INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 2 (2025) 164-176
https://publisher.uthm.edu.my/ojs/index.php/ijie

Enhancing Short-Term Solar Generation Forecasting: The Superiority of LSTM Over SVR

Hanis Nasuha Amer[1], Nofri Yenita Dahlan[1,3]*, Azlin Mohd Azmi[2,3], Mohd Fuad Abdul Latip[1]

[1] *School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, 40450, MALAYSIA*
[2] *School of Mechanical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, 40450, MALAYSIA*
[3] *Solar Research Institute (SRI), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, 40450, MALAYSIA*

*Corresponding Author: nofriyenita012@uitm.edu.my
DOI: https://doi.org/10.30880/ijie.2025.17.02.013

Article Info

Received: 14 August 2024
Accepted: 24 June 2025
Available online: 28 July 2025

Keywords

Forecasting, solar generation, time series, long short-term memory, support vector regression

Abstract

The global surge in solar power systems and Artificial Intelligence (AI)-driven technology has spurred the development of precise solar generation forecasts. These are particularly essential in Malaysia's tropical climate with abundant sunlight. Traditional forecasting methods like ARIMA, SARIMA, and ANN are inadequate in modelling solar energy systems' inherent complexity and non-linearity. This study addresses the challenge of forecasting solar generation in a tropical region. The objective is to develop a 30-minute-ahead solar generation forecasting model using advanced techniques, specifically Long Short-Term Memory (LSTM), and compare its performance with Support Vector Regression (SVR). Utilising data from February 2022 to April 2023 collected from a rooftop solar system in Pulau Pinang, Malaysia, this approach effectively handles sudden changes in solar output, known as "ramping events," caused by cloud movement and unpredictable weather. The results reveal LSTM's superiority, with an nRMSE of 6.75%, outperforming SVR (nRMSE of 7.28%). This pattern recognition capability of LSTM holds promise for larger datasets, offering precise forecasts beneficial for weather prediction and power management. Implementing this technique in more solar PV systems can enhance power reliability and promote sustainable energy practices, showcasing LSTM's potential for optimising solar generation forecasting.

1. Introduction

As the global energy demand continues to rise, the demand for renewable energy sources becomes more apparent. Solar energy arises as a promising alternative, especially in regions such as Malaysia, which is endowed with abundant daily direct sunlight due to its tropical climate. The tropical climate assures a continuous supply of sunlight, and accurate solar energy forecasts are key to unlocking sustainable power production and meeting the future's escalating energy demands. As conventional energy sources struggle to keep up with demand, solar power stands out as a renewable solution. Solar energy forecasting, equivalent to other renewable energy sources such as the wind speed forecasting model in the wind energy domain, plays an essential part in harnessing this ample

solar potential [1], [2]. The goal is to forecast solar generation using available solar resources. Solar energy forecasting is crucial for bridging the gap between solar power's potential and its effective utilization.

Classification of solar generation forecasting models can be based on the forecast horizon, which plays an essential role in aligning predictions with real-world applications [3]. Each forecast horizon has applications in significant sectors, such as managing load capacities, reserve allocation, operational security enhancement, and cost optimization. Short-term forecasting comprising the range of 30 minutes to 1 hour to 1 week is indispensable for operational precision and efficacy. Within this context, the ability of this study to forecast solar generation a half hour ahead of time arises as an important strength. This forecast enables immediate actions, including automatic generation control, unit dispatching, load balancing, and real-time plant management. These forecasts provide invaluable information for optimizing power smoothing and real-time power dispatching by reducing the forecast horizon to a half-hour interval. Accurate short-term forecasts have the potential to enhance unit commitment, load balancing, and overall system scheduling, thereby facilitating the seamless integration of solar energy into the power grid, as outlined in a prior comprehensive review [4].

Solar energy exhibits limitations arising from its reliance on external factors, despite its appearance as an abundant and readily available energy source [5]. The availability of solar energy varies based on circumstances such as the time of day, the cycle of the seasons, and the geographical area. In the context of Malaysia, solar energy has significant potential as a renewable energy source despite its limitations. For example, solar energy production is limited to daylight hours and can be hindered by cloudy or poor weather. In addition, the output power of photovoltaic (PV) solar systems has a non-linear relationship with meteorological conditions such as regional climate, wind pressure, humidity, solar radiation, ambient and PV module temperatures, and device efficiency. These constraints highlight the need for effective forecasting techniques, such as Long Short-Term Memory (LSTM), to optimize solar energy utilization and navigate its intermittency.

The framework for data mining is generally recognized as a methodology that employs multiple techniques to extract useful information from large amounts of data. Building upon the application of data mining techniques for solar radiation forecasting, this study effectively extends the same approach to the collected datasets, revealing valuable patterns and relationships to enhance solar generation forecasting performance [6]. In this study, the Pearson Correlation Coefficient (PCC) was employed as a feature selection technique, as prior research has shown its efficacy in enhancing the accuracy and robustness of the forecasting model compared to when PCC was not implemented [7], [8]. The PCC identified the most significant meteorological features correlated with solar generation, as using all meteorological features or inappropriate ones could compromise the accuracy of the models. This study's extensive data preprocessing journey includes data collection, data understanding, data visualization for insights, meticulous data cleaning, feature selection, normalizing data for consistency, constructing forecasting models using advanced techniques, and rigorously evaluating these models to ensure their accuracy for forecasting.

Alongside recent advancements in computer technology and machine learning, forecasting models for solar energy generation have incorporated AI algorithms with a continuous implementation strategy. Numerous authors have proposed various forecasting methods, each with its own conceptual design, implementation, and application based on the required tools and data, as seen in review articles [9]–[13]. Long Short-Term Memory (LSTM), a type of Recurrent Neural Network (RNN) under deep learning network, is being aggressively researched as a precise forecasting method in solar energy due to its memory and self-learning capabilities. LSTM can develop sequence-to-sequence mapping by adding cyclic connections between neurons, where information flows in varying directions.

Zhou et al. developed a hybrid ensemble deep learning model comprising two LSTM neural networks and an attention mechanism for forecasting short-term solar energy output [14]. Their method prioritizes significant input characteristics via the attention mechanism, thereby facilitating precise predictions. LSTM was used to extract long-term dependencies from solar power time series data, and their model ALSTM outperformed ARIMAX, MLP, and traditional LSTM. In contrast, Jebli et al. used machine learning and deep learning techniques to estimate solar energy using data from Errachidia, a semi-desert climate region in Morocco [7]. They used the Pearson Correlation Coefficient to select features and compare models, including LSTM, for real-time and short-term forecasts. Notably, LSTM demonstrated outstanding performance, outperforming other ML models. It is essential to note that this study differs from the work of Jebli et al. in that it concentrates on a tropical climate region where the abundance of solar energy poses unique forecasting challenges.

This study aimed to develop a 30-minute-ahead solar generation forecasting model capable of capturing the intricacies of solar energy fluctuations, particularly those heavily influenced by weather patterns. Traditional forecasting methods often fall short in capturing the complex temporal relationships and non-linearities inherent in dynamic systems like solar energy generation [15]. In contrast, our model, leveraging advanced techniques such as Long Short-Term Memory (LSTM) and compared against the Support Vector Regression (SVR) model, seeks to bridge this gap by effectively modeling the complexity of solar generation curves. Forecasting solar generation every 30 minutes allows for effective management of abrupt fluctuations known as "ramping events" caused by cloud movements or dynamic weather changes. The comprehensive data preprocessing journey exemplifies the

Int. Journal of Integrated Engineering Vol. 17 No. 2 (2025) p. 164-176

meticulous approach taken in this study, highlighting the dependability and efficacy of the resultant forecasting models through the application of data mining techniques. The superiority of LSTM over SVR is also demonstrated through a comparison of the two methods. The proposed forecasting model contributes to accurate predictions of renewable energy and supports sustainable renewable energy initiatives, especially in tropical regions such as Malaysia.

2. Research Method

This research evaluates two methodologies: LSTM and SVR. While these procedures have certain similarities, they differ in their approaches. The process of extracting information from patterns within large datasets by employing a variety of techniques commonly known as data mining, has gained significant attention in recent years due to its extensive application [16]. The initial phase involves data preprocessing, which entails removing incomplete or missing records from the database. The subsequent selection and transformation of data culminates in extracting a refined dataset. This dataset is then subjected to data mining procedures to extract patterns, which are then evaluated and analyzed to reveal knowledge. Knowledge discovery within diverse datasets is an iterative process characterized by forward or backward adjustments. Data mining has applications in solar energy for forecasting solar output power, solar irradiance, energy storage management, power optimization, identification of PV module faults, and control of solar PV panel installation [17].

The methodologies proposed include essential phases such as raw data collection, comprehensive data preprocessing, forecasting models and performance evaluation. It is essential to adhere to these stages to develop a reliable, effective, and accurate solar generation forecasting model. Both LSTM and SVR approaches begin with collecting unprocessed data, which is then carefully examined to determine if preprocessing is required. If the data is not readily applicable, preprocessing techniques are used to make it suitable for the needs. The methodologies require a subset of the dataset's parameters, necessitating preprocessing that includes feature selection. After the completion of preprocessing, the resulting dataset is polished and ready for applying forecasting algorithms. These algorithms make it possible to create distinctive models that can forecast solar output. Various evaluation metrics are used to determine the precision of both algorithms.

Fig. 1 illustrates the system architecture of the fundamental methodology employed by LSTM and SVR for solar generation forecasting. Each component of the system architecture is described in the following manner.

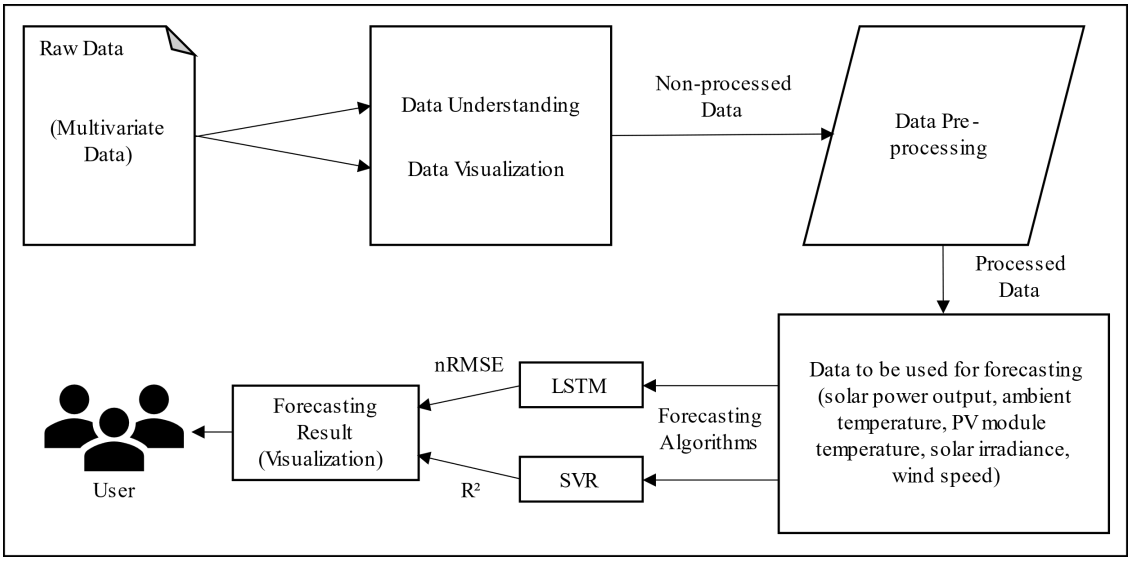


Fig. 1 *System architecture*

2.1 Data Collection

Unprocessed information that has not been extracted is referred to as raw data. It is often referred to as source data, which has not been altered by manual or algorithmic processes. The historical dataset utilized in this study was derived from the rooftop solar system on Pulau Pinang, a Malaysian island located off the northwest coast of Peninsular Malaysia, as shown in Fig. 2. This raw data set spans from January 1, 2022, to April 13, 2023, and contains 22,513 sets of 30-minute intervals. The historical solar output power (kW) data was collected by means of a submeter station within the case study setup. In contrast, meteorological data is collected from a nearby weather station, which is routinely monitored by the facility's department. This set contains seven important variables: total global horizontal irradiance (W/m^2), total horizontal irradiation (Wh/m^2), global irradiance on

the module plane (W/m^2), total slope irradiance (Wh/m^2), ambient temperature (°C), PV module temperature (°C), and wind speed (m/s).

Fig. 2 *UiTM Permatang Pauh, Pulau Pinang, Malaysia*

2.2 Data Pre-processing

A thorough data preprocessing phase is required prior to model development. This includes understanding the data, visualizing patterns, cleaning, selecting relevant features, and normalizing the data. Python-based plotting libraries were utilized for data visualisation, enabling the depiction of trends over time and the relationship between solar generation and other parameters. This visual analysis demonstrated temporal trends and the complex relationships between solar output and multiple features. Excluding data from 6:05 p.m. to 6:55 a.m. improved the performance of the model by eliminating undesirable data, such as incorrect or redundant values and average nighttime output power. Alternately, time series interpolation reduced the voids caused by absent values. Examining the time series plot following data cleaning ensured the reliability of the actual data post-preprocessing.

The primary objective of the preprocessing phase is to select significant data for input features in forecasting models. Feature selection has reduced input variables, resulting in shorter computation time. This phase eliminates unnecessary or duplicate data, enhancing the input quality for machine learning. More historical data and input variables improve model accuracy and increase model complexity. Consequently, determining the most suitable amount of input parameters is crucial, given their strong correlation with PV output power. The Pearson Correlation Coefficient (PCC) measures the influence of significant input variables on PV output power. PCC assesses linear dependence between random variables and is commonly employed in AI for feature selection [18]. Eq. (1) describes the linear correlation coefficient, a metric for measuring the linear correlation between two continuous variables.

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

Where $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$ represents the mean of x, $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} y_i$ represents the mean of y, and r_{xy} represents the coefficient. Eq. (1) ranges between -1 and 1, invariant by linear transformations of the two variables.

- $r_{xy} = 1$, x and y have a positive correlation.
- $r_{xy} = 0$, x and y have no linear relationship.
- $r_{xy} = -1$, x and y have a negative correlation.

During the learning phase of the model, parameters were regarded as acceptable or excluded based on threshold values of 0.5 or higher. The outcome was driven by the ranking of PCC values for all meteorological factors that influence solar generation.

2.3 Forecasting Algorithms

This phase entails the execution of appropriate forecasting algorithms. Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) are used as data forecasting methods in this work. Distinct forecasts are produced by various algorithms. The outcomes of these forecasted results are used in graphic data representations and evaluated in terms of the nRMSE and the R-squared values.

Int. Journal of Integrated Engineering Vol. 17 No. 2 (2025) p. 164-176

2.3.1 LSTM-based Methodology

Long Short-Term Memory (LSTM) is a prominent neural network architecture in deep learning designed to address challenges associated with temporal data sequences. LSTM, which was developed as an improvement to traditional Recurrent Neural Networks (RNNs), has received widespread acclaim for its proficiency in capturing long-range dependencies and temporal dynamics, as can be seen in a review paper by Nassif et al. [19], making it particularly appropriate for time series forecasting tasks such as solar generation forecasting. At its core, LSTM holds a unique memory cell that retains information over extended time steps, enabling it to effectively manage sequential data while mitigating the vanishing gradient problem encountered by conventional RNNs. The architecture includes several gates, including the input gate, output gate, and forget gate, each of which regulates the flow of data into and out of the memory cell. The gating logic architecture of an LSTM model, as described by Liu et al. [20], is depicted in Fig. 3, which consists of neural network layers activated by sigmoid and tanh activation functions.

This phase entails the execution of appropriate forecasting algorithms. Long Short-Term Memory (LSTM) and Support Vector Regression (SVR) are used as data forecasting methods in this work. Distinct forecasts are produced by various algorithms. The outcomes of these forecasted results are used in graphic data representations and evaluated in terms of the nRMSE and the R-squared values.

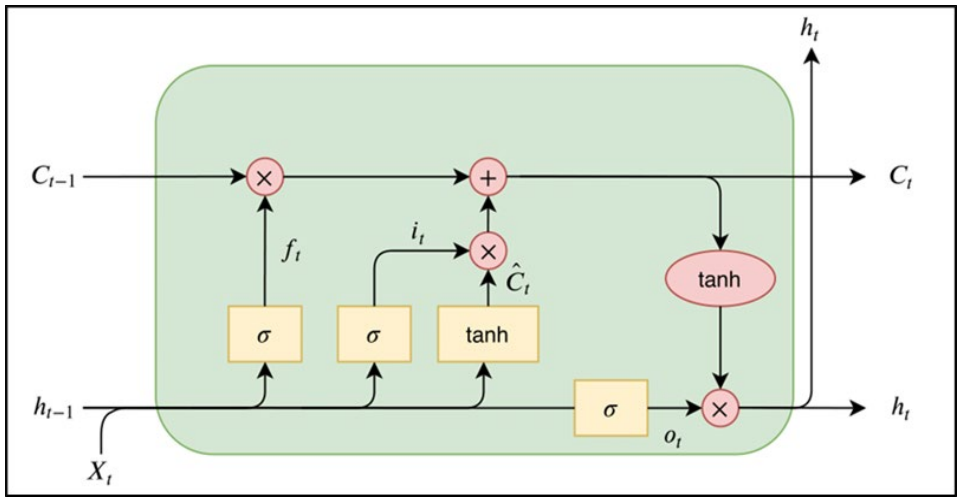


Fig. 3 *LSTM basic cell architecture*

This gate mechanism enables LSTM networks to determine which data should be retained and which can be discarded, thereby facilitating the modelling of complex temporal relationships within data sequences, and it has been proven in practice conducted by Hassan et al. [21]. Given the inherent variability in solar energy production due to varying weather conditions and diurnal cycles, the ability of LSTM to capture both short-term fluctuations and longer-term patterns in solar power output is particularly advantageous in the context of solar generation forecasting. This method's primary objective is pattern recognition over a lengthy amount of time, which is quite useful for forecasting or prediction approaches. Consequently, the error in prediction is determined by comparing the actual and forecasted values.

2.3.2 SVR-based Methodology

Support Vector Regression (SVR) is an adaptable machine learning algorithm that has risen to prominence in numerous forecasting tasks, including the prediction of solar generation, as proven in reviews written by Zendehboudi et al. [22]. SVR is based on support vector machine principles, originally developed for classification but has since been extended to regression problems. SVR functions by locating the optimal hyperplane in a high-dimensional space that best represents the relationship between the input features and the objective variable, which in this case is solar power output. This hyperplane, known as the decision boundary, attempts to minimize the difference between predicted and actual solar generation values, considering the epsilon-insensitive tube's tolerance margin.

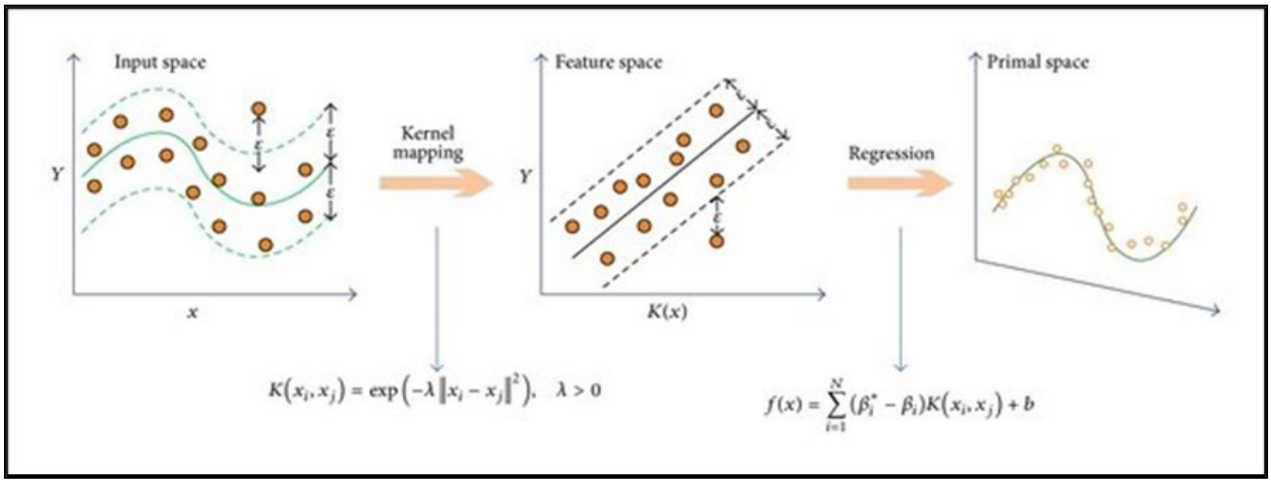


Fig. 4 *SVR architecture*

SVR's strength in the context of solar generation forecasting is its ability to capture nonlinear relationships between meteorological factors and solar energy generation, as illustrated in Fig. 4. SVR models utilize a kernel function that maps input features to a higher-dimensional space where non-linear patterns can be distinguished more effectively, as implemented by several authors in forecasting solar output power [23]. This quality is especially advantageous when dealing with the inherently nonlinear behaviour of solar power output, which is influenced by dynamic factors such as varying weather conditions and daily variations in sunlight. To optimize model performance, SVR-based methods for predicting solar generation frequently entail selecting appropriate kernel functions, tuning hyperparameters, and performing feature engineering. Although SVR does not inherently account for temporal dependencies like LSTM, it can still provide valuable insights into the non-linear relationships between input variables and solar generation, making it a powerful tool in solar energy forecasting.

2.4 Evaluation Metrics

In general, evaluating the effectiveness of a model involves calculating the difference between its forecasts and actual observations. Numerous evaluation metrics, such as normalized Root Mean Square Error (nRMSE) and R-squared (R^2), are frequently used to evaluate the outcomes of forecasting models. These evaluation metrics were also implemented by several authors in previous studies [24], [25]. The equations are as follows:

$$nRMSE = \frac{1}{P_C}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{pred} - x_{act})^2} \times 100\% \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_{pred} - x_{act})^2}{\sum_{i=1}^{N}(\bar{x}_{act} - x_{act})^2} \tag{3}$$

Where, x_{pred} represents the value predicted by the algorithm, x_{act} represents the actual value, $\bar{x}_{act}$ represents the average value of the actual solar generation data, P_C is the installation capacity of the corresponding rooftop solar PV system, and N is the number of samples. The normalized Root Mean Square Error (nRMSE) is calculated using Eq. (2). A lower nRMSE implies superior efficiency. The density of a forecasting model also influences its accuracy, as illustrated in Eq. (3). The number of errors varies with the density of the forecasted solar generation data. When the forecasted values are dense, the error between actual and predicted values is somewhat less. When the density is low, an inaccuracy becomes more visible. Thus, R-squared quantifies the correlation between the predicted and actual values. R-squared has a value between 0 and 1. A value closer to 0 implies relatively low accuracy, while a value closer to 1 implies greater forecasting model accuracy.

Int. Journal of Integrated Engineering Vol. 17 No. 2 (2025) p. 164-176

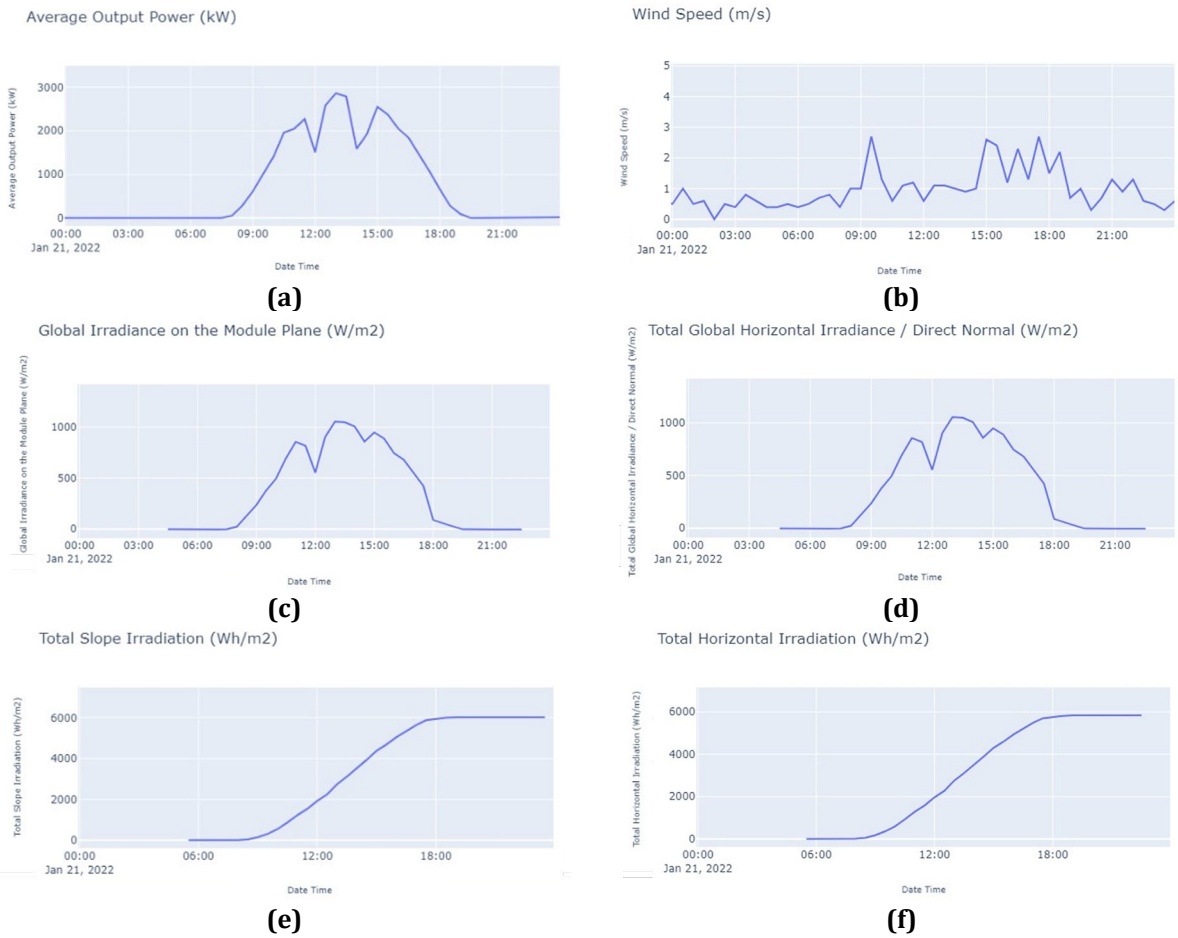
2.5 Data Visualization

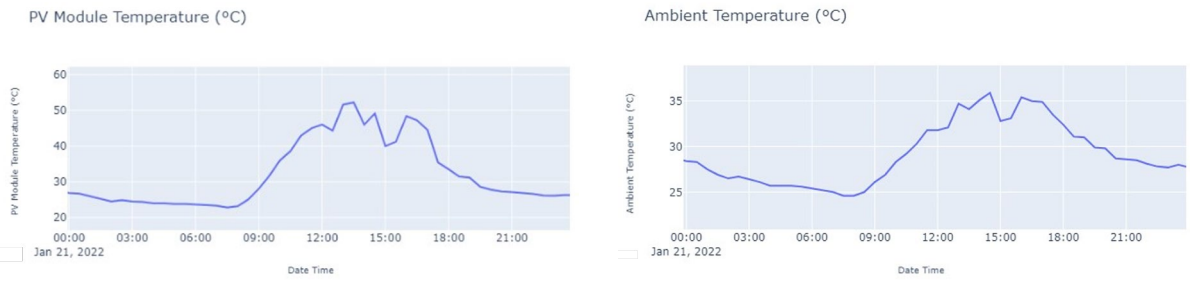
The analysis of acquired outcomes involves visualization. During this phase, the forecasted solar data is visualized, and the results presented in terms of nRMSE and R-squared are thoroughly evaluated to generate the forecasting plots.

3. Results and Discussion

3.1 Data Understanding

In this study, LSTM and SVR techniques were applied to the same dataset to forecast solar generation for a 30-minute horizon using LSTM and SVR, respectively. The dataset spanned one year and three months, with approximately 85% of the data (from February 12, 2022, to February 11, 2023) designated for training and validation and the remaining 15% (from February 12, 2023, to April 13, 2023) designated for testing. For interpretive purposes, Fig. 5 depicts the patterns of solar power generation and meteorological variables over the course of one day. As seen in Fig. 5, the peak generation of solar power is prominently observed between 1:00 PM and 2:00 PM. During this time frame, several meteorological variables exhibit concurrent peaks, suggesting a strong association with the peak generation moments. Specifically, total global horizontal irradiance, global irradiance on the module plane, ambient temperature, and PV module temperature all coincides with the peak in solar power generation. This synchronization of peaks indicates a significant correlation between these meteorological conditions and the optimal generation of solar energy, underlining the importance of favourable weather conditions and increased irradiance in driving peak solar power production. In contrast, total horizontal and total slope irradiance showed no significant correlations with solar PV production, while wind speed showed no correlations at all. This data understanding is essential to ensure that only relevant features are selected for the forecasting model to reduce errors.





(g) **(h)**

Fig. 5 *A day patterns of solar power generation and several meteorological variables (a) Average output power (kW); (b) Wind speed (m/s); (c) Global irradiance on the module plane (W/m^2); (d) Total global horizontal irradiance / direct normal (W/m^2); (e) Total slope irradiation (Wh/m^2); (f) Total horizontal irradiation (Wh/m^2); (g) PV module temperature (°C); (h) Ambient temperature (°C)*

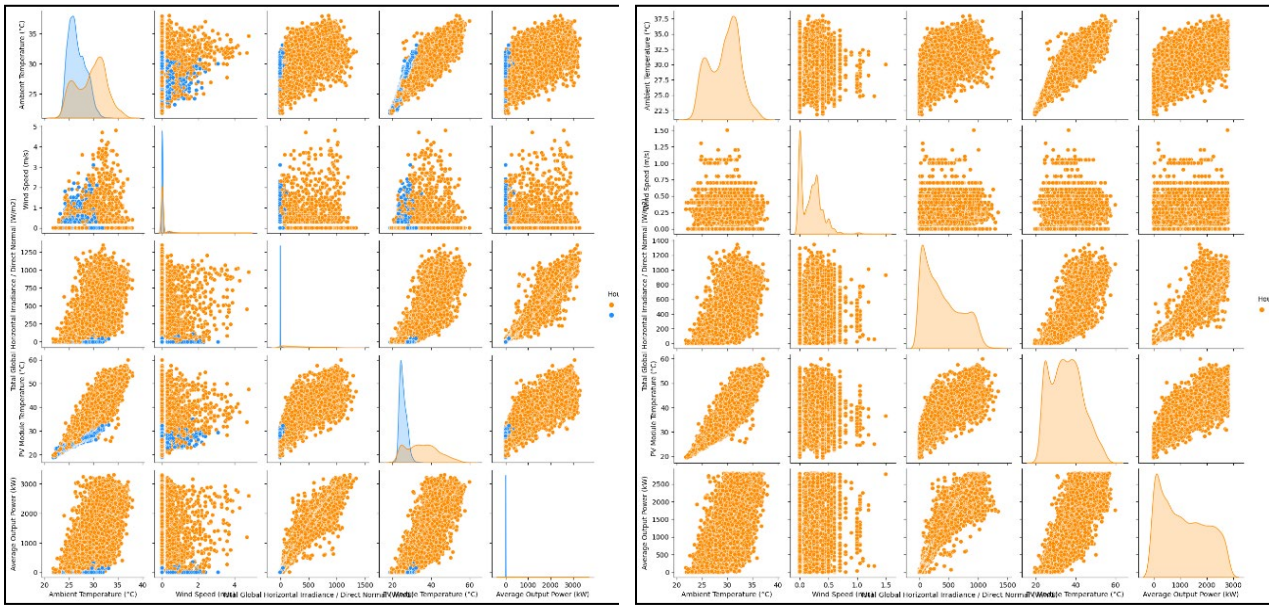
3.2 Data Pre-processing

Data cleaning is the process of removing unneeded information from a dataset, such as error values, missing values, redundant values, and outliers, which can affect the accuracy of a forecasting model. The deletion of rows containing null values, which represent data gaps. Fig. 6 shows the solar generation pattern across the day, helping filter out nighttime data to focus the forecast on daylight hours.

Fig. 6 *Solar PV generation patterns across the day*

According to the data, a significant portion of average output power was generated between 7:00 AM and 7:00 PM. Since nighttime hours do not contribute to solar PV generation forecasting, zero nighttime solar PV generation instances were excluded from the initial dataset. The data distribution is depicted in the figures below, separated into two discrete time intervals: Group 1 in orange represents data between 7:00 AM and 6:55 PM, whereas Group 2 in blue represents data between 7:00 PM and 6:55 AM. Fig. 7 (a) illustrates the dataset before any data cleaning performed, whereas Fig. 7 (b) shows the same dataset after data cleaning procedure completed. In the data cleaning process, nighttime data was deliberately excluded from consideration, as it holds no relevance for solar PV generation forecasting due to the absence of generation during those hours. Subsequently, any outliers, identified as data points falling outside the valid range for each variable, were systematically replaced with "NaN" values to ensure data integrity. Further, these "NaN" values, along with any missing data points, were meticulously addressed through a time-based interpolation method, effectively restoring completeness to the dataset while maintaining data coherence for subsequent analysis and modelling. The figure illustrates the data cleaning process, where removing outliers, as described, is crucial for enhancing the model's accuracy in forecasting solar generation.

Int. Journal of Integrated Engineering Vol. 17 No. 2 (2025) p. 164-176



(a) **(b)**

Fig. 7 *Data distribution of (a) Pre-data cleaning process; (b) Post-data cleaning process*

The selection of features is a crucial aspect of developing a forecasting model. Instead of relying solely on data visualization, quantitative analysis employing PCC values, as shown in Table 1, is used to evaluate the relationship between input and output variables.

Table 1 *PCC values between solar output and meteorological variables*

Input Variables	Pearson Correlation Coefficient
Total Global Horizontal Irradiance (W/m^2)	0.96
Total Horizontal Irradiation (Wh/m^2)	0.15
Global Irradiance on the Module Plane (W/m^2)	0.96
Total Slope Irradiation (Wh/m^2)	0.15
Ambient Temperature (°C)	0.69
PV Module Temperature (°C)	0.88
Wind Speed (m/s)	0.51

Total global horizontal irradiance and global irradiance on the module plane both have remarkable PCC values of 0.96 in relation to solar PV generation. However, these factors are mostly the same and only vary in how they are measured, which depends on whether the solar panels are placed horizontally or at an angle. As a result, only total global horizontal irradiance is selected as the input feature. The PCCs for the PV module temperature, the ambient temperature, and the wind speed are 0.88, 0.69, and 0.51, respectively. These variables, with PCC values equal to or exceeding the 0.5 threshold, emerge as the forecasting model's marginally significant input features. In contrast, the PCC between total horizontal irradiation and total slope irradiation for solar PV generation is 0.15, making it inappropriate as an input feature due to its near-zero value, indicating a weak linear relationship.

Considering these findings, the four main features that exhibit a significant correlation with solar PV generation are global horizontal irradiance (W/m^2), PV module temperature (°C), ambient temperature (°C), and wind speed (m/s).

3.3 Forecasting Result

Fig. 8 provides a detailed representation of the training and testing loss generated throughout the data training process as a crucial aspect of the LSTM methodology implementation. The observed learning curve demonstrates a well-fit trajectory, indicating a strong convergence between the training and testing loss, as monitored using Mean Absolute Error (MAE). This observation signifies the effectiveness of our model's learning process and its ability to generalize well to unseen data.

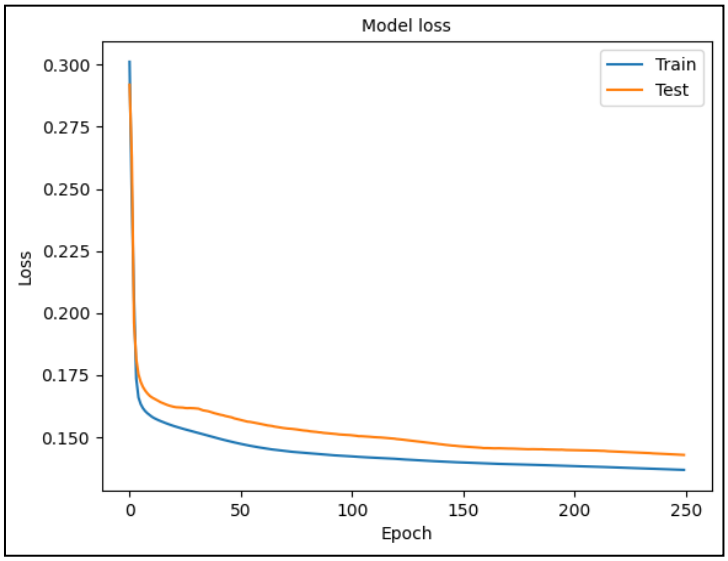


Fig. 8 *Learning curve*

Fig. 9 shows the comparison between the actual and predicted solar generation using the LSTM method on the testing dataset. In contrast, Fig. 10 displays the differences between actual and predicted solar generation when forecasting with the SVR algorithm on the same testing dataset. These graphs plainly depict irregular fluctuations in the predicted versus actual solar generation, which can be attributed to data occurrence variations during training and testing. Such variations are caused predominantly by complex algorithmic issues. Since both LSTM and SVR utilize the same datasets, the graphics depicting the forecasting results are remarkably comparable. Nonetheless, performance evaluation metrics, specifically the nRMSE and R-squared values, reveal differences between the two approaches.

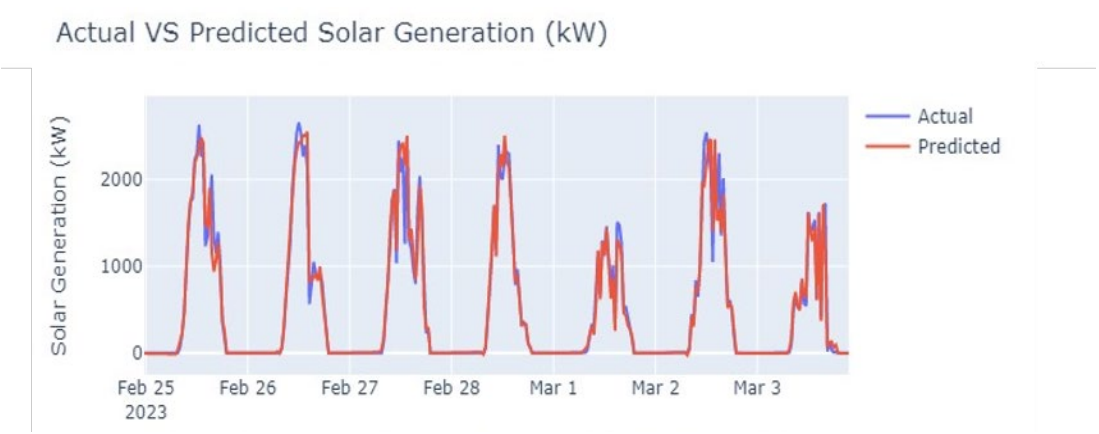


Fig. 9 *Actual and predicted testing dataset for LSTM-based solar generation forecasting*

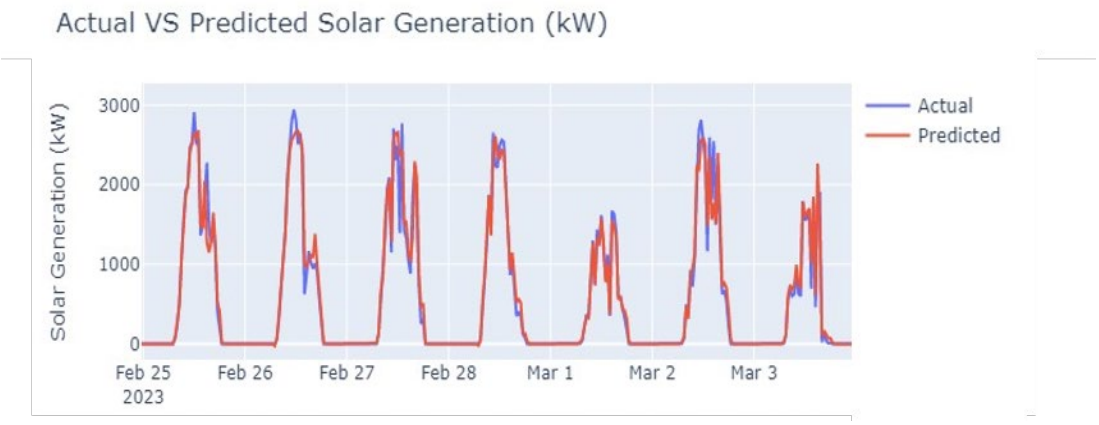


Fig. 10 *Actual and predicted testing dataset for SVR-based solar generation forecasting*

Int. Journal of Integrated Engineering Vol. 17 No. 2 (2025) p. 164-176

As indicated in Table 2, the performance evaluation of LSTM and SVR models is based on two key metrics: nRMSE and R^2. These metrics offer valuable insights into the accuracy of the forecasting models. LSTM demonstrates its effectiveness with a low nRMSE of 6.75%, signifying its ability to make precise predictions of solar generation. Additionally, LSTM achieves a high R^2 value of 0.957, indicating a strong correlation between its predictions and the actual solar generation data. Conversely, SVR presents a higher nRMSE at 7.28%, suggesting a comparatively larger prediction error. The SVR model, while showing good performance, obtains an R^2 value of 0.951, which is slightly lower than that of LSTM, highlighting a slightly weaker correlation between its predictions and the observed solar generation values. This discrepancy between the two models can be attributed to their inherent characteristics. LSTM, being a recurrent neural network, excels at capturing temporal dependencies and patterns within the data, allowing it to outperform SVR in forecasting solar energy generation.

To delve further into the impact of forecasting model parameters, it is crucial to recognize the significance of LSTM's architecture. The model's ability to remember and utilize past data points is influenced by hyper-parameters such as the number of LSTM units, the learning rate, and the sequence length. In the case of LSTM, a Bayesian optimization approach was employed to efficiently search for the optimal hyperparameter configuration, ensuring that the model could make accurate predictions and capture fluctuation in solar generation patterns due to shifting weather. In contrast, SVR's performance relies on parameter settings such as the choice of kernel function and the regularization parameter. A k-fold cross-validation method was applied in this work to assess and fine-tune these parameters, enabling the establishment of intricate relationships between input features and solar generation.

Table 2 *Evaluation metric results of LSTM and SVR*

Input Variables	Evaluation Metric	Value
LSTM-based	nRMSE (%)	6.750
	R^2	0.957
SVR-based	nRMSE (%)	7.280
	R^2	0.951

4. Conclusion

In contrast to previous works that primarily focus on forecasting models, this study integrates a comprehensive data mining framework to enhance forecasting accuracy. By applying advanced techniques such as the PCC for feature selection, this research not only improves the solar generation forecasting process but also ensures that the most relevant meteorological variables are considered, further distinguishing it from existing studies. On the other hand, it is undeniable that LSTM in terms of forecasting accuracy outperforms SVR based on a comprehensive comparison of the results for both approaches. LSTM has a lower error rate, with an nRMSE of 6.75%, making it a better choice for forecasting approaches than SVR, which has an nRMSE of 7.28%. LSTM's effectiveness is enhanced by its deep learning capabilities, which enable it to retain patterns for longer periods of time. The study highlights the importance of LSTM in forecasting methods, particularly for solar generation. The complex pattern recognition capability of LSTM holds great promise for larger datasets, providing highly precise forecasts that can benefit various organizations, including enhanced weather forecasting and balancing power generation and consumption. With an nRMSE of 6.75%, the results surpass expectations and are suitable for simulations in real time. Implementing this technique in more applicants' rooftop solar PV systems in Southeast Asia, particularly Malaysia, can lead to more reliable power output and contribute to sustainable energy practices. This research highlights the potential of advanced techniques such as LSTM for optimizing solar energy generation, effectively achieving the research objective of enhancing solar generation forecasting models.

Acknowledgement

We would like to thank the rooftop solar plant system facility for using their property as a study site. We also thank our colleagues and collaborators for their support and assistance throughout the research process.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Prof. Ir. Dr. Nofri Yenita Dahlan, Ts. Mohd Fuad Bin Abdul Latip;* ***data collection:*** *Hanis Nasuha Amer;* ***analysis and interpretation of results:*** *Hanis Nasuha Amer, Ts. Mohd Fuad Bin Abdul Latip, Prof. Ir. Dr. Nofri Yenita Dahlan;* ***draft manuscript preparation:*** *Assoc. Prof. Ir. Dr. Azlin Mohd Azmi, Hanis Nasuha Amer. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] D. Panjwani, S. Barhate, R. Rane, A. Pandey, and F. Kazi, "Short-Term Solar and Wind Generation Forecasting for the Western Region of India," *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, vol. 2021-November, 2021, doi: 10.1109/APPEEC50844.2021.9687683.

[2] L. P. Plant, M. Gao, J. Li, F. Hong, and D. Long, "Short-Term Forecasting of Power Production in a Large-Scale Photovoltaic Plant Based on LSTM," 2019.

[3] I. Kaaya, J. Ascencio-Vásquez, I. Kaaya, and J. Ascencio-Vásquez, "Photovoltaic Power Forecasting Methods," *Solar Radiation - Measurement, Modeling and Forecasting Techniques for Photovoltaic Solar Energy Applications*, Apr. 2021, doi: 10.5772/INTECHOPEN.97049.

[4] J. Ma and X. Ma, "A review of forecasting algorithms and energy management strategies for microgrids," *Systems Science and Control Engineering*, vol. 6, no. 1, pp. 237–248, Jan. 2018, doi: 10.1080/21642583.2018.1480979.

[5] Z. Usman, J. Tah, H. Abanda, and C. Nche, "A critical appraisal of pv-systems' performance," *Buildings*, vol. 10, no. 11, pp. 1–23, 2020, doi: 10.3390/buildings10110192.

[6] H. Mori and A. Takahashi, "A data mining method for selecting input variables for forecasting model of global solar radiation," *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, pp. 1–6, 2012, doi: 10.1109/TDC.2012.6281569.

[7] I. Jebli, F. Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, p. 120109, Jun. 2021, doi: 10.1016/J.ENERGY.2021.120109.

[8] H. Fraihat, A. A. Almbaideen, A. Al-Odienat, B. Al-Naami, R. De Fazio, and P. Visconti, "Solar Radiation Forecasting by Pearson Correlation Using LSTM Neural Network and ANFIS Method: Application in the West-Central Jordan," *Future Internet 2022, Vol. 14, Page 79*, vol. 14, no. 3, p. 79, Mar. 2022, doi: 10.3390/FI14030079.

[9] G. Alkhayat and R. Mehmood, "A review and taxonomy of wind and solar energy forecasting methods based on deep learning," *Energy and AI*, vol. 4, p. 100060, 2021, doi: 10.1016/j.egyai.2021.100060.

[10] R. A. Rajagukguk, R. A. A. Ramadhan, and H. J. Lee, "A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power," *Energies (Basel)*, vol. 13, no. 24, 2020, doi: 10.3390/en13246623.

[11] K. J. Iheanetu, "Solar Photovoltaic Power Forecasting: A Review," *Sustainability (Switzerland)*, vol. 14, no. 24, 2022, doi: 10.3390/su142417005.

[12] R. A. Rajagukguk, R. A. A. Ramadhan, and H. J. Lee, "A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power," *Energies (Basel)*, vol. 13, no. 24, 2020, doi: 10.3390/en13246623.

[13] D. K. Dhaked, S. Dadhich, and D. Birla, "Power output forecasting of solar photovoltaic plant using LSTM," *Green Energy and Intelligent Transportation*, vol. 2, no. 5, 2023, doi: 10.1016/j.geits.2023.100113.

[14] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-Term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019, doi: 10.1109/ACCESS.2019.2923006.

[15] Chibuzor Nkemdilim Obiora, "Deep Learning Approach to Forecasting Hourly Solar Irradiance ," 2020. Accessed: Jan. 27, 2023. [Online]. Available: https://www.proquest.com/pqdtglobal/docview/2458984059

[16] M. Yesilbudak, M. Colak, and R. Bayindir, "A review of data mining and solar power prediction," *2016 IEEE International Conference on Renewable Energy Research and Applications, ICRERA 2016*, vol. 0, pp. 1117–1121, 2016, doi: 10.1109/ICRERA.2016.7884507.

[17] S. Sumathi, *Introduction to data mining & its applications (Studies in computational intelligence, Vol. 29)*. 2006. [Online]. Available: http://www.lavoisier.fr/livre/notice.asp?id=2AKWO2AK2AOOWB

[18] Y. S. Kim, H. Y. Joo, J. W. Kim, S. Y. Jeong, and J. H. Moon, "Use of a big data analysis in regression of solar power generation on meteorological variables for a Korean solar power plant," *Applied Sciences (Switzerland)*, vol. 11, no. 4, pp. 1–10, 2021, doi: 10.3390/app11041776.

[19] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019, doi: 10.1109/ACCESS.2019.2896880.

[20] C. H. Liu, J. C. Gu, and M. T. Yang, "A Simplified LSTM Neural Networks for One Day-Ahead Solar Power Forecasting," *IEEE Access*, vol. 9, pp. 17174–17195, 2021, doi: 10.1109/ACCESS.2021.3053638.

[21] A. Hassan and A. Mahmood, "Deep Learning approach for sentiment analysis of short texts," in *2017 3rd International Conference on Control, Automation and Robotics, ICCAR 2017*, Institute of Electrical and Electronics Engineers Inc., Jun. 2017, pp. 705–710. doi: 10.1109/ICCAR.2017.7942788.

[22] A. Zendehboudi, M. A. Baseer, and R. Saidur, "Application of support vector machine models for forecasting solar and wind energy resources: A review," *J Clean Prod*, vol. 199, pp. 272–285, 2018, doi: 10.1016/j.jclepro.2018.07.164.

[23] K. P. Lin and P. F. Pai, "Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression," *J Clean Prod*, vol. 134, no. Part B, pp. 456–462, 2016, doi: 10.1016/j.jclepro.2015.08.099.

[24] R. J. Wai and P. X. Lai, "Design of Intelligent Solar PV Power Generation Forecasting Mechanism Combined with Weather Information under Lack of Real-Time Power Generation Data," *Energies (Basel)*, vol. 15, no. 10, 2022, doi: 10.3390/en15103838.

[25] J. Zhang *et al.*, "A suite of metrics for assessing the performance of solar power forecasting," *Solar Energy*, vol. 111, pp. 157–175, 2015, doi: 10.1016/j.solener.2014.10.016.