

Machine Learning-Based Prediction of Solar Power Generation Using Weather Data

Fatima khalid¹, Maryam Shabbir², Saeed Rasheed³

¹ Department of Physics, University of Agriculture, Faisalabad, Pakistan,
Email: fatimakhalidch22@gmail.com

² Department of Physics, University of Agriculture, Faisalabad, Pakistan,
Email: mariyamsabbir786@gmail.com

³ Department of Computer Science, University of Agriculture, Faisalabad, Pakistan,
Email: saeed.rasheed0211@gmail.com

DOI: <https://doi.org/10.63163/jpehss.v4i2.1439>

Abstract

Solar energy has become one of the most important renewable energy resources due to the increasing global demand for clean and sustainable energy systems. However, the intermittent and weather-dependent nature of photovoltaic (PV) power generation creates significant challenges for reliable energy management and smart grid operation. Accurate solar power forecasting is therefore essential for improving energy scheduling, reducing operational uncertainty, and enhancing the integration of renewable energy into modern power systems. In recent years, Machine Learning techniques have gained considerable attention for solar power prediction because of their ability to model complex and nonlinear relationships between weather parameters and photovoltaic output. This paper presents a review of machine learning-based approaches for predicting solar power generation using weather data. Various machine learning techniques including Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machines (GBM), XGBoost, and Long Short-Term Memory (LSTM) networks are discussed and compared. The study also examines important weather parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric pressure that influence forecasting accuracy. Furthermore, commonly used evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2 -score) are analyzed. The reviewed literature demonstrates that machine learning models significantly improve forecasting accuracy compared to traditional statistical methods, while hybrid and ensemble learning approaches provide superior performance in renewable energy forecasting applications. Finally, the paper discusses current challenges and future research directions for developing more accurate and intelligent solar power forecasting systems.

Keywords: Solar Power Forecasting, Machine Learning, Renewable Energy, Weather Data, Artificial Neural Network, Random Forest, Deep Learning, Photovoltaic Systems,

Introduction

The rapid growth in global energy demand and increasing environmental concerns have accelerated the adoption of renewable energy resources such as solar and wind power. Among these renewable sources, solar energy has emerged as one of the most promising and sustainable alternatives to fossil fuels due to its abundance, cleanliness, and environmental benefits. However, the power generated by photovoltaic (PV) systems is highly dependent on weather conditions, making solar energy generation inherently variable and difficult to predict accurately. Factors such as solar irradiance, temperature, humidity, wind speed, and cloud cover significantly

influence photovoltaic output power and create challenges in maintaining stable energy production. Accurate forecasting of solar power generation plays a crucial role in modern smart grid systems, energy management, and power system stability. Reliable prediction models help energy providers optimize grid operation, reduce energy imbalance, minimize operational costs, and improve the integration of renewable energy into existing electrical systems. Traditional statistical forecasting methods often struggle to handle the nonlinear and complex relationships between weather variables and solar power generation. Therefore, advanced computational approaches are required to improve forecasting accuracy.

Recently, Machine Learning techniques have gained significant attention in renewable energy forecasting due to their ability to learn complex patterns from large datasets. Machine learning algorithms such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machines (GBM), and Long Short-Term Memory (LSTM) networks have demonstrated promising performance in predicting solar power output using historical weather data. These techniques can efficiently model nonlinear relationships and adapt to changing environmental conditions.

The relationship between solar power generation and weather parameters can generally be expressed as:

$$P_{solar} = f(I, T, H, W, P_a)$$

where P_{solar} represents solar power output, I denotes solar irradiance, T represents temperature, H indicates humidity, W denotes wind speed, and P_a represents atmospheric pressure.

Several researchers have proposed different machine learning-based approaches for solar energy forecasting. Artificial Neural Networks have been widely used because of their capability to capture nonlinear relationships between weather variables and power output. Ensemble learning techniques such as Random Forest and XGBoost have also shown strong predictive performance by combining multiple learning models. Furthermore, hybrid and deep learning models are increasingly being explored to enhance forecasting accuracy and reduce prediction errors.

This paper presents a review of machine learning-based techniques for predicting solar power generation using weather data. The study discusses commonly used machine learning algorithms, important weather parameters, evaluation metrics, and forecasting methodologies employed in recent research. In addition, the paper analyzes the advantages and limitations of different forecasting approaches and highlights future research directions for improving renewable energy prediction systems.

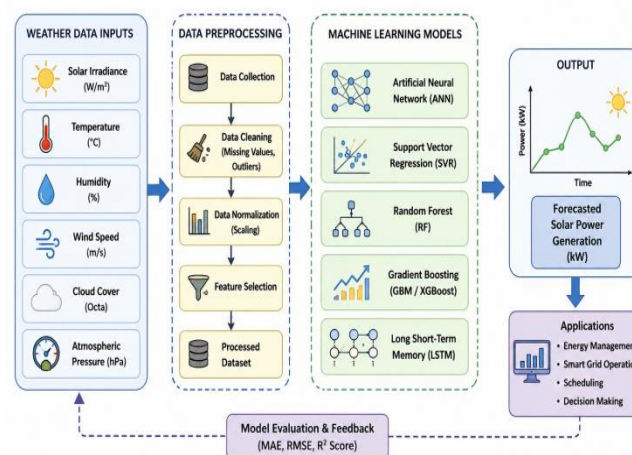


Fig. 1. Machine learning-based solar power forecasting framework.

Literature Review

Recent advancements in Machine Learning have significantly improved the prediction accuracy of solar power generation systems. Researchers have applied various machine learning and deep learning techniques to forecast photovoltaic (PV) output using meteorological and historical

energy datasets. These studies demonstrate that weather parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric pressure strongly influence solar energy production.

Rupesh et al. proposed a solar power prediction model based on Artificial Neural Networks (ANN) using a Feed-Forward Back-Propagation (FFBP) algorithm. The study utilized weather parameters including irradiance, temperature, humidity, and wind speed collected from a photovoltaic system in India. The neural network model achieved approximately 99.4% prediction accuracy, demonstrating the effectiveness of ANN models for renewable energy forecasting.[1]

Another study focused on solar power prediction using multiple machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting, Decision Tree, and Deep Neural Networks (DNN). The researchers incorporated preprocessing and feature engineering techniques to improve forecasting accuracy. The proposed framework achieved nearly 99% AUC accuracy and highlighted the importance of ensemble learning methods in renewable energy forecasting.[2]

Rana et al. introduced a weather-type clustering approach combined with ensembles of neural networks for next-day solar photovoltaic forecasting. The researchers classified weather conditions using K-means clustering and developed separate neural network models for each weather category. The forecasting system used meteorological variables such as solar irradiance, temperature, humidity, and wind speed. Experimental results showed improved forecasting performance with a Mean Relative Error (MRE) of 6.88%, outperforming traditional forecasting methods. [3]

The relationship between weather variables and photovoltaic power generation can generally be represented as:

$$P_{solar} = f(I, T, H, W)$$

where **I** denotes solar irradiance, **T** represents temperature, **H** indicates humidity, and **W** represents wind speed.

Rhafes et al. presented a systematic literature review on machine learning models used in renewable energy forecasting. The study analyzed 21 research papers published between 2019 and 2023 and concluded that Decision Tree-based models, particularly Random Forest, Gradient Boosting Machine (GBM), and XGBoost, are among the most effective techniques for renewable energy forecasting. The review also identified commonly used features such as global horizontal irradiation, temperature, atmospheric pressure, humidity, and wind speed. Furthermore, the authors emphasized that no single machine learning model performs best under all forecasting conditions and recommended hybrid learning approaches for future research. [4]

Several studies have compared different machine learning models for photovoltaic power prediction. Random Forest models have shown strong performance in handling nonlinear relationships and reducing overfitting through ensemble learning techniques. Similarly, Support Vector Regression (SVR) has demonstrated reliable forecasting capability with lower computational complexity. Deep learning models such as Multilayer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks have also gained popularity because of their ability to capture temporal dependencies in time-series weather data. [4]

Evaluation metrics play a significant role in measuring forecast performance. Commonly used metrics in renewable energy forecasting include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and coefficient of determination (R²-score). Among these metrics, MAE and RMSE are the most widely used for evaluating solar forecasting models because they provide accurate information about prediction errors.[4]

The RMSE metric is mathematically expressed as:

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

where y_i represents actual solar power values and \hat{y}_i represents predicted values.

Although machine learning techniques have significantly improved solar power forecasting, several challenges still remain. These include missing weather data, high computational requirements, overfitting problems, and difficulties in long-term forecasting. Researchers suggest that hybrid and ensemble learning models may further improve forecasting accuracy and reliability in future renewable energy systems.

III. MACHINE LEARNING TECHNIQUES FOR SOLAR POWER FORECASTING

Machine learning techniques play a significant role in improving the accuracy of solar power generation forecasting. These techniques analyze historical weather and photovoltaic (PV) data to identify patterns and relationships between environmental parameters and solar energy output. Compared to traditional statistical approaches, machine learning models are more effective in handling nonlinear and complex datasets, making them highly suitable for renewable energy applications.

A. *Artificial Neural Network (ANN)*

Artificial Neural Network (ANN) is one of the most widely used machine learning techniques in solar power forecasting. ANN models are inspired by the structure of the human brain and consist of interconnected neurons organized into input, hidden, and output layers. These networks learn patterns from historical data and can model nonlinear relationships between weather variables and solar power output.

In solar forecasting applications, ANN models commonly use:

temperature, solar irradiance, Humidity and wind speed as input features to predict photovoltaic power generation.

The ANN forecasting process can be represented as:

$$P_{solar} = ANN(I, T, H, W)$$

where:

- I = solar irradiance
- H = humidity
- T=temperature
- W = wind speed

ANN models provide high forecasting accuracy but may require large datasets and longer training times.

B. *Support Vector Regression (SVR)*

Support Vector Regression (SVR) is a machine learning method derived from Support Vector Machines (SVM). SVR is commonly used for regression tasks where continuous numerical values must be predicted. In solar energy forecasting, SVR identifies an optimal regression hyperplane that minimizes forecasting errors while maintaining good generalization performance.

SVR is effective for:

- nonlinear forecasting
- limited datasets
- high-dimensional weather data

The regression function can be expressed as:

$$f(x) = w^T x + b$$

where:

- w represents weight parameters
- x represents input features
- b is the bias term

SVR models are known for stable performance and reduced overfitting; however, selecting appropriate kernel functions and parameters remains challenging.

C. *Random Forest (RF)*

Random Forest is an ensemble learning technique that combines multiple decision trees to improve forecasting accuracy and reduce overfitting. Each decision tree is trained on different subsets of the data, and the final prediction is obtained by averaging the outputs of all trees.

- strong nonlinear modeling capability
- high prediction accuracy
- reduced overfitting
- robustness against noise

According to recent renewable energy forecasting studies, Random Forest is among the most effective machine learning models for solar power prediction. [4]

The Random Forest prediction process can be represented as:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N T_i(x)$$

where:

- $T_i(x)$ represents predictions from individual trees
- N is the total number of trees

D. *Gradient Boosting and XGBoost*

Gradient Boosting Machine (GBM) and Extreme Gradient Boosting (XGBoost) are advanced ensemble learning methods that sequentially improve prediction performance by correcting errors from previous models. These techniques are highly effective in handling large weather datasets and capturing complex nonlinear relationships.

XGBoost has become popular in renewable energy forecasting because of:

- fast training speed
- high computational efficiency
- improved accuracy
- regularization capability

Recent studies indicate that XGBoost often outperforms conventional machine learning models in renewable energy forecasting tasks.[4]

E. *Long Short-Term Memory (LSTM)*

Long Short-Term Memory (LSTM) is a deep learning model designed for time-series forecasting. LSTM networks are a specialized type of Recurrent Neural Network (RNN) capable of learning long-term temporal dependencies from sequential data.

Solar power forecasting involves time-dependent weather information; therefore, LSTM models are highly suitable for:

- hourly forecasting
- day-ahead prediction
- sequential weather analysis

The internal memory mechanism of LSTM enables the network to retain important information from previous time steps while ignoring irrelevant data.

A simplified representation of time-series forecasting using LSTM is:

$$h_t = f(x_t, h_{t-1})$$

where:

- x_t represents current input data
- h_{t-1} represents previous hidden state

LSTM models generally achieve high forecasting accuracy but require larger datasets and higher computational resources.

F. Ensemble and Hybrid Models

Ensemble and hybrid learning approaches combine multiple machine learning techniques to improve prediction accuracy and forecasting stability. Hybrid models utilize the strengths of different algorithms while reducing their individual limitations.

Common hybrid approaches include:

- ANN + Optimization algorithms
- CNN-LSTM models
- RF + XGBoost combinations
- Clustering + Neural Networks

Research studies show that hybrid models often outperform single machine learning algorithms in renewable energy forecasting applications. [4]

Overall, machine learning techniques have significantly enhanced solar power forecasting performance by improving prediction accuracy, reducing forecasting errors, and enabling efficient renewable energy management systems.

IV. WEATHER PARAMETERS AND DATA PREPROCESSING

Weather parameters are among the most important factors affecting solar photovoltaic (PV) power generation. Since solar energy production depends heavily on environmental and atmospheric conditions, accurate weather data is essential for developing reliable machine learning forecasting models. In most solar forecasting systems, meteorological variables are collected from weather stations, photovoltaic monitoring systems, and satellite-based measurements. The efficiency and output power of solar panels vary continuously with changing weather conditions. Therefore, selecting appropriate weather features significantly improves prediction accuracy and model performance. Researchers commonly use multiple environmental parameters to train machine learning models for solar power forecasting. Important weather parameters include solar irradiance, temperature, humidity, wind speed, cloud cover, and atmospheric pressure. Among these variables, solar irradiance is considered the most influential factor because photovoltaic systems directly convert sunlight into electrical energy.

Data preprocessing is an essential step before applying machine learning algorithms. Raw weather datasets often contain missing values, noisy measurements, duplicated records, and inconsistent formats. Therefore, preprocessing techniques such as data cleaning, normalization, and feature selection are required to improve model performance and forecasting reliability. Feature selection helps identify the most important weather variables affecting solar power generation. Proper feature selection reduces computational complexity, improves forecasting accuracy, and prevents overfitting. Data normalization is also commonly applied because weather parameters often have different numerical ranges. Feature engineering techniques such as moving averages, lagged weather values, and seasonal indicators further improve forecasting capability. Despite significant advancements, several challenges remain including missing data, sensor failures, sudden weather changes, and seasonal variability. Overall, accurate weather data collection and preprocessing significantly improve the performance of machine learning-based solar power forecasting systems.

Evaluation Metrics and Performance Analysis

Evaluation metrics are essential for measuring the accuracy and effectiveness of machine learning models used in solar power forecasting. These metrics help researchers compare different forecasting techniques and determine how closely predicted solar power values match actual photovoltaic output. Commonly used evaluation metrics include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and coefficient of determination (R^2 -score). Lower values of MAE, RMSE, and MSE indicate better forecasting performance, while higher R^2 -score values represent improved model fitting and prediction capability. Different machine learning algorithms show different forecasting performances depending on dataset quality, weather conditions, and forecasting horizon. Ensemble learning

and hybrid machine learning models often outperform traditional statistical forecasting methods because they effectively handle nonlinear weather relationships and reduce forecasting errors. Proper evaluation of forecasting models provides several benefits including improved forecasting reliability, better renewable energy management, optimized grid operation, and reduced electricity uncertainty. However, no single evaluation metric is sufficient to fully evaluate forecasting quality. Therefore, researchers often combine multiple evaluation metrics for comprehensive performance analysis. Despite significant progress, several challenges still affect evaluation performance such as incomplete datasets, weather uncertainty, overfitting problems, and computational complexity. Accurate evaluation remains an important component of machine learning-based solar forecasting systems.

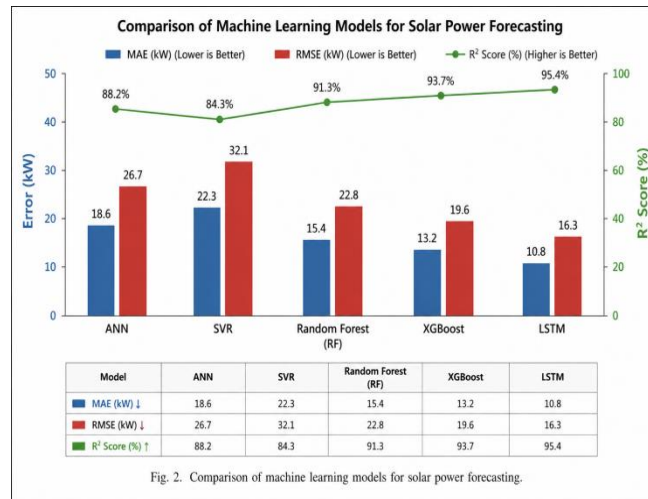


Fig. 2. Comparison of machine learning models for solar power forecasting.

Challenges and Future Research Directions

Although machine learning techniques have significantly improved solar power forecasting accuracy, several challenges still exist in renewable energy prediction systems. Solar energy generation is highly dependent on dynamic environmental conditions, making accurate forecasting a complex research problem. One of the major challenges is rapid weather variability. Environmental conditions such as cloud movement, rainfall, atmospheric pressure changes, and seasonal variations strongly affect photovoltaic power generation. Machine learning models also require large amounts of accurate weather data; however, real-world datasets often contain missing records, sensor failures, and noisy measurements. Complex machine learning models, especially deep learning networks, may suffer from overfitting problems where the model performs well on training data but poorly on unseen testing data. In addition, advanced deep learning algorithms such as LSTM and hybrid learning systems require high computational power and long training times. Future research is expected to focus on hybrid machine learning models, deep learning architectures, real-time forecasting systems, Internet of Things (IoT)-based monitoring, and smart grid integration. Hybrid systems that combine multiple machine learning techniques may further improve forecasting accuracy and reliability. Researchers are also exploring advanced deep learning approaches such as transformer models and attention mechanisms for renewable energy forecasting. Overall, future intelligent forecasting systems are expected to improve renewable energy management, reduce forecasting uncertainty, and support sustainable energy development.

Conclusion

The increasing demand for renewable energy and environmental sustainability has accelerated the development of intelligent solar power forecasting systems. Accurate prediction of photovoltaic power generation is essential for maintaining power system stability, improving energy management, and supporting the integration of renewable energy into modern smart grids.

This paper reviewed various machine learning techniques used for solar power generation prediction using weather data. Different algorithms including Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machines (GBM), XGBoost, and Long Short-Term Memory (LSTM) networks were analyzed and compared.

The literature review demonstrated that weather parameters such as solar irradiance, temperature, humidity, wind speed, and atmospheric pressure significantly influence solar forecasting accuracy. Proper preprocessing techniques including data cleaning, normalization, and feature selection also improve machine learning performance. Among the reviewed approaches, ensemble learning and deep learning models demonstrated superior forecasting performance compared to traditional statistical methods. However, several challenges still remain including weather uncertainty, missing data, computational complexity, and long-term forecasting limitations. In conclusion, machine learning-based solar power forecasting has become an important research area in renewable energy applications. Intelligent forecasting systems can significantly improve photovoltaic power prediction accuracy and support the development of efficient, reliable, and sustainable energy systems.

References

- [1] R. Rupesh et al., "Prediction of Solar Power Using Machine Learning Algorithm," *International Journal of Engineering Research and Technology*, 2023.
- [2] S. Subramanian et al., "Solar Power Prediction Using Machine Learning," *International Journal of Scientific Research in Engineering and Management*, 2024.
- [3] M. Rana and A. Rahman, "Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks," *IEEE Conference Proceedings*, 2020.
- [4] M. Y. Rhafes et al., "Machine Learning Models in Renewable Energy Forecasting: A Systematic Literature Review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 3, pp. 1874–1886, 2025.
- [5] Gupta, M., Arya, A., Varshney, U., Mittal, J., & Tomar, A. *A review of PV power forecasting using machine learning techniques*. *Progress in Engineering Science*, 2(1), 100058.
- [6] Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review. *Energies*, 18(8), 2108.
- [7] A review on data-driven methods for solar energy forecasting. *Applied Energy*.
- [8] Gaboitaolelwe, J., et al. *Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison*. *IEEE Access*, 11, 40820–40845.
- [9] Trends and gaps in photovoltaic power forecasting with machine learning. *Energy Reports*, 9, 447–471.
- [10] Gupta, P., & Singh, R. *PV power forecasting based on data-driven models: A review*. *International Journal of Sustainable Engineering*, 14(6), 1733–1755.
- [11] Alkabbani, H., Ahmadian, A., Zhu, Q., & Elkamel, A. *Machine Learning and Metaheuristic Methods for Renewable Power Forecasting: A Recent Review*. *Frontiers in Chemical Engineering*
- [12] Tina, G. M., Ventura, C., Ferlito, S., & De Vito, S. *A State-of-Art Review on Machine-Learning Based Methods for PV*. *Applied Sciences*, 11(16), 7550.