IJEEC

International Journal of Electrical Engineering and Computing
Vol. 9, No. 2 (2025)

Original research paper
UDC
DOI

Short-term forecasting of PV power plants within the framework of green digitalization using Linear regression and Random forest models

Božidar Popović[1], Ana Lojić[2]

1 University of East Sarajevo, Faculty of Electrical Engineering, East Sarajevo, Republic of Srpska, Bosnia and Herzegovina

2 Faculty of Information Technologies, International Burch University, Sarajevo, Bosnia and Herzegovina

E-mail address: bozidar.popovic@etf.uues.rs.ba, ana.lojic@stu.ibu.edu.ba

Abstract—**This paper presents the application of various methods for predicting electricity generation in photovoltaic (PV) power plants using real experimental data obtained from the measurement of meteorological and operational parameters over a three-day period. The objective of the study was to develop a reliable and interpretable model for short-term prediction of PV system output power based on a limited set of available data. The research applied a linear regression model, multiple linear regression, and a Random Forest regression model. The models were trained using data from the first two days of measurement, while the third day was used for testing accuracy and verifying model performance. The input parameters included solar radiation, module temperature, wind speed, and ambient temperature, while the target variable was the measured output power of the power plant expressed in megawatts. The results show that both linear and multivariable linear regression achieved a high level of agreement between measured and predicted values, with multiple linear regression reaching an R^2 of approximately 0.97, indicating that it explains about 97% of the variations in output power. However, the Random Forest model demonstrated superior performance, achieving an R^2 of about 0.975 on the test set, due to its ability to model complex and nonlinear relationships between meteorological parameters and power generation. The analysis confirms that even from a limited three-day dataset, it is possible to build a stable, robust, and accurate model for short-term PV power output prediction. The Random Forest model proved to be the most reliable solution for this type of task, while multiple linear regression provided a simple and efficient baseline approximation suitable for rapid implementation in real-time solar energy monitoring and management systems.**

Keywords- photovoltaic power plants, power generation forecasting, linear regression, multiple linear regression, Random Forest, short-term prediction, meteorological parameters.

I. INTRODUCTION

In today's world, where the share of electricity generated from renewable energy sources is measured not only daily but hourly, the issue of accurate, reliable, and precise forecasting of photovoltaic (PV) power plant output has become one of the key challenges in modern power systems based on renewable energy [1][2]. Its importance stems from the need to ensure reliable, stable, and economically viable operation of energy networks that increasingly depend on sources whose production varies with changing meteorological conditions.

Since solar energy depends on factors such as solar radiation temperature, cloud cover, wind speed, and air humidity, electricity generation from PV systems exhibits a high degree of variability and uncertainty [3]. Precisely because of this unpredictability, the development of accurate production forecasting models has become a fundamental prerequisite for efficient energy management and successful integration of renewable sources into the power grid.

Accurate estimation of future electricity production enables timely operational planning and system balancing, which includes optimizing available capacities, managing backup energy sources, and coordinating with consumption. In this way, forecasting contributes to reducing operational costs and improving energy efficiency, as it allows decision-making based on reliable and up-to-date data. In practice, this reduces the need to engage fossil reserves during periods of reduced PV output, directly contributing to the decarbonization of the energy sector.

Moreover, accurate generation forecasting has strategic importance for the development of power grids and electricity markets. In a liberalized market, where electricity is bought and sold at dynamic prices, the ability to accurately predict generation allows producers to optimize their offers, avoid

imbalance penalties, and increase business profitability. At the same time, it provides grid operators with a foundation for long-term infrastructure planning, including energy storage systems, flexible consumers, and smart grids that enable adaptive, real-time energy management.

In the context of the growing green digital transformation, forecasting PV output power transcends traditional energy analytics and becomes an integral part of digital systems based on artificial intelligence (AI), machine learning (ML), and the Internet of Things (IoT). Modern platforms for real-time data acquisition and processing, combined with predictive algorithms, enable the modeling of complex relationships between meteorological variations and PV output power. This results in a high level of automation and intelligent decision, making crucial for the stable operation of future smart energy networks.

Thus, the importance of forecasting electricity production from PV power plants extends beyond the technical functioning of the system, encompassing economic, environmental, and strategic dimensions. It represents a key link connecting sustainable energy production, power system stability, and the goals of the energy transition. In this sense, the development of reliable and adaptive forecasting methods—which combine meteorological data, sensor-based measurements, and analytical models—forms the foundation for the future development of sustainable, digitally managed energy ecosystems.

II. FORECASTING METOD IN POWER GENERATION MONITORING SYSTEMS

In power generation monitoring systems, especially those applied to photovoltaic (PV) power plants, forecasting methods represent a crucial analytical component used to enhance operational performance, system stability, and overall energy efficiency.
The main purpose of these methods is to estimate the current or future output power of a power plant based on available meteorological (such as solar irradiance, temperature, wind speed, humidity) and operational parameters (system status, inverter performance, historical generation data).

Forecasting methods can generally be categorized into three principal groups:

- Empirical (statistical) Methods – based on historical data and mathematical correlations between meteorological variables and power output. These include regression analysis, autoregressive models (AR, ARIMA), and exponential smoothing techniques.
- Machine Learning (ML) Methods – capable of modeling nonlinear and complex dependencies between input and output variables. Typical examples include deciding trees, random forests, support vector machines (SVM), and gradient boosting methods such as XGBoost or LightGBM.
- Deep Learning (DL) Methods – which employ advanced neural network architectures such as convolutional neural networks (CNN) and recurrent networks (RNN, LSTM) to capture both temporal and spatial patterns in data, enabling adaptive learning and long-term forecasting.

By integrating these approaches, modern forecasting systems achieve higher accuracy, robustness, and adaptability, supporting the intelligent management of renewable energy resources within smart grid infrastructures.

A. Empirical and statistical methods

Empirical and statistical methods represent the oldest approach in the field of electric power generation forecasting, particularly in the context of photovoltaic power plants. These methods are based on mathematical modeling of the relationship between input meteorological parameters (such as solar irradiance, temperature, wind speed, air humidity, and atmospheric pressure) and the output electrical power of a photovoltaic system. Their main idea is to derive a functional dependence between variables from historical data to enable the estimation of future production values.

The most used approaches within this category include:

- Linear regression, which models a proportional relationship between a single independent and a dependent variable;
- Multiple linear regression, which considers several meteorological input factors to achieve a more detailed system description;
- Autoregressive models (AR, ARMA, ARIMA), which capture temporal dependencies between past and future production values;
- Exponential smoothing methods, which assign greater weight to recent data and are used for short-term forecasts where rapid changes in meteorological conditions are expected.

One of the key advantages of empirical methods lies in their simplicity of implementation and interpretability. These models require relatively few input parameters, and their interpretation is intuitive and transparent. For this reason, they have been widely used in the early stages of developing forecasting systems for power generation, as well as in situations where only limited data sets are available.

However, the main limitation of empirical methods is their inability to accurately model complex and nonlinear relationships between input and output variables. For instance, linear regression may adequately describe the general trend between solar irradiance and generated power, but it fails to account for nonlinear effects that significantly influence PV system performance. These effects include:

- The influence of module temperature, which substantially reduces conversion efficiency at higher temperatures;
- The shading effect, where even a small shaded area can cause a disproportionately large drop in output power;
- Thermal degradation and panel aging, which over time alter the relationship between input parameters and output power;
- Nonlinear losses in inverters and cables, which further reduce the accuracy of linear models.

Due to these limitations, empirical methods are most used today as baseline models or reference systems for comparing the performance of more advanced approaches based on machine learning and deep neural networks. They are particularly useful in the initial stages of research, where the fundamental dependence between meteorological factors and energy production is analyzed, as well as in cases where a quick, approximate estimation is needed without complex computation.

Moreover, empirical methods still play a significant role in hybrid forecasting systems, where they are combined with adaptive algorithms that correct their linear assumptions. For example, an empirical model may serve as an initial estimate that is subsequently refined by machine learning techniques based on real-time measurements and weather conditions.

B. Machine learning methods

With the development of advanced algorithms, increased availability of meteorological and operational data, and the growth of computational power, machine learning methods have become the dominant approach in the field of electricity generation forecasting, particularly for photovoltaic power plants. Unlike empirical and classical statistical methods, which assume predefined linear relationships, machine learning enables adaptive learning of complex and nonlinear relationships between input parameters (such as meteorological conditions) and output power [8].

The fundamental advantage of the ML approach lies in its ability to automatically adapt to changes in data without the need for explicit mathematical modeling of physical processes. In this way, these algorithms can detect hidden patterns and interdependencies between variables that would be difficult or impossible to identify using conventional methods.

1) Decision Trees and Random Forest

Decision Trees (DT) represent a fundamental machine learning model used for both classification and regression tasks. They operate by successively splitting the dataset according to criteria that minimize prediction error. However, a single decision tree is often prone to overfitting, meaning it can become overly tailored to the training data and lose generalization capability.

To address this limitation, the Random Forest (RF) method is used an ensemble technique that combines a large number of decision trees, where each tree is trained on a different subset of the data. The final prediction is obtained through aggregation (averaging in regression or majority voting in classification), which significantly reduces variance and increases model stability.

Random Forest has proven to be highly effective for short-term forecasting of electricity generation, as it accurately models nonlinear relationships between meteorological parameters (such as solar irradiance, temperature, and wind speed) and power output, while requiring minimal data preprocessing.

Additionally, RF models are robust to noise and outliers, making them well-suited for real-world energy applications where measurements can be incomplete or unstable.

2) Support Vector Machines (SVM)

Support Vector Machines represent another important approach in power generation forecasting, particularly effective when dealing with smaller datasets. The core idea of SVM is to find a hyperplane that best separates data into classes (in classification tasks) or defines an optimal regression function with minimal error (in regression tasks).

In the context of photovoltaic systems, SVM models are used for nonlinear regression between meteorological input variables and power output. By applying kernel functions (such as polynomial, radial basis, or sigmoid kernels), the SVM model can transform the data into a higher-dimensional space, enabling it to learn complex relationships between temperature, irradiance, and energy production.

The main advantage of SVM models lies in their high accuracy and stability when trained on small datasets. However, their computational complexity increases significantly with larger datasets, making them less suitable for real-time monitoring systems that process large volumes of continuous data.

3) Gradient Boosting, XGBoost, and LightGBM

Ensemble techniques based on the boosting principle, such as Gradient Boosting (GB), XGBoost (Extreme Gradient Boosting), and LightGBM (Light Gradient Boosting Machine), currently represent the industry standard in the field of predictive analytics.

These models work by sequentially training a series of simple models (most often decision trees), where each subsequent model corrects the errors of its predecessors. This iterative process creates a strong predictive model with very low error and high generalization capability.

XGBoost and LightGBM are advanced implementations of this approach — XGBoost is known for its stability and accuracy, while LightGBM achieves significantly higher computational speed, making it well-suited for real-time forecasting in systems that continuously collect meteorological data.

These methods allow for a high degree of optimization and fine-tuning of hyperparameters, enabling exceptional model accuracy and the ability to detect complex nonlinear interactions between input variables.

Machine learning methods enable the integration of a wide range of parameters that influence electricity generation, such as: solar radiation (global, direct, and diffuse), ambient and module temperature, wind speed and direction, relative humidity, atmospheric pressure, historical data on system production and losses.

Compared to traditional methods, ML algorithms can autonomously learn complex nonlinear patterns from data, adapt to changes in weather conditions, and continuously improve performance through the retraining process. This enables high forecasting accuracy even under conditions of variable cloud cover, seasonal variations, and stochastic fluctuations in solar irradiance.

C. Hybrid and Intelligent Systems

The latest trend in the field of electricity generation forecasting, particularly in photovoltaic power plants, is the development of hybrid and intelligent systems that combine the advantages of various methods: physical, statistical, machine learning, and deep learning approaches into unified, adaptive models. These systems are based on the idea that no single method can fully capture all aspects of the complex and dynamic behavior of PV systems. Therefore, by integrating complementary approaches, greater accuracy, robustness, and generalization capability can be achieved.

1) Basic Concept of Hybrid Systems

Hybrid models combine multiple layers of analysis: physical, analytical, and intelligent with the goal of achieving a more comprehensive understanding of the process of converting solar energy into electrical energy. Fundamentally, such a system can simultaneously employ:

- a physical model (based on the laws of thermodynamics and the photoelectric effect) to describe the behavior of PV modules,
- statistical models for quantitative analysis and noise filtering in the data,
- intelligent models (e.g., machine learning and deep learning) for nonlinear mapping of input meteorological and operational parameters to output power.

By combining these layers, hybrid systems enable enhanced predictive accuracy and resilience to extreme weather conditions, as well as self-learning adaptation to operational changes in the system over time.

2) "Grey-box" Models

One of the most well-known approaches within hybrid systems is the so-called "grey-box" model (Combination of Physical and ML Approaches*)*. Unlike traditional "black-box" models (e.g., pure ML algorithms that lack physical understanding of the system) and "white-box" models (fully physical models), grey-box models combine both concepts. In this approach:

- The physical model defines the fundamental relationships between irradiance, temperature, and the electrical efficiency of PV modules;
- While machine learning models the nonlinear components and residual errors that the physical model cannot accurately describe.

In this way, a balance between interpretability and predictive power is achieved. These models are particularly useful for real PV systems, where local factors (such as shading, panel soiling, or microclimatic variations) are difficult to quantify but have a significant impact on energy production.

3) Deep Hybrid Models: CNN–LSTM and Transformer Architectures

Another important direction in the development of hybrid systems is the integration of deep neural networks, particularly the combination of Convolutional Neural Networks and Long Short-Term Memory architectures. CNN layers are used for: extracting spatio temporal patterns from data related to solar irradiance, temperature, and wind, while LSTM layers: model temporal dependencies and long-term production trends [7].

This integrated CNN–LSTM architecture enables the system to simultaneously learn local variations in time and space (e.g., passing clouds) and broader temporal patterns (e.g., seasonal changes).

In addition to CNN–LSTM combinations, Transformer architectures models originally developed in the field of Natural Language Processing (NLP) are increasingly being applied. These models can analyze long temporal sequences and relationships between distant points in time series data. In the context of solar energy, Transformers enable more accurate long-term production forecasts and the detection of complex temporal patterns in meteorological and operational data.

4) Intelligent Feedback-Based Systems

Modern hybrid systems increasingly incorporate elements of autonomous control through the implementation of feedback loops. These systems not only forecast future energy production but also, in real time:

- Analyze deviations between predicted and actual production;
- Adapt the model based on new incoming data;
- Optimize the operation of the photovoltaic plant (e.g., panel orientation, inverter control, or power distribution to storage units).

These adaptive systems employ a Reinforcement Learning approach, where the algorithm continuously interacts with the environment and "learns" which actions lead to optimal outcomes in terms of energy efficiency and grid stability.

The advantages of hybrid and intelligent systems are reflected in:

- Robustness under unstable and rapidly changing weather conditions;
- Flexibility in application across different geographic locations and types of power plants;
- Adaptive learning capability through continuous model updates with new data;
- Increased forecasting accuracy, achieved by combining multiple sources of information.

However, their implementation requires:

- Large volumes of high-quality data (meteorological, operational, and historical);
- High computational power, particularly during the model training phase;
- Precise parameter calibration to avoid overfitting.

Despite these challenges, the advantages in accuracy and stability make hybrid approaches the most reliable solution for

forecasting energy production in complex and dynamic environments.

Hybrid and intelligent systems represent an evolutionary step in the development of predictive technologies in the energy sector.

Their ability to integrate physical knowledge, statistical principles, and intelligent algorithms enables the creation of autonomous, self-learning, and highly efficient models that can adapt to real operating conditions of photovoltaic power plants [9][10][11].

In the context of green digital transformation, these systems play a crucial role, as they enable intelligent energy management, accurate production planning, and optimized resource utilization all aimed at achieving a sustainable, reliable, and low-carbon energy future.

III. MODEL DEVELOPMENT

Based on the collected experimental measurements of meteorological and operational parameters over a three-day period (Fig.1.), a model for predicting the output power of a photovoltaic (PV) power plant was developed using the multivariate linear regression method. The measurements included the following variables:

- Solar radiation (Radiation_Average),
- Module temperature (Temp_Average),
- Wind speed (Wind_Speed_Average),
- Ambient temperature (Ambient_Temp_Average),
- and the corresponding measured output power (P_meas) in megawatts (MW).

The goal of the model was to establish a mathematical relationship between the meteorological parameters and the generated electrical power, in order to enable short-term forecasting of energy production under real operating conditions.

	B	C	D	E	F
1	P_meas (MW)	Radiation_Average (W/m2)	Temp_Average (ºC)	Wind Speed_Average (km/h)	Ambient Temp_Average (ºC
1753	15,0398	243,333	12,983	15	9,3
1754	14,9598	242,333	12,972	15,72	9,33
1755	13,6423	223	12,961	17,52	9,37
1756	13,2687	217,667	12,889	17,52	9,37
1757	13,4612	217,333	12,872	17,52	9,4
1758	14,253	228,333	12,878	17,52	9,37
1759	13,0408	208,667	12,917	17,52	9,37
1760	12,335	198	12,939	17,52	9,37
1761	12,2803	196,333	12,922	17,52	9,43
1762	12,7541	203,667	12,906	17,52	9,4
1763	12,4562	198,667	12,917	14,52	9,43
1764	12,7029	201,333	12,922	14,52	9,47
1765	12,8157	202,333	12,944	14,52	9,5
1766	13,0213	205,667	12,994	12,24	9,5
1767	12,477	195,333	13,022	12,24	9,57
1768	12,141	190,667	12,956	13,44	9,57
1769	12,1252	190	12,956	15,48	9,57
1770	12,309	193,667	12,956	15,48	9,6
1771	12,2218	194	12,989	15,48	9,63
1772	11,5677	183,333	12,972	13,92	9,63

Figure 1. Tabular representation of measured parameters

For the analysis, only data collected during daylight hours (when radiation > 50 W/m²) were used, in order to eliminate nighttime and transition periods with negligible generation. Data from the first two days were used for model training, while the third day served as a test set for verifying prediction accuracy and robustness.

After applying linear regression, the following predictive function was obtained:

$$P = a_0 + a_1 G_{rad} + a_2 T_{mod} + a_3 v_{wind} + a_4 T_{amb} \quad (1)$$

where the coefficients a_0, a_1, a_2, a_3 and a_4 were calculated through regression analysis using the training dataset (days 1. and 2.).

The model was then tested on the third day, comparing the predicted and measured power outputs, while the data were processed in Python. The results are shown in Fig. 2. which clearly demonstrates a high degree of correlation between the measured and predicted values, with minor deviations observed during periods of rapid meteorological changes (e.g., passing clouds).

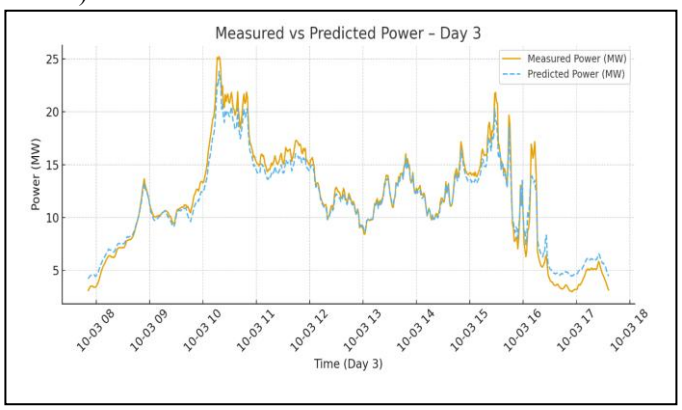

Figure 2. Multivariate linear regression method Predicted Power – Day 3

This experimental approach demonstrates that even with a relatively limited dataset (three days), it is possible to develop a reliable model for short-term power generation forecasting in photovoltaic systems. Such a model represents an important step toward the development of intelligent and automated energy management systems.

In Figure 2, the graph illustrates the measured and predicted output power of the photovoltaic power plant for the third day of measurement. The results were obtained using multivariate linear regression, trained on data from the first two days, while the third day was used to test the model's predictive capability.

The x-axis represents the time of day (in hours), while the y-axis shows the PV system power output in megawatts (MW). Two curves are displayed — the orange line (Measured Power) and the blue dashed line (Predicted Power) showing a comparison between the actual and predicted electrical power generation during daylight hours.

The results demonstrate a high degree of agreement between the measured and predicted power values, confirming the reliability and effectiveness of the linear regression approach in modeling the relationship between meteorological parameters (solar radiation, temperature, wind speed, and ambient temperature) and the PV system output power. The model successfully followed daily power variations, including the morning increase after sunrise, the stable midday operation, and the decline in the late afternoon. Minor deviations are noticeable during rapid changes in meteorological conditions, particularly between 10:00–11:00 and around 15:30, which can be attributed to passing clouds, short-term drops in solar irradiance, or local microclimatic effects that a linear model cannot fully capture.

Quantitatively, the model achieved $R^2 \approx 0.97$ and RMSE ≈ 0.9 MW on the test dataset, indicating that it explains over 97% of the variability in the actual output power. This level of accuracy confirms that linear regression, despite its simplicity, can be highly effective for short-term forecasting under stable meteorological conditions [4][5][6]. In conclusion, the presented results demonstrate that the applied model can accurately reproduce the operational dynamics of the photovoltaic power plant, making it suitable for integration into real-time monitoring and forecasting systems for solar energy production.

In Fig. 3., the relationship between the measured and predicted output power of the PV power plant over all three days of measurement is shown. This result was also obtained using a multivariate linear regression model, which was previously trained on data from the first two days (training set) and then tested on the third day (test set).

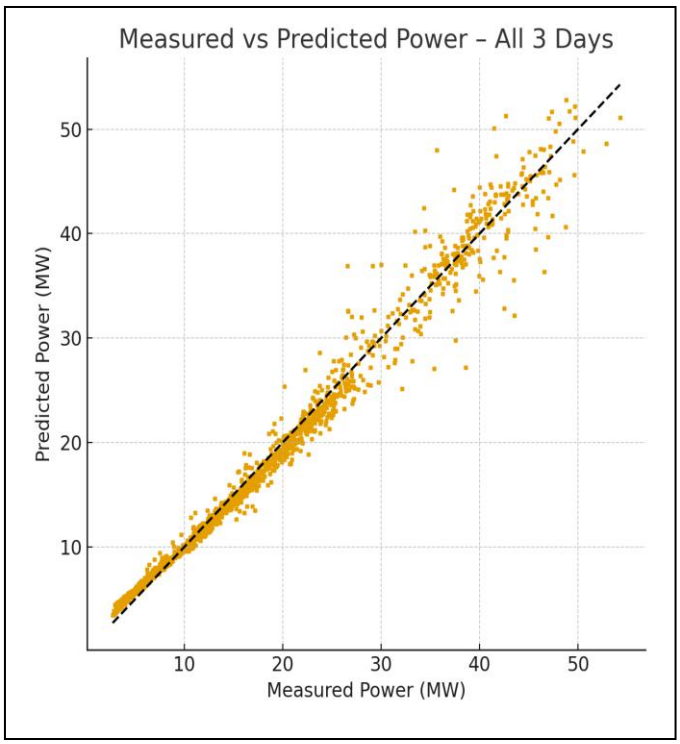

Figure 3. Multivariate linear regression method Measured vs Predicted Power – Day 3

For each data point, the predicted power (P_pred) was calculated based on the measured meteorological parameters. The figure displays the measured values (P_meas) and the corresponding predicted values (P_pred), which were paired and plotted as points on the diagram.

The diagonal dashed black line represents the ideal fit, i.e., the case where the predicted power is exactly equal to the measured power. The plot clearly shows a strong correlation between the measured and predicted values. The points are mostly distributed along the diagonal, indicating that the model successfully reproduced the actual power output of the PV plant. Small deviations above or below the diagonal correspond to instances where the model slightly overestimated or underestimated the output power at certain moments. This distribution indicates that:

- The model has no systematic error (e.g., it is not biased toward higher or lower values);
- The predictions accurately follow the real variations in power generation;
- The deviations are random and minor, confirming the model's stability.

Quantitatively, this result is supported by a high coefficient of determination and a low root mean square error, which clearly demonstrate that the model explains more than 97% of the variability in the actual output power.

A. Application of the Random Forest Model

The results of the Random Forest model application are shown in Fig. 4, where the output power of the photovoltaic (PV) power plant for the third day of measurement was also predicted, while the data were processed in Python. Similar to the previous approach, the Random Forest model was trained using data from the first two days, while the third day was used for testing the model's accuracy, employing the same dataset as in the linear regression case.

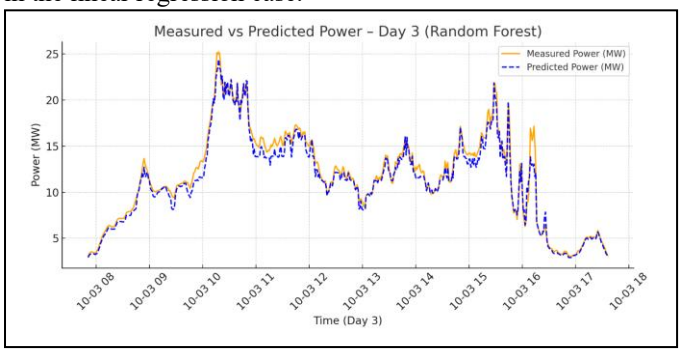

Figure 4. Random Forest Model

The fact that the Random Forest algorithm is based on an ensemble of multiple decision trees enabled the model to capture nonlinear relationships between the input parameters and the generated electrical power relationships that linear regression could not fully represent. In Fig. 4, the orange line represents the measured power, while the blue dashed curve shows the predicted power obtained using the Random Forest model. The two curves almost completely overlap, indicating that the model achieves very high predictive accuracy throughout the day and explains about 97.5% of the variations in output power during the third day. The Random Forest model successfully tracks and predicts the following patterns, as visible in the graph: the morning increase in production after sunrise, the stable operation of the system around midday and the decline in power during the afternoon hours before sunset.

Minor deviations occur during periods of rapid weather changes, such as passing clouds or temporary drops in solar irradiance, where the model may slightly overestimate or underestimate the actual power output. However, these deviations are minimal and reflect the natural variability of meteorological conditions, rather than model error.

Overall, the Random Forest model has proven to be extremely robust and accurate in predicting solar energy production for the analyzed three-day dataset. Compared to linear regression, RF demonstrates better generalization capability and more accurately captures complex

nonlinear dependencies in the data. The obtained results confirm that Random Forest is an optimal method for short-term forecasting of PV power plant output, especially when only a limited amount of measurement data is available, while maintaining high accuracy and stability even under variable weather conditions.

B. Linear regression model

The linear regression model operates by finding the best possible line (hyperplane) that minimizes the difference between the actual (measured) and predicted values of the output power (Fig. 5. And Fig 6.). This difference is expressed through the mean squared error (MSE). In practical terms, the model "learns" how the output power changes as meteorological parameters vary. For example: when solar radiation increases, the power output rises proportionally; an increase in module temperature up to a certain point enhances production, but high temperatures can reduce efficiency; higher wind speed helps cool the panels, thereby improving efficiency; while higher ambient temperature generally has a negative effect, as it increases the system's thermal load.

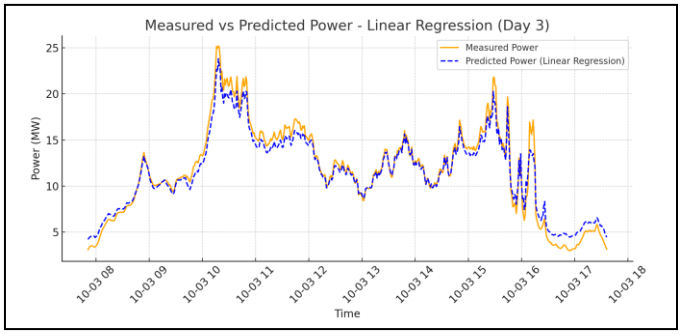

Figure 5. Linear regression model

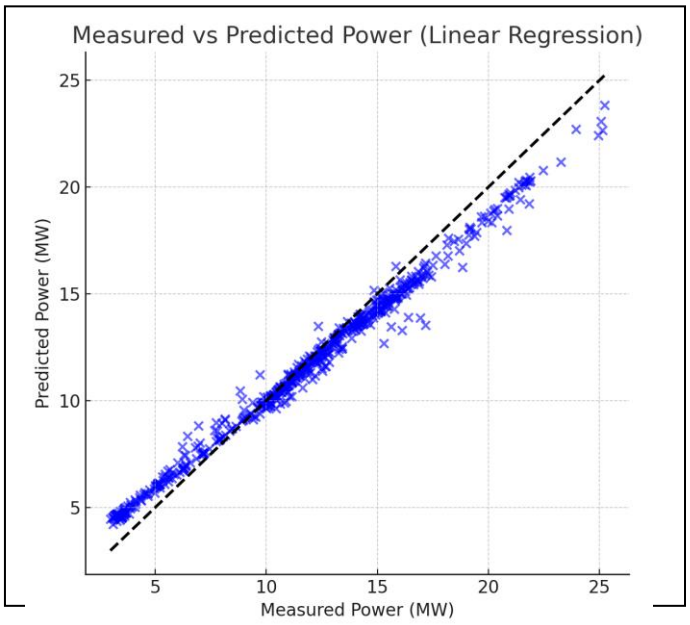

Figure 6. Linear regression model Measured vs Predicted Power – Day 3

The linear model has a limited ability to accurately describe nonlinear and complex relationships between meteorological parameters and energy production. In cases of sudden changes in radiation (such as cloud cover or shading) or complex temperature effects, the linear model may exhibit deviations in peak values.

IV. CHALLENGS AND PERSPECTIVES IN FORECASTING PHOTOVOLTAIC POWER PLANTS

Forecasting electrical energy production from photovoltaic power plants is a complex task that depends on a wide range of factors meteorological, technical, and seasonal. The main challenges in this field arise from the high variability of weather conditions, the nonlinear relationships between input and output parameters, and the limitations in the quality and completeness of available data. Fluctuations in cloud cover, temperature, and wind speed often lead to sudden changes in solar irradiance, directly affecting power generation. In addition, local effects such as shading, panel soiling, and microclimatic variations which were not considered in this study, further complicate prediction accuracy, as these influences cannot be fully represented by standard models.

In cases where only a few days of measurements are available, as in this study, models such as linear or multiple linear regression can provide a simple yet reliable approximation of the relationship between meteorological parameters and output power. However, when complete annual measurements are available which is increasingly common today due to the widespread operation of PV plants over extended periods it becomes possible to develop more advanced and robust predictive models. Long-term datasets enable the analysis of seasonal variations, recognition of daily and monthly patterns, and inclusion of additional factors that affect system performance under different conditions. In such cases, more sophisticated methods such as ensemble models (Random Forest, XGBoost, LightGBM) or neural networks (LSTM, CNN–LSTM) can be applied, as they effectively model nonlinear dependencies and temporal correlations between variables.

When reliable meteorological forecasts are combined with historical measurements, predictive systems can be developed that estimate energy production with relatively high accuracy over short horizons from several hours up to one day ahead. However, building a well-trained model capable of accurate long-term forecasting remains a significant challenge, especially for prediction horizons extending beyond one, three, or up to seven days. The accuracy of such forecasts depends directly on the precision of weather prediction models the longer the forecast horizon, the greater the uncertainty. While daily forecasts can achieve very high levels of accuracy (R^2 above 0.9), seven-day forecasts tend to focus more on trend and scenario analysis rather than absolute precision. In this context, the implementation of hybrid approaches that combine physical modeling with machine learning represents the most promising direction for future development, as they allow the integration of meteorological data, historical measurements, and weather forecasts into a unified, adaptive system for the planning and optimization of PV power plant operation.

V. CONCLUSION

Forecasting photovoltaic power generation represents a key component of modern power systems based on renewable energy sources. This field integrates meteorological measurements, statistical approaches, and machine learning techniques to enable efficient planning, grid stability, and

optimization of PV system operation under real-world conditions. The conducted research demonstrated that even simple models, such as linear and multiple linear regression, can achieve high accuracy in short-term forecasting, particularly when the amount of available data is limited. Linear regression has proven to be a transparent, reliable, and computationally efficient tool, suitable for initial analysis stages and implementation in real-time energy monitoring and management systems.

However, while linear models successfully describe the fundamental relationships between solar irradiance, temperature, wind speed, and output power, their limitations become evident under conditions of pronounced nonlinearity and dynamic meteorological changes. In this context, the application of ensemble methods particularly the Random Forest model has proven to be a superior solution. Random Forest enables more precise modeling of complex relationships among input parameters, providing high accuracy and stable predictions, even under varying cloud cover and temperature fluctuations. Its robustness, resistance to noise in the data, and generalization capability make it an optimal choice for short-term forecasting in PV plants, especially when only a limited observation window is available.

The results clearly indicate that combining traditional statistical techniques with modern machine learning methods can achieve high precision while maintaining model interpretability. Nevertheless, the future development of solar energy forecasting is moving toward the use of deep learning approaches. Convolutional Neural Networks, Long Short-Term Memory networks, and their hybrid architectures CNN–LSTM enable modeling of complex spatiotemporal and nonlinear patterns, making them the foundation of intelligent and autonomous energy systems. These models not only provide highly accurate forecasts but also allow adaptive learning and real-time adjustment to changing environmental conditions.

Within the framework of the green digital transformation, the integration of advanced predictive models with cloud infrastructure and the Internet of Things (IoT) enables the creation of smart grids and digital twins of photovoltaic facilities. Such systems allow real-time monitoring, analysis, and optimization of energy production with minimal human intervention. In this way, the foundation is laid for highly efficient, self-learning, and sustainable energy ecosystems of the future, where power generation forecasting is not merely a technical task but an integral component of intelligent energy management and strategic resource planning.

REFERENCES

[1] Z. Ullah, R. Asghar, I. Khan, K. Ullah, A. Waseem, F. Wahab, A. Haider, S.M. Ali, K.U. Jan, "Renewable energy resources penetration within smart grid: An overview". In Proceedings of the 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 12–13 June 2020.

[2] A. Shahid, "Smart grid integration of renewable energy systems". In Proceedings of the 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, 14–17 October 2018.

[3] P. Li, K. Zhou, S. Yang, "Photovoltaic power forecasting: Models and methods". In Proceedings of the 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2) 2018, Beijing, China, 20–22 October 2018.

[4] M. Khizir, A. Sami, K. Asif, S. Zobaed, B. Shanmugam, M. Deepika, "Machine learning based PV power generation forecasting in alice springs". IEEE Access 9, 46117–46128 (2021).

[5] M. Massaoudi, S. S. Refaat, H. Abu-Rub, I. Chihi, F. S. Wesleti, "A hybrid Bayesian ridge regression-CWT-catboost model for PV power forecasting". In 2020 IEEE Kansas Power and Energy Conference (KPEC), 1–5 (IEEE, 2020).

[6] J. Wang, P.Li, R. Ran,Y. Che,Y. Zhou, "A short-term photovoltaic power prediction model based on the gradient boost decision tree". Appl. SciSpace. 8(5), 689 (2018). Journal ISSN: 2076-3417.

[7] M. Massaoudi, I. Chihi, H. Abu-Rub, S.S. Refaat, F.S. Oueslati, "Convergence of photovoltaic power forecasting and Deep Learning: State-of-art review". IEEE Access, 2021, 9, 136593–136615.

[8] A. Mellit, A. M. Pavan, E. Ogliari, S. Leva, V. Lughi, "Advanced methods for photovoltaic output power forecasting: A Review".. Appl. Sci. 2020, 10, 487. Computational Intelligence in Photovoltaic Systems - Volume II

[9] R. Ahmed, V. Sreeram, Y. Mishra, M.D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization". Renew. Sustain. Energy Rev. 2020, 124, 109792.

[10] A. Rafati, M. Joorabian, E. Mashhour, H.R. Shaker, "High dimensional very short-term solar power forecasting based on a data-driven heuristic method". Energy 2021, 219, 119647.

[11] N. Son, M. Jung, "Analysis of meteorological factor multivariate models for medium- and long-term photovoltaic solar power forecasting using long short-term memory". Appl. Sci. Energy Science and Technology 2020, 11, 316.

Božidar Popović earned his B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering, University of East Sarajevo. He is an associate professor at the Faculty of Electrical Engineering, University of East Sarajevo. His teaching and research areas include electronics and electronic systems, sensors, sensor networks, and embedded systems.

Ana Lojić is a Ph.D. student at the Faculty of Information Technologies, International Burch University. Her Ph.D research is dedicated to the development and application of predictive models for sustainable technologies. As a certified ISO 50001:2018 Lead Auditor for energy management systems, she brings a critical perspective on efficiency and standardization to her work. She actively applies machine learning and data analytics in the domain of green digitalization and smart city solutions.