

Machine Learning Based Solar Energy Prediction for Smart Grid Stability

Gunjal Sahil Dattatray¹, Prof. Y. G. Suryavanshi², Dr. H. B. Jadhav³

^{1,2,3}Department of Computer Engineering

Adsul's Technical Campus, Chas, Ahilyanagar, Maharashtra, India

¹sahilgunjal750@gmail.com, ²yogitasuryavanshi2018@gmail.com, ³hbm3577@gmail.com

Abstract: The increasing integration of solar photovoltaic (PV) systems into modern power grids has created a need for accurate solar energy prediction to support efficient energy management and grid stability. Solar power generation is highly dependent on changing weather conditions and the position of the sun, making reliable forecasting a challenging task. This paper presents a machine-learning-based approach for short-term solar energy prediction using historical PV generation data, meteorological parameters, and solar-geometric features. The study considers variables such as temperature, relative humidity, cloud cover, zenith angle, azimuth angle, and angle of incidence to improve prediction accuracy. Several machine-learning algorithms are considered, including Multiple Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, Support Vector Regression, Random Forest, AdaBoost, Bagging, and Gradient Boosting Regressor (GBR). Model performance is evaluated using the coefficient of determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The reported results indicate that Gradient Boosting Regressor achieved the strongest performance in the main model comparison, with an R^2 of 0.827, RMSE of 399.44, MSE of 158559.33, and MAE of 253.62. Hyperparameter tuning using grid search and five-fold cross-validation further reduced the reported test RMSE from 425.20 to 406.47. The findings highlight the importance of combining weather information with solar-geometric features for solar power forecasting. The proposed framework can support electricity generation scheduling, energy storage coordination, and operational planning in smart-grid environments. However, further evaluation using time-based data and multiple geographical locations is required before practical deployment.

Keywords: Solar Energy Prediction, Machine Learning, Photovoltaic Power Generation, Smart Grid Stability, Gradient Boosting Regressor, Feature Engineering, Renewable Energy Forecasting, Hyperparameter Tuning

I. INTRODUCTION

The increasing demand for electricity and the growing need for clean and sustainable energy sources have encouraged the widespread adoption of renewable energy technologies. Among these technologies, solar photovoltaic (PV) systems have become an important source of electricity because they convert sunlight directly into electrical energy. Solar energy is widely considered a promising alternative to conventional energy sources because it is renewable and can contribute to reducing dependence on fossil fuels. However, the integration of solar power into modern electricity networks presents several operational challenges, particularly because its generation varies throughout the day and depends on changing environmental conditions. Therefore, accurate solar energy prediction has become an important requirement for efficient power system operation and smart-grid management.

Solar photovoltaic power generation is influenced by several factors, including solar radiation, temperature, relative humidity, cloud cover, atmospheric conditions, and the position of the sun. Changes in these parameters can produce variations in the amount of electricity generated by a PV installation. For example, increased cloud cover may reduce



the sunlight reaching solar panels, while changes in the sun's position affect the angle at which solar radiation reaches the panel surface. These variations make solar power generation less predictable than conventional electricity generation under certain operating conditions. Consequently, electricity grid operators require reliable forecasting methods to estimate the expected solar output and prepare for changes in renewable energy availability.

Smart grids represent an important development in modern electricity infrastructure because they support improved monitoring, communication, energy management, and coordination between electricity generation and consumption. The increasing integration of renewable energy sources into smart grids requires better forecasting techniques to manage fluctuations in power production. When solar generation differs significantly from the expected output, grid operators may need to adjust other generation sources, manage energy storage systems, or modify electricity distribution schedules. Reliable short-term solar energy prediction can provide useful information for these decisions and help reduce uncertainty in renewable energy management.

II. PROBLEM STATEMENT

Solar photovoltaic power generation varies with weather conditions, solar radiation, and the position of the sun, making accurate prediction difficult. These fluctuations create challenges in balancing electricity generation and demand, managing energy storage, and maintaining reliable smart-grid operations. Therefore, an effective machine-learning-based prediction system is required to analyze historical solar power generation and meteorological data, identify complex patterns, and provide reliable short-term solar energy forecasts to support efficient energy management and smart-grid stability.

III. OBJECTIVES

1. To develop a machine-learning framework for solar energy prediction.
2. To use meteorological and solar-geometric information as model inputs.
3. To compare classical regression and ensemble models using a common evaluation process.
4. To tune the strongest model to improve prediction accuracy.
5. To examine how solar energy forecasting can support smart-grid energy management and stability.

IV. LITERATURE SURVEY

1. A. R. V. Babu, N. Bharath Kumar, R. P. Narasipuram, S. Periyannan, A. Hosseinpour and A. Flah (2025)

Babu et al. (2025) presented a study titled "Solar Energy Forecasting Using Machine Learning Techniques for Enhanced Grid Stability." The research focused on the application of machine-learning techniques to predict solar energy generation and improve the management of renewable energy in power grids. The study highlighted the importance of accurate forecasting for addressing variations in solar power output and supporting reliable grid operations. Their work provides a relevant foundation for developing predictive models that can assist electricity generation planning, energy management, and smart-grid stability.

2. G. Tziolis, A. Livera, J. Montes-Romero, S. Theocharides, G. Makrides and G. E. Georghiou (2023)

Tziolis et al. (2023) proposed a machine-learning-based approach for direct short-term net-load forecasting in solar-integrated microgrids. Their research addressed the forecasting challenges that arise when solar energy generation is combined with electricity demand in small-scale power systems. The study emphasized the importance of accurate net-load prediction for balancing electricity supply and demand. Their work is relevant to smart-grid energy management because reliable forecasts can help operators plan generation, coordinate available resources, and manage fluctuations associated with renewable energy integration.

3. M. Mohamed, F. E. Mahmood, M. A. Abd, A. Chandra and B. Singh (2022)

Mohamed et al. (2022) investigated dynamic forecasting of solar energy microgrid systems using feature engineering. Their research focused on improving solar energy prediction by selecting and preparing relevant input variables for forecasting models. Feature engineering plays an important role in identifying relationships between environmental



conditions and photovoltaic power generation. The study is relevant to the present research because meteorological and solar-geometric variables, such as temperature, humidity, cloud cover, and solar position, can provide useful information for estimating solar output and supporting microgrid energy management.

4. U. K. Das et al. (2018)

Das et al. (2018) presented a review of photovoltaic power generation forecasting and model optimization in Renewable and Sustainable Energy Reviews. The researchers examined forecasting approaches and the importance of selecting suitable models for predicting solar power output. Their work highlighted the role of model optimization and performance evaluation in improving forecasting quality. The review provides a useful background for comparing different regression and machine-learning algorithms. It supports the need to evaluate models using standard statistical measures and select an appropriate forecasting method for renewable energy applications. Source: Reference [4] in the supplied PDF.

5. Cyril Voyant, Gilles Notton, Soteris A. Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte and Alexis Fouilloy (2017)

Voyant et al. (2017) conducted a comprehensive review titled “Machine Learning Methods for Solar Radiation Forecasting: A Review.” The researchers discussed machine-learning approaches used for predicting solar radiation, including artificial neural networks, support vector regression, regression trees, random forests, and ensemble techniques. Their study explained that forecasting performance can vary depending on the dataset, prediction interval, forecasting horizon, and evaluation metrics. The review also highlighted the potential benefits of combining multiple prediction methods. This research provides a theoretical foundation for comparing machine-learning algorithms for solar energy prediction.

Comparison Table

Author & Year	Method Used	Advantages	Limitations
A. R. V. Babu et al. (2025)	Machine-learning techniques for solar energy forecasting	Supports renewable energy forecasting and smart-grid management.	Detailed limitations are not specified in the supplied PDF.
G. Tziolis et al. (2023)	Machine learning for short-term net-load forecasting in solar-integrated microgrids	Helps predict net load and supports electricity supply-demand balancing.	Detailed limitations are not specified in the supplied PDF.
M. Mohamed et al. (2022)	Dynamic solar energy forecasting using feature engineering	Improves the representation of relevant inputs for solar forecasting.	Detailed limitations are not specified in the supplied PDF.
U. K. Das et al. (2018)	Review of PV power forecasting and model optimization	Provides an overview of forecasting approaches and model optimization.	A review does not itself establish the performance of one specific model on the current dataset.
C. Voyant et al. (2017)	Review of machine-learning methods for solar radiation forecasting	Explains different machine-learning approaches for solar forecasting.	Forecasting performance depends on data, prediction horizon, and evaluation method.



V. WORKING OF SYSTEM

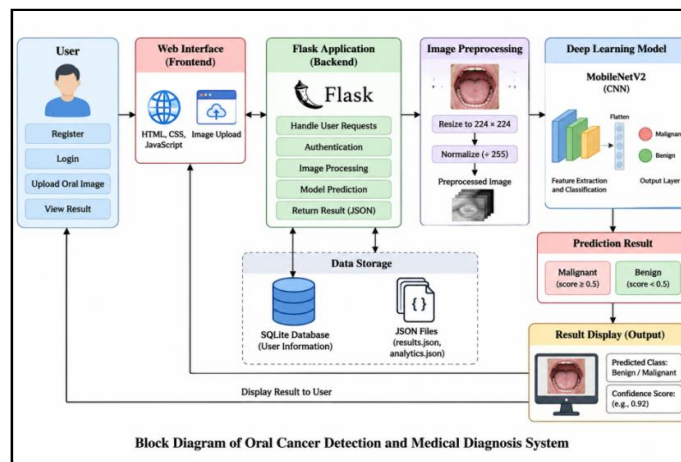


Fig 1: Design of the system

The working of the proposed system, “Machine Learning Based Solar Energy Prediction for Smart Grid Stability,” involves collecting historical solar photovoltaic (PV) generation data and weather-related information, processing the data, training machine-learning models, and predicting solar power output. The system compares different regression and ensemble learning algorithms and evaluates their performance using standard statistical metrics. The following steps explain the working process based on the provided PDF.

1. Data Collection

The system begins by collecting historical solar PV power generation data and associated meteorological information from an open-source Solar Energy Power Generation Dataset. The data includes temperature, relative humidity, cloud cover, and other weather-related variables. These observations provide the necessary information for understanding the relationship between environmental conditions and solar electricity generation.

2. Data Preprocessing and Cleaning

After data collection, the available observations are cleaned and prepared for model training. Missing or noisy entries are treated, and the historical PV generation records are aligned with the corresponding weather observations. This process helps maintain data consistency and prepares the dataset for reliable analysis.

3. Feature Engineering

In this stage, relevant input features are selected or calculated to improve solar energy prediction. Along with meteorological variables, the system considers solar-geometric features such as zenith angle, azimuth angle, and angle of incidence. These variables describe the position of the sun and the direction of incoming sunlight relative to the photovoltaic system.

4. Training and Testing Data Division

The prepared dataset is divided into training and testing subsets. The training data is used to develop the machine-learning models, while the testing data is used to evaluate their predictive performance. This separation helps assess how accurately the models predict solar output for observations not directly used during training.

5. Machine-Learning Model Training

The system trains and compares several machine-learning algorithms, including Multiple Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, Support Vector Regression, Random Forest, AdaBoost, Bagging, and Gradient Boosting Regressor. Each algorithm learns relationships between the input variables and solar power output. Comparing these models helps identify the approaches with better reported prediction performance.



6. Model Evaluation

The trained models are evaluated using four statistical measures: coefficient of determination ((R^2)), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics measure the model's ability to explain variations in solar output and quantify prediction errors. According to the PDF, Gradient Boosting Regressor achieved the strongest reported performance in the main model comparison.

7. Hyperparameter Tuning

The Gradient Boosting Regressor is further optimized using grid search and five-fold cross-validation. The reported configuration uses a learning rate of 0.05, a maximum tree depth of 5, 150 estimators, and a subsample value of 0.8. The source reports that the test RMSE decreased from 425.20 for the default model to 406.47 after tuning, although the tuned model's reported (R^2) was 0.819.

VI. SYSTEM DESIGN

The system design of “Machine Learning Based Solar Energy Prediction for Smart Grid Stability” describes the overall architecture and the sequence of operations used to predict solar photovoltaic (PV) power generation. The proposed system combines historical solar generation data, meteorological parameters, solar-geometric features, machine-learning algorithms, and performance evaluation techniques. The predicted solar output can then support smart-grid energy management, generation scheduling, and storage coordination. The design follows the workflow presented in the supplied PDF.

Solar_Energy_Prediction_Email_Updated.pdf

1. System Architecture

The system architecture consists of data collection, preprocessing, feature engineering, model training, model evaluation, hyperparameter tuning, and solar energy prediction. Historical PV generation and weather information are used as inputs. After data preparation, multiple machine-learning algorithms are trained and compared to identify the strongest reported model. The resulting prediction is intended to provide useful information to the smart-grid energy-management layer.

2. Data Collection Module

The data collection module obtains historical PV power generation and related meteorological information from the open-source dataset described in the PDF. The input variables include temperature, relative humidity, cloud cover, and pressure- or irradiance-related weather information. These data help the system identify patterns between environmental conditions and solar power generation.

3. Data Preprocessing Module

The preprocessing module prepares the collected data for machine-learning analysis. It involves cleaning the dataset, handling missing or noisy entries, and aligning the PV generation records with the corresponding weather observations. Proper data preparation improves consistency and helps prevent data-quality problems from affecting model development.

4. Feature Engineering Module

The feature engineering module selects and prepares the variables required for solar energy prediction. In addition to meteorological parameters, it incorporates solar-geometric features such as zenith angle, azimuth angle, and angle of incidence. These features represent the sun's position relative to the PV installation and help the model learn relationships between solar geometry, weather conditions, and generated power.

5. Machine-Learning Model Module

This module trains and compares nine algorithms: Multiple Linear Regression, Ridge Regression, Lasso Regression, Decision Tree Regression, Support Vector Regression, Random Forest Regression, AdaBoost Regressor, Bagging Regressor, and Gradient Boosting Regressor. Linear models provide baseline predictions, while tree-based and ensemble methods can capture nonlinear relationships between the input variables and PV output.

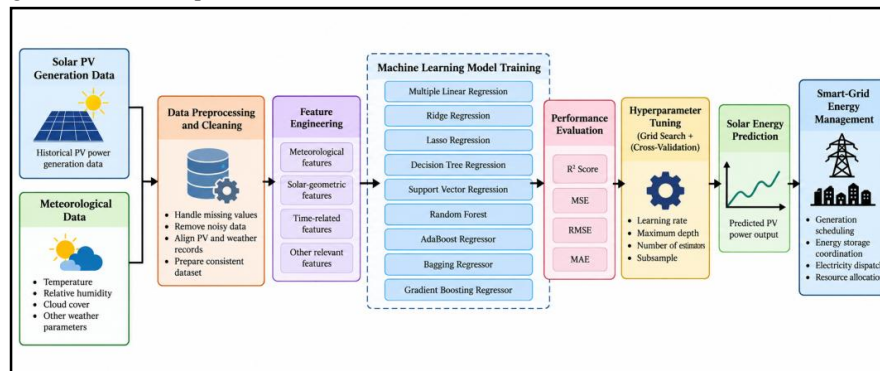


6. Model Evaluation Module

The evaluation module measures the predictive performance of each model using (R^2) , MSE, RMSE, and MAE. These metrics help determine how closely predicted values match actual solar power observations. According to the supplied PDF, Gradient Boosting Regressor achieved the strongest performance in the main model comparison, with $(R^2=0.827)$, RMSE of 399.44, MSE of 158559.33, and MAE of 253.62.

7. Overall System Design Workflow

The overall design follows this sequence:



VII. RESULTS

The results of the study on “Machine Learning Based Solar Energy Prediction for Smart Grid Stability” demonstrate the effectiveness of machine-learning techniques in predicting solar photovoltaic (PV) power generation. The study compares multiple regression and ensemble learning algorithms using historical PV generation data, meteorological parameters, and solar-geometric features. The performance of the models is evaluated using the coefficient of determination ((R^2)), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). According to the supplied PDF, the Gradient Boosting Regressor (GBR) achieved the strongest reported performance in the main model comparison.

Solar_Energy_Prediction_Email_Updated.pdf

1. Performance Comparison of Machine-Learning Models

Sr. No.	Machine-Learning Model	Reported (R^2)
1	Multiple Linear Regression	≈ 0.67
2	Ridge Regression	≈ 0.73
3	Lasso Regression	≈ 0.73
4	Decision Tree Regression	≈ 0.76
5	Support Vector Regression	≈ 0.39
6	Gradient Boosting Regressor	0.827

Table 1: Reported model performance comparison.

The results indicate that Gradient Boosting Regressor achieved the highest reported (R^2) among the models with numerical results in the source comparison. Random Forest, AdaBoost, and Bagging were also included in the study, but their individual numerical results were not provided in the main comparison table.

Solar_Energy_Prediction_Email_Updated.pdf

2. Performance Metrics of Gradient Boosting Regressor

Sr. No.	Evaluation Metric	Reported Value
1	(R^2) Score	0.827
2	Mean Squared Error (MSE)	158559.33



3	Root Mean Squared Error (RMSE)	399.44
4	Mean Absolute Error (MAE)	253.62

Table 2: Reported performance of GBR in the main model comparison.

The reported (R^2) value of 0.827 indicates that the model explains approximately 82.7% of the variance in the target values for the evaluated dataset. The RMSE and MAE represent the magnitude of prediction errors, while MSE penalizes larger errors more heavily. These values are specific to the source study's evaluation and depend on the dataset and the scale of the target variable.

3. Results of Hyperparameter Tuning

The Gradient Boosting Regressor was further optimized using grid search and five-fold cross-validation. The reported results are summarized below.

Parameter	Reported Value
Learning Rate	0.05
Maximum Tree Depth	5
Number of Estimators	150
Subsample	0.8
Default Model Test RMSE	425.20
Tuned Model Test RMSE	406.47
Tuned Model MAE	261.25
Tuned Model (R^2)	0.819
Cross-Validation RMSE	(422.12 ± 16.92)

Table 3: Reported hyperparameter tuning results.

The test RMSE decreased from 425.20 to 406.47 after tuning, a reduction of approximately 4.41%. This indicates an improvement in the reported test error. However, the tuned model's reported (R^2) was 0.819, compared with 0.827 in the main model comparison; these figures represent separate reported evaluations and should not be treated as identical test results.

Solar_Energy_Prediction_Email_Updated.pdf

4. Important Features Identified

The feature-importance analysis reported the following important predictors of solar energy generation:

- Angle of incidence: Represents the angle at which sunlight reaches the PV panel.
- Azimuth angle: Represents the horizontal direction of the sun.
- Zenith angle: Represents the sun's angular position relative to the vertical.
- Cloud cover: Influences the sunlight available for electricity generation.
- Relative humidity: Provides information about atmospheric conditions.
- Shortwave radiation: Represents incoming shortwave solar energy.

The source study identifies angle of incidence, azimuth, and zenith as leading predictors. Their importance highlights the value of combining solar-geometric features with meteorological data when developing solar energy forecasting models.

5. Application to Smart-Grid Energy Management

The predicted solar output can support smart-grid energy-management activities, including:

1. Generation scheduling based on expected solar power availability.
2. Energy storage coordination to help manage variations in renewable generation.
3. Electricity dispatch and resource allocation.
4. Operational planning for solar-integrated microgrids.
5. Potential integration with SCADA systems and microgrid dashboards.



These are intended applications of the forecasting framework. The supplied study does not report a live-grid deployment or directly measured improvement in grid stability.

6. Overall Result

The study reports that Gradient Boosting Regressor achieved the strongest performance in the main model comparison, with an (R^2) of 0.827 and an RMSE of 399.44. Hyperparameter tuning further reduced the reported test RMSE to 406.47 from the default model's 425.20. The results support the potential of combining meteorological information, solar-geometric features, and machine-learning techniques for short-term solar energy prediction. Further testing with time-based data and multiple geographical locations is necessary to establish the model's reliability under real-world operating conditions.

VIII. CONCLUSION

The study on “Machine Learning Based Solar Energy Prediction for Smart Grid Stability” demonstrates the potential of machine-learning techniques for predicting solar photovoltaic power generation using historical generation data, meteorological parameters, and solar-geometric features. Among the evaluated models, Gradient Boosting Regressor achieved the strongest reported performance in the main comparison, with an (R^2) value of 0.827, RMSE of 399.44, MSE of 158559.33, and MAE of 253.62. Hyperparameter tuning further reduced the reported test RMSE from 425.20 to 406.47. The findings highlight the importance of weather conditions, solar position, and model optimization in solar energy forecasting. The predicted output can support generation scheduling, energy storage coordination, and smart-grid energy management. However, further testing across different geographical locations and time periods is required to establish the model's reliability for practical deployment and its potential contribution to grid stability.

IX. FUTURE SCOPE

The future scope of Machine Learning Based Solar Energy Prediction for Smart Grid Stability includes improving forecasting accuracy through the integration of additional data sources, advanced machine-learning algorithms, and real-time environmental information. Future research can incorporate numerical weather prediction data, IoT-based sensors, and weather APIs to update solar energy forecasts as atmospheric conditions change. Advanced techniques such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), ARIMA, and Exponential Smoothing (ETS) can be explored and compared with the existing Gradient Boosting Regressor to identify suitable forecasting approaches. The system can also be extended to handle larger datasets and evaluate prediction performance across different geographical locations, climatic conditions, and seasons. Time-based validation can be introduced to provide a more realistic assessment of forecasting performance under practical operating conditions. Furthermore, integration with Supervisory Control and Data Acquisition (SCADA) systems and microgrid control dashboards could help grid operators use predicted solar output for generation scheduling, energy storage coordination, electricity dispatch, and resource allocation. Digital-twin technology may also be explored to simulate smart-grid behaviour and evaluate forecasting strategies before implementation in real environments.

REFERENCES

1. Babu, A. R. V., Kumar, N. B., Narasipuram, R. P., Periyannan, S., Hosseinpour, A., & Flah, A. (2025). “Solar Energy Forecasting Using Machine Learning Techniques for Enhanced Grid Stability.” IEEE Access, 13. <https://doi.org/10.1109/ACCESS.2025.3574093>
2. Tziolis, G., Livera, A., Montes-Romero, J., Theocharides, S., Makrides, G., & Georghiou, G. E. (2023). “Direct Short-Term Net Load Forecasting Based on Machine Learning Principles for Solar-Integrated Microgrids.” IEEE Access, 11, 102038–102049.
3. Mohamed, M., Mahmood, F. E., Abd, M. A., Chandra, A., & Singh, B. (2022). “Dynamic Forecasting of Solar Energy Microgrid Systems Using Feature Engineering.” IEEE Transactions on Industry Applications, 58(6), 7857–7869.



4. Das, U. K., et al. (2018). "Forecasting of Photovoltaic Power Generation and Model Optimization: A Review." *Renewable and Sustainable Energy Reviews*, 81, 912–928.
5. Voyant, C., Notton, G., Kalogirou, S. A., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). "Machine Learning Methods for Solar Radiation Forecasting: A Review." *Renewable Energy*, 105, 569–582. <https://doi.org/10.1016/j.renene.2016.12.095>
6. Gaboitaolelwe, J., Zungeru, A. M., Yahya, A., Lebekwe, C. K., Vinod, D. N., & Salau, A. O. (2023). "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison." *IEEE Access*, 11, 40820–40845. <https://doi.org/10.1109/ACCESS.2023.3270041>.
7. Mathur, K., & Mittal, S. (2022). "A Review on Forecasting the Photovoltaic Power Using Machine Learning." *Journal of Physics: Conference Series*, 2286, 012010. <https://doi.org/10.1088/1742-6596/2286/1/012010>
8. Tian, J., Ooka, R., & Lee, D. (2023). "Multi-Scale Solar Radiation and Photovoltaic Power Forecasting with Machine Learning Algorithms in Urban Environment: A State-of-the-Art Review." *Journal of Cleaner Production*, 426, 139040. <https://doi.org/10.1016/j.jclepro.2023.139040>.
9. Voyant, C., Muselli, M., Paoli, C., & Nivet, M.-L. (2012). "Numerical Weather Prediction (NWP) and Hybrid ARMA/ANN Model to Predict Global Radiation." *Energy*, 39(1), 341–355.
10. Mellit, A., & Pavan, A. M. (2010). "A 24-h Forecast of Solar Irradiance Using Artificial Neural Network: Application for Performance Prediction of a Grid-Connected PV Plant at Trieste, Italy." *Solar Energy*, 84(5), 807–821.
11. Pedro, H. T. C., & Coimbra, C. F. M. (2012). "Assessment of Forecasting Techniques for Solar Power Production with No Exogenous Inputs." *Solar Energy*, 86(7), 2017–2028.
12. Inman, R. H., Pedro, H. T. C., & Coimbra, C. F. M. (2013). "Solar Forecasting Methods for Renewable Energy Integration." *Progress in Energy and Combustion Science*, 39(6), 535–576.
13. Diagne, M., David, M., Lauret, P., Boland, J., & Schmutz, N. (2013). "Review of Solar Irradiance Forecasting Methods and a Proposition for Small-Scale Insular Grids." *Renewable and Sustainable Energy Reviews*, 27, 65–76.
14. Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F. J., & Antonanzas-Torres, F. (2016). "Review of Photovoltaic Power Forecasting." *Solar Energy*, 136, 78–111.
15. Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., & Fouilloy, A. (2017). "Machine Learning Methods for Solar Radiation Forecasting: A Review." *Renewable Energy*, 105, 569–582.

