Biosystems Engineering and Renewable Energies

Journal homepage: https://bere.gau.ac.ir

GUASNR

ISAMEM

Original Research Paper

Smart modeling of photovoltaic energy production based on meteorological data and production capacity: Utilizing advanced machine learning algorithms

Mohammad Hajian, Tayyeb Nazghelichi*

Department of Biosystems Engineering, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran

Biosystems Engineering and Renewable Energies 2025, 1 (1): 70-74

KEYWORDS

Forecasting
Photovoltaic systems
Machine learning
Random forest algorithm
Support vector regression
Light gradient boosting regression

*** Corresponding author**
nazghelichi@gau.ac.ir

Article history
Received: 2025-01-11
Revised: 2025-03-12
Accepted: 2025-05-15

ABSTRACT

This study investigates the prediction of photovoltaic (PV) energy production using advanced machine learning algorithms, leveraging meteorological data and production capacity from 300 residential PV plants in Sydney, Australia. The dataset was processed into daily values to account for weather variability, and three machine learning models, i.e., random forest regression (RFR), support vector regression (SVR), and light gradient boosting regression (LightGBR), were implemented. Following rigorous preprocessing and hyperparameter optimization, LightGBR exhibited superior predictive performance, achieving a coefficient of determination (R^2) of 0.9020, a mean absolute error (MAE) of 3.1621, and a mean squared error (MSE) of 0.1005. Compared to previous studies, the optimized LightGBR model demonstrated enhanced accuracy in PV energy forecasting, underscoring its potential for improving predictive modeling in this domain. These findings have significant implications for optimizing energy distribution, enhancing smart grid integration, and supporting decision-making in energy management systems. Accurate forecasting of PV energy output is essential for improving operational efficiency, minimizing energy waste, and advancing sustainability objectives in renewable energy management.

1. Introduction

With the advancement of technology and the increase in population, energy consumption has significantly risen. According to reports published by British Petroleum (2023), the average energy consumption per person reached 57.6 kWh. The total global energy consumption in 2022 was reported as 64.04 EJ, showing a 1.08% increase compared to the 59.741 EJ of 2021. Fossil fuels account for 81.8%, renewable energy 7.5%, hydroelectric 6.7%, and nuclear 4%. The International Energy Agency reports indicate fossil fuels dominate global energy consumption, while non-carbon sources gradually increase. Despite the importance of fossil fuels in the global energy supply, these resources have several negative effects, including the production of greenhouse gases such as CO2, which are recognized as the primary cause of climate change (Karakurt and Aydin, 2023).

Concerns about climate change, environmental damage, and the limitations of renewable energy resources have made using renewable energy technologies, such as geothermal, solar, wind, and biofuels, increasingly attractive and rapidly expanding. In addition to reducing greenhouse gas emissions and offering limitless production sources, it can provide energy to rural and remote areas (Khezri et al., 2022). These technologies create extensive employment opportunities in production, installation, and regular consumption of energy (Giri et al., 2024).

Photovoltaic (PV) and solar thermal technologies are among the most widely adopted renewable energy solutions due to their ability to harness solar radiation with minimal environmental impact. PV systems, in particular, have evolved through multiple generations, from crystalline silicon cells to advanced multi-junction and nanomaterial-based designs, improving efficiency and cost-effectiveness. These systems can be categorized as centralized or decentralized, with decentralized rooftop PV installations gaining popularity for their scalability and direct energy consumption benefits. Additionally, concentrated solar power (CSP) systems offer dispatchable power generation by integrating thermal energy storage, making them suitable for large-scale applications in high-irradiance regions (Osman and Qureshi, 2025). One common application of PV systems is their rooftop installation, which has rapidly grown due to reduced installation costs and government incentive packages. These systems allow for direct consumption of the generated electricity, and they can also inject the produced electricity into the power grid to earn revenue. Therefore, rooftop PV systems are considered a suitable and economical solution for consumers (Le and Benjapolakul, 2019).

In recent years, organizations and various systems have been recorded. The increased attention to data mining and the use of recorded data stems from several factors: reduced data storage costs and easier data collection through networks, advancements in strong and efficient machine learning algorithms for data processing, and decreased computational costs that enable the use of complex computational methods for data analysis (Mitchell, 1999). Today, the installed capacity of PV systems is rapidly increasing, and many of these systems continuously publish their electricity production data. The use and access to these data are crucial for modeling PV networks or improving solar energy production predictions. Accurate energy production forecasting from renewable sources ensures better grid integration and reduces the fluctuations' impact. Solar power

© Published by Gorgan University of Agricultural Sciences and Natural Resources. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-ncnd/4.0/).

https://doi.org/10.22069/bere.2025.23125.1014

producers can achieve greater efficiency through precise forecasting (Bright et al., 2019).

Recent solar power forecasting studies have focused on developing advanced machine learning and deep learning models to improve prediction accuracy and reliability. According to Ahmed et al. (2024), a hybrid model incorporating support vector regression (SVR), random forest regression (RFR), gradient boosting regression (GBR), and multivariate adaptive regression splines (MARS) was developed for solar power prediction. The results showed that the hybrid-MARS model outperformed standalone models, with the MARS-SVR model demonstrating the highest reliability. In research by Tahir et al. (2024), Bayesian optimization and random search methods were utilized to enhance machine learning models such as artificial neural networks (ANN) and Gaussian process regression (GPR). This study highlighted the importance of hyperparameter optimization in improving model accuracy. Sulaiman et al. (2024) explored short-term solar power forecasting for rooftop systems, finding that the ANN model, outperformed other models such as long short-term memory (LSTM) and gated recurrent unit (GRU), leveraging data-driven features. According to Syauqi et al. (2024), combining physical and machine-learning models improved performance under both data-scarce and standard conditions. This model achieved increased prediction accuracy across diverse scenarios. Rao et al. (2024) utilized parallel BiLSTM networks combined with DNN to enhance prediction accuracy significantly. This approach achieved satisfactory results by mitigating errors associated with data quality. The hybrid model proposed by Zhao et al. (2024) integrated the whale optimization algorithm, variational mode decomposition, and SCINet, improving short-term solar power prediction performance under seasonal and variable conditions. Wang et al. (2024) developed a seasonal grey prediction model considering time lag and interactive effects, increasing forecasting accuracy compared to baseline models. According to Zhang et al. (2024), a deep reinforcement learning-based framework for solar power prediction reduced errors and enhanced model interpretability through key feature analysis.

Solar power forecasting faces several critical challenges. First, most existing studies primarily focus on large-scale PV plants, often overlooking the unique challenges associated with small, distributed solar systems (Jamil et al., 2023). These decentralized systems are more susceptible to local environmental variations and require more adaptive forecasting models. Second, many conventional approaches rely solely on either historical energy production data or meteorological variables, limiting their adaptability to dynamic climate conditions (Alaraj et al., 2021). This reliance reduces the robustness of predictions, especially in regions with highly variable weather patterns. Additionally, commonly used models such as ANNs, despite their widespread adoption, often suffer from computational inefficiencies and struggle to generalize across different seasonal variations (Wan et al., 2024). Given the high sensitivity of solar power generation to environmental factors, models that fail to explicitly incorporate seasonal dependencies may lack the accuracy required for short-term predictions.

To address these limitations, this study introduces a hybrid approach based on the LightGBM, a GBR model known for its high computational efficiency and predictive accuracy. Unlike traditional methods, the proposed model integrates both historical energy production data and meteorological variables, enhancing the adaptability of predictions under varying weather conditions. This study specifically focuses on 300 small-scale residential PV plants during the summer season, providing a comprehensive analysis of seasonal variations in energy generation. Furthermore, by optimizing feature selection and leveraging the scalability of LightGBR, the proposed approach improves model generalization while reducing computational costs. The findings of this research contribute to enhancing forecasting accuracy, integrating predictive models with smart grids, and optimizing solar energy management strategies, ultimately promoting the efficient utilization of solar resources in residential-scale applications. In this research, we will explore various machine learning algorithms based on meteorological data and solar panel production capacity to develop an optimal model for accurately predicting energy production while considering environmental changes. A comparative analysis of different regression algorithms, including LightGBR, SVR, and ensemble methods such as RFR, will be conducted to evaluate their performance and accuracy. Given the rapid expansion of PV technologies and the significance of precise energy forecasting for smart grids and energy distribution systems, the findings of this study could enhance operational decision-making in energy management and improve the efficiency of PV systems. Ultimately, this research aims to develop an intelligent model that not only ensures high accuracy in PV energy forecasting but also adapts to diverse environmental and climatic variations.

2. Materials and Methods

The data used in this study was extracted from the official IPART website. This dataset contains real-world PV energy production data. For the necessary analyses, the data for PV energy production during the summer was separated using the Pandas and NumPy libraries, and then the data was converted to daily values. Subsequently, the meteorological data for this period was integrated into the main dataset, obtained from Visual Crossing (VC), a specialized weather data and forecasting platform.

2.1. Dataset mapping

This study analyzed hourly data from a single season of PV energy production from 300 home PV power plants in Sydney. The PV systems used in this study consist of residential rooftop solar panels with varying generator capacities. The dataset was converted to daily data using the Pandas and NumPy libraries in Python. Additionally, meteorological data, including temperature, precipitation, cloud cover, solar radiation, solar energy, and UV index, corresponding to the recorded production dates, were incorporated into the dataset.

2.2. Methodology

The methodology of this study follows a structured process, as illustrated in Figure 1. The workflow begins with the collection of initial datasets, followed by data mapping and integration with meteorological data to incorporate relevant environmental factors. Next, data preprocessing is performed to normalize and clean the dataset. Various machine learning algorithms are then applied to predict PV energy generation. Finally, the best-performing algorithm is selected based on accuracy and error analysis, leading to the final modeling phase. Figure 1 visually represents the step-by-step procedure followed in this research.

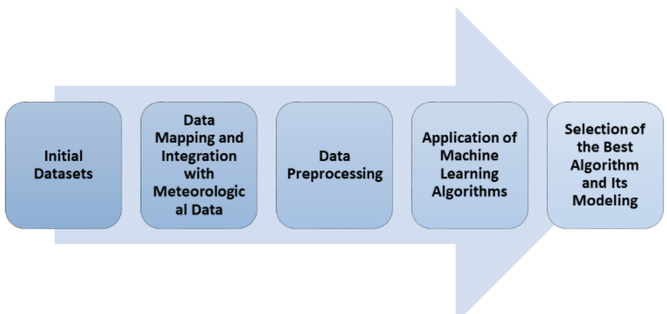

Figure 1. Steps of the PV energy production forecasting model

2.3. Data preprocessing

This study used the StandardScaler for data preprocessing, a common method for scaling data. This tool, available in the Scikit-Learn library in Python, transforms the data so that the mean of each feature is shifted to zero, and its variance is scaled to one. The criteria used in this method are the mean and standard deviation, such that the mean of each feature becomes zero and its standard deviation equals one (Sharma, 2022). The standardization formula is as Eq. (1)

$$z = ((\mathrm{x} - \mu))/\sigma \tag{1}$$

where x is the original data value, μ is the mean, and σ is the standard deviation. After the standardization process, the machine learning models were trained with the preprocessed data. This process enhances the final model's performance and results in better predictions.

2.4. Application of machine learning algorithms

In this study, three machine learning algorithms, i.e., SVR, RFR, and LightGBR, were used for forecasting PV energy production. RFR is an ensemble learning algorithm that combines multiple decision trees to make a final prediction. Each decision tree is independently trained on random data samples, and its predictions are combined using majority voting. The decision trees form the base of the RFR, which uses ensemble learning techniques to improve accuracy and reduce the likelihood of overfitting (Smith et al., 2013). Key parameters for the RFR include: the number of trees (n_estimators), maximum depth (max_depth), and minimum number of samples required to split a node (min_samples_split).

The SVR utilizes the concept of support vectors to solve regression problems. It aims to find a function with the least error within a defined margin while minimizing the model's complexity. Its objective is to find a function that maximizes the margin between the support vectors and the data points outside the margin (Das et al., 2017). Key parameters for the SVR include: the regularization parameter that controls the trade-off between error tolerance and model complexity (C), kernel type that transforms data into a feature space, and epsilon (ε), which is the maximum error allowed within the margin. Common kernel types include linear, polynomial, radial basis function (RBF), and sigmoid.

LightGBR is a boosting algorithm based on decision trees, known for its high speed and accuracy. The algorithm incrementally improves decision trees using boosting techniques. LightGBR uses boosting techniques to refine base models and utilizes decision tree optimization methods, resulting in high speed and accuracy. It is particularly popular due to its efficiency in processing large and complex datasets (Fan et al., 2019). Key parameters for the LightGBR algorithm include: the number of boosting iterations (n_estimators), learning rate (learning_rate), and maximum number of leaves each tree can have (num_leaves).

The following ranges of parameters were tested for each model. For the RFR: n_estimators: [50, 100, 150, 200], max_depth: [5, 10, 15, 20], and min_samples_split: [2, 5, 10, 20], for the SVR: C: [0.1, 1, 10], kernel type: ['linear', 'poly', 'rbf'], and epsilon: [0.01, 0.1, 0.2], and for LightGBR: n_estimators: [50, 100, 200, 300], learning_rate: [0.05, 0.1, 0.2], and num_leaves: [10, 15, 20, 25]. Grid search with cross-validation was used to optimize the model parameters.

2.5. Evaluation criteria

To assess the performance of the forecasting models, as well as in grid search to find optimal machines, three common evaluation metrics were used: mean squared error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2). These metrics provide insights into the accuracy and reliability of the predictions by quantifying the error between predicted and actual values. MSE measures the average squared difference between actual and predicted values. A lower MSE value indicates better model performance, as it penalizes larger errors more significantly. It is defined as Eq. (2)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

where $\hat{y}$ represents the actual values, $\hat{y}_i$ is the predicted value, and n is the total number of observations. MAE calculates the average absolute difference between actual and predicted values. Unlike MSE, it does not square the errors, making it less sensitive to outliers. It is defined as Eq. (3).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3}$$

Lower MAE values indicate a model with fewer overall prediction errors. The R^2 measures the proportion of variance in the actual values that the model explains. It ranges from negative values (poor model performance) to 1 (perfect prediction). It is defined as Eq. (4)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

where $\bar{y}$ represents the mean of the actual values. A higher R^2 score indicates a better fit between the model and the data. To select the best algorithm and optimize parameters, the 'joblib' library was used to store the optimal models.

3. Results and Discussion

In this study, by adjusting various parameters in the machine learning algorithms and comparing their performance, the model with the highest accuracy and the lowest error was selected for each algorithm and subsequently compared with other algorithms. As shown in Table 1, the optimal values of the maximum depth of the tree, the minimum samples for splitting, and the number of trees for the RFR model are 10, 2, 100, respectively. Upon training the model and applying it for predictions, it is observed that, as depicted in Figure 2, the comparison chart between the predicted and actual data aligns well with the trend line (y = x). This alignment indicates the model's high accuracy in forecasting values. However, the dispersion of some points from the trend line suggests the presence of some error in the predictions, which could be improved through alternative methods.

According to Table 2, the optimal values of regularization parameter, the maximum allowable error, and the kernel type for the SVR model were determined to be 10, 0.01, and 'rbf', respectively. After training the model and predicting the values, the comparison chart in Figure 3 displays the real and predicted data. Although the predicted values are close to the trend line (y = x) in most cases, indicating the model's accuracy, the greater dispersion of points compared to the Random Forest regression model suggests a relatively weaker performance of this model.

As shown in Table 3, the optimal values for learning rate, the number of boosting iterations, and the maximum number of leaves for the LightGBR were identified as 0.1, 200, 15, respectively. After training the model and predicting the values, the comparison chart in Figure 4 shows the real and predicted data. Unlike the previous models, the predicted values in this model are more consistently aligned with the trend line (y = x), indicating higher accuracy and lower error. This better alignment suggests that the LightGBR outperforms the SVR and RFR models, providing more accurate prediction results.

For PV energy production modeling and prediction, the joblib library was used, and the LightGBR model was identified as the superior model. Future PV energy production can be predicted and controlled using meteorological data, including production capacity, temperature, precipitation, cloud cover, solar radiation, solar energy, UV index, and production capacity data. This approach facilitates the management and optimization of solar energy production, significantly improving prediction accuracy and error reduction in this field. Thus, this model can be utilized to enhance the efficiency of solar energy production systems and optimize energy resources.

Table 1. Results of the RFR model

Parameters	MSE	MAE	R^2
MD=10, MSS=2, NE=100 *	1.1383	3.4441	0.8933
MD=5, MSS=5, NE=50	1.2307	4.1025	0.8754
MD=15, MSS=20, NE=150	1.2024	3.8123	0.8841
MD=20, MSS=10, NE=200	1.2156	3.9502	0.8789

* MD represents the maximum depth of the tree, MSS stands for the minimum samples for splitting, and NE refers to the number of trees.

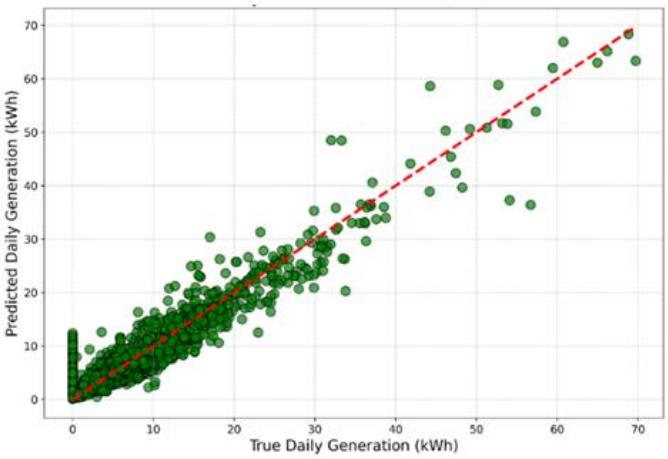

Figure 2. Comparison between actual and predicted data using the RFR

Table 2. Results of the SVR model

Parameters	MSE	MAE	R^2
C=10, E=0.01, K=rbf *	1.2477	4.6715	0.8553
C=1, E=0.1, K=rbf	1.3085	5.4012	0.8123
C=0.1, E=0.2, K=poly	1.2689	4.9204	0.7431
C=0.1, E=0.01, K=rbf	1.2593	4.8950	0.8262

* C represents the regularization parameter, E stands for the maximum allowable error, and K refers to the kernel type.

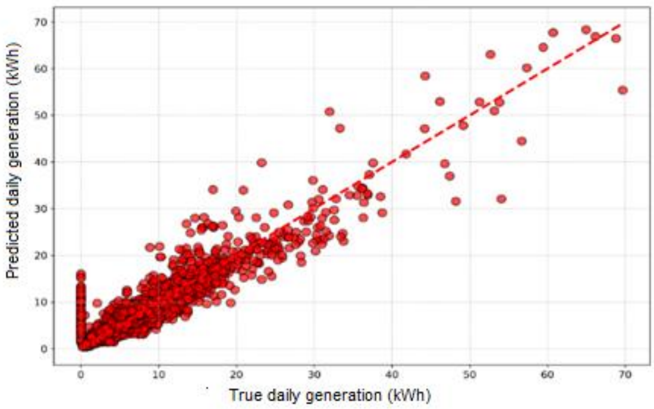

Figure 3. Comparison between actual and predicted data using the SVR

Table 3. Results of the LightGBR model

Parameters	MSE	MAE	R^2
LR=0.1, NE=200, NL=15 *	0.1005	3.1621	0.9020
LR=0.05, NE=100, NL=10	0.8874	3.5567	0.8874
LR=0.2, NE=300, NL=25	1.1453	3.4015	0.8948
LR=0.1, NE=50, NL=20	1.1302	3.3044	0.8992

* LR represents the learning rate, NE stands for the number of boosting iterations, and NL refers to the maximum number of leaves.

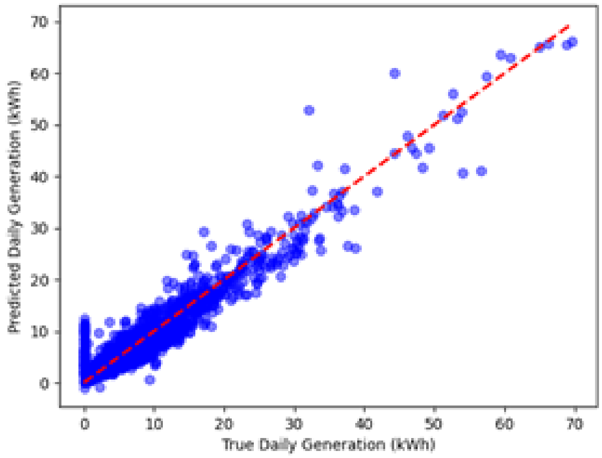

Figure 4. Comparison between actual and predicted data using the LightGBR

Table ***4*** presents a summary of previous studies on energy consumption forecasting using various machine learning models. The table includes different models along with their forecasting errors, allowing for a comparative assessment of their performance. Lima et al. employed an ANN for energy consumption forecasting. Their results indicated that the model achieved an MSE of 20.95, an RMSE of 29.48, and an R^2 of 0.88. While this model provided reasonably accurate predictions, the error values remained relatively high. Olatomiwa et al. utilized the SVM for energy consumption forecasting. Their evaluation showed an RMSE of 1.8661 and an R^2 of 0.7280. Although the model demonstrated a fair level of accuracy, the relatively low R^2 value suggests limitations in capturing the variability of actual data. Amrouche et al. applied a combination of spatial modeling and ANNs to forecast energy consumption. Their study reported an MSE of 16.4593 and an RMSE of 33.10. While the incorporation of neural networks enhanced predictive performance, the relatively high RMSE indicates that the model exhibited lower accuracy in certain instances. Javier Huertas Tato et al. employed the RFR algorithm for energy consumption forecasting. The model's performance was assessed with an R^2 of 0.752 and an RMSE of 26.94. These results suggest that RF provided a moderate level of accuracy, though the RMSE remains relatively high. Christophe Paoli et al. implemented a multilayer perceptron (MLP) ANN for energy consumption forecasting. Their findings indicated an R^2 of 0.801, an RMSE of 3.59, and an MAE of 2.65. These results suggest that the MLP model performed well compared to some other approaches, yet further optimization could potentially reduce prediction errors.

Table 4. Comparison of energy production forecasting models based on error metrics

Reference	Model	Forecasting error
Lima et al.	ANN	MSE = 20.95, RMSE = 29.48, and R^2 = 0.88
Olatomiwa et al.	Support Vector Machine	RMSE = 1.8661, R^2 = 0.7280
Amrouche et al.	Spatial modeling and Artificial Neural Networks	MSE = 16.4593, RMSE = 33.10
Javier Huertas Tato	RF	R^2 = 0.752, and RMSE = 26.94
Paoli et al.	MLP	R^2 = 0.801, RMSE = 3.59, and MAE = 2.65
This work	LighGBR	R^2 = 0.902, MSE = 0.101, and MAE = 3.162

4. Conclusion

This study focused on predicting PV energy production using machine learning models, incorporating weather data and the production capacity of residential solar systems. The results show that among the tested models, LightGBR performed the best, achieving an R^2 of 0.9020, MAE of 3.1621, and MSE of 0.1005, making it a highly effective method for solar energy forecasting. These findings are important for renewable energy management, optimizing energy distribution, and integrating smart grids, as accurate predictions can improve grid stability, reduce energy waste, and enhance decision-making in energy markets. Additionally, this research supports sustainability goals and contributes to the efficient utilization of solar resources. Future studies are recommended to explore deep learning models and more comprehensive meteorological datasets to further improve forecasting accuracy.

References

Ahmed, S. I., Bhuiyan, K. A., Rahman, I., Salehfar, H., & Selvaraj, D. F. (2024). Reliability of regression based hybrid machine learning models for the prediction of solar photovoltaics power generation. *Energy Reports*, *12*, 5009–5023.

Alaraj, M., Kumar, A., Alsaidan, I., Rizwan, M., & Jamil, M. (2021). Energy Production Forecasting from Solar Photovoltaic Plants Based on Meteorological Parameters for Qassim Region, Saudi Arabia. *IEEE Access*, *9*, 83241–83251.

Bright, J. M., Killinger, S., & Engerer, N. A. (2019). Data article: Distributed PV power data for three cities in Australia. Journal of Renewable and *Sustainable Energy*, *11*(3).

Das, U. K., Tey, K. S., Seyedmahmoudian, M., Idna Idris, M. Y., Mekhilef, S., Horan, B., & Stojcevski, A. (2017). SVR-based model to forecast PV power generation under differentweather conditions. *Energies*, *10*(7), 1–17.

Fan, J., Ma, X., Wu, L., Zhang, F., Yu, X., & Zeng, W. (2019). Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agricultural Water Management*, *225*, 105758.

Giri, P., Paul, S., & Debnath, B. K. (2024). A fuzzy Graph Theory and Matrix Approach (fuzzy GTMA) to select the best renewable energy alternative in India. *Applied Energy*, *358*, 122582.

Jamil, I., Lucheng, H., Iqbal, S., Aurangzaib, M., Jamil, R., Kotb, H., Alkuhayli, A., & AboRas, K. M. (2023). Predictive evaluation of solar energy variables for a large-scale solar power plant based on triple deep learning forecast models. *Alexandria Engineering Journal*, *76*, 51–73.

Karakurt, I., & Aydin, G. (2023). Development of regression models to forecast the CO_2 emissions from fossil fuels in the BRICS and MINT countries. *Energy*, *263*, 125650.

Khezri, M., Karimi, M. S., Mamkhezri, J., Ghazal, R., & Blank, L. (2022). Assessing the impact of selected determinants on renewable energy sources in the electricity mix: The case of ASEAN countries. *Energies*, *15*(13), 4604.

Le, N. T., & Benjapolakul, W. (2019). Comparative electrical energy yield performance of micro-inverter pv systems using a machine learning approach based on a mixed-effect model of real datasets. *IEEE Access*, *7*, 175126–175134.

Mitchell, T. M. (1999). Machine learning and data mining. *Communications of the ACM*, *42*(11), 30–36.

Osman, M., & Qureshi, I. (2025). Review of photovoltaic and concentrated solar technologies including their performance, reliability, efficiency and storage. *Results in Engineering*, *25*, 104424.

Paoli, C., Voyant, C., Muselli, M., & Nivet, M. L. (2010). Forecasting of preprocessed daily solar radiation time series using neural networks. *Solar Energy*, *84*(12), 2146–2160.

Rao, Z., Yang, Z., Li, J., Li, L., & Wan, S. (2024). Prediction of photovoltaic power generation based on parallel bidirectional long short-term memory networks. *Energy Reports*, *12*, 3620–3629.

Sharma, V. (2022). A study on data scaling methods for machine learning. *International Journal For Global Academic & Scientific Research*, *1*(1).

Smith, P. F., Ganesh, S., & Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of Neuroscience Methods*, *220*(1), 85–91.

Sulaiman, M. H., Jadin, M. S., Mustaffa, Z., Daniyal, H., & Mohd Azlan, M. N. (2024). Short-term forecasting of rooftop retrofitted photovoltaic power generation using machine learning. *Journal of Building Engineering*, *94*, 109948.

Syauqi, A., Pavian Eldi, G., Andika, R., & Lim, H. (2024). Reducing data requirement for accurate photovoltaic power prediction using hybrid machine learning-physical model on diverse dataset. *Solar Energy*, *279*, 112814.

Tahir, M. F., Yousaf, M. Z., Tzes, A., El Moursi, M. S., & El-Fouly, T. H. M. (2024). Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization. Renewable and *Sustainable Energy Reviews*, *200*, 114581.

Visual Crossing. (n.d.). Weather Query Builder - Sydney. Retrieved March 11, 2025, from https://www.visualcrossing.com/weather-query-builder/sydney/?v=wiza

Wan, M. J., Phuang, Z. X., Hoy, Z. X., Dahlan, N. Y., Azmi, A. M., & Woon, K. S. (2024). Forecasting meteorological impacts on the environmental sustainability of a large-scale solar plant via artificial intelligence-based life cycle assessment. *Science of the Total Environment*, *912*, 168779.

Wang, J., Ye, L., Ding, X., & Dang, Y. (2024). A novel seasonal grey prediction model with time-lag and interactive effects for forecasting the photovoltaic power generation. *Energy*, *304*, 131939.

Zhang, R., Bu, S., Zhou, M., Li, G., Zhan, B., & Zhang, Z. (2024). Deep reinforcement learning based interpretable photovoltaic power prediction framework. *Sustainable Energy Technologies and Assessments*, *67*, 103830.

Zhao, Y., Peng, X., Tu, T., Li, Z., Yan, P., & Li, C. (2024). WOA-VMD-SCINet: Hybrid model for accurate prediction of ultra-short-term Photovoltaic generation power considering seasonal variations. *Energy Reports*, *12*, 3470–3487.