Proceedings of the 8th IEOM Bangladesh International Conference on Industrial Engineering and Operations Management, Dhaka, Bangladesh, December 20-21, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/BA08.20250477
Published: December 20, 2025

A Machine Learning Based Study on Climatic Influences Affecting Solar PV Power Output

Mithun Pal
Department of Electrical and Electronic Engineering
European University of Bangladesh
Dhaka, Bangladesh
mmithunpal101@gmail.com

Abstract

Solar power generation is directly dependent on weather conditions. Temperature, humidity, cloud density, solar radiation and wind speed cause significant changes in PV system performance. But existing studies have typically analyzed one or two climate components separately, leaving an unclear understanding of the combined effects. This results in uncertainty in short-term solar power forecasts, which creates problems in grid planning and energy management. Current research has limitations in analyzing the combined effects of climatic factors and applying machine learning based forecasting models. In particular, the simultaneous analysis of multiple climatic features using real datasets and reliable short-term forecasting with data-driven models has been scarce. This research fills that gap. Analyzing the combined effect of different climatic factors using a real solar dataset and providing reliable PV power forecasting through Random Forest regression model and the model achieved reliable forecasting accuracy ($R^2 \approx 0.80$). This research will help in grid planning, energy determination, and solar plant operations by knowing the short-term solar power forecast.

Keywords

Solar PV Forecasting, Machine Learning for Solar Model, Climatic Factors, Weather-Based Power Prediction, Random Forest Regression for Renewable Energy.

1. Introduction

Solar photovoltaic (PV) systems are one of the fastest growing sources of renewable energy, but their power generation is dependent on climatic and geometrical factors. Irradiance, temperature, humidity, cloud cover, wind speed, and solar angles (incidence, zenith, azimuth) directly influence power generation. Because of this variability, accurate forecasting is essential for grid stability, energy scheduling, and solar plant operation.

Previous studies have shown that both physical models and machine learning methods can be used in PV prediction (Gaboitaolelwe et al., 2023). In particular, machine learning and deep learning methods are highly effective in capturing nonlinear relationships and time series patterns (Arya & Sharma, 2024). Cloud cover estimation and irradiance prediction using ML models have also proven important (Park et al., 2021). Geometric parameters such as azimuth and incidence angle strongly affect annual energy production (Dhimish & Silvestre, 2019)(Jenčo, 2025). Together, these studies highlight the need for integrated forecasting frameworks. My research combines climatological correlation and geometric optimization into a single and transparent model, which will make forecasting more reproducible and practical for real-world solar planning.

1.1 Objectives

The objective of this study is to analyze how climatic factors such as irradiance, cloud cover, humidity, temperature and solar geometry affect the power output of a solar PV system and to develop a machine learning model that is able to reliably predict solar power output using these climatic factors.

2. Literature Review

Solar PV forecasting has studied through a number of methods, ranging from physical models to advanced machine learning techniques. Physical models basically rely on panel performance, incidence angle, and meteorological input, but their accuracy is largely limited by the accuracy of weather forecasts (Gaboitaolelwe et al., 2023). Later, data-driven methods such as random forests, support vector machines, and gradient boosting have been used to capture nonlinear relationships in solar datasets. Deep learning models, such as LSTM and hybrid CNN-LSTM architectures, have demonstrated strong performance in time series forecasting (Arya & Sharma, 2024)(Zhang & Wang, 2020). These works show the need of ML and DL to increase PV prediction accuracy, however most studies have mainly focused on algorithmic performance and have not incorporated meteorology and solar geometry. My research incorporates climatological factors and solar geometry into a framework, which makes forecasting more interpretable and practical.

Cloud cover estimation utilising aerial cameras and ML segmentation algorithms has revealed that cloud proportion is strongly connected to irradiance and PV output (Park et al., 2021)(Park et al., 2021). However, these algorithms generally rely on image collections and have restrictions in foggy environments. Other studies have shown that humidity, wind speed and atmospheric pressure cause stochastic variability in PV output (Jenčo, 2025)(Arya & Sharma, 2024). Dust accumulation has also been identified as an important factor, showing that untreated panels lose significant efficiency compared to cleaned panels (Hossain et al., 2020). Although these works highlight the importance of climate and environmental data, they have often treated the variables separately. My research investigated numerous climate attributes: radiation, humidity, temperature, cloud cover, and wind speed, together, and presented their combined effects using correlation heatmaps and scatter plots.

Geometrical and environmental parameters are another important aspect of PV prediction. Studies on azimuth and tilt angle have shown that installations between −4° and 2° azimuth produce the most annual power, while extreme orientations significantly reduce output (Dhimish & Silvestre, 2019). Incident angle modeling using equivalent surfaces has provided new ways of calculating sunrise, sunset and maximum irradiance (Jenčo, 2025). Dust effect studies in Bangladesh have also shown that local conditions such as pollution and dry winter weather, if panels are not cleaned regularly, can reduce PV efficiency by more than 20% (Hossain et al., 2020). These investigations largely focused on geometry or dust particles, but did not mix them with climatic circumstances. These investigations largely focused on geometry or dust particles, but did not mix them with climatic circumstances. My research combines geometric characteristics (incidence, zenith, azimuth) with climate factors in a machine learning model, which shows how both dimensions together impact PV output.

3. Methods

In this research, I used a solar dataset received from Kaggle. First I loaded the dataset and cleaned the data by removing missing and inconsistent values. I then constructed additional characteristics to enhance the study, including solar geometry parameters (inclination angle, zenith, azimuth) and time-based features. These technical elements assist capture genuine behavior of sunlight and atmospheric interactions.

After dataset preparation we conducted exploratory data analysis (EDA) using heatmap, scatter plot and pairplot to understand the relationship between climate variables and solar energy output. This step helped identify strong, weak and nonlinear patterns and provided the basis for model selection and subsequent testing.

Based on correlations and visual patterns, key predictive features are selected: radiation, cloud cover, humidity, temperature, wind speed, and solar geometry variables. Using these features we trained a Random Forest regression model, selected for its ability to handle nonlinear relationships and interactions between multiple climate variables. Trained the model on the cleaned and processed dataset and evaluated its performance by standard metrics such as standard deviation (MAE), root mean square error (RMSE) and R^2 score. Additionally, we analyzed the importance of attributes to determine which climatic factors have the strongest influence on solar PV energy production.

After achieving a high performance model, we developed a forecasting module capable of predicting the next day's solar power output. Predicted values of tomorrow's weather (such as radiation, humidity and cloud cover) are provided as input to the model. Based on this the model generates a reliable estimate of the expected power generation in kilowatts. Next, all analysis, visualization and forecasting results are explained, to understand how different weather conditions affect PV performance. This approach provides a complete workflow. From raw data to predictions. which is effectively applicable for real-world energy planning, grid management and performance optimization of solar power plants (Figure 1).

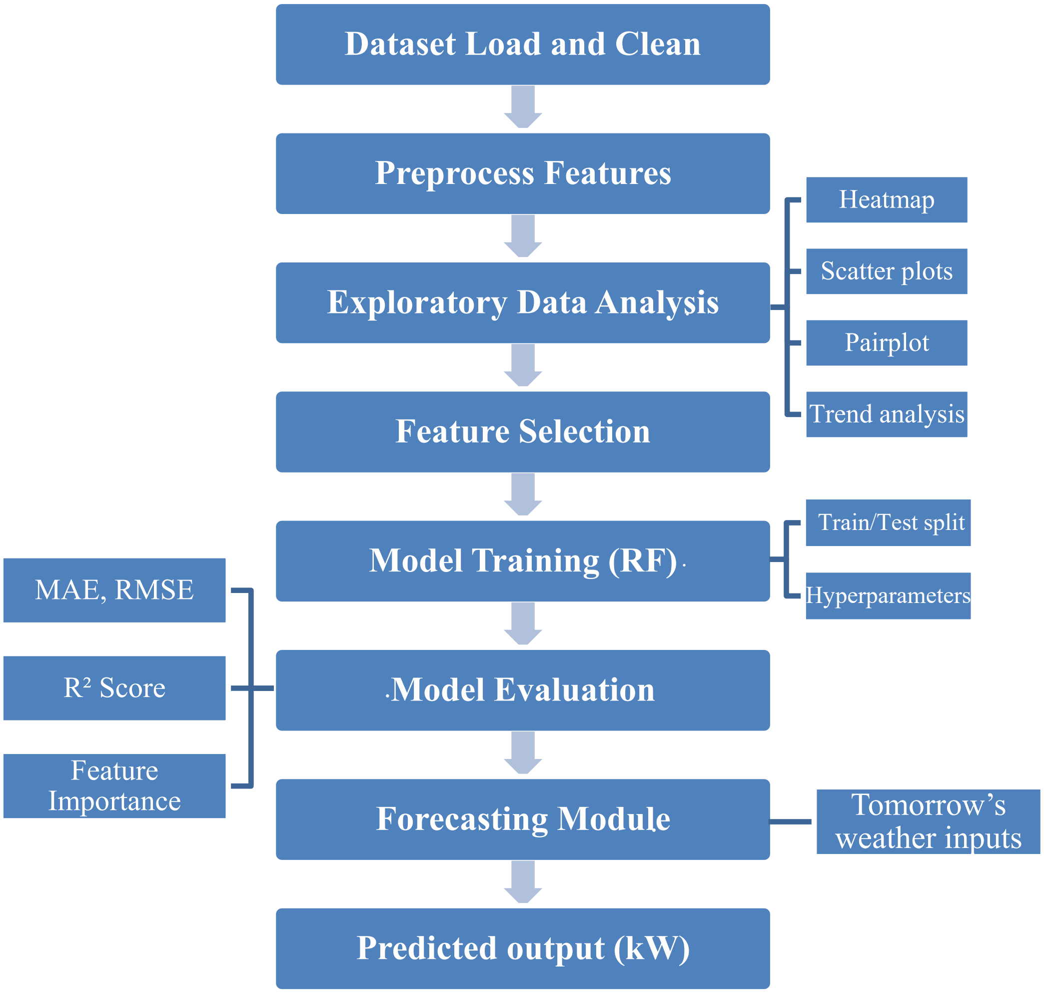

Figure 1. Workflow of Climate-Driven Solar Power Forecasting

4. Results and Discussion

Correlation Heatmap of Climatic Variables and Power Output

The heatmap in Figure 2 shows the relationship between climatic factors and solar energy production. Shortwave radiation has the strongest positive correlation with electricity generation, which is the main driver of PV output. On the other hand, cloud cover and humidity show a negative correlation, i.e. they reduce solar radiation. Temperature shows a slight positive correlation, implying that not only does increasing temperature increase power generation, but sufficient radiation is also required. This visual analysis establishes basic climate dependencies that then influence machine learning models.

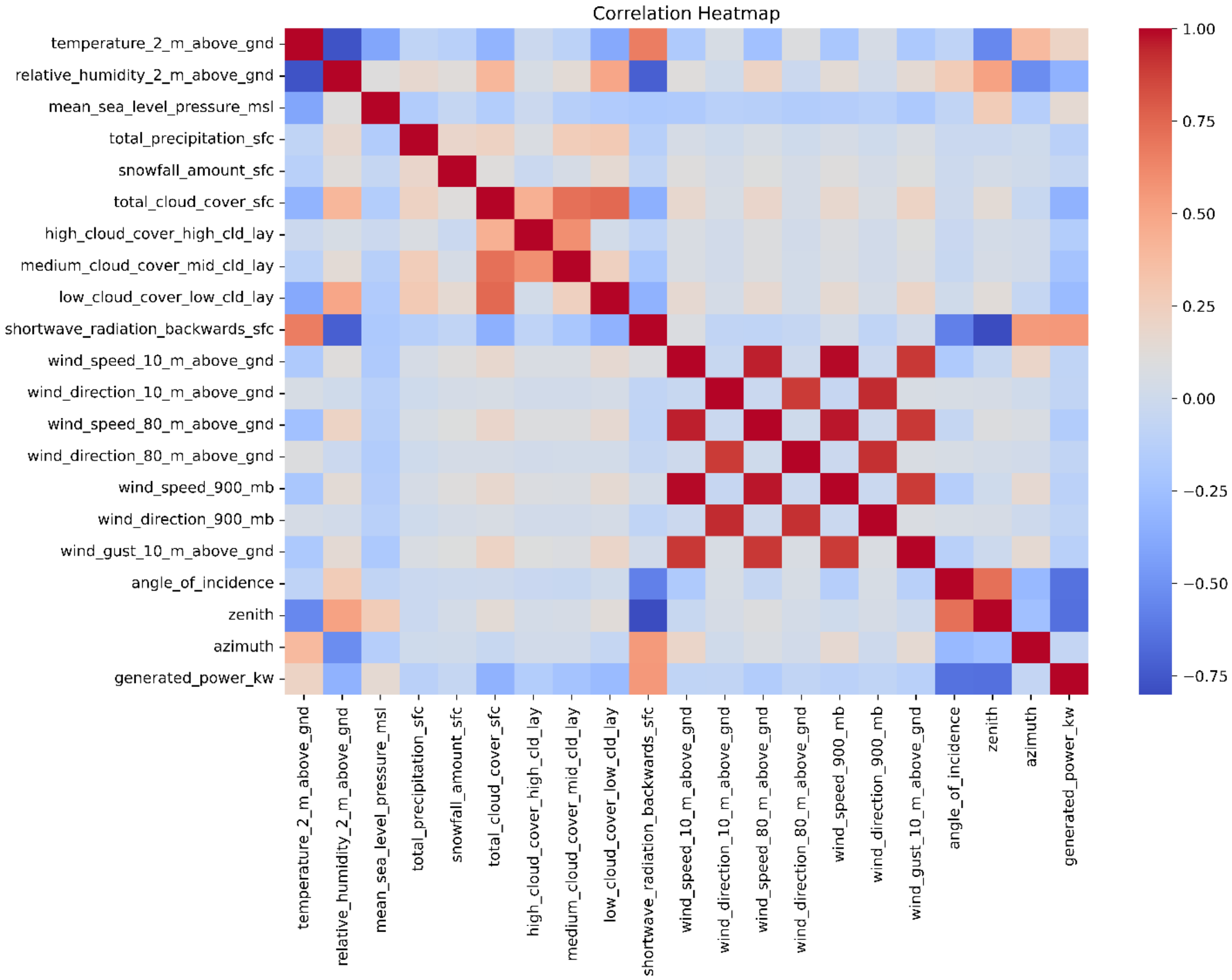

Figure 2. Correlation Heatmap of Climatic Variables and Power Output

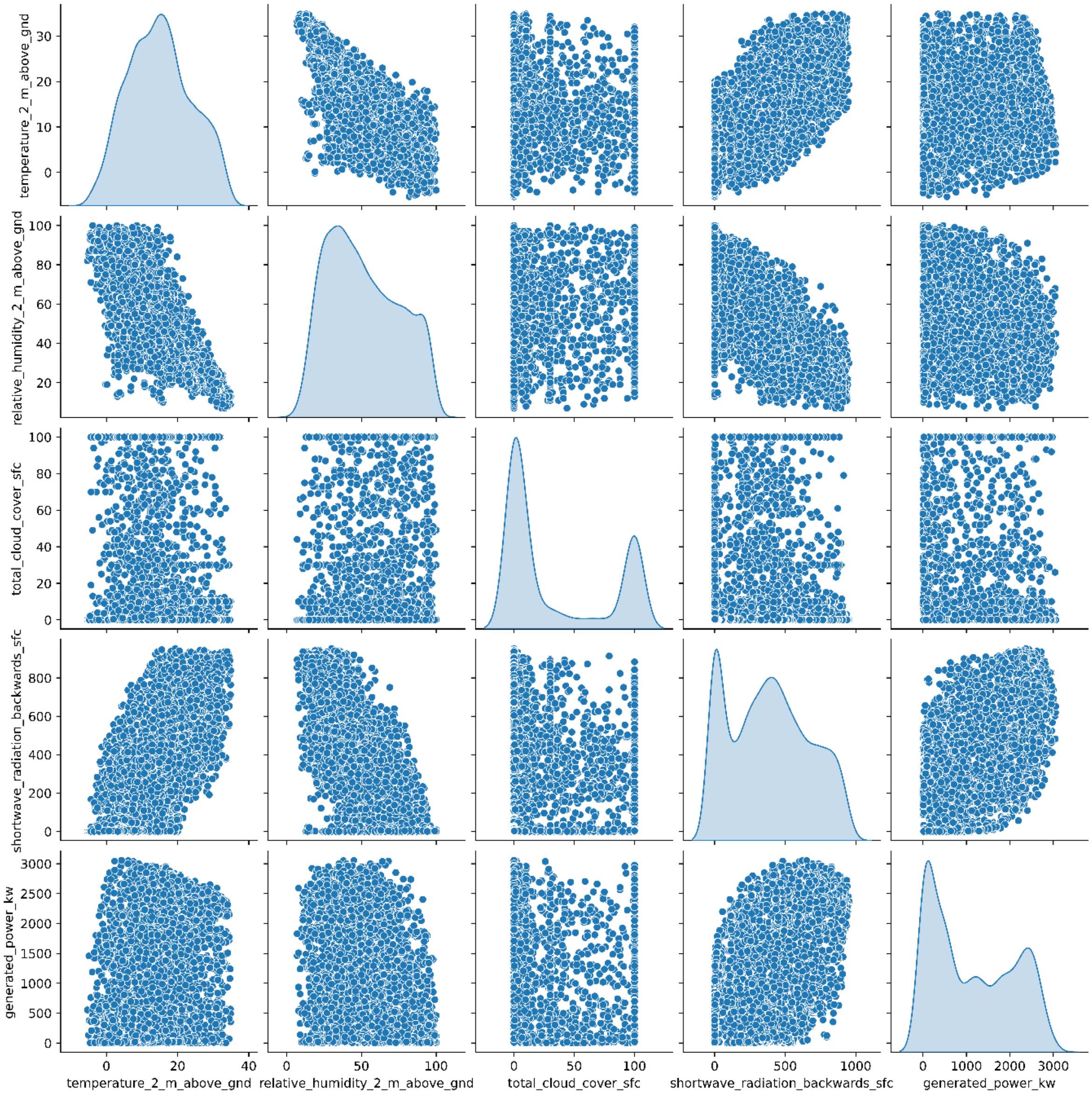

Figure 3. Pairplot of Selected Solar and Weather Features

Pairplot of Selected Solar and Weather Features

Figure 3 presents the pairwise relationship between key predictors and output variables. A clear monotonic increasing trend is observed between irradiance and energy, whereas cloud cover and humidity show opposite trends. The distribution plots along the diagonal indicate that both radiation and energy are right-skewed, reflecting typical diurnal production behavior. These patterns confirm the need for nonlinear regression models, which are able to capture such asymmetric relationships.

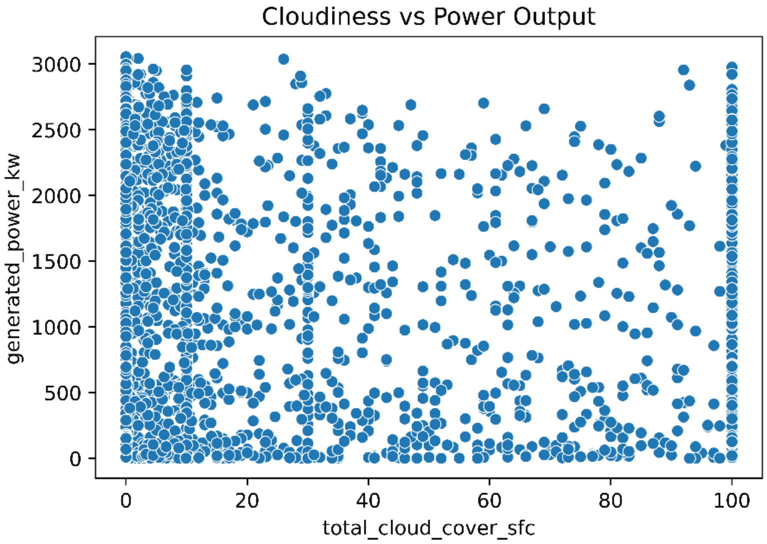

Figure 4. Cloudiness vs Power Output

Cloudiness vs Power Output

Figure 4 shows the effect of cloud cover on power output. A clear downward trend is observed: as cloudiness increases, solar radiation decreases and power generation decreases significantly. High cloud cover (>70%) corresponds to the lowest power band, clarifying the suppressive effect of atmospheric scattering.

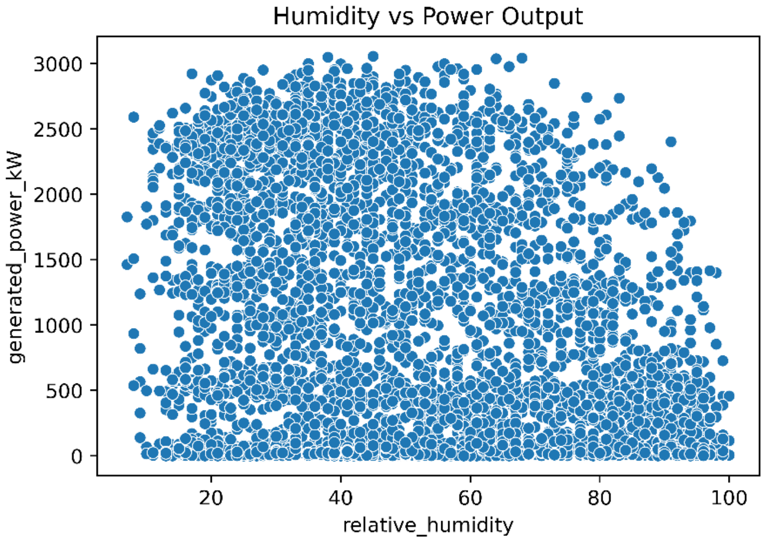

Figure 5. Humidity vs Power Output

Humidity vs Power Output

Figure 5 illustrates the relationship between relative humidity and PV power. A slight but consistent negative trend is visible, where increasing humidity decreases the effective transmission of sunlight. Although humidity does not reduce energy as drastically as cloud cover, its cumulative effect becomes more pronounced under partly cloudy conditions. This supports the conclusion that atmospheric humidity indirectly limits PV performance.

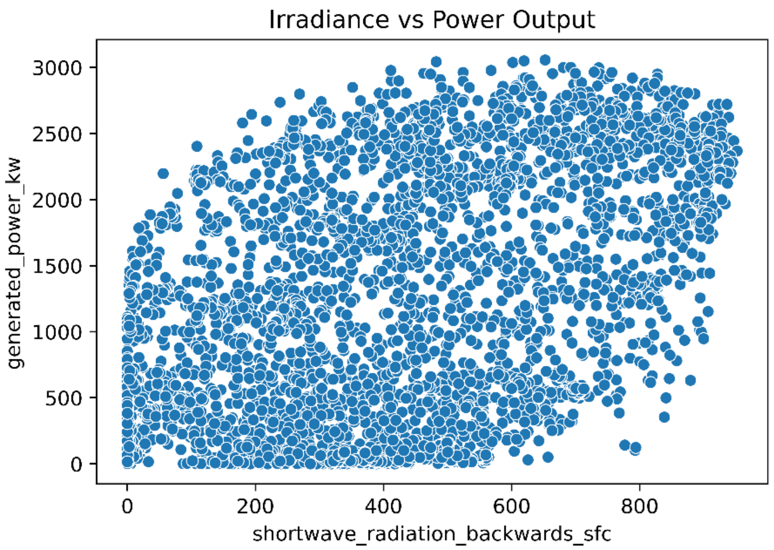

Figure 6. Irradiance vs Power Output

Irradiance vs Power Output

Figure 6 presents a scatter plot of radiation versus generated power. The relationship is visibly nonlinear, with low power output during low irradiance periods and a steep increase after irradiance exceeds a threshold. Scattered partial clouds in the mid-radiation region indicate variability and transient weather conditions. This behavior justifies the adoption of ensemble-based models to handle stochastic fluctuations.

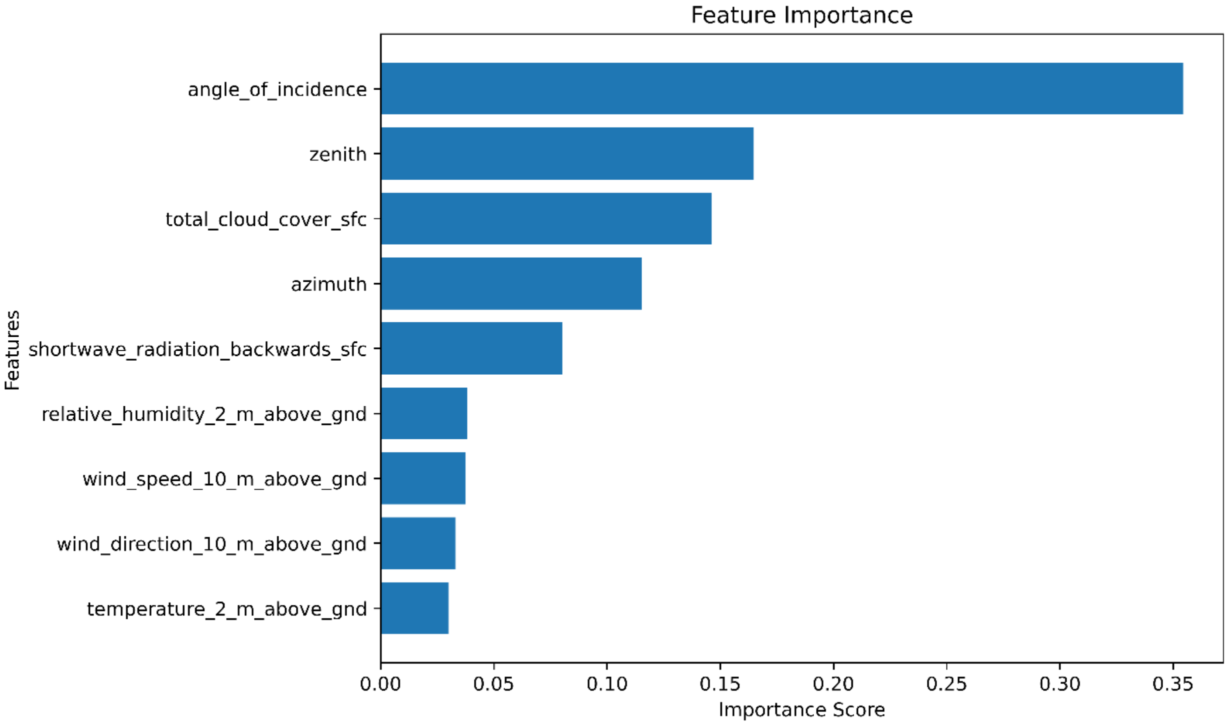

Figure 7. Feature importance from the Random Forest model

Feature importance from the Random Forest model

Figure 7 shows the attribute importance values obtained from the random forest model. Incidence angle and solar zenith angle emerge as the most important features, followed by cloud cover and shortwave radiation. This confirms that solar geometry, combined with atmospheric precipitation, mainly shapes the PV output. Low-ranked attributes such as temperature and wind parameters contribute little, consistent with their low correlations observed previously.fluctuations.

5. Conclusion

This research has shown that climatic factors a strong and measurable impact on solar PV energy production. Humidity and temperature contribute to moderate variation, and irradiance and cloud cover largely control solar PV output. A clear and practical understanding of weather effects on real PV systems is gained through visual analysis and Random Forest regression model. The model achieved high forecast accuracy ($R^2 \approx 0.80$), which proves that it is possible to effectively forecast solar energy using climate data. These results highlight the importance of climate-aware machine learning models and can be helpful in improving the planning, scheduling and energy management of renewable power systems.

References

J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwed, D. N. Vinod, and A. O. Salau, “Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison,” *IEEE Access*, vol. 11, pp. 40820–40840, Apr. 2023, doi: 10.1109/ACCESS.2023.3270041.

T. Arya and A. Sharma, “Solar Power Forecasting Using Machine Learning and Deep Learning,” *International Journal of Creative Research Thoughts (IJCRT)*, vol. 12, no. 12, pp. 952–959, Dec. 2024.

S. Park, Y. Kim, N. J. Ferrier, S. M. Collis, R. Sankaran, and P. H. Beckman, “Prediction of Solar Irradiance and Photovoltaic Solar Energy Product Based on Cloud Coverage Estimation Using Machine Learning Methods,” *Atmosphere*, vol. 12, no. 3, p. 395, Mar. 2021, doi: 10.3390/atmos12030395.

M. Dhimish and S. Silvestre, “Estimating the Impact of Azimuth-Angle Variations on Photovoltaic Annual Energy Production,” *Clean Energy*, vol. 3, no. 1, pp. 47–58, Feb. 2019, doi: 10.1093/ce/zky022.

M. Jenčo, “Determination of the Solar Angle of Incidence Using an Equivalent Surface and the Possibility of Applying This Approach in Geosciences and Engineering,” *ISPRS International Journal of Geo-Information*, vol. 14, no. 9, p. 323, Aug. 2025, doi: 10.3390/ijgi14090323.

L. Zhang and Y. Wang, “Solar Irradiance Prediction for Power Generation Using Deep Learning,” *Renewable Energy Systems Journal*, vol. 22, no. 3, pp. 85–93, 2020.

M. Hossain, F. Rabbi, S. Zaman, F. Rahman, and M. S. Islam, “Study of Dust Effect on PV Panel’s Efficiency During Winter for Dhaka, Bangladesh,” in *Proc. 2020 2nd Int. Conf. Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, Dec. 2020, pp. 1–6, doi: 10.1109/STI50764.2020.9350391.