MLP-Based PV Power Output Estimation Using Weather Variables and SHAP Analysis

Juan David Garcia-Racines
Graduate School of System Informatics
Kobe University
Kobe, Japan
juandsa@es4.eedept.kobe-u.ac.jp

Masahide Nakamura
Graduate School of System Informatics
Kobe University
Kobe, Japan
masa-n@cmds.kobe-u.ac.jp

Sinan Chen
Graduate School of System Informatics
Kobe University
Kobe, Japan
chensinan@gold.kobe-u.ac.jp

***Abstract*—Global energy consumption is increasing due to population growth, urbanization, and industrialization. As reliance on fossil fuels decreases and their environmental impact becomes more evident, renewable energy sources like solar power are becoming increasingly important for sustainable development and climate initiatives. Solar energy is especially valued for its scalability and cost-effectiveness, but weather variability poses challenges, requiring precise modeling for system planning. Traditional PV prediction techniques rely on historical data or irradiance measurements, which provide high short-term accuracy but may be limited in regions lacking extensive data. Weather-based models often emphasize seasonal trends, neglecting intra-day changes. Using weather inputs alone can enhance PV deployment in areas with unreliable solar data, especially in underserved regions. In this study, an MLP regression model trained on weather data and theoretical irradiance was used to predict PV output for a 40kW system in Kobe, Japan, with an RMSE of 3.79 and an R^2 of 0.88. SHAP analysis identified Humidity and Temperature as the most influential factors, followed by Gust direction and Wind gust. The findings suggest that on-site irradiance measurements are unnecessary; weather data alone can reliably model PV performance.**

***Index Terms*—PV, Solar power, Weather impact, Forecast, MLP regressor, SHAP analysis.**

I. Introduction

Over the past few decades, global energy use has consistently increased due to population growth, urbanization, and industrial expansion. As traditional fossil fuels become scarcer and their environmental impacts worsen, renewable energy sources have become increasingly important as sustainable alternatives to meet growing energy needs and mitigate climate change [1].

Renewable technologies are increasingly entering the global energy market. According to the International Renewable Energy Agency (IRENA), renewables accounted for over 43% of the globally installed power capacity in 2023, with solar photovoltaic (PV) installations increasing by 345 GW [2]. Japan exemplifies this trend: by the end of 2024, solar power generation accounted for 11.4%, the highest among renewable energies, compared to 9.3% in 2021 [3].

Among the various renewable energy options, solar power is notable for its scalability and decreasing costs; however, its output remains inherently uncertain due to variations in solar irradiance. Therefore, precise estimation models are crucial for system planning, grid integration, reliability, and the successful deployment of PV projects [4].

Previous research on PV power estimation has primarily focused on using historical output data or irradiance measurements to achieve high accuracy for short-term forecasts. However, these methods are limited in regions lacking detailed solar or irradiance data or where PV installations are absent, highlighting the need for alternative approaches that work with limited data. Some studies have explored the use of weather variables for prediction, comparing various models and examining the impact of weather on PV performance across different seasons, although many lack intra-day analysis to capture short-term variations.

This work addresses two interconnected challenges: developing a reliable model to estimate PV power output solely from weather-based inputs, and evaluating how weather variability affects power generation. By integrating weather observations, such as temperature, humidity, and precipitation, with machine learning, the research aims to enhance the accuracy of solar production forecasts, facilitating more resilient renewable energy planning. This weather-based PV estimation framework not only fills a critical gap in regions lacking historical solar output data or where such data is unreliable, but also supports the expansion of PV installations in underrepresented areas by eliminating the need for on-site irradiance measurements. Consequently, it enables stakeholders to assess solar energy generation potential without extensive data collection, while also providing the opportunity to better understand the influence of weather on PV performance.

II. Background

Previous studies have generally focused on predicting PV power generation by utilizing historical PV power output data or irradiance measurements. These predictive methods typically provide forecasts for one day ahead and tend to achieve high accuracy, using various techniques [5], [6].

Nonetheless, such approaches have limits in their applicability, especially in regions where irradiance data are scarce or nonexistent, or where PV installations are absent. This limits their usefulness in areas lacking comprehensive solar resource data or active PV infrastructure, underscoring the need for alternative prediction strategies that can operate effectively under data-limited conditions.

Previous studies have explored estimating PV power generation using only weather variables. For instance, [7] compares different models and their accuracy, but does not delve into how weather variables specifically influence the results. Similarly, [8] includes a large dataset of weather information but primarily concentrates on comparing various models and their performance. Conversely, [9] investigated the impact of weather on PV performance and identified distinct models for each season. However, these models have a daily resolution, so intraday behavior was not captured. [10] It also studies the impact of weather, specifically wind speed and ambient air temperature, but lacks an intra-day analysis.

III. Methodology

Our goal is to develop a model capable of predicting solar power generation from PV systems relying solely on weather data and theoretical information. This approach eliminates the need for on-site irradiance measurements or historical PV output data. We propose using an MLP regressor with six weather variables and theoretical irradiance as inputs, aiming to predict the PV system's solar power output. The process includes post-processing to improve accuracy and evaluating the model's performance with four standard metrics. We then analyze the results by visually inspecting the model's behavior across different days of the year. Lastly, we apply SHapley Additive exPlanations (SHAP) analysis to understand how weather variables affect power output, identifying the most influential factors.

A. Input variables

1) Weather variables: The Japan Meteorological Agency provides 10-minute weather data for all Japanese cities [11], including surface and sea-level pressure, precipitation, temperature, humidity, wind speed and direction, wind gusts, and sunshine duration. Since we model solar generation based on weather and theoretical inputs, we did not include sunshine duration. Wind speed and direction were converted from categorical data to continuous measurements (degrees) for modeling purposes. We assessed redundancy using Pearson's correlation (see Table I); variables with values greater than 0.85 were excluded—specifically, surface pressure, wind speed, and wind direction. As a result, our final set includes six weather variables: sea-level pressure (hPa), Precipitation (mm), Temperature (°C), Humidity (%), Wind gust (m/s), and Gust direction (°). Table II describes these variables.

2) Theoretical irradiance: To avoid on-site irradiance measurement, we used the Perez-Driesse model [12] within the PVlib Python toolbox [13] to compute theoretical irradiance. This model estimates total irradiance (W/m2) at the PV system's location by combining direct, diffuse, and reflected components. While the model can account for atmospheric conditions, we chose not to incorporate them and instead calculated irradiance under clear sky conditions every 10 minutes.

B. Target variable: Power generated

The PV system analyzed in this study is located at Kobe University in Kobe, Japan. It consists of 240 crystalline silicon solar panels with a combined capacity exceeding 40 kW. Installed on the rooftop of a university building, the system has an inclination tilt of 20° and an azimuth angle of 154°.

The system's power generation is recorded every 10 minutes throughout the day. For this study, we used five years of power generation data from April 2020 to March 2025 as the target variable for model training.

C. MLP regressor model

The Multi-Layer Perceptron (MLP) regressor is frequently employed in solar power modeling due to its ability to capture complex, nonlinear relationships among input features. This neural network type has been successfully used to forecast solar power output by analyzing large datasets that include irradiance, temperature, and cloud cover [14]. Research indicates that MLP regressor models often surpass traditional linear models in prediction accuracy [15]. In this study, we used scikit-learn [16] in Python to train an MLP regressor. Based on an encoder architecture, we defined a neural network with three hidden layers containing 100, 50, and 25 neurons, respectively. Fig. 1 illustrates the proposed regressor model.

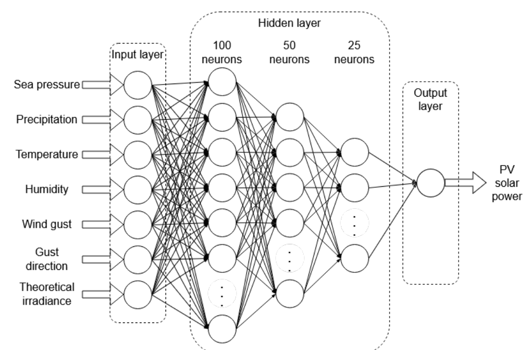

Fig. 1: MLP regressor model used.

The input data were normalized to improve MLP regressor training. This step prevents features with larger scales from exerting undue influence, leading to better learning and improved model performance. The data were randomly split into two parts: 70% for training and 30% for testing, before being used as input.

D. Post-processing

To complement the regressor model's results, we set its output to zero in two specific cases: when the estimated output is negative and when the sun's inclination is below -18 ° relative to the horizon. Our analysis of the PV system's historical data revealed that it produces a small amount of power during twilight, so we chose this particular sun inclination because it indicates the end of night and the start of astronomical twilight.

TABLE I: Pearson correlation between weather variables

	Surface pressure	**Sea-level pressure**	**Precipi-tation**	**Tempera-ture**	**Humidity**	**Wind speed**	**Wind gust**	**Sunshine minutes**	**Wind direction**	**Gust direction**
Surface pressure	1.00	1.00	0.11	0.60	0.38	0.12	0.13	0.09	0.06	0.06
Sea-level pressure	1.00	1.00	0.11	0.61	0.38	0.12	0.13	0.09	0.06	0.06
Precipitation	0.11	0.11	1.00	0.03	0.19	0.06	0.06	0.06	0.04	0.04
Temperature	0.60	0.61	0.03	1.00	0.25	0.03	0.05	0.11	0.07	0.07
Humidity	0.38	0.38	0.19	0.25	1.00	0.16	0.21	0.38	0.07	0.08
Wind speed	0.12	0.12	0.06	0.03	0.16	1.00	0.95	0.11	0.10	0.11
Wind gust	0.13	0.13	0.06	0.05	0.21	0.95	1.00	0.14	0.02	0.02
Sunshine minutes	0.09	0.09	0.06	0.11	0.38	0.11	0.14	1.00	0.02	0.03
Wind direction	0.06	0.06	0.04	0.07	0.07	0.10	0.02	0.02	1.00	0.85
Gust direction	0.06	0.06	0.04	0.07	0.08	0.11	0.02	0.03	0.85	1.00

Correlation coefficient equal to or higher than 0.85 highlighted.

TABLE II: Description of weather variables

Variable	**Min**	**Median**	**Max**	**Std deviation**
Sea-level pressure (hPa)	985.0	1015.3	1036.1	6.8
Precipitation (mm)	0.0	0.0	23.5	0.3
Temperature (°C)	-3.0	18.1	37.4	8.5
Humidity (%)	12.0	66.0	100.0	14.9
Wind Gust (m/s)	0.0	4.9	35.7	2.9
Gust direction (°)	0.0	225.0	337.5	101.9

E. Validation: Performance metrics

To evaluate the model's performance, we use the following 4 metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Median Absolute Error (MedAE), and the Coefficient of Determination (R2). These metrics are defined in Equations 1-4, respectively.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$MedAE = median(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, ..., |y_i - \hat{y}_i|) \tag{3}$$

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

Where, y_i is the actual value, $\hat{y}_i$ is the estimated value, $\bar{y}$ is the mean of the actual values.

F. SHapley Additive exPlanations (SHAP)

SHAP is a method that measures the influence of each input on the output by calculating Shapley values from cooperative game theory [17]. It treats the output as the combined marginal effects across all input subsets, ensuring fairness and consistency. By combining these influences, SHAP generates additive explanations that are easy to visualize, helping users understand the model's behavior.

IV. Results and Discussion

The MLP model was trained using six weather variables and theoretical irradiance as inputs, and solar power generated as the target. After training the model on PV system power generation data, we validate its performance using the previously mentioned metrics. The results are presented in Table III.

TABLE III: Model performance metrics

RMSE	**MAE**	**MedAE**	**R^2**
3.79	1.98	0.33	0.88

The MLP model's performance in estimating solar power generation demonstrates that, on average, predictions are approximately 2kW away from the actual values, as reflected by the MAE. This signifies a low error when compared to the 40 kW PV capacity. Moreover, the MedAE shows that half of the estimates are within 0.33 kW of actual values, highlighting its reliability. Additionally, the MLP successfully captures key weather patterns that influence solar power, as evidenced by an R^2 score of 0.88. Examining the results visually, we verify that the model correctly captures the primary pattern of power generation. Fig. 2 illustrates the variation in power generation values over a week in June 2024 and how the model estimations follow that variation.

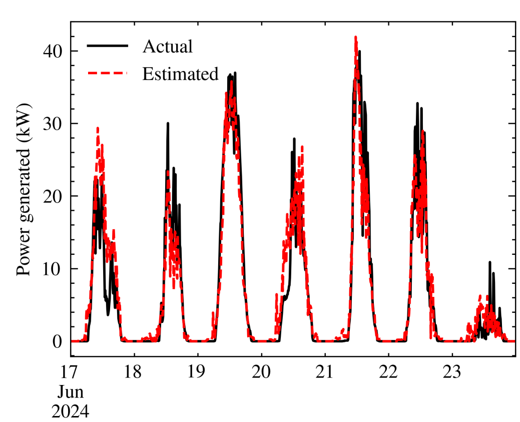

Fig. 2: A week of actual and estimated power generation.

As shown in Fig. 3, on June 21, 2024, power generation started increasing shortly before 9:00, reaching about 38 kW at noon, then declining around 14:30. The model accurately captures the rise and fall in power, closely matching the observed behavior based on weather data, although it does not provide an exact estimate.

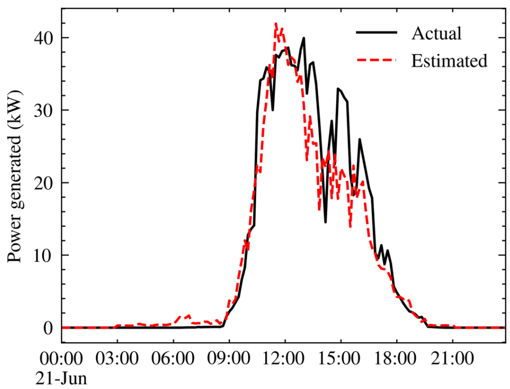

Fig. 3: Actual and estimated power generation on June 21, 2024.

The model's performance was also clear on June 23, 2024, when power generation was notably lower than on other days. Fig. 4 shows how the model accurately estimates the low generation throughout the day, with an average of 3 kW.

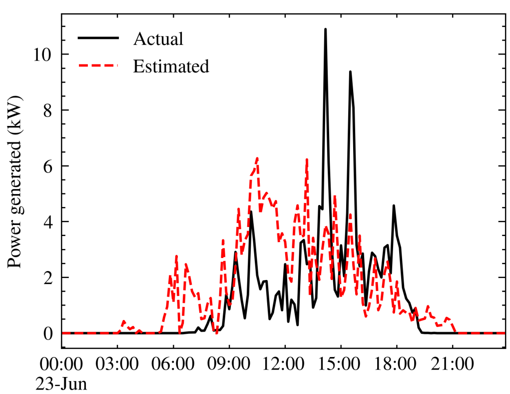

Fig. 4: Actual and estimated power generation on June 23, 2024.

Based on these findings, the MLP model reliably predicts solar power output by integrating multiple weather parameters to accurately estimate how fluctuations affect generation. It correctly reflects both increases and decreases, typically estimating production relative to system capacity with errors that are only a small fraction of the PV size—half of the estimates are very close to observed values. The model effectively captures key weather-driven patterns, explaining most of the variation in generation. Visual checks over a week in June show it closely follows the main daily generation pattern, including the strong mid-day rise and afternoon decline on clear days, and correctly reflects days with much lower output, though it doesn't reproduce every exact value. To better understand the influence of weather factors, SHAP analysis was performed.

SHAP values help interpret how specific weather factors influence a PV system's predicted power output. They turn complex forecasts into straightforward, physically meaningful explanations of energy variability. By identifying weather conditions that most significantly increase or decrease power, and revealing interactions between factors, SHAP supports decision-making. This includes evaluating site suitability based on weather sensitivity, optimizing panel placement, planning operations, and improving grid management, as well as stakeholder assessments of solar potential. In the SHAP analysis shown in Fig. 5, each input variable is shown in color, with blue indicating low values and red indicating high values.

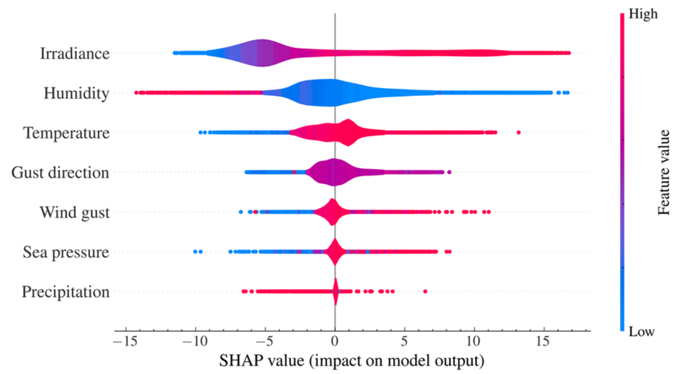

Fig. 5: SHAP value of input variables.

Irradiance has the greatest effect, with values ranging from -13kW to 17kW. Humidity exhibits a similar range, from -15kW to 17kW, indicating that lower humidity tends to enhance power generation. On the other hand, higher temperatures are associated with increased power output. Wind gusts, Sea-level pressure, and Precipitation exert less influence on the output, as their values cluster around the center of the graph where the SHAP value is zero.

The Partial SHAP dependence analysis provides more detail on how each input variable affects the estimated power output. Fig.6 to Fig.12 show the results for each variable. The light blue area in the figures indicates the potential output values (y-axis) for the input values (x-axis). The dark blue line shows the average output. The light grey bars at the bottom represent the histogram of the input variable. In this case, the input values are presented in a standardized format because that is how the model receives them.

Irradiance exhibits a direct correlation, as depicted in Fig. 6 by the dark blue crescent line. Conversely, the light blue area narrows at low irradiance levels, as expected, because no power is generated at zero or near-zero irradiance. However, as irradiance increases, output variability also rises; even at high irradiance, the output can be near zero. This suggests that weather and other factors can lead to zero power generation.

Humidity mainly shows an inverse relationship with power generation, as illustrated in Fig. 7. When humidity is low, power generation variability increases, whereas high humidity

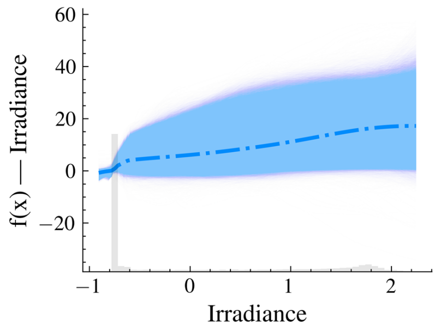

Fig. 6: SHAP dependence results for Irradiance.

decreases variability and results in lower output levels. This indicates that high humidity hampers PV system performance regardless of other weather factors.

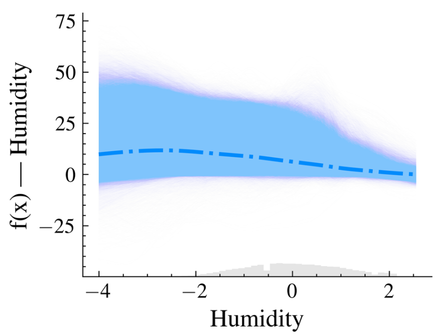

Fig. 7: SHAP dependence results for Humidity.

Fig. 8 displays the results for Temperature, which generally correlates directly with power generation. Although its impact is less pronounced than irradiance or humidity, as evidenced by the high variability in output, the data show that at lower temperatures, the power output does not reach the same high values seen at higher temperatures.

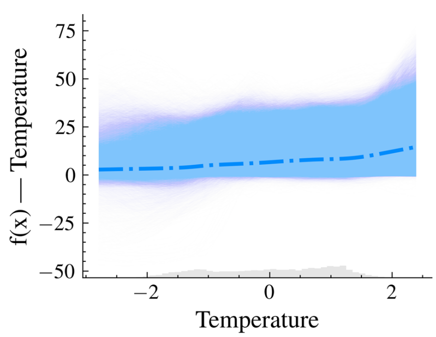

Fig. 8: SHAP dependence results for Temperature.

Gust direction appears to have a minimal impact on the output, as shown in Fig. 9. The output's variability remains mostly constant, with a slight increase in the middle. As shown in Fig. 5, this variable may help estimate power in some cases; however, there is generally no clear correlation. A similar pattern is observed with Wind gust in Fig. 10, where the variability stays mostly constant, as indicated by the histogram. Nonetheless, as Wind gust values rise, the maximum output values tend to increase. However, this finding should be interpreted cautiously, as high wind gust readings are infrequent in this dataset.

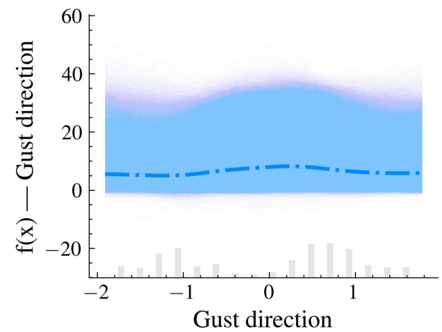

Fig. 9: SHAP dependence results for Gust direction.

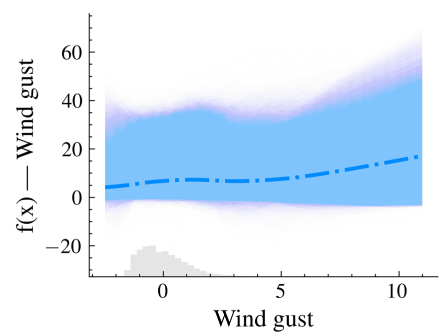

Fig. 10: SHAP dependence results for Wind Gust.

Judging from the histogram in Fig. 11, most Sea-level pressure values appear to have a limited influence on power generation. The figure shows a low correlation, despite a significant drop in output at low Sea-level pressure values and a slight increase at high sea-level pressure values. However, these instances are rare and might be misleading.

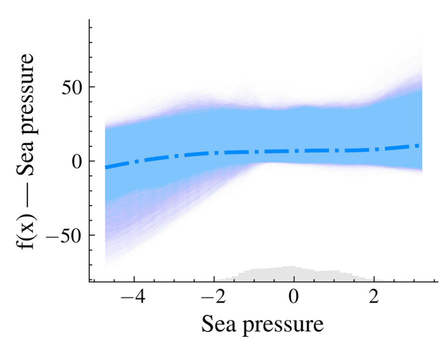

Fig. 11: SHAP dependence results for Sea-level pressure.

As illustrated in Fig. 12, most Precipitation values are zero since Kobe experiences rain only in a few seasons. However, the results indicate that, except for a few instances, Precipitation has a limited impact on power generation compared to other variables.

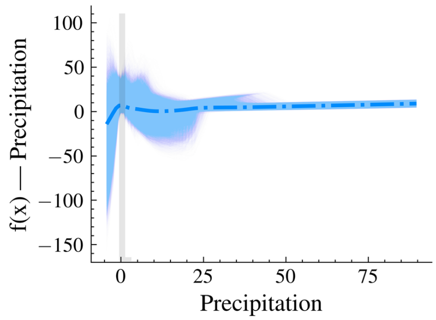

Fig. 12: SHAP dependence results for Precipitation.

V. Conclusion

This study presents a weather-based PV estimation framework designed to improve energy planning, grid integration, and solar deployment in regions with limited data availability. By using standard meteorological data rather than local irradiance sensors, the framework provides a practical solution for areas with sparse sensing infrastructure. We developed an MLP regressor to predict the power output of a PV system in Kobe, Japan, using six weather variables along with theoretical irradiance as inputs. This approach enables stakeholders to make rapid, well-informed decisions, supporting scalable, sustainable PV deployment even under constrained sensing conditions. It also facilitates effective site selection, capacity planning, and regional operational forecasting without requiring on-site irradiance measurements or historical PV output data.

The MLP model is effective for estimating PV power generation, demonstrating that weather variables are sufficiently accurate predictors of PV power fluctuations. It is non-recurrent and does not require actual PV generation data for predictions, unlike earlier approaches. Typically, predictions differ by about 2 kW from actual values, which is small compared to the 40 kW PV capacity. The MedAE shows that half of the predictions are within 0.33 kW of true values, indicating reliability. The MLP consistently identifies key weather-related features affecting solar output and accurately reproduces the main daily generation pattern, including increases, plateaus, and declines at appropriate times and with similar magnitudes. While it closely follows observed fluctuations and daily averages, deviations from exact values are present.

SHAP analysis indicates that irradiance has the greatest impact on power, ranging from approximately -13 kW to 17 kW, whereas humidity shows a similar range, with lower humidity generally increasing energy production. Temperature is positively correlated with higher output, but its effect is less significant than that of irradiance or humidity. Irradiance correlates directly with output, so weather or other factors become evident when there is near-zero output despite high irradiance. Humidity exhibits an inverse relationship, implying that high humidity generally hampers PV performance regardless of other weather conditions. Wind gusts and Gust direction have minimal influence, although maximum outputs tend to peak during rare high gust events. Sea-level pressure and Precipitation, aside from a few cases, have a limited impact on generation compared to other variables.

Acknowledgment

This research is supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Postgraduate Scholarship, JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP24K02765, JP24K02774, JP23K17006, JP23H03401, JP23K28383.

References

[1] IEA, "World Energy Outlook 2023," 2023.

[2] IRENA. , "Renewable capacity statistics 2024."

[3] ISEP, "2024 Share of Electricity from Renewable Energy Resources in Japan | Institute for Sustainable Energy Policies," Aug. 2025.

[4] N. Jannah, T. S. Gunawan, S. H. Yusoff, M. S. A. Hanifah, and S. N. M. Sapihie, "Recent Advances and Future Challenges of Solar Power Generation Forecasting," *IEEE Access*, vol. 12, pp. 168 904–168 924, 2024.

[5] K. J. Iheanetu, "Solar Photovoltaic Power Forecasting: A Review," *Sustainability*, vol. 14, no. 24, p. 17005, Jan. 2022.

[6] P. Gupta and R. Singh, "PV power forecasting based on data-driven models: A review," *International Journal of Sustainable Engineering*, vol. 14, no. 6, pp. 1733–1755, Nov. 2021.

[7] D. Rangelov, M. Boerger, N. Tcholtchev, P. Lämmel, and M. Hauswirth, "Design and Development of a Short-Term Photovoltaic Power Output Forecasting Method Based on Random Forest, Deep Neural Network and LSTM Using Readily Available Weather Features," *IEEE Access*, vol. 11, pp. 41 578–41 595, 2023.

[8] S. Ghimire, R. C. Deo, D. Casillas-Pérez, S. Salcedo-Sanz, E. Sharma, and M. Ali, "Deep learning CNN-LSTM-MLP hybrid fusion model for feature optimizations and daily solar radiation prediction," *Measurement*, vol. 202, p. 111759, Oct. 2022.

[9] A. Gopi, K. Sudhakar, N. W. Keng, A. R. Krishnan, and S. S. Priya, "Performance Modeling of the Weather Impact on a Utility-Scale PV Power Plant in a Tropical Region," *International Journal of Photoenergy*, vol. 2021, no. 1, p. 5551014, Jan. 2021.

[10] A. Gopi, P. Sharma, K. Sudhakar, W. K. Ngui, I. Kirpichnikova, and E. Cuce, "Weather Impact on Solar Farm Performance: A Comparative Analysis of Machine Learning Techniques," *Sustainability*, vol. 15, no. 1, p. 439, Jan. 2023.

[11] "Japan Meteorological Agency | Historical Weather Data," https://www.data.jma.go.jp/stats/etrn/index.php.

[12] A. Driesse, A. R. Jensen, and R. Perez, "A continuous form of the Perez diffuse sky model for forward and reverse transposition," *Solar Energy*, vol. 267, p. 112093, Jan. 2024.

[13] K. S. Anderson, C. W. Hansen, W. F. Holmgren, A. R. Jensen, M. A. Mikofski, and A. Driesse, "Pvlib python: 2023 project update," *Journal of Open Source Software*, vol. 8, no. 92, p. 5994, Dec. 2023.

[14] M. Abdelsattar, M. A. Ismeil, M. M. A. A. Zayed, A. Abdelmoety, and A. Emad-Eldeen, "Assessing Machine Learning Approaches for Photovoltaic Energy Prediction in Sustainable Energy Systems," *IEEE Access*, vol. 12, pp. 107 599–107 615, 2024.

[15] G. Sahin, G. Isik, and W. van Sark, "Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression," *Energy Reports*, vol. 10, pp. 2837–2849, Nov. 2023.

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] S. Hart, "Shapley Value," in *Game Theory*, J. Eatwell, M. Milgate, and P. Newman, Eds. London: Palgrave Macmillan UK, 1989, pp. 210–216.