

Machine Learning-Based Medium-Term Power Forecasting of a Grid-Tied Photovoltaic Plant

Amal A. Hassan*, Doaa M. Atia, Hanaa T. El-Madany

Electronics Research Institute, Cairo, Egypt Email: *amal.elramly@eri.sci.eg

How to cite this paper: Hassan, A.A., Atia, D.M. and El-Madany, H.T. (2024) Machine Learning-Based Medium-Term Power Forecasting of a Grid-Tied Photovoltaic Plant. *Smart Grid and Renewable Energy*, **15**, 289-306.

https://doi.org/10.4236/sgre.2024.1512017

Received: November 16, 2024 Accepted: December 10, 2024 Published: December 18, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Abstract

Due to the variability and unpredictability of solar power, which relies heavily on weather variables such as solar irradiance and temperature, precise forecasting of photovoltaic (PV) energy production is crucial for effectively planning and operating power systems incorporating solar technology. Several machine learning algorithms (MLAs) have recently been developed for PV energy forecasting. This paper discusses various machine learning (ML) techniques for predicting the power output of a PV plant connected to the grid. Multiple algorithms, including linear regression (LR), neural networks (NNs), deep learning (DL), and k-nearest neighbors (k-NNs), are evaluated. The models use real-time data collected from various weather sensors and electrical output over a year, including solar irradiance, ambient temperature, wind speed, and cell temperature, to forecast PV power generation. Over a medium-term horizon, forecasting accuracy is assessed using datasets covering an entire week. The models are analyzed based on multiple performance metrics, such as absolute error (AE), root mean square error (RMSE), normalized absolute error (NAE), relative error (RE), relative root square error (RRSE), and correlation coefficient (R). The results indicate that the deep learning algorithm achieves the highest accuracy, with an RMSE of 0.026, an AE of 0.014, an NAE of 0.064, and an R of 99.7% for the weekly forecast validation. These precise forecasts produced in this research could assist grid operators in managing the variability of PV power output and planning to integrate fluctuating PV energy into the grid.

Keywords

Photovoltaic, Grid-Tied PV Systems, Power Forecasting, Machine Learning, Deep Learning

1. Introduction

Renewable energy sources (RESs), particularly solar photovoltaic (PV) power, have

gained significant attention in recent years because of their potential for long-term use without damaging the environment. However, integrating various but unpredictable RES is essential due to the irregularity of the energy market and the intermittent nature of these sources [1]. Utilizing energy forecasting can assist governmental bodies in developing, implementing, and modifying energy policies. Consequently, effectively incorporating and managing these resources into the current infrastructure has become a significant responsibility for the energy sector, especially in regions heavily dependent on climate-influenced energy supplies [1].

The variability in PV generation presents significant challenges for managing the existing power infrastructure [2]. As the integration of solar energy increases, this complexity and cost escalate, impacting decision-making regarding backup resources, scheduling, storage, and long-term strategies. The unpredictable output from solar panels complicates the alignment with electricity demand, leading to difficulties sustaining an immediate balance between supply and demand. Thus, achieving high levels of solar PV penetration requires substantial modifications in managing power systems and electricity trading [3] [4].

Accurate forecasting of solar power is crucial for tackling these challenges. Techniques such as numerical weather predictions, image analysis, statistical approaches, and hybrid neural network methods are employed to estimate solar irradiance and PV output [4]. Advancements in technology, mainly through artificial intelligence (AI) and machine learning (ML), are enhancing the precision and effectiveness of solar forecasting. AI and ML support power grid managers in making informed choices, planning operations, and optimizing energy market strategies, which helps to minimize the costs associated with managing the fluctuations of intermittent solar power [5] [6].

The relation between weather parameters and the output of PV systems is influenced by the specific location, shaped by the area's geographical and climatic characteristics. As a result, the extent of the relationship between weather conditions and PV energy production varies from one location to another. Nevertheless, the effectiveness of a forecasting model relies on the nature of the correlation between inputs and outputs, emphasizing the need for a customized approach for each site [7]. Various factors influence the precision of PV forecasts, making it a complex task. This complexity is affected by elements such as the range of forecasting, the inputs used in the forecast model, and the performance evaluation [8].

Several previous studies have established PV power forecasting models utilizing ML methods. In [9], the author employed deep learning (DL) techniques to estimate the generation of residential PV systems. Actual data was used to assess the forecasting accuracy of long- and short-term memory (LSTM), convolutional, and hybrid convolutional-LSTM networks over various horizons, compared with Prophet, based on MAE, RMSE, and NRMSE error metrics. In [10], the authors introduced a five-layer CNN-LSTM framework for PV power predictions using actual data from a site in Mexico. The findings indicated that the hybrid neural

network model produced superior predictions. In [11], the authors proposed a forecasting method reliant on particle swarm optimization. Evaluations conducted over four forecasting resolutions and time horizons demonstrated that the suggested method lowered the MASE. Authors in [12] developed a single ML model to predict the power output of a large distributed solar array comprising numerous PV systems. In [13], the authors employed support vector machine (SVM) models based on kernel techniques to evaluate rooftop PV potential across urban areas in Switzerland. In [14], the authors concentrated on hybrid renewable energy power forecasting by applying the SVM algorithm using actual sensor data. Their findings confirmed that big data analysis improves predictive accuracy for renewable energy sources by analyzing large datasets. Authors in [15] examined the accuracy, stability, and computational cost of two tree-based models, specifically extra trees (ETs) and random forest (RF), for hourly PV power predictions.

Authors in [16] analyzed different ML approaches with varying degrees of complexity to assess whether more complex models yield improved forecasting accuracy. They also examined the techniques of both online and offline training methods to find the most efficient strategy. In [17], the authors investigated and compared the efficacy of individual, ensemble, and hybrid ML models in forecasting solar PV output power over four distinct time frames (one day, one week, two weeks, and one month in advance). In [18], authors introduced an ensemble stacked ML model for hourly predictions of two photovoltaic systems that vary in size and age. The research evaluated three ML techniques, random forest, gradient boosting, and multiple linear regressions, against a baseline linear regression (LR) model and a reference model for predicting PV power.

Most prior studies focus on short-term forecasting, with limited attention to medium-term horizons crucial for operational planning and grid stability. Moreover, the influence of site-specific climatic conditions, such as those in arid regions like Cairo, still needs to be explored.

This paper introduces ML forecasting models to estimate the output power of a PV array in a GTPV plant. To evaluate this forecasting approach, different machine learning algorithms (MLAs) are utilized, such as LR, neural networks (NNs), DL, and the k-nearest neighbor (k-NN) method. Input data consists of meteorological variables like solar irradiance, ambient temperature, cell temperature, and wind speed, gathered on-site through a weather station linked to the PV array. The actual output power from the PV array, collected through data loggers at the GTPV plant, serves as a label for the ML models. All algorithms are trained on a comprehensive dataset that spans an entire year of high-frequency data captured at five-minute intervals, thus covering various seasonal conditions. Subsequently, the ML models are tested for medium-term predictions extending one week into the future. All models are implemented and tested using the RapidMiner software platform.

The main contribution of this paper is to evaluate the performance of different ML models for medium-term forecasting using a comprehensive dataset collected

from a grid-tied PV plant in Cairo, Egypt. By using a full year of high-resolution weather and power data, the study provides valuable insights into the impact of seasonal variability of meteorological parameters on PV power forecasting. Moreover, it proposes practical forecasting solutions for regions with similar climates.

The paper is organized into four sections: Section 2 describes the GTPV plant. Section 3 outlines the methodology, which includes a description of data collection, pre-processing, and the evaluation criteria used to assess the performance of the MLAs. Section 4 provides a thorough analysis and discussion of the findings. Finally, Section 5 summarizes the key insights and conclusions.

2. Grid-Tied PV Power Plant

The GTPV system is installed on the roof of the ERI building located in Cairo, Egypt, with geographic coordinates of 30°7'49.44"N latitude and 31°22'48"E longitude, oriented south (zero azimuth) at a fixed tilt angle of 26 degrees. The PV array system has an overall capacity of 30.26 kW, divided into two sub-arrays, each with a capacity of 15.13 kW. Each sub-array comprises 34 modules arranged in two parallel strings, with 17 modules connected in series. The PV module has a maximum power output of 445 W and has an efficiency rating of 20.4% under STC. These photovoltaic modules are linked to a three-phase inverter with a capacity of 27.6 kW; further details regarding the electrical characteristics of the photovoltaic modules, inverter, and plant layout can be found in Reference [19]. Meteorological information gathered from an on-site weather station, along with measurements of electrical parameters, was precisely recorded using a Solar-Log Base 100 data logger. This data acquisition unit regularly documented measurements, serving as the main monitoring device that connects the photovoltaic station to the weather sensors. High time-based resolution data was collected from both the PV system and weather monitors, sampled at intervals of 5 minutes. The parameters gathered included solar irradiance (W/m²), ambient temperature (°C), wind speed (m/s), cell temperature (°C), and the output power of the PV array.

3. Problem Formulation and Methodology

This research uses ML techniques to create medium-term PV power generation predictions for a GTPV power plant. Various categories of ML techniques and datasets are suggested to apply the forecasting model to guarantee their accuracy and dependability. The identified ML algorithms include neural-based techniques such as NN, and DL, regression-based approaches such as LR, and lazy-learning methods, k-NN. The modeling process is implemented using RapidMiner Studio's graphical user interface software. RapidMiner is a data analytics platform established in 2001 [20]. It provides a variety of operators and repositories for activities such as data preparation, transformation, modeling, and assessment. Its specific features include tools for process management and utilities, access to repositories, data importing/exporting, manipulation, and model construction. Its extensive toolset assists the complete data science workflow, including parameter tuning, model training validation, and performance assessment. **Figure 1** represents the proposed PV power forecasting framework, which integrates environmental data and advanced computational techniques to enhance prediction accuracy. The framework begins by collecting key meteorological inputs, including solar irradiation, wind speed, ambient temperature, and cell temperature. These inputs are processed using a neural network model comprising an input layer, hidden layers, and an output layer to extract patterns and relationships in the data. Weights and mathematical operations are applied within the network to generate preliminary forecasts. A subsequent decision-making module refines the predictions by employing distance calculation, value sorting, and selecting the optimal neural network model. The final output of the framework is a reliable and precise forecast of PV power aimed at optimizing system performance and supporting energy management strategies.



Figure 1. The proposed PV power forecasting framework.

Figure 2 illustrates the workflow utilized in this research, which consists of input, training, and forecasting stages. The input stage involves gathering data, which includes environmental factors and plant output, followed by data pre-processing. This critical step enhances the inputs before modeling to boost accuracy. MLAs are created during training based on the pre-processed historical input-output datasets. The models identify the correlations between inputs and targets to make forecasts. The forecasting phase enables the prediction of future PV power generation. With daily collection of new, real-time data, models are consistently updated and employed to provide medium-term power generation predictions. Once the raw data has been pre-processed, feature selection methods such as feature importance ranking and Pearson correlation analysis are employed to determine the most predictive independent variables for inclusion in the models. The processed data is split into training, validation, and test subsets as 70% training and 30% testing. MLAs are trained on the training dataset to identify patterns and fine-tune model parameters over multiple iterations iteratively. At the same time, hyperparameter tuning using the validation dataset aids in configuring elements of the model structure that are not adjusted during training, including the number of hidden layers in a neural network. The parameters of the MLAs are also tuned to enhance predictive precision. Finally, various error and correlation metrics are computed to assess and contrast the forecasting quality of the MLAs. These metrics included absolute error (AE), root mean square error (RMSE), normalized absolute error (NAE), relative error (RE), relative root square error (RRSE), and correlation coefficient (R).



Figure 2. Workflow of the proposed ML-based power forecasting of PV power.

The steps in the workflow are outlined as follows:

1) Data from the PV plant is gathered using the weather station and data logger, encompassing all meteorological data and the plant's output measurements over one year at 5-minute intervals.

2) RapidMiner software is utilized to implement the MLAs, and the data is retrieved for training and testing.

3) Outlier values are eliminated through data filtering.

4) Data normalization is conducted using the min-max scaling technique.

5) The processed data is divided into training and testing subsets, comprising 70% for training and 30% for testing.

6) ML models are employed to predict PV power output.

7) To confirm the model's accuracy, a validation model is applied to a new dataset with a medium-term horizon.

8) The model's performance metrics are assessed and displayed.

3.1. Feature Selection

Meteorological factors are crucial in determining the effectiveness of PV power forecasting since they directly affect energy generation. The primary input is solar irradiance, closely followed by ambient temperature, as both strongly correlate with PV output. The movement of clouds also causes sudden and significant fluctuations in PV power production. This research includes all essential meteorological factors that adequately represent site conditions as input features. Solar irradiance is a critical parameter due to its direct relationship with energy production. Ambient temperature influences efficiency even more at elevated levels.



Figure 3. Meteorological data recordings over a year: (a) Solar irradiance; (b) Ambient temperature; (c) Cell temperature.

This study's meteorological data encompasses an entire year, from October 2022 to September 2023, covering all seasons. The dataset consists of eight variables: date, time, radiation, ambient temperature, wind speed, cell temperature, and PV power. Measurements were taken using a logger at five-minute intervals, resulting in over 105,000 samples for each parameter. The original dataset is split into 70% training and 30% testing subsets. In the analysis, all meteorological factors affecting PV output are selected as input features; solar irradiance is identified as the critical variable, followed by ambient temperature, which has a notable impact on

performance at elevated temperatures, cell temperature, and wind speed. Figure 3 depicts the variations in meteorological parameters and cell temperature throughout the studied year. Solar irradiance ranges from 0 to 1000 W/m², with occasional slight spikes reaching up to 1200 W/m² during clear weather, particularly noticeable in April and May. Ambient temperature fluctuates between 10 and 44°C, with night-time lows around 10°C and daytime highs approaching 44°C. Cell temperature exhibits a range from 10 to 68°C. With consistently low values, wind speed ranges from 1 to 9 m/s, as Cairo typically experiences moderate to low wind conditions. Winter shows significant fluctuations in solar radiation and increased challenges in forecasting. Before modeling, this dataset undergoes pre-processing to improve the model's accuracy. This pre-processing involves two critical steps: outlier removal and data normalization. The outlier removal process eliminates abnormal readings, while data normalization standardizes the variables to uniform measurement units, making the data suitable for practical training. The appropriate data normalization method was selected based on the chosen ML technique.

3.2. Data Pre-Processing—Filtration

The input data for solar power generation and meteorological information necessitate pre-processing to enhance model accuracy and computational efficiency. The goal of pre-processing is to refine datasets before the development of ML models. A sequence of filtering procedures was implemented for both photovoltaic power output and meteorological data. Initially, implausible values such as negative figures, null entries resulting from sensor errors, or missing sensor data were eliminated. It streamlines the datasets to mitigate improper training complications and the computational costs associated with irregularities, outliers, and irrelevant inputs. Supplying complete high-resolution datasets, which include null PV values during nighttime, could decrease training and accuracy due to data sparsity. Filters were utilized to eliminate unlikely outliers and days with extensive missing data.

In this research, the approach involved discarding readings that exceeded the maximum rated capacity of the PV system. Furthermore, days that lacked sufficient data records were also filtered out. This approach aimed to lessen potential complications during training and the computational load caused by incomplete or improper inputs by meticulously cleaning and conditioning the datasets before the model development phase. The pre-processing steps focused on removing faulty inputs, tackling sparsity, isolating consistent patterns, and adjusting dataset characteristics for optimal ML.

Additionally, a filtration method was employed to eliminate negative irradiance values. The frequency of night-time sampling was reduced rather than entirely discarded. The sparse data during the night could result in ineffective training and reduced accuracy; since PV power is zero at night, the dataset for this period was diminished but not removed, as these points still signify the cyclical nature of solar irradiation absence.

3.3. Data Pre-Processing—Normalization

Data normalization is crucial for standardizing input features and the output label to consistent values. This study utilized range transformation to normalize data to a fixed scale of 0 to 1. The range normalization method involves rescaling the input data to fit within a smaller standard range derived from its original broader range. Precisely, all variables are adjusted to lie between 0 and 1. This method minimizes forecasting errors by limiting values to a narrow range while preserving the correlations between input parameters. It ensures that variables with naturally high values stay within those with smaller magnitudes. Normalizing data in this manner also enables MLAs to treat all features equally during training, enhancing the training speed and convergence. Variables that have been standardized to the same scale can then be directly compared regarding their relative significance. Mathematically, the min-max scalar modifies data following the formula [21]:

$$x_{s} = (x_{i} - x_{\min}) / (x_{\max} - x_{\min})$$
(1)

where x_s is the scaled value, x_i is the measured value, and x_{max} and x_{min} are the maximum and the minimum values of the dataset.

3.4. Machine Learning Models

3.4.1. Linear Regression (LR)

Regression analysis is a statistical technique employed for numerical forecasting. It measures the connection between a dependent variable (also known as the target variable) and various independent variables (regular attributes) [22] [23]. LR fits a straight line to represent the linear association between the dependent and independent variables. It describes the observed data using a linear equation [22] [23].

3.4.2. Neural Networks (NNs)

The architecture of the ML-based NN model utilized in this research is a feedforward multi-layer perceptron NN, which is trained using a back-propagation algorithm. A standard sigmoid activation function is applied to the nodes in this configuration. Furthermore, the output node employs a sigmoid or linear activation function based on the specific problem type. In this case, sigmoid output activation was chosen since the task involves forecasting or classifying a variable as a classification problem. Conversely, a linear activation would be more suitable for numeric regression tasks where the goal is to predict an exact target value [24]-[26].

3.4.3. Deep Learning (DL)

In RapidMiner, the DL algorithm utilizes H_2O optimization [20]. It is a multi-layer feed-forward artificial neural network, which is trained through stochastic gradient descent using back-propagation [27]-[29]. The network architecture has numerous hidden layers that feature neurons using tanh, rectifier, and max-out activation functions.

3.4.4. k-Nearest Neighbor (k-NN)

The k-NN algorithm makes predictions by analyzing the nearest training instances in a feature space. During the training phase, k-NN merely retains the feature vectors and their associated target values without making explicit generalizations [30]. In the case of regression, it calculates the average of the target variable values from the nearest neighbors to predict a continuous outcome. Closer neighbors may be given greater weight in influencing the prediction. Feature values are often normalized before calculating distances. One of the benefits of k-NN is its straightforward implementation and comprehension. A drawback, however, is that it requires significant memory to store all training examples and has high computational costs for classification [17] [31]-[34].

3.5. Performance Indices

Several statistical metrics are used to assess the effectiveness of the MLAs utilized in this study and evaluate their predictive performance. Among these metrics are absolute error (AE), root mean square error (RMSE), normalized absolute error (NAE), relative root square error (RRSE), and correlation coefficient (R). The Absolute Error is the RapidMiner software's standard function [35].

$$RMSE = \sqrt{\left(\frac{1}{n\sum\left(\left(\frac{yp-yi}{y}\right)}{yi}^{2}\right)}$$
(2)

$$AE = y_p - y_i \tag{3}$$

$$NAE = 1/n \sum |((yp - yi))/yi|$$
(4)

$$RE = \left(\left(yp - yi \right) \right) / yi \tag{5}$$

$$RRSE = \sqrt{\left(\left(\sum(yp - yi)^2\right) / \left(\sum(yi - \overline{y})^2\right)\right)}$$
(6)

where *yi* is the measured power value, and *yp* is the predicted power value.

4. Results and Discussion

4.1. ML Models Train and Test

This research employed weather data and cell temperature to train various MLAs for forecasting PV power. Our primary objective was to reduce the absolute error as the main metric. The model was validated using actual PV power measurements from October 1, 2022, to September 30, 2023 (from a 30.26 kW GTPV plant). **Figure 4** illustrates the evaluation metrics for the tested MLAs. This bar chart compares the performance of four ML models: LR, NN, DL, and k-NN— across various error metrics: RMSE, RE, AE, RRSE, and NAE. RMSE reflects the average magnitude of prediction errors. LR has the highest RMSE, indicating it performs the worst in minimizing prediction errors. DL shows the lowest RMSE, highlighting it as the most accurate model for predicting PV power. RE is the ratio of error to the actual value. NN exhibits a significantly higher RE than other models, indicating poor performance. DL and k-NN maintain lower RE values,

indicating better consistency. AE is the absolute difference between predicted and actual values. DL achieves the lowest AE, emphasizing its substantial predictive accuracy. LR and NN have higher AE values, demonstrating weaker performance in capturing precise predictions. RRSE normalizes RMSE relative to the variance of the data. LR has the highest RRSE, confirming its inability to generalize across the dataset effectively. DL again performs the best, demonstrating robust generalization. NAE quantifies the absolute error relative to the scale of the data. DL and NN achieve similar, low NAE values, suggesting strong predictive capabilities concerning data scaling. LR and k-NN show higher NAE values, implying less reliable performance in handling variability. DL consistently outperforms the other methods across most metrics, emphasizing its ability to handle complex relationships in PV power forecasting. LR consistently ranks lowest across all metrics, likely due to its linear nature, which is insufficient for capturing non-linear patterns in PV power generation. NN performs well in RMSE and AE but falters with a high RE, indicating it might struggle with certain aspects of error normalization or data variability. k-NN offers moderate accuracy but is less effective than DL, particularly in metrics like AE and NAE. The figure shows that the DL algorithm recorded the lowest values for RMSE, 0.022, AE, 0.015, and RRSE, 0.076. The results highlight the importance of selecting advanced models (like DL) for PV power forecasting due to their superior ability to handle non-linearity and complexity in data.

Figure 5 compares the predicted annual PV and AC power output with actual values, including environmental factors. This figure pertains to the DL algorithm, which demonstrated the best performance. The figure demonstrates a positive correlation between solar irradiation (W/m^2) and PV power (W). As solar radiation increases, the PV power output also increases linearly, which is expected given that solar radiation is the primary driver of PV power generation. The predicted PV powers closely follow the actual values, indicating that the forecasting model accurately captures the relationship between solar radiation and PV power. The correlation is 99.7%, indicating high predictive accuracy.



Figure 4. Performance indices of the MLAs.



Figure 5. Actual and predicted PV power based on DL algorithm vs solar irradiance and ambient temperature.

The relationship between ambient temperature (°C) and PV power output is more complex than solar radiation. PV power increases with ambient temperature initially but declines beyond a certain point. This behavior reflects the combined effects of temperature on PV module efficiency (e.g. higher temperatures reducing efficiency). Predicted PV power values broadly follow the trend of actual values, but the scatter is more pronounced, particularly at higher temperatures. This increased scatter indicates that the model may not fully capture the non-linear effects of temperature on PV power output.

4.2. Weekly Forecasting Validation

This section examines the effectiveness of various MLAs (LR, NN, DL, and k-NN) for medium-term forecasting. The model utilizes validation data from a week during January. **Figure 6** illustrates the meteorological data for the chosen week. The solar irradiance follows a daily pattern, peaking during midday and decreasing to zero at night. The curve demonstrates consistent cyclic behavior across seven days during January 2024, indicating clear-sky conditions with minimal variability. Fluctuations within the peaks may correspond to transient weather events, such as clouds representing the winter climate in Cairo. Solar irradiance is the primary input for PV power generation, and this plot confirms the strong dependence on time and weather patterns. Two temperature profiles are shown: ambient temperature and

cell temperature. Both follow a similar daily pattern (24-hour), rising during the day and decreasing at night. The cell temperature consistently exceeds the ambient temperature, with a more pronounced peak during midday. The ambient temperature curve exhibits less variation compared to the cell temperature curve, suggesting that PV module heating is highly influenced by solar intensity. Wind speed values fluctuate significantly over time (**Figure 5(c)**), ranging from 0 to 6 m/s. The pattern does not show a clear periodic or regular trend; rather, it highlights irregular fluctuations typical of natural wind patterns.

Meanwhile, **Figure 7** presents all models' forecasted PV power outputs over a week, representing the medium-term forecasting timeframe. The prediction dataset utilized in this evaluation comprised 2016 data points at 5-minute intervals. All models show a nearly linear relationship between solar irradiance and PV power, aligning with photovoltaic systems' fundamental physics. The data points predicted by all models closely follow the actual trend of PV power with solar irradiance, indicating good model performance. The clustering of points along the diagonal line suggests minimal deviation between predicted and actual power values, especially at higher irradiance levels.



Figure 6. Weather data for the selected week: (a) Solar irradiation; (b) Temperature (cell and ambient); (c) Wind speed.

Figure 7. PV power forecasting for a week of different MLAs.

Figure 8 presents the average evaluation metrics for the MLAs based on the weekly validation data. The evaluation of the ML models utilized for weekly aggregated data showed remarkable enhancements in prediction accuracy. Among the models assessed, the DL model performance exceeded the performance of the others, attaining a remarkable level of precision with the following metrics: an RMSE of 0.026, an AE of 0.014, an RRSE of 0.102, an NAE of 0.064, and a correlation coefficient of 99.7%. These findings highlight the DL model's capability to effectively manage the non-linear and intricate relationships between meteorological factors and PV power output. The RMSE of 0.026 indicates a very low average deviation, while the low AE of 0.014 reflects the model's proficiency in matching predictions with actual data. The RRSE and NAE values (10.2% and 6.4%, respectively) also illustrate the DL model's strength, surpassing traditional models such as LR and k-NN. The correlation coefficient of 99.7% signifies an almost flawless correspondence between predicted and actual values, further highlighting the model's accuracy.

Figure 8. Performance metrics of the ML models for a week.

Figure 9 compares the actual and predicted PV power output generated by the

DL model. The figure shows a repeating daily pattern corresponding to solar power generation cycles, with peak values occurring during midday and tapering off towards the evening and early morning. The close alignment between the actual and predicted values demonstrates the high accuracy of the deep learning model in forecasting PV power. The model successfully captures the variability and trends of solar power generation across multiple days. Minor discrepancies may exist during rapid transitions, likely due to environmental factors. Overall, the figure highlights the effectiveness of the proposed forecasting framework in predicting PV power outputs.

Figure 9. Actual and predicted PV power of DL model.

5. Conclusion

This research established forecasting models to predict the power output from a GTPV plant by employing various ML techniques alongside meteorological data for medium-term forecasting. The models were trained using a high-frequency dataset encompassing one year of actual performance data, deployed and evaluated within the RapidMiner platform, and validated through application to weekly data measurements. Key findings highlight that both solar irradiance and ambient temperature significantly impact the power output of PV systems. The DL model achieved the highest forecasting accuracy, recording the lowest RMSE of 0.022, AE of 0.015, RRSE of 0.076, NAE of 0.064, and R of 99.7% for the weekly forecasting validation. Validation for medium-term forecasting confirmed the accuracy of the models; when utilizing weekly data, the models exhibited good precision, with the DL model attaining the highest performance, achieving an RMSE of 0.026, AE of 0.014, RRSE of 0.102, NAE of 0.064, and R of 99.7%. Overall, the models displayed effective performance across varying weather scenarios. These findings highlight the potential of DL for medium-term PV power forecasting. Its high accuracy makes it particularly suitable for operational decision-making in grid management and energy trading, where reliable predictions are critical for balancing supply and demand. Moreover, it can help grid operators prepare for fluctuations in PV power output, thus enabling better integration of variable solar energy into the power grid.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Cabello-López, T., Carranza-García, M., Riquelme, J.C. and García-Gutiérrez, J. (2023) Forecasting Solar Energy Production in Spain: A Comparison of Univariate and Multivariate Models at the National Level. *Applied Energy*, **350**, Article ID: 121645. <u>https://doi.org/10.1016/j.apenergy.2023.121645</u>
- [2] Cantillo-Luna, S., Moreno-Chuquen, R., Celeita, D. and Anders, G. (2023) Deep and Machine Learning Models to Forecast Photovoltaic Power Generation. *Energies*, 16, Article 4097. <u>https://doi.org/10.3390/en16104097</u>
- [3] Dawan, P., Sriprapha, K., Kittisontirak, S., Boonraksa, T., Junhuathon, N., Titiroongruang, W., et al. (2020) Comparison of Power Output Forecasting on the Photovoltaic System Using Adaptive Neuro-Fuzzy Inference Systems and Particle Swarm Optimization-Artificial Neural Network Model. *Energies*, 13, Article 351. https://doi.org/10.3390/en13020351
- [4] Dimd, B.D., Voller, S., Cali, U. and Midtgard, O. (2022) A Review of Machine Learning-Based Photovoltaic Output Power Forecasting: Nordic Context. *IEEE Access*, 10, 26404-26425. <u>https://doi.org/10.1109/access.2022.3156942</u>
- [5] Munawar, U. and Wang, Z. (2020) A Framework of Using Machine Learning Approaches for Short-Term Solar Power Forecasting. *Journal of Electrical Engineering & Technol*ogy, 15, 561-569. <u>https://doi.org/10.1007/s42835-020-00346-4</u>
- [6] Vennila, C., Titus, A., Sudha, T.S., Sreenivasulu, U., Reddy, N.P.R., Jamal, K., et al. (2022) Forecasting Solar Energy Production Using Machine Learning. *International Journal* of Photoenergy, 2022, Article ID: 7797488. <u>https://doi.org/10.1155/2022/7797488</u>
- [7] Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., et al. (2018) Forecasting of Photovoltaic Power Generation and Model Optimization: A Review. *Renewable and Sustainable Energy Reviews*, 81, 912-928. <u>https://doi.org/10.1016/j.rser.2017.08.017</u>
- [8] Ahmed, R., Sreeram, V., Mishra, Y. and Arif, M.D. (2020) A Review and Evaluation of the State-Of-The-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renewable and Sustainable Energy Reviews*, **124**, Article ID: 109792. <u>https://doi.org/10.1016/j.rser.2020.109792</u>
- [9] Costa, R.L.D.C. (2022) Convolutional-lstm Networks and Generalization in Forecasting of Household Photovoltaic Generation. *Engineering Applications of Artificial Intelligence*, **116**, Article ID: 105458. <u>https://doi.org/10.1016/j.engappai.2022.105458</u>
- [10] Tovar, M., Robles, M. and Rashid, F. (2020) PV Power Prediction, Using CNN-LSTM Hybrid Neural Network Model. Case of Study: Temixco-Morelos, México. *Energies*, 13, Article 6512. <u>https://doi.org/10.3390/en13246512</u>
- [11] Perera, M., De Hoog, J., Bandara, K. and Halgamuge, S. (2022) Multi-Resolution, Multi-Horizon Distributed Solar PV Power Forecasting with Forecast Combinations. *Expert Systems with Applications*, **205**, Article ID: 117690. https://doi.org/10.1016/j.eswa.2022.117690
- [12] Grzebyk, D., Alcañiz, A., Donker, J.C.B., Zeman, M., Ziar, H. and Isabella, O. (2023) Individual Yield Nowcasting for Residential PV Systems. *Solar Energy*, 251, 325-336.

https://doi.org/10.1016/j.solener.2023.01.036

- [13] Assouline, D., Mohajeri, N. and Scartezzini, J. (2017) Quantifying Rooftop Photovoltaic Solar Energy Potential: A Machine Learning Approach. *Solar Energy*, 141, 278-296. <u>https://doi.org/10.1016/j.solener.2016.11.045</u>
- [14] Preda, S., Oprea, S., Bâra, A. and Belciu (Velicanu), A. (2018) PV Forecasting Using Support Vector Machine Learning in a Big Data Analytics Context. *Symmetry*, 10, Article 748. <u>https://doi.org/10.3390/sym10120748</u>
- [15] Ahmad, M.W., Mourshed, M. and Rezgui, Y. (2018) Tree-Based Ensemble Methods for Predicting PV Power Generation and Their Comparison with Support Vector Regression. *Energy*, **164**, 465-474. <u>https://doi.org/10.1016/j.energy.2018.08.207</u>
- [16] Ferlito, S., Adinolfi, G. and Graditi, G. (2017) Comparative Analysis of Data-Driven Methods Online and Offline Trained to the Forecasting of Grid-Connected Photovoltaic Plant Production. *Applied Energy*, 205, 116-129. <u>https://doi.org/10.1016/j.apenergy.2017.07.124</u>
- [17] Asiedu, S.T., Nyarko, F.K.A., Boahen, S., Effah, F.B. and Asaaga, B.A. (2024) Machine Learning Forecasting of Solar PV Production Using Single and Hybrid Models over Different Time Horizons. *Heliyon*, 10, e28898. <u>https://doi.org/10.1016/j.heliyon.2024.e28898</u>
- [18] Abdelmoula, I.A., Elhamaoui, S., Elalani, O., Ghennioui, A. and Aroussi, M.E. (2022) A Photovoltaic Power Prediction Approach Enhanced by Feature Engineering and Stacked Machine Learning Model. *Energy Reports*, 8, 1288-1300. https://doi.org/10.1016/j.egyr.2022.07.082
- [19] Hassan, A.A., Atia, D.M., El-Madany, H.T. and Eliwa, A.Y. (2024) Performance Assessment of a 30.26 Kw Grid-Connected Photovoltaic Plant in Egypt. *Clean Energy*, 8, 120-133. <u>https://doi.org/10.1093/ce/zkae074</u>
- [20] Ralf, K. and Mierswa, I. (2014) RapidMiner Studio Manual. https://docs.rapidminer.com/downloads/RapidMiner-v6-user-manual.pdf
- [21] Luo, X., Zhang, D. and Zhu, X. (2021) Deep Learning Based Forecasting of Photovoltaic Power Generation by Incorporating Domain Knowledge. *Energy*, 225, Article ID: 120240. <u>https://doi.org/10.1016/j.energy.2021.120240</u>
- [22] Alizamir, M., Kim, S., Kisi, O. and Zounemat-Kermani, M. (2020) A Comparative Study of Several Machine Learning Based Non-Linear Regression Methods in Estimating Solar Radiation: Case Studies of the USA and Turkey Regions. *Energy*, **197**, Article ID: 117239. <u>https://doi.org/10.1016/j.energy.2020.117239</u>
- [23] Thaker, J. and Höller, R. (2024) Hybrid Model for Intra-Day Probabilistic PV Power Forecast. *Renewable Energy*, 232, Article ID: 121057. https://doi.org/10.1016/j.renene.2024.121057
- [24] Talayero, A.P., Melero, J.J., Llombart, A. and Yürüşen, N.Y. (2023) Machine Learning Models for the Estimation of the Production of Large Utility-Scale Photovoltaic Plants. *Solar Energy*, 254, 88-101. <u>https://doi.org/10.1016/j.solener.2023.03.007</u>
- [25] Huang, Q. and Wei, S. (2020) Improved Quantile Convolutional Neural Network with Two-Stage Training for Daily-Ahead Probabilistic Forecasting of Photovoltaic Power. *Energy Conversion and Management*, **220**, Article ID: 113085. <u>https://doi.org/10.1016/j.enconman.2020.113085</u>
- [26] du Plessis, A.A., Strauss, J.M. and Rix, A.J. (2021) Short-Term Solar Power Forecasting: Investigating the Ability of Deep Learning Models to Capture Low-Level Utility-Scale Photovoltaic System Behaviour. *Applied Energy*, 285, Article ID: 116395. <u>https://doi.org/10.1016/j.apenergy.2020.116395</u>

- [27] Zhou, H., Zheng, P., Dong, J., Liu, J. and Nakanishi, Y. (2024) Interpretable Feature Selection and Deep Learning for Short-Term Probabilistic PV Power Forecasting in Buildings Using Local Monitoring Data. *Applied Energy*, **376**, Article ID: 124271. <u>https://doi.org/10.1016/j.apenergy.2024.124271</u>
- [28] Shabbir, N., Kütt, L., Astapov, V., Daniel, K., Jawad, M., Husev, O., *et al.* (2024) Enhancing PV Hosting Capacity and Mitigating Congestion in Distribution Networks with Deep Learning Based PV Forecasting and Battery Management. *Applied Energy*, **372**, Article ID: 123770. <u>https://doi.org/10.1016/j.apenergy.2024.123770</u>
- [29] Nguyen Trong, T., Vu Xuan Son, H., Do Dinh, H., Takano, H. and Nguyen Duc, T. (2023) Short-Term PV Power Forecast Using Hybrid Deep Learning Model and Variational Mode Decomposition. *Energy Reports*, 9, 712-717. <u>https://doi.org/10.1016/j.egyr.2023.05.154</u>
- [30] Dong, Y., Ma, X. and Fu, T. (2021) Electrical Load Forecasting: A Deep Learning Approach Based on K-Nearest Neighbors. *Applied Soft Computing*, 99, Article ID: 106900. https://doi.org/10.1016/j.asoc.2020.106900
- [31] Shijer, S.S., Jassim, A.H., Al-Haddad, L.A. and Abbas, T.T. (2024) Evaluating Electrical Power Yield of Photovoltaic Solar Cells with K-Nearest Neighbors: A Machine Learning Statistical Analysis Approach. *e-Prime—Advances in Electrical Engineering, Electronics and Energy*, 9, Article ID: 100674. https://doi.org/10.1016/j.prime.2024.100674
- [32] Al-Dahidi, S., Hammad, B., Alrbai, M. and Al-Abed, M. (2024) A Novel Dynamic/Adaptive K-Nearest Neighbor Model for the Prediction of Solar Photovoltaic Systems' Performance. *Results in Engineering*, 22, Article ID: 102141. https://doi.org/10.1016/j.rineng.2024.102141
- [33] Abubakar Mas'ud, A. (2022) Comparison of Three Machine Learning Models for the Prediction of Hourly PV Output Power in Saudi Arabia. *Ain Shams Engineering Journal*, **13**, Article ID: 101648. <u>https://doi.org/10.1016/j.asej.2021.11.017</u>
- [34] Liu, W., Shen, Y., Aungkulanon, P., Ghalandari, M., Le, B.N., Alviz-Meza, A., et al. (2023) Machine Learning Applications for Photovoltaic System Optimization in Zero Green Energy Buildings. Energy Reports, 9, 2787-2796. <u>https://doi.org/10.1016/j.egyr.2023.01.114</u>
- [35] Kuo, W., Chen, C., Hua, S. and Wang, C. (2022) Assessment of Different Deep Learning Methods of Power Generation Forecasting for Solar PV System. *Applied Sciences*, 12, Article 7529. <u>https://doi.org/10.3390/app12157529</u>