning models, specifically the LSTM-RF (Long Short-Term Memory—Random Forest) and CNN-LSTM-RF (Convolutional Neural Network—Long Short-Term Memory—Random Forest). These models combine the strengths of different architectures to harness complementary features and enhance overall performance.

The LSTM-RF model integrates the temporal dependencies captured by Long Short-Term Memory networks with the robustness of Random

Table 3

Forecasting test result for Deep Learning model.

	LSTM	Bi-LSTM	RNN	GRU
RMSE (kW)	0.10	<b>0.10</b>	0.10	0.10
MAE (kW)	0.007	<b>0.06</b>	0.06	0.06
R <sup>2</sup>	0.898	<b>0.9027</b>	0.895	0.894
Time training (second)	18.10	<b>18.20</b>	26.24	19.94
Memory (MB)	605.61	<b>602.83</b>	752.55	657.21

Forests. This fusion aims to leverage the sequential information within the data while harnessing the ensemble learning capabilities of Random Forests for improved accuracy and generalization. Building upon the

LSTM-RF framework, the CNN-LSTM-RF model incorporates convolutional neural networks (CNNs) to extract hierarchical features from input data. CNNs excel at capturing spatial patterns, and when combined with LSTM and RF components, they form a powerful triad for handling complex and multi-dimensional data.

During the experimentation phase, the Adam optimizer emerged as the optimal choice for fine-tuning our hybrid models. The hyperparameters were carefully tuned, revealing that employing a configuration with two hidden layers, a learning rate of 0.01, and a batch size of 32 yielded the most promising results. Furthermore, a total of 50 epochs were found to strike a balance between model convergence and computational efficiency. The decision to adopt the Adam optimizer at this specific configuration stems from its adaptive learning rate capabilities, making it well-suited for training complex hybrid architectures. The amalgamation of LSTM, CNN, and RF components under the guidance of Adam Optimization forms the backbone of our exploration into hybrid machine learning for the given task.

Fig. 25 illustrates the actual and expected outcomes of forecast SPG with the LSTM-RF hybrid model. Where the test outcome values were obtained from the test, and the Adam optimizer had two hidden layers. The LSTM-RF model has the greatest value,  $R^2 = 0.89$ . It was determined that the LSTM-RF model had the best accuracy and the least error rate. This shows that the LSTM-RF model may be slightly more accurate in forecasting real SPG values, with results comparable to the DL and RF models.

Fig. 26 depicts both the real and anticipated outcomes of forecasting SPG with the CNN-LSTM-RF hybrid model. Where the test result values were obtained from the test, and the Adam optimizer had two hidden layers. The CNN-LSTM-RF model has the greatest  $R^2$  score, 0.89. It was determined that the CNN-LSTM-RF model had the best accuracy and least error rate. This shows that the CNN-LSTM-RF model may be marginally better at forecasting real SPG values, with results comparable to those of the DL and RF models. Furthermore, it is the best model among conventional and DL models, with CNN-LSTM-RF achieving the greatest accuracy and smallest error.

Table 4 shows the forecasting test results for four hybrid ML models: LSTM-RF and CNN-LSTM-RF. The metrics used to evaluate the models are RMSE, MAE, and R-square. The CNN-LSTM-RF model has the lowest RMSE and MAE values, at 0.07 and 0.05, respectively. This suggests that

the CNN-LSTM-RF model is the most accurate among the two hybrid models in terms of predicting the absolute error. The CNN-LSTM-RF model has the highest  $R^2$  value, at 0.92. This indicates that the CNN-LSTM-RF model has the best fit for the data among the two hybrid models.

Overall, the CNN-LSTM-RF model appears to be the best-performing model for this specific task based on both RMSE and MAE metrics. Moreover, the training time for the proposed model is 0.58 second, and the used memory in training is 10,145.66 MB.

#### 4.4. Comparing the best result for each classical, deep, and hybrid models

In this section, the comparison of the best-performing models from three different categories: classical machine learning, deep learning, and hybrid models, will be conducted. The evaluation of the models will be based on the following metrics: RMSE, MAE, and R-square. For each category, the model with the best performance based on the chosen metrics will be identified. Subsequently, a comparison of the best-performing models from each category will be conducted to determine which type of model (classical, deep, or hybrid) is best suited for this task.

Table 5 shows the results of a forecasting test for the best three ML models used for SPG forecasting: RF, Bi-LSTM, and CNN-LSTM-RF. The metrics used to evaluate the models are RMSE, MAE, and  $R^2$ . The CNN-LSTM-RF hybrid model has the lowest RMSE and MAE values, at 0.07 and 0.05, respectively. This suggests that the CNN-LSTM-RF model is the most accurate among the ML models in terms of predicting the absolute error in SPG. The CNN-LSTM-RF model has the highest  $R^2$  value, at 0.92. This indicates that the CNN-LSTM-RF model has the best fit for the data among the three models. In other words, the CNN-LSTM-RF model is the best at capturing the underlying relationships in the data that affect SPG. Therefore, the CNN-LSTM-RF is the best and the most accurate forecasting model for forecasting SPG.

Fig. 27 shows the comparison of model performance for three models (RF, Bi-LSTM, and CNN-LSTM-RF) using RMSE and MAE metrics. The CNN-LSTM-RF model appears to be the best-performing model overall. It has the lowest RMSE and MAE values, indicating that it has the smallest average absolute difference between its predictions and the actual values. All three models perform relatively well, with RMSE

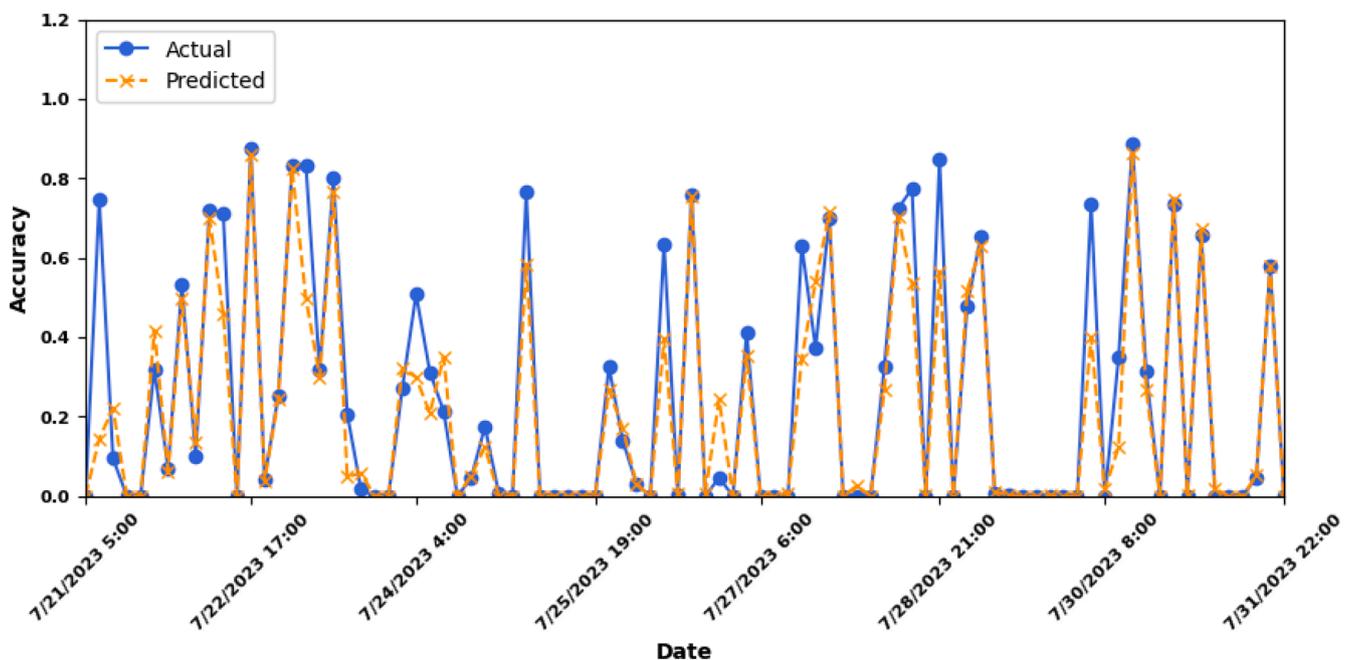


Fig. 25. LSTM-RF test results for forecasting SPG.

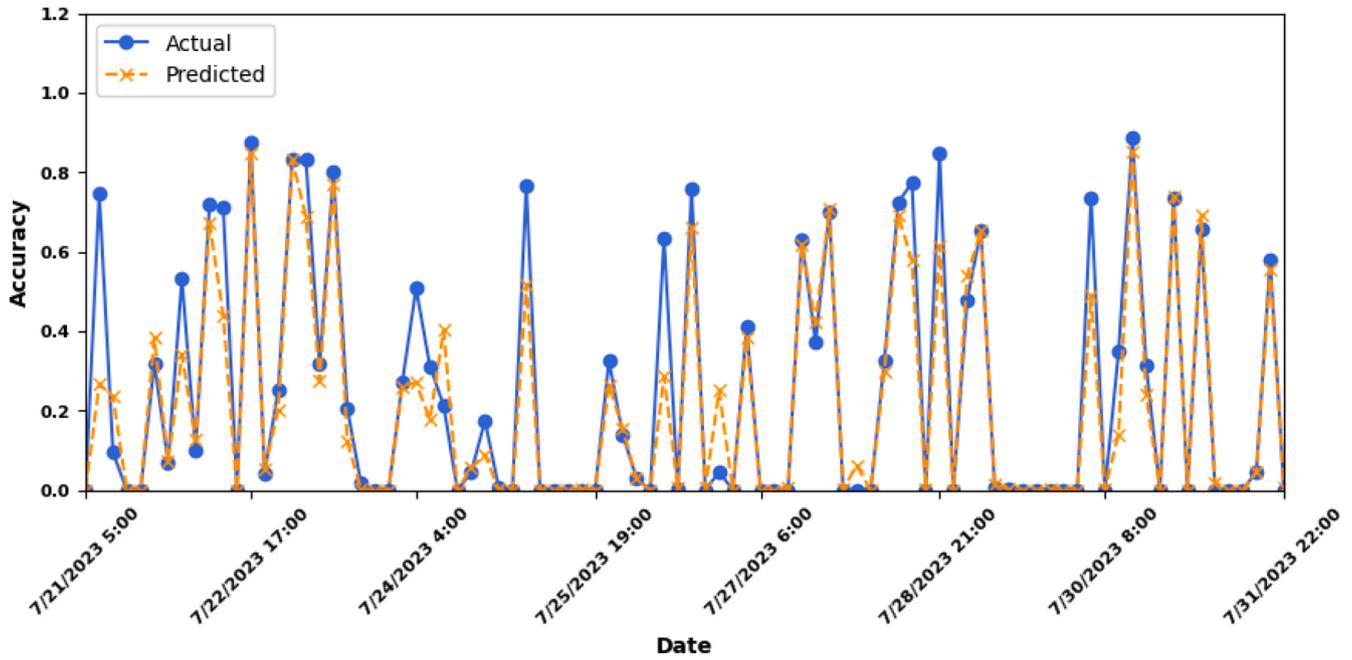


Fig. 26. CNN-LSTM-RF test results for forecasting SPG.

**Table 4**  
Forecasting test result for Hybrid ML model.

	LSTM-RF	CNN-LSTM-RF
RMSE (kW)	0.09	<b>0.07</b>
MAE (kW)	0.05	<b>0.05</b>
R <sup>2</sup>	0.89	<b>0.92</b>
Time training (second)	28.62	<b>49.89</b>
Memory training (MB)	694.25	<b>783.33</b>

**Table 5**  
Comparing forecasting test result for ML model.

	RMSE (kW)	MAE (kW)	R <sup>2</sup>
RF	0.09	0.05	0.89
Bi-LSTM	0.10	0.06	0.9027
CNN-LSTM-RF	<b>0.07</b>	<b>0.05</b>	<b>0.92</b>

values below 0.1 and MAE values below 0.06. This suggests that they are all capable of making reasonably accurate SPG forecasts. The Bi-LSTM model has a slightly higher RMSE than the other two models, but its MAE is similar to the CNN-LSTM-RF model. This suggests that the Bi-LSTM model may be more prone to occasional large errors. The RF model has the highest RMSE and MAE values of the three models. Then the CNN-LSTM-RF hybrid model is the preferred choice for forecasting solar power generation because it has the lowest error and high accuracy compared with RF and Bi-LSTM.

Table 6 reveals that the results of each study vary due to several factors, including the dataset, specific features (such as solar radiation, wind speed, atmospheric pressure, and temperature), the algorithms implemented, and their tuning parameters. Our research demonstrates improved performance utilizing the combined CNN-LSTM-RF model, achieving an RMSE of 0.07, an MAE of 0.05, and an R-squared value of 0.92. This improvement is attributed to the adjustment of eight hyper-parameters and the incorporation of additional features not previously considered, optimizing the model’s performance while preventing overfitting.

While our hybrid CNN-LSTM-RF model showed promising results in forecasting solar power generation, it’s important to recognize certain limitations encountered during the study. These may include challenges are the inherent variability and unpredictability of weather conditions. Solar power generation is heavily influenced by factors such as cloud cover, atmospheric conditions, and seasonal changes, which can be challenging to accurately predict over extended timeframes. Additionally, fluctuations in solar panel efficiency and degradation over time can introduce uncertainties into the forecasting process. These factors may lead to inaccuracies in the model predictions, particularly for long-term forecasts.

**5. Conclusions and future work**

Forecasting solar power generation is essential for the efficient development and planning of power systems, contributing to improved technical performance and financial efficiency. This study aimed to develop accurate forecasting models by analyzing a real dataset of solar power generation. The developed models offer valuable insights into future solar power production, enabling optimization of resource

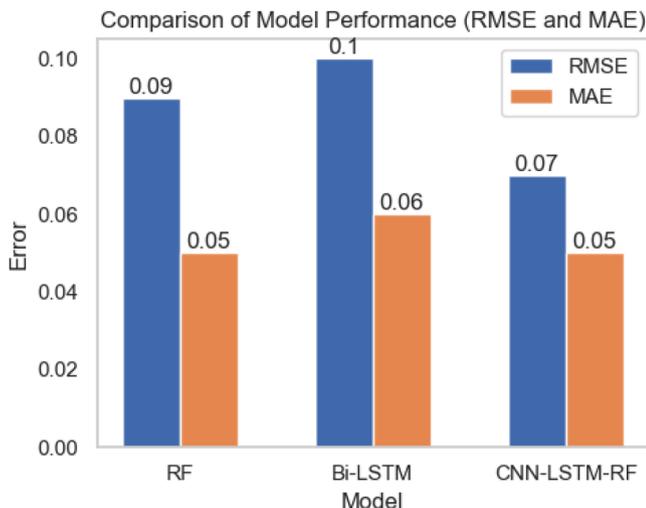


Fig. 27. Comparison of Model Performance (RMSE and MAE).

**Table 6**  
Results from previous studies.

Ref	Algorithms	Result	Location
[26]	Deep Belief Networks, Auto Encoder, and LSTM	RMSE of 0.0713, MAE of 0.0366, absolute deviation of 0.2765.	Kassel, Germany
[27]	LSTM, MLP	MAE = 676.34, RMSE=883.5, $R^2=0.745$	Canada
[36]	ANN, SVR, LR, RF	RF achieved MAE=2.64e-05, MSE=9.93e-0, and RMSE = 0.0009	Morocco
[37]	SP,DBN, SVM, RF	RMSE=0.0343, MAE= 0.01286	Hambantota, Sri Lanka
[38]	ANN, NN, LR, SVM	ANN achieved an RMSE = 0.998	National Solar Radiation Database (NSRD)
[49]	CNN, LSTM, CNN, MLP, WPD-LSTM, ESNCNN	ESNCNN achieved RMSE = 0.1432, MBE = 0.0048	Shaoxing, China
[39]	SVR,ANN,DT,RF,XGBOOST,GAM	ANN achieved RMSE = 2.6e-08, MAE = 0.00013	Morocco
[29]	ANN, LSTM, Bagging, DSE-XGB	DSE-XGB achieved RMSE = 0.78, MAE = 0.59	Public
[40]	ANN, KNN, SVM, LR	ANN achieved RMSE = 86.4, MAE = 8.4	Medellín, Colombia
[41]	Holt-Winters, Multivariate Linear Regression, ARIMA, SARIMA, ARIMAX, SARIMAX, and LSTM	LSTM achieved $R^2 = 0.943$ , MAPE= 5.79	Ansan city, South Korea
[42]	ANN, KNN, GBRT	RMSE for ANN is 0.113, GBRT = 0.112, and KNN = 0.14	Sweden
[43]	SVM, and GBR	Results show that the non-linear methods obtain lower errors than the linear one.	public data from Kaggle for the State of Oklahoma
[44]	SVM, and LR	The results show that SVM more accurate and less error than LR	National Weather Service (NWS)
[45]	SVM, GPR, LR, and DT	DT achieved $R^2 = 95.91$ , MAPE= 5.83	Malaysia
[46]	SVM, RF, and LR	RF achieved MAE = 12.45, RMSE= 27.32	India
[30]	LSTM, GRU, Auto-LSTM, and Auto-GRU	LSTM gives RMSE and MAE are 0.0739, 0.0176, respectively.	Shagaya in Kuwait and Cocoa in the USA
[49]	LSTM, GRU, RNN, MLP, WPD-LSTM, and hybrid ESNCNN	ESNCNN gives RMSE and MBE are 0.1432,0.0048, respectively.	Australia Solar Centre
[47]	SVM, and the Hidden Markov	Accuracy for SVM is 94 %, and Hidden Markov Model 61.8 % in sunny	Australia
[50]	CNN, and the ANN	ANN gives high accuracy and less error.	Stanford University
[51]	CNN, Multi-headed-CNN, CNN-LSTM, ARMA, and MLR	CNN-LSTM achieved MAE =0.051511, RMSE= 0.065213	Public
[28]	RNN, LSTM, and GRU	RNN achieved MAE =1.83, RMSE= 2.92	Errachidia province
[31]	MLFFNN, RBFNN, SVR, FIS, and ANFIS	Results demonstrated that for the SVR and MLFFNN models have the maximum performance to predict the solar irradiance with $R = 0.9999$ and $0.9795$ , respectively.	Iran
[52]	CNN, LSTM, MLP, hybrid deep learning(Prop)	Prop achieved average for RMSE = 1.46.	Shaoxing, China
[53]	RNN+ LSTM	RMSE of 7.416 % and an MAPE of 10.805 %	South Korea
[54]	Ensemble of ANN, SVR, ELM	EA achieved MSE is 0.1295, and MAE =0.1936	India
[55]	CNN and LSTM	MAPE of 4.58 on sunny days and 7.06 on cloudy days	Busan, Korea
[32]	DL and SVR	DL outperformed SVR, with a MAPE of 7.9 % for DL and 8.52 % for SVR	Spain
[33]	LSTM, and ANFIS	LSTM model gives the best results with RMSE, MAE, and R equal to 60.66 kWh, 30.47 kWh, and 0.9777, respectively.	Turkey
[34]	ANN, CNN, and RNN, polynomial regression, SVR, and RF.	RNN achieved $R^2$ of 0.9546 and an RMSE/MAE of 82.22 W/m <sup>2</sup> / 36.52 W/m <sup>2</sup> .	Nigeria
[48]	RBFNN, K-nearest neighbor, K-means clustering	The enhanced model outperforms traditional RBFNN and Multilayer Perceptron Neural Network methods, exhibiting a higher level of precision with a low mean square error, particularly when employing a relatively few neurons on the hidden layer	Jericho, Palestine
[35]	MFNNBP	achieved an average RMSE of 0.034955333	Palestine

utilization and cost reduction. Utilizing various classical ML, DL, and hybrid ML techniques, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Recurrent Neural Networks (RNN), Random Forest (RF), Support Vector Regression (SVR), Bi-directional LSTM (Bi-LSTM), and Convolutional Neural Network (CNN), this study explored different approaches to solar power forecasting. Among these techniques, the hybrid model combining CNN-LSTM-RF demonstrated superior accuracy, achieving an impressive R-squared value of 92 %, with a Root Mean Square Error (RMSE) of 0.07 kW and a Mean Absolute Error (MAE) of 0.05 kW.

The findings highlight the effectiveness of the hybrid machine learning model in accurately forecasting solar power generation. Future research directions could include developing web interfaces for showcasing anticipated solar power generation, specifically tailored for relevant departments. Additionally, there should be a focus on predicting long-term SPG outcomes to further enhance the applicability and effectiveness of the forecasting models. Additionally, investigating the integration of cloud cover data into forecasting models to enhance accuracy in predicting solar power generation.

#### CRediT authorship contribution statement

**Mobarak Abumohsen:** Software, Methodology, Formal analysis,

Data curation, Conceptualization. **Amani Yousef Owda:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis. **Majdi Owda:** Writing – review & editing, Visualization, Methodology, Formal analysis. **Ahmad Abumihsan:** Writing – review & editing, Software, Formal analysis, Data curation.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Supplementary material

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.prime.2024.100636](https://doi.org/10.1016/j.prime.2024.100636).

## References

- [1] A.R. bin Mohd Yusoff, D. Kim, H.P. Kim, F.K. Shneider, W.J. Da Silva, J. Jang, A high efficiency solution processed polymer inverted triple-junction solar cell exhibiting a power conversion efficiency of 11.83%, *Energy Environ. Sci.* 8 (1) (2015) 303–316, <https://doi.org/10.1039/C4EE03078E>.
- [2] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renew. Sustain. Energy Rev.* 27 (2013) 65–76, <https://doi.org/10.1016/j.rser.2013.06.034>.
- [3] S. Oueslati, et al., Physical and electrical characterization of high-performance Cu<sub>2</sub>ZnSnSe<sub>4</sub> based thin film solar cells, *Thin Solid Films* 582 (2015) 224–228, <https://doi.org/10.1016/j.tsf.2014.11.067>.
- [4] S. Guo, et al., Investigation of the short-circuit current increase for PV modules using halved silicon wafer solar cells, *Sol. Energy Mater. Sol. Cells* 133 (2015) 240–247, <https://doi.org/10.1016/j.solmat.2014.10.015>.
- [5] E. Lorenz, J. Hurka, D. Heinemann, H.G. Beyer, Irradiance forecasting for the power prediction of grid-connected photovoltaic systems, *IEEE J. Sel. Top. Appl. earth Obs. Remote Sens.* 2 (1) (2009) 2–10, <https://doi.org/10.1109/JSTARS.2009.2027445>.
- [6] A. Khaligh, O.C. Onar, *Energy harvesting: solar, wind, and Ocean Energy Conversion Systems*, CRC press, 2017, <https://doi.org/10.1201/9781315226432>.
- [7] A. Mellit, A.M. Pavan, A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy, *Sol. Energy* 84 (5) (2010) 807–821, <https://doi.org/10.1016/j.solener.2010.02.009>.
- [8] R. Ahmed, V. Sreeram, Y. Mishra, M.D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization, *Renew. Sustain. Energy Rev.* 124 (2020) 109792, <https://doi.org/10.1016/j.rser.2020.109792>.
- [9] M.L. Sørensen, P. Nystrup, M.B. Bjerregård, J.K. Møller, P. Bacher, H. Madsen, Recent developments in multivariate wind and solar power forecasting, *Wiley Interdiscip. Rev. Energy Environ.* 12 (2) (2023) e465, <https://doi.org/10.1002/wene.465>.
- [10] A.T.T.U. Balal, Y.P.T.T.U. Jafarabadi, A.T.T.U. Demir, M.T.T.U. Igene, M.T.T.U. Giesselmann, and S.T.T.U. Bayne, “Forecasting solar power generation utilizing machine learning models in lubbock,” 2023, <https://10.28991/ESJ-2023-07-04-02>.
- [11] M. Abumohsen, A.Y. Owda, M. Owda, Electrical load forecasting using LSTM, GRU, and RNN algorithms, *Energies. (Basel)* 16 (5) (2023) 2283, <https://doi.org/10.3390/en16052283>.
- [12] M. Abumohsen, A.Y. Owda, M. Owda, Electrical load forecasting based on random forest, XGBoost, and linear regression algorithms, in: 2023 International Conference on Information Technology (ICIT), 2023, pp. 25–31, <https://doi.org/10.1109/ICIT58056.2023.10225968>.
- [13] R. Tawn, J. Browell, A review of very short-term wind and solar power forecasting, *Renew. Sustain. Energy Rev.* 153 (2022) 111758, <https://doi.org/10.1016/j.rser.2021.111758>.
- [14] A. Abumihsan, A. Awad, M. Jubran, A. Qaroush, I. Tumar, A.S. Alshra, Multipath TCP for short flows supported by SDN in heterogeneous networks, in: 2023 8th International Conference on Computer and Communication Systems (ICCCS), 2023, pp. 532–538, <https://doi.org/10.1109/ICCCS57501.2023.10150613>.
- [15] O. Nooruldeen, S. Alturki, M.R. Baker, A. Ghareeb, Time series forecasting for decision making on city-wide energy demand: a comparative study, in: 2022 International Conference on Decision Aid Sciences and Applications (DASA), 2022, pp. 1706–1710, <https://doi.org/10.1109/DASA54658.2022.9765193>.
- [16] Y. Cui, Z. Chen, Y. He, X. Xiong, F. Li, An algorithm for forecasting day-ahead wind power via novel long short-term memory and wind power ramp events, *Energy* 263 (2023) 125888, <https://doi.org/10.1016/j.energy.2022.125888>.
- [17] M.R. Baker, et al., Uncertainty management in electricity demand forecasting with machine learning and ensemble learning: case studies of COVID-19 in the US metropolitans, *Eng. Appl. Artif. Intell.* 123 (2023) 106350, <https://doi.org/10.1016/j.engappai.2023.106350>.
- [18] M. Stankovic, L. Jovanovic, M. Antonijevic, A. Bozovic, N. Bacanin, M. Zivkovic, Univariate Individual household energy forecasting by tuned long short-term memory network, in: *Inventive Systems and Control: Proceedings of ICISC 2023*, Springer, 2023, pp. 403–417.
- [19] S. Garg, R. Krishnamurthi, A survey of long short term memory and its associated models in sustainable wind energy predictive analytics, *Artif. Intell. Rev.* 56 (Suppl 1) (2023) 1149–1198.
- [20] O. Nooruldeen, M.R. Baker, A.M. Aleesa, A. Ghareeb, E.H. Shaker, Strategies for predictive power: machine learning models in city-scale load forecasting, *e-Prime - Adv. Electr. Eng. Electron. Energy* 6 (2023) 100392, <https://doi.org/10.1016/j.prime.2023.100392>.
- [21] A. Khan, R. Bhatnagar, V. Masrani, V.B. Lobo, A comparative study on solar power forecasting using ensemble learning, in: 2020 4th international conference on trends in electronics and informatics (ICOEI)(48184), 2020, pp. 224–231, <https://doi.org/10.1109/ICOEI48184.2020.9142884>.
- [22] P.W. Khan, Y.-C. Byun, S.-J. Lee, D.-H. Kang, J.-Y. Kang, H.-S. Park, Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources, *Energies. (Basel)* 13 (18) (2020) 4870, <https://doi.org/10.3390/en13184870>.
- [23] D.-F. Li, et al., *International Journal of Fuzzy System Applications*, Int. J. 9 (1) (2020).
- [24] Z. Li, S.M.M. Rahman, R. Vega, B. Dong, A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting, *Energies. (Basel)* 9 (1) (2016) 55, <https://doi.org/10.3390/en9010055>.
- [25] Y. Wang, G. Cao, S. Mao, R.M. Nelms, Analysis of solar generation and weather data in smart grid with simultaneous inference of nonlinear time series, in: 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 2015, pp. 600–605, <https://doi.org/10.1109/INFOCOMW.2015.7179451>.
- [26] A. Gensler, J. Henze, B. Sick, N. Raabe, Deep Learning for solar power forecasting—an approach using AutoEncoder and LSTM Neural Networks, in: 2016 IEEE international conference on systems, man, and cybernetics (SMC), 2016, pp. 2858–2865, <https://doi.org/10.1109/SMC.2016.7844673>.
- [27] M. Elsaraiti, A. Merabet, Solar power forecasting using deep learning techniques, *IEEE Access.* 10 (2022) 31692–31698, <https://doi.org/10.1109/ACCESS.2022.3160484>.
- [28] I. Jebli, F.-Z. Belouadha, M.I. Kabbaj, A. Tilioua, Deep learning based models for solar energy prediction, *Adv. Sci. Technol. Eng. Syst. J.* 6 (1) (2021) 349–355, <https://doi.org/10.25046/aj060140>.
- [29] W. Khan, S. Walker, W. Zeiler, Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach, *Energy* 240 (2022) 122812, <https://doi.org/10.1016/j.energy.2021.122812>.
- [30] M. AlKandari, I. Ahmad, Solar power generation forecasting using ensemble approach based on deep learning and statistical methods, *Appl. Comput. Inf.* (2020), <https://doi.org/10.1016/j.aci.2019.11.002>.
- [31] A. Khosravi, R.N.N. Koury, L. Machado, J.J.G. Pabon, Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms, *J. Clean. Prod.* 176 (2018) 63–75, <https://doi.org/10.1016/j.jclepro.2017.12.065>.
- [32] M.A.F.B. Lima, L.M. Fernández Ram\`{a}n, P.C.M. Carvalho, J.G. Batista, D. M. Freitas, A comparison between deep learning and support vector regression techniques applied to solar forecast in Spain, *J. Sol. Energy Eng.* 144 (1) (2022) 10802, <https://doi.org/10.1115/1.4051949>.
- [33] A. Ozbek, A. Yildirim, M. Bilgili, Deep learning approach for one-hour ahead forecasting of energy production in a solar-PV plant, *Energy Sourc. Part A Recover. Util. Environ. Eff.* 44 (4) (2022) 10465–10480, <https://doi.org/10.1080/15567036.2021.1924316>.
- [34] O. Bamisile, A. Oluwasanmi, C. Ejayi, N. Yimen, S. Obiora, Q. Huang, Comparison of machine learning and deep learning algorithms for hourly global/diffuse solar radiation predictions, *Int. J. Energy Res.* 46 (8) (2022) 10052–10073, <https://doi.org/10.1002/er.6529>.
- [35] I. Qasrawi, M. Awad, Prediction of the power output of solar cells using neural networks: solar cells energy sector in Palestine, *Int. J. Comput. Sci. Secur.* 9 (6) (2015) 280.
- [36] I. Jebli, F.-Z. Belouadha, M.I. Kabbaj, A. Tilioua, Prediction of solar energy guided by pearson correlation using machine learning, *Energy* 224 (2021) 120109, <https://doi.org/10.1016/j.energy.2021.120109>.
- [37] P.A.G.M. Amarasinghe, S.K. Abeygunawardane, Application of machine learning algorithms for solar power forecasting in Sri Lanka, in: 2018 2nd International Conference On Electrical Engineering (EECon), 2018, pp. 87–92, <https://doi.org/10.1109/EECon.2018.8541017>.
- [38] F. Jawaid, K. NazirJunejo, Predicting daily mean solar power using machine learning regression techniques, in: 2016 Sixth International Conference on Innovative Computing Technology (INTECH), 2016, pp. 355–360, <https://doi.org/10.1109/INTECH.2016.7845051>.
- [39] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, A. Chehri, Forecasting solar energy production: a comparative study of machine learning algorithms, *Energy Reports* 10 (2023) 1004–1012, <https://doi.org/10.1016/j.egy.2023.07.042>.
- [40] L. Gutiérrez, J. Patiño, E. Duque-Grisales, A comparison of the performance of supervised learning algorithms for solar power prediction, *Energies. (Basel)* 14 (15) (2021) 4424, <https://doi.org/10.3390/en14154424>.
- [41] E. Kim, M.S. Akhtar, O.-B. Yang, Designing solar power generation output forecasting methods using time series algorithms, *Electr. Power Syst. Res.* 216 (2023) 109073, <https://doi.org/10.1016/j.epr.2022.109073>.
- [42] E. Isaksson and M. Karpe Conde, “Solar power forecasting with machine learning techniques,” 2018, <https://diva2:1215661>.
- [43] R. Aler, R. Mart\`{a}n, J.M. Valls, I.M. Galván, A Study of Machine Learning Techniques for Daily Solar Energy Forecasting Using Numerical Weather Models, *Intelligent Distributed Computing VIII*, 2015, pp. 269–278, [https://doi.org/10.1007/978-3-319-10422-5\\_29](https://doi.org/10.1007/978-3-319-10422-5_29).
- [44] N. Sharma, P. Sharma, D. Irwin, P. Shenoy, Predicting solar generation from weather forecasts using machine learning, in: 2011 IEEE international conference on smart grid communications (SmartGridComm), 2011, pp. 528–533, <https://doi.org/10.1109/SmartGridComm.2011.6102379>.
- [45] Z. Zulkifly, K.A. Baharin, C.K.I.M. Gan, Improved machine learning model selection techniques for solar energy forecasting applications, *Int. J. Renew. Energy Res.* 11 (1) (2021) 308–319.
- [46] K. Anuradha, D. Erlapally, G. Karuna, V. Srilakshmi, K. Adilakshmi, Analysis of solar power generation forecasting using machine learning techniques, in: *E3S Web of Conferences*, 2021, p. 1163, <https://doi.org/10.1051/e3sconf/202130901163>.
- [47] J. Li, J.K. Ward, J. Tong, L. Collins, G. Platt, Machine learning for solar irradiance forecasting of photovoltaic system, *Renew. Energy* 90 (2016) 542–553, <https://doi.org/10.1016/j.renene.2015.12.069>.
- [48] M. Awad, I. Qasrawi, Enhanced RBF neural network model for time series prediction of solar cells panel depending on climate conditions (temperature and irradiance), *Neural Comput. Appl.* 30 (2018) 1757–1768.
- [49] Z.A. Khan, T. Hussain, I.U. Haq, F.U.M. Ullah, S.W. Baik, Towards efficient and effective renewable energy prediction via deep learning, *Energy Rep.* 8 (2022) 10230–10243, <https://doi.org/10.1016/j.egy.2022.08.009>.

- [50] Y. Sun, V. Venugopal, A.R. Brandt, Short-term solar power forecast with deep learning: exploring optimal input and output configuration, *Sol. Energy* 188 (2019) 730–741, <https://doi.org/10.1016/j.solener.2019.06.041>.
- [51] A.M. Alam, I.Asif Razeed Nahid-Al-Masood, M. Zunaed, Solar PV power forecasting using traditional methods and machine learning techniques, in: 2021 IEEE Kansas Power and Energy Conference (KPEC), 2021, pp. 1–5, <https://doi.org/10.1109/KPEC51835.2021.9446199>.
- [52] H. Zhou, Q. Liu, K. Yan, Y. Du, Deep learning enhanced solar energy forecasting with AI-driven IoT, *Wirel. Commun. Mob. Comput.* 2021 (2021) 1–11, <https://doi.org/10.1155/2021/9249387>.
- [53] Y. Jung, J. Jung, B. Kim, S. Han, Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: case study of South Korea, *J. Clean. Prod.* 250 (2020) 119476, <https://doi.org/10.1016/j.jclepro.2019.119476>.
- [54] A. Nayak, L. Heistrene, Hybrid machine learning model for forecasting solar power generation, in: 2020 International Conference on Smart Grids and Energy Systems (SGES), 2020, pp. 910–915, <https://doi.org/10.1109/SGES51519.2020.00167>.
- [55] Z. Chen, I. Koprinska, Ensemble methods for solar power forecasting, in: 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1–8. <https://power.larc.nasa.gov/data-access-viewer/>.
- [56] U. Javed, et al., Exploratory data analysis based short-term electrical load forecasting: a comprehensive analysis, *Energies*. (Basel) 14 (17) (2021) 5510, <https://doi.org/10.3390/en14175510>.
- [57] A.G. Asuero, A. Sayago, A.G. González, The correlation coefficient: an overview, *Crit. Rev. Anal. Chem.* 36 (1) (2006) 41–59, Jan.
- [58] B. Yuan, B. He, J. Yan, J. Jiang, Z. Wei, X. Shen, Short-term electricity consumption forecasting method based on empirical mode decomposition of long-short term memory network, *IOP Conf. Ser. Earth Environ. Sci.* 983 (1) (2022) 012004, <https://doi.org/10.1088/1755-1315/983/1/012004>, Feb.
- [59] C. C. Aggarwal and Others, *Data Mining: the Textbook*, 1, Springer, 2015.
- [60] P. Punyani, R. Gupta, A. Kumar, A multimodal biometric system using match score and decision level fusion, *Int. J. Inf. Technol.* 14 (2) (2022) 725–730, <https://doi.org/10.1007/s41870-021-00843-3>.
- [61] H. Vafaie, K.A. De Jong, Genetic Algorithms as a Tool for Feature Selection in Machine Learning, *ICTAI*, 2018, pp. 200–203.
- [62] I. Okpala, C. Nnaji, A.A. Karakhan, Utilizing emerging technologies for construction safety risk mitigation, *Pract. Period. Struct. Des. Constr.* 25 (2) (2020) 1–13, [https://doi.org/10.1061/\(asce\)sc.1943-5576.0000468](https://doi.org/10.1061/(asce)sc.1943-5576.0000468).
- [63] F.A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: continual prediction with LSTM, *Neural Comput.* 12 (10) (2020) 2451–2471.
- [64] M. Tami and A.Y. Owda, "Efficient commodity price forecasting using long short-term memory model," *Int. J. Artif. Intell.* ISSN, vol. 2252, no. 8938, p. 8938, <http://doi.org/10.11591/ijai.v13.i1.pp994-1004>.
- [65] W. Kong, Z. Dong, Y. Jia, D. Hill, Y. Xu, Short-term residential load forecasting based on LSTM recurrent neural network, *IEEE Trans. Smart Grid* 195 (2017) 841–851.
- [66] L. Burgueno, J. Cabot, S. Li, S. Gérard, A generic LSTM neural network architecture to infer heterogeneous model transformations, *Softw. Syst. Model.* 21 (1) (2022) 139–156, <https://doi.org/10.1007/s10270-021-00893-y>.
- [67] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT press, 2016.
- [68] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, *Appl. Energy* 236 (2019) 700–710.
- [69] L.G. Gannon, C.R. Marshall, A recurrent neural network model for structural response to underwater shock, *Ocean Eng.* 287 (2023) 115898, <https://doi.org/10.1016/j.oceaneng.2023.115898>.
- [70] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the Properties of Neural Machine Translation: Encoder-Decoder approaches, 12, *arXiv Prepr. arXiv1409.1259*, 2019, pp. 125–196.
- [71] D. Britz, Recurrent neural network tutorial, part 4 implementing a gru/lstm rnn with python and theano, *Inf. Syst. E-bus. Manag.* 256 (2015) 560–587.
- [72] M. Ravanelli, P. Brakel, M. Omologo, Y. Bengio, Light gated recurrent units for speech recognition, *IEEe Trans. Emerg. Top. Comput. Intell.* 2 (2) (2018) 92–102.
- [73] Y. Su, C.-C.J. Kuo, On extended long short-term memory and dependent bidirectional recurrent neural network, *Neurocomputing* 356 (2019) 151–161.
- [74] N. Gruber, A. Jockisch, Are GRU cells more specific and LSTM cells more sensitive in motive classification of text? *Front. Artif. Intell.* 3 (2020) 40.
- [75] I. Drosouli, A. Voulodimos, P. Mastorocostas, G. Miaoulis, D. Ghazanfarpour, A spatial-temporal graph convolutional recurrent network for transportation flow estimation, *Sensors* 23 (17) (2023) 7534.
- [76] M.S. Başarslan, F. Kayaalp, MBI-GRUMCONV: a novel Multi Bi-GRU and Multi CNN-Based deep learning model for social media sentiment analysis, *J. Cloud Comput.* 12 (1) (2023) 5.
- [77] G.-H. Kwak, C.-W. Park, H.-Y. Ahn, S.-I. Na, K.-D. Lee, N.-W. Park, Potential of bidirectional long short-term memory networks for crop classification with multitemporal remote sensing images, *Korean J. Remote Sens.* 36 (4) (2020) 515–525.
- [78] K.A. Althelaya, E.-S.M. El-Alfy, S. Mohammed, Evaluation of bidirectional LSTM for short-and long-term stock market prediction, in: 2018 9th international conference on information and communication systems (ICICS), 2018, pp. 151–156.
- [79] M.J. Hamayel, A.Y. Owda, A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms, *AI 2 (4)* (2021) 477–496.
- [80] K. Fukushima, Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biol. Cybern.* 36 (4) (1980) 193–202.
- [81] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [82] V.H. Phung, E.J. Rhee, A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets, *Appl. Sci.* 9 (21) (2019) 4500.
- [83] Y. Tian, Artificial intelligence image recognition method based on convolutional neural network algorithm, *IEEe Access.* 8 (2020) 125731–125744, <https://doi.org/10.1109/ACCESS.2020.3006097>.
- [84] M. Muthukumar, D. Mohan, M. Rajendran, Optimization of mix proportions of mineral aggregates using Box Behnken design of experiments, *Cem. Concr. Compos.* 25 (7) (2003) 751–758.
- [85] R. Taghizadeh-Mehrjardi, R. Neupane, K. Sood, S. Kumar, Artificial bee colony feature selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA, *Carbon. Manage* 8 (3) (2017) 277–291.
- [86] G. Biau, E. Scornet, A random forest guided tour, *Test* 25 (2016) 197–227.
- [87] A.M. Aburbeian, H.I. Ashqar, Credit card fraud detection using enhanced random forest classifier for imbalanced data, in: International Conference on Advances in Computing Research, 2023, pp. 605–616, [https://doi.org/10.1007/978-3-031-33743-7\\_48](https://doi.org/10.1007/978-3-031-33743-7_48).
- [88] A. Afzal, et al., Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows, *Aerosp. Sci. Technol.* 107 (2020) 106318, <https://doi.org/10.1016/j.ast.2020.106318>.
- [89] V. Plevris, G. Solorzano, N.P. Bakas, M.E.A. Ben Seghier, Investigation of performance metrics in regression analysis and machine learning-based prediction models, *IEEe Trans. Emerg. Top. Comput. Intell.* 13 (6) (2021) 1–40, <https://doi.org/10.23967/eccomas.2022.155>.



Mobarak Abumohsen: Computer system engineer in the Tubas Electricity Company. I received a B.S. in computer system engineering from Palestine Technical University, Tulkarem, Palestine, in 2018, and an M.S. in data science from Arab American University, Palestine, in 2023. He serves as a development engineer at the Tubas Electricity Company, in Palestine. His research focuses on machine learning, deep learning, and data visualization. He can be contacted using the following emails: [m.abumohsen@student.aau.edu](mailto:m.abumohsen@student.aau.edu) or [mobarak.daraghme96@gmail.com](mailto:mobarak.daraghme96@gmail.com).



Amani Yousef Owda: Assistant Professor in Computer Engineering and Data Science in the Faculty of Graduate Studies at the Arab American University in Palestine. She worked as a head of department of Natural, Engineering, and Technology Sciences in the Faculty of Graduate Studies at Arab American University from 2022 –2023. She worked as a research associate in the Faculty of Engineering at the University of Manchester from 2019 –2020. In addition, she worked in the School of Engineering at Manchester Metropolitan University from 2015 – 2019. She worked at Birzeit University from 2007–2011. She received her MSc. degree (Hons.) from The University of Manchester, UK in 2013, and her Ph.D. degree in Computer Engineering from Manchester Metropolitan University, UK in 2018. Since 2018, she leads research in multi-disciplinary fields with a focus on artificial intelligence, machine learning, decision support systems, image processing, medical applications of microwave and millimeter-wave imaging, security screening, and anomaly detection. She has published more than 43 articles in well reputable journals. She is a reviewer in many well-known Journals, and she is supervising MSc and PhD students. She can be contacted using the following emails: [Amani.Owda@aaup.edu](mailto:Amani.Owda@aaup.edu) or [amaniaubaha@gmail.com](mailto:amaniaubaha@gmail.com)



Majdi Owda: Associate Professor in Computer Science and Dean of Faculty in Data Science at the Arab American University in Palestine. In addition, he is a UNESCO Chair for Data Science for Sustainable Development. Worked as a head of the Department of Natural, Engineering, and Technology Sciences in the Faculty of Graduate Studies at Arab American University from 2020 –2022. Worked in the School of Computing, Mathematics, and Digital Technology at Manchester Metropolitan University from 2009 to 2020. He gained a BSc in Computer Science from the Arab American University in 2004, and an MSc by research in Computer Science with distinction from Manchester Metropolitan University in 2005 and a Ph.D. in Computer Science in 2011. His main research interests are AI

Techniques for Natural Language Interfaces to Relational Databases, Data Science, Conversational Informatics, Conversational Agents, Knowledge Trees, Knowledge Engineering, Planning, Information Extraction, AI Techniques for the Help of Society, Web/Data/Text Mining, Digital Forensics Processes and Frameworks, Digital Forensics Artefacts, Information Retrieval from Large Data Sources, Internet of Things Frameworks and Internet of Things Digital Forensics Artefacts and Security. He can be contacted using the following email: [Majdi.Owda@aaup.edu](mailto:Majdi.Owda@aaup.edu)



Ahmad Abumihsan: Received a B.S. degree in electronic engineering from Al-Quds University, Abu-Dis, Palestine, in 2013, and an M.S. degree in electrical engineering through a joint program from Birzeit University and Palestine Polytechnic University, Palestine, in 2021. He serves as a network engineer at the Higher Council for Youth and Sport in Ramallah, Palestine. His research focuses on SDN, computer network security, as well as machine learning, and deep learning. He can be contacted using the following email: [ahmaddaraghme413@gmail.com](mailto:ahmaddaraghme413@gmail.com)