

BENCHMARKING MACHINE LEARNING ALGORITHMS FOR HIGH-FIDELITY POWER FORECASTING IN UTILITY-SCALE PV PLANTS

Ahmed Saidi¹, Abdelghani Draoui², Touhami Abdelouahed³

¹Ahmed Draia University, Department of Electrical Engineering, Adrar, Algeria, ²University of Tindouf, Department of Computer Science, Tindouf, Algeria, ³M'Hamed Bougara University, Faculty of Technology, Boumerdes, Algeria

Abstract. Accurate prediction of photovoltaic (PV) power is essential for effective monitoring, control, and grid integration of large-scale solar systems. This study uses high-resolution data from a 20 MW utility-scale PV plant to benchmark four advanced machine learning (ML) models: Support Vector Machine (SVM), Random Forest (RF), Gaussian Process Regression (GPR), and Artificial Neural Network (ANN). Model performance was evaluated using R^2 , RMSE, and MAE. Additionally, a thorough analysis of the three-dimensional (3D) predicted power surfaces was conducted. The findings indicate that GPR and ANN outperform SVM and RF, achieving near-perfect accuracy ($R^2 = 1.0000$) with minimal error. These methods yield smooth, physically consistent surfaces that accurately represent PV system behaviour. These findings establish GPR and ANN as the most reliable machine learning (ML) approaches for high-fidelity photovoltaic (PV) power forecasting and energy management applications.

Keywords: photovoltaic systems, power forecasting, machine learning, model benchmarking, artificial neural network (ANN), Gaussian process regression (GPR)

PORÓWNANIE ALGORYTMÓW UCZENIA MASZYNOWEGO DO WYSOKO PRECYZYJNEGO PROGNOZOWANIA MOCY ELEKTROWNI FOTOWOLTAICZNYCH NA SKALĘ PRZEMYSŁOWĄ

Streszczenie. Dokładne prognozowanie mocy fotowoltaicznej (PV) ma zasadnicze znaczenie dla skutecznego monitorowania, sterowania i integracji z siecią energetyczną wielkoskalowych instalacji słonecznych. W niniejszym badaniu wykorzystano dane o wysokiej rozdzielczości pochodzące z elektrowni fotowoltaicznej o mocy 20 MW, działającej na skalę przemysłową, w celu porównania czterech zaawansowanych modeli uczenia maszynowego (ML): maszyny wektorów nośnych (SVM), lasu losowego (RF), regresji procesów gaussowskich (GPR) oraz sztucznej sieci neuronowej (ANN). Wydajność modeli oceniono za pomocą wskaźników R^2 , RMSE i MAE. Ponadto przeprowadzono szczegółową analizę trójwymiarowych (3D) powierzchni prognozowanej mocy. Wyniki wskazują, że metody GPR i ANN przewyższają SVM i RF, osiągając niemal idealną dokładność ($R^2 = 1,0000$) przy minimalnym błędzie. Metody te dają gładkie, fizycznie spójne powierzchnie, które dokładnie odzwierciedlają zachowanie systemu fotowoltaicznego. Wyniki te potwierdzają, że GPR i ANN są najbardziej wiarygodnymi metodami uczenia maszynowego (ML) w zastosowaniach związanych z wysoko precyzyjnymi prognozami mocy fotowoltaicznej (PV) oraz zarządzaniem energią.

Słowa kluczowe: systemy fotowoltaiczne, prognozowanie mocy, uczenie maszynowe, analiza porównawcza modeli, sztuczna sieć neuronowa (ANN), regresja procesów gaussowskich (GPR)

Introduction

The global energy landscape is undergoing a profound transformation, with photovoltaic (PV) systems playing a leading role in the transition to renewable energy sources. The rapid proliferation of utility-scale photovoltaic (PV) plants is essential to achieving decarbonisation goals. However, this development also introduces substantial challenges for power grid operators [16, 19]. The primary challenge stems from the inherent variability of solar generation, driven by fluctuating meteorological conditions such as solar irradiance and ambient temperature. This intermittency poses significant challenges to grid stability, energy dispatch, and economic planning. The reliable integration of solar power into the energy landscape thus becomes a complex engineering problem [13].

To address these issues, accurately forecasting photovoltaic power output has become a cornerstone of modern energy management. High-fidelity power forecasts are essential for critical applications, including maintaining supply-demand balance, optimizing energy storage systems, and facilitating participation in electricity markets [1, 11]. Moreover, precise forecasts underpin advanced monitoring and control strategies, such as performance diagnostics and predictive maintenance, which are vital for maximizing the operational efficiency and profitability of PV assets [6].

This study directly addresses these gaps by presenting a comprehensive benchmark of four prominent machine learning (ML) algorithms [17]: Support Vector Machine (SVM), Random Forest (RF), Gaussian Process Regression (GPR), and Artificial Neural Network (ANN). A dual evaluation approach was employed, using a high-resolution dataset from a 20-MW utility-scale PV plant. The evaluation encompasses both a quantitative examination of the models using standard regression metrics (R^2 , RMSE, and MAE) and a qualitative analysis of the 3D predicted power surfaces they generate. The objective of this study

is to provide a clear, evidence-based recommendation on the most suitable and reliable algorithm for high-fidelity PV power forecasting in practical applications.

Although deep learning architectures, such as Long Short-Term Memory (LSTM) networks, are widely used for time-series forecasting because of their ability to capture temporal dependencies, the present study focuses on the instantaneous mapping capabilities of regression models. Given the strong physical correlation between the input features (irradiance, temperature, voltage) and the output power, a feed-forward Artificial Neural Network (ANN) was selected to represent the neural paradigm. This choice enables a direct comparison of the models' capacity to approximate the non-linear $P = f(V, I, G, T)$ transfer function, while avoiding the computational complexity associated with recurrent layers.

1. Related work

The field of photovoltaic power forecasting is well-established, with research organized around three primary methodologies: physical models, statistical time-series models, and machine learning models. Physical models rely on comprehensive meteorological data and the semiconductor physics of PV cells to predict output [5, 9]. While these systems are often precise, they require extensive site-specific parameters that are not always available. Conventional statistical models, including ARIMA, have been employed; however, they frequently struggle to capture the complex and volatile nature of solar power generation, which is strongly influenced by stochastic weather patterns [10, 17].

Consequently, machine learning has emerged as the state-of-the-art methodology, demonstrating superior performance in capturing complex relationships between atmospheric variables and power output. A substantial body of literature explores the applications of various machine learning algorithms for this



task. Artificial neural networks, including deep learning variants, have become a popular subject of study due to their demonstrated efficacy as universal function approximators. This characteristic renders them well-suited for modelling the non-linear dynamics of photovoltaic systems [2, 12]. Kernel-based methods, such as SVR, have also gained widespread adoption due to their demonstrated proficiency in achieving effective generalization in high-dimensional feature spaces. In a similar vein, ensemble methods such as Random Forest are frequently employed [15]. These methods are lauded for their resilience against overfitting and their capacity to manage intricate feature interactions without the need for extensive tuning. In recent years, probabilistic models such as GPR have garnered significant attention due to their high accuracy and their inherent capacity to provide uncertainty estimates with their predictions. This capacity is a critical feature for risk assessment in grid operations [8].

Despite the prevalence of research that employs these models, a significant lacuna persists in the extant literature. A significant number of studies have been conducted that focus on the performance of a single model or that compare two similar models. However, only a small number of studies have conducted a rigorous, head-to-head benchmark of multiple, structurally distinct machine learning paradigms on the same utility-scale, real-world dataset [7]. The paradigms in question include kernel-based, ensemble, neural, and probabilistic models. The absence of a standardized comparison framework poses a significant challenge for researchers and industry practitioners, who are unable to make informed decisions due to the lack of a unified metric. Moreover, the evaluation in most extant works is confined to quantitative performance metrics such as RMSE and MAE. While these metrics are undoubtedly beneficial, they lack the capacity to elucidate the factors that lead to the success or failure of a model, and they do not provide an assessment of the physical consistency of a model's learned behaviour [3, 4].

The present paper addresses these limitations. A comprehensive benchmark was conducted on high-resolution data from a 20 MW utility-scale plant, encompassing SVM, RF, GPR, and ANN. This study provides a direct comparison that is essential to the field. A qualitative analysis of the learned 3D power surfaces is included, enabling the evaluation of physically plausible relationships. The objective of this work is to furnish a definitive, evidence-based recommendation on the most suitable algorithms for deploying high-fidelity PV power forecasting solutions in real-world applications.

2. Methodology

This section outlines the systematic approach used in this study, covering data acquisition, preprocessing, training, and evaluation of machine learning models.

2.1. Data source and description

The dataset was obtained from a 20 MW grid-connected PV power plant in Adrar, Algeria. The region's defining characteristics are high solar irradiance and arid climatic conditions, making it a formidable test case for evaluating model performance under real-world operational stress. Data were collected at 10-minute intervals from July 1st to July 28th, 2024, via the plant's Supervisory Control and Data Acquisition (SCADA) system. The dataset comprises five critical time-series variables:

- DC voltage (V),
- DC current (A),
- DC power (W),
- solar irradiance (W/m²),
- ambient temperature (°C).

The target variable for prediction is DC Power, while the other four variables serve as input features for the models.

Figure 1 presents the correlation matrix of the collected variables. As anticipated, a robust positive correlation (0.91) is evident between voltage and power. Power exhibited a moderate positive correlation with irradiance (0.42) and current (0.30).

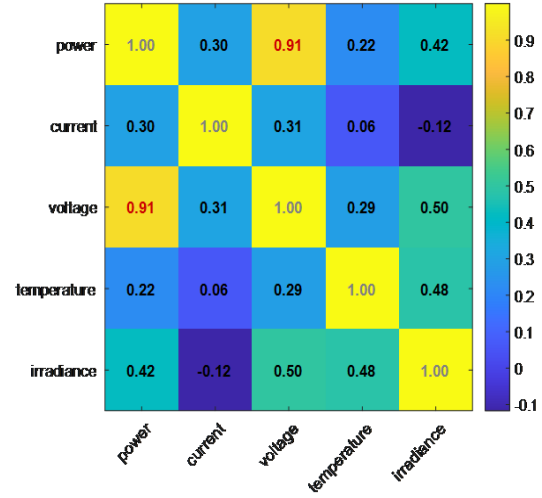


Fig. 1. Correlation matrix of the collected variables

2.2. Data preprocessing and partitioning

In order to ensure optimal model performance and prevent biases, the raw data underwent two key steps:

The standardization of features is a critical aspect of the system's functionality. The z-score normalization technique was employed to standardize all input features, including voltage, current, irradiance, and temperature. This process involves rescaling the features to have a mean of 0 and a standard deviation of 1, a critical step for algorithms that are sensitive to the scale of input data, such as SVM and ANN.

Data Partitioning: The dataset was partitioned using a standard hold-out validation strategy. Seventy-five percent of the data was allocated for training the models, while the remaining 25% was reserved as an unseen test set to evaluate their final predictive performance. This rigorous separation ensures an accurate evaluation of the model's capacity to generalize to novel data.

2.3. Machine learning models

Four distinct machine learning paradigms were implemented and subsequently benchmarked.

Support vector machine (SVM)

The objective is to identify a function that best fits the data while deviating from the target values by no more than a specified margin, denoted by ε . This makes the model robust to outliers.

The objective is to identify a function $f(x) = w^T * x + b$ that exhibits minimal variation. This objective is realised by minimising the norm of the weight vector, i.e. the square of the magnitude of w , denoted as $\|w\|^2$. The problem can be formulated as a convex optimization problem as follows [7]:

$$\text{Minimize: } \left(\frac{1}{2}\right) * \|w\|^2 + C * \sum(\xi_i + \xi_{i*}) \quad (1)$$

Subject to the constraints:

$$P_i - (w^T * X_i + b) \leq \varepsilon + \xi_i$$

$$(w^T * X_i + b) - P_i \leq \varepsilon + \xi_{i*}$$

$$\xi_i, \xi_{i*} \geq 0$$

In this context, $C > 0$ is a regularisation parameter that governs the trade-off between the flatness of $f(x)$ and the extent to which deviations greater than ε are permissible. The slack variables, termed ξ_i and ξ_{i*} , denote errors occurring within the ε -insensitive tube, i.e. points lying outside this tube [14].

Random forest (RF)

Random forest is an ensemble learning method that operates by constructing a multitude of decision trees during the training phase. The final prediction is the average of the predictions from all individual trees, a process which reduces variance and mitigates overfitting.

A Random Forest model consists of an ensemble of B individual decision trees, denoted $h_b(X)$. Each tree is trained on a different bootstrap sample of the original dataset. For a given input X , the final prediction $\hat{P}(t)$ is the average of all tree predictions [13].

$$\hat{P}(t) = f_{RF}(X) = \left(\frac{1}{B}\right) * \sum_{b=1 \text{ to } B} h_b(X) \quad (2)$$

In order to decorrelate the trees, at each node during the construction of the tree, only a random subset of the input features is considered. This process serves to enhance the model's generalisation capability [18].

Gaussian process regression (GPR)

GPR is a non-parametric, probabilistic model that defines a prior distribution over functions. Rather than learning specific parameters, it directly models the underlying function, thereby yielding not only a point prediction but also a quantification of uncertainty.

A Gaussian process is based on the assumption that the output $P(t)$ for a given set of inputs $X(t)$ can be modelled as follows:

$$P(t) = f(X(t)) + \varepsilon \quad (3)$$

Where $f(X(t))$ is a function drawn from a Gaussian process, and ε is Gaussian noise with variance σ_n^2 .

The GP is fully specified by its mean function $m(X)$ and its covariance (kernel) function $k(X, X')$

$$f(x) \sim GP(m(x), k(x, x')) \quad (4)$$

The kernel function $k(X, X')$ is defined as the similarity between two data points. The Radial Basis Function (RBF) kernel is a frequently employed option.

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{i=1}^d \frac{(x_i - x'_i)^2}{l_i^2}\right) \quad (5)$$

It is important to note that σ_f^2 denotes the signal variance, while 1 denotes the characteristic length-scale. Both are hyperparameters that are optimized during the training process. The posterior predictive distribution for a new input X^* is also Gaussian, with a mean representing the power prediction $\hat{P}(t)$ and a variance representing its uncertainty [19].

Artificial neural network (ANN)

An artificial neural network is a computational model inspired by the structure and function of biological neural networks. The model under consideration is designed to emulate intricate, non-linear correlations between a set of input features and an output by employing a network of interconnected artificial neurons, or nodes.

The specific architecture used in this study is a multi-layer feedforward network. The following elements comprise the system:

- The system under consideration includes an input layer that receives the feature vector, which comprises the voltage, current, irradiance, and temperature.
- The system under consideration contains two hidden layers, each comprising 10 neurons. These neurons process inputs through weighted connections.
- The final power prediction is produced by an output layer.

The output of a single neuron, designated as j , is calculated as follows:

$$y_i = \phi\left(\sum_{i=1}^n w_{ij}x_i + b_j\right) \quad (6)$$

Where x_i are the inputs, w_{ij} are the connection weights, b_j is the bias term, and ϕ is a non-linear activation function. The network's weights and biases are optimized during a training phase to minimize a loss function, typically the Mean Squared

Error (MSE), between the predicted and actual outputs. This optimization is performed using the back-propagation algorithm, specifically the efficient Levenberg-Marquardt variant [15].

Performance evaluation metrics

In order to rigorously evaluate and compare the performance of each model, the following standard metrics were calculated. Let $P(t)$ denote the actual power value, $\hat{P}(t)$ denote the predicted power value, $\bar{P}$ denote the mean of the actual power values, and n denote the total number of data points in the test set.

Root Mean Square Error (RMSE): the standard deviation of the prediction errors is measured. The system imposes significant penalties for larger errors [19].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P(t) - \hat{P}(t))^2} \quad (7)$$

Mean Absolute Error (MAE): this process involves the calculation of the mean magnitude of the errors in a set of predictions, without taking into account their direction.

$$MAE = \frac{\sum_{i=1}^n |P(t) - \hat{P}(t)|}{n} \quad (8)$$

Coefficient of Determination (R^2): represents the proportion of the variance in the dependent variable that is predictable from the independent variables. A value of 1 indicates a perfect fit [11].

$$R^2 = 1 - \frac{\sum_{i=1}^n (P(t) - \hat{P}(t))^2}{\sum_{i=1}^n (P(t) - \bar{P})^2} \quad (9)$$

3. Results and discussion

The four trained machine learning models were evaluated on the unseen test dataset to ascertain their predictive performance. The results, which are presented both numerically in Table 1 and through a series of visualizations, reveal a stark performance difference between the models. Gaussian process regression and the artificial neural network demonstrate superior accuracy and robustness.

3.1. Quantitative performance evaluation

The core performance of each model was quantified using the root mean square error, the mean absolute error, and the coefficient of determination (R^2). The results of this study are summarized in Table 1.

Table 1. Quantitative performance metrics of the four ML models

Model	R^2	RMSE (W)	MAE (W)
SVM	0.98987	13.98532	12.78912
RF	0.99972	2.31820	1.10945
GPR	0.99999	0.21378	0.09030
ANN	0.99999	0.22642	0.11025

The data unequivocally demonstrate that the GPR and ANN models perform significantly better than the other two models. It has been demonstrated that both of these variables achieve near-perfect R^2 .

A value of approximately 0.99999 indicates that the model explains nearly all of the variance in the data. The proposed model's RMSE and MAE are consistently lower than those of SVM and RF, indicating that the prediction error is negligible. While RF performs well, the SVM model demonstrates a notably higher error rate, making it the least suitable for this high-fidelity forecasting task.

3.2. Visual analysis of predictive performance

A visual examination of the models' predictions further substantiates the quantitative findings. As illustrated in Figure 2, the models' dynamic tracking performance is evaluated against the true power signal. The predictions from the GPR and ANN models are visually indistinguishable from the true power, and both models capture the sharp peaks and troughs observed.

In contrast, the SVM predictions show notable deviations, particularly at the peaks.

The prediction residuals are shown in Figure 3. This plot provides a clear visual summary of the error magnitudes, which are summarized in Table 1. The errors for the GPR and ANN models are tightly clustered around zero, whereas those for the SVM model are substantially larger and more erratic.

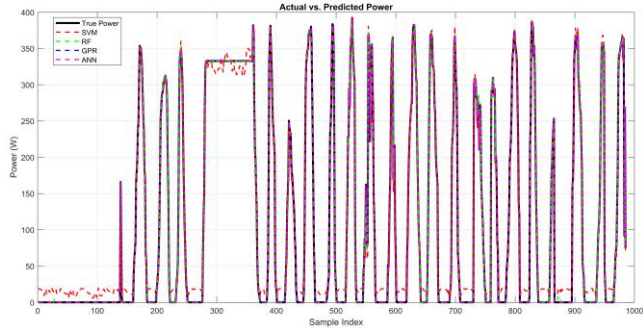


Fig. 2. Time-series comparison of actual power versus the predictions from the four ML models

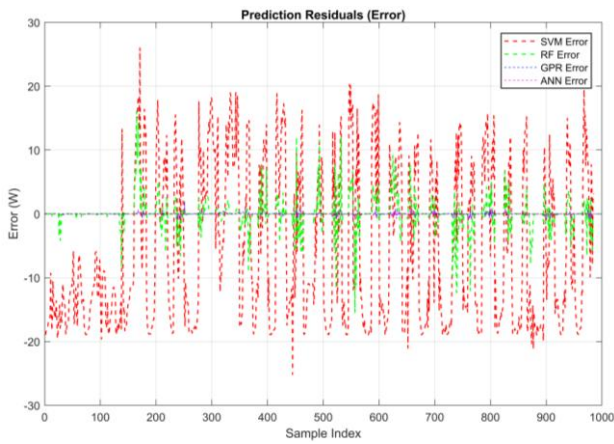


Fig. 3. Prediction residuals (actual – predicted) for each model

3.3. Deeper model insights

To understand the reasons for the observed performance disparities, a thorough analysis of the models' learned behaviours and internal structures was conducted. As illustrated in Figure 4, the three-dimensional power surfaces obtained by each model are presented as a function of voltage and irradiance.

- The GPR and ANN models demonstrated the capacity to learn smooth, physically consistent surfaces, suggesting a robust comprehension of the underlying system physics.
- The RF model yielded a characteristic "stepped" surface, an artefact of its decision-tree-based structure.
- The SVM model produced a distorted surface with anomalous dips, suggesting that it has difficulty generalizing and may be producing non-physical predictions in specific regions of the feature space.

The error heatmaps in Figure 5 offer additional insight into where each model's predictions deviate from the GPR reference surface. The substantial, colored regions in the SVM and RF plots indicate significant, structured errors, whereas the uniformly blue ANN plot confirms that its predictions are nearly identical to those of the GPR model.

The diagnostic checks confirm the high quality of the GPR and ANN models. The Quantile-Quantile (QQ) plots in Figure 6 show that the residuals for both the GPR and the ANN closely follow a normal distribution, indicating a well-fitted model in which random noise is the only significant factor. Additionally, GPR offers a distinct advantage by quantifying its own uncertainty, as shown in Figure 7. The narrow 95% confidence interval indicates the model's high degree of certainty in its predictions.

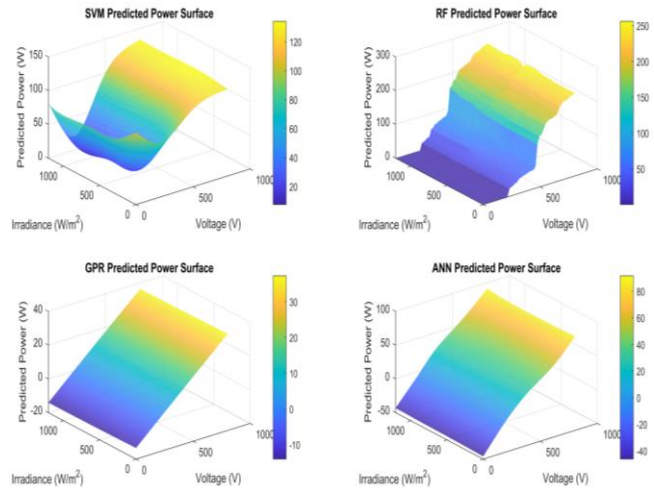


Fig. 4. 3D predicted power surfaces learned by each model

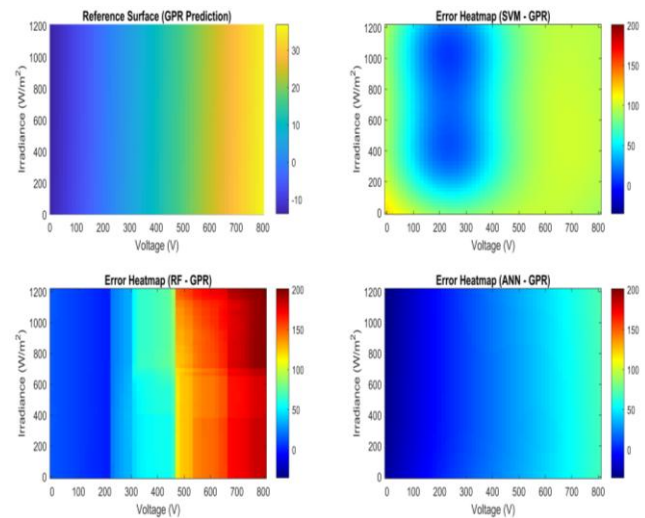


Fig. 5. Error heatmaps showing the deviation of each model's predicted surface from the GPR reference surface

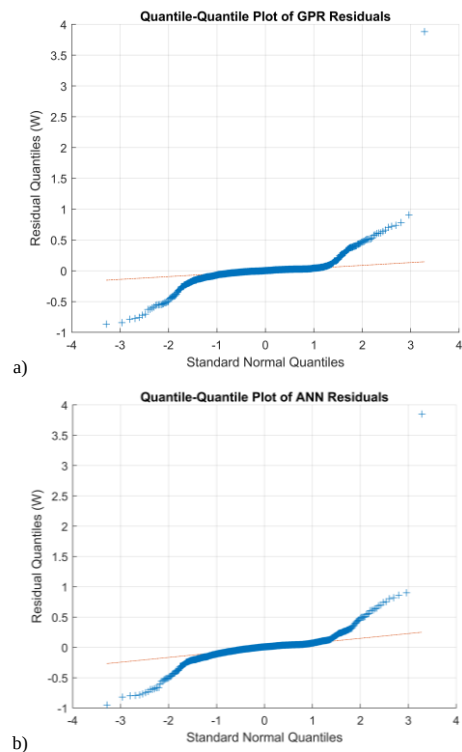


Fig. 6. QQ-plots for GPR (a) and ANN (b) residuals

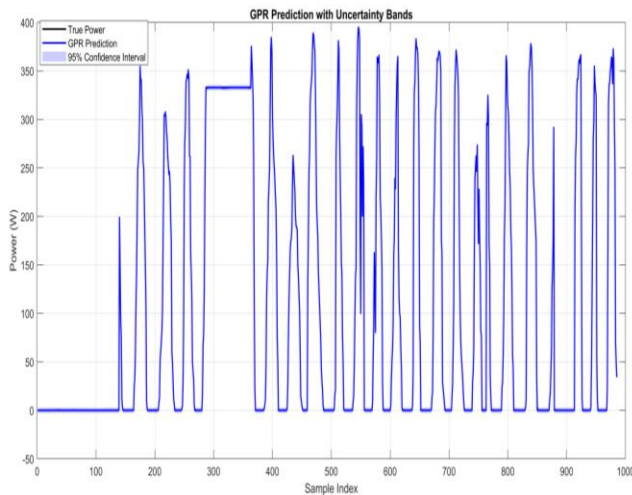


Fig. 7. GPR prediction shown with its 95% confidence interval

The empirical results from this study present a clear and decisive conclusion: Gaussian process regression and artificial neural networks outperform support vector machines and random forests in high-fidelity photovoltaic power forecasting. This conclusion is supported not only by near-perfect quantitative metrics ($R^2 \approx 1.0$) but also by the models' performance, which further strengthens the hypothesis. The GPR and ANN models demonstrated an exceptional capacity to track the true power output, producing smooth and physically consistent 3D power surfaces. In stark contrast, the SVM model exhibited significant systematic biases and generated a distorted surface, while the RF model produced a characteristic "blocky" approximation, highlighting their limitations in accurately representing the underlying system physics.

The superior performance of GPR and ANN can be attributed to their inherent capacity to model highly complex, nonlinear systems. Their architectural flexibility makes them well-suited to capturing the intricate relationships between environmental inputs and the electrical behaviour of a PV array. For engineers and researchers developing data-driven tools for PV systems, this study provides a clear directive: for applications requiring high precision, such as grid-level dispatch, energy trading, or as a baseline for anomaly detection, GPR and ANN should be the primary candidates. The systematic errors observed in the SVM model, such as the "clipping" of power peaks, underscore that a model with seemingly good overall metrics can still falter at operational extremes, potentially leading to flawed decision-making in downstream applications.

4. Discussion and limitations

Despite the superior predictive performance of the GPR and ANN models, it is imperative to acknowledge this study's limitations to ensure a balanced interpretation of the results.

Diversity and generalizability of datasets

The benchmarking analysis used high-resolution SCADA data from a single utility-scale PV plant in Algeria's Adrar region. Although the site provides a robust test case given its harsh arid climate and high irradiance, the lack of multi-site validation limits the immediate generalizability of the findings to other climatic zones (e.g., tropical or temperate regions). Subsequent efforts will prioritize cross-site validation to confirm the transferability of the GPR and ANN models.

The issue of model overfitting and physical consistency

The models have achieved near-perfect accuracy ($R^2 \approx 1.0$), raising valid concerns about potential overfitting. However, their consistent performance on the unseen test set (25% hold-out) suggests that the models have captured the deterministic physical

relationship between irradiance and power output rather than noise. To provide further validation, subsequent iterations of this work will incorporate k -fold cross-validation and explicit physical constraints (e.g., monotonicity checks) to ensure that the models adhere to semiconductor physics principles.

Computational complexity

While GPR provides uncertainty quantification, its computational cost scales cubically with the number of data points ($O(N^3)$), making it computationally expensive for very large datasets compared with the relatively efficient inference of ANN and Random Forest models. In real-time embedded systems, where computational resources are limited, the ANN model can offer an optimal balance between high fidelity and computational speed.

5. Conclusion

This paper conducted a rigorous, data-driven benchmark of four machine learning paradigms – SVM, RF, GPR, and ANN – for photovoltaic power forecasting. The empirical evaluation, based on a comprehensive set of quantitative metrics and qualitative visual analyses, yielded a clear outcome. The GPR and ANN models demonstrated overwhelming superiority, attaining near-perfect predictive accuracy ($R^2 \approx 1.0$) and learning smooth, physically consistent power surfaces. Conversely, SVM and RF exhibited substantial errors and structural artefacts, making them less suitable for high-fidelity applications.

The findings establish GPR and ANN as the most accurate and reliable models for this application, offering a strong, evidence-based recommendation for researchers and engineers. Beyond this primary conclusion, this work underscores the need to move beyond standard accuracy metrics and incorporate model diagnostics and interpretability. Even meticulously designed "black-box" models' ability to detect non-physical behaviours underscores the imperative to validate their internal logic. Future research should prioritize validating these findings across a broader range of datasets and climates. Additionally, there is a need to formally integrate interpretability methods, such as SHAP, into the research framework. Furthermore, analyzing the computational trade-offs for real-world deployment is essential to enhance the practical applications of these findings. This research contributes to the development of more robust, transparent, and intelligent tools for managing the next generation of solar energy assets.

References

- [1] Aldabbagh, G. M. T., & Hasoon, S. O. (2024). Defect severity code prediction based on ensemble learning. *Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska*, 14(4), 146–153. <https://doi.org/10.35784/iapgos.6393>
- [2] Algafri, M., Sayad, M., Abdel-Aal, M. A. M., & Attia, A. M. (2025). A Comparative Study of Machine Learning Models for Accurate E-Waste Prediction. *Results in Engineering*, 26, 105388. <https://doi.org/10.1016/j.rineng.2025.105388>
- [3] Durganjali, C. S., Avinash, G., Megha, K., Ponnalagu, R. N., Goel, S., & Radhika, S. (2023). Prediction of PV cell parameters at different temperatures via ML algorithms and comparative performance analysis in Multiphysics environment. *Energy Conversion and Management*, 282, 116881. <https://doi.org/10.1016/j.enconman.2023.116881>
- [4] Gupta, M., Arya, A., Varshney, U., Mittal, J., & Tomar, A. (2025). A review of PV power forecasting using machine learning techniques. *Progress in Engineering Science*, 2(1), 100058. <https://doi.org/10.1016/j.pes.2025.100058>
- [5] Harrou, F., Zeroual, A., Kadri, F., & Sun, Y. (2024). Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. *Results in Engineering*, 23, 102342. <https://doi.org/10.1016/j.rineng.2024.102342>
- [6] Hong, Y.-Y., & Pula, R. A. (2023). Diagnosis of PV faults using digital twin and convolutional mixer with LoRa notification system. *Energy Reports*, 9, 1963–1976. <https://doi.org/10.1016/j.egyr.2023.01.011>
- [7] Jobayer, M., Shaikat, M. A. H., Naimur Rashid, M., & Hasan, M. R. (2023). A systematic review on predicting PV system parameters using machine learning. *Heliyon*, 9(6), e16815. <https://doi.org/10.1016/j.heliyon.2023.e16815>
- [8] Karthick, B., & Sudhakaran, M. (2025). Mitigating power loss in PV arrays using Four Pyramid Sudoku under Partial Shading. *Ain Shams Engineering Journal*, 16(10), 103599. <https://doi.org/10.1016/j.asej.2025.103599>
- [9] Khan, K., Rashid, S., Mansoor, M., Khan, A., Raza, H., Zafar, M. H., & Akhtar, N. (2023). Data-driven green energy extraction: Machine learning-based MPPT control with efficient fault detection method

- for the hybrid PV-TEG system. *Energy Reports*, 9, 3604–3623. <https://doi.org/10.1016/j.egyr.2023.02.047>
- [10] Khouili, O., Hanine, M., Louzazni, M., Flores, M. A. L., Villena, E. G., & Ashraf, I. (2025). Evaluating the impact of deep learning approaches on solar and photovoltaic power forecasting: A systematic review. *Energy Strategy Reviews*, 59, 101735. <https://doi.org/10.1016/j.esr.2025.101735>
- [11] Labrador Rivas, A. E., Da Silva, N., & Abrão, T. (2020). Adaptive current harmonic estimation under fault conditions for smart grid systems. *Electric Power Systems Research*, 183, 106276. <https://doi.org/10.1016/j.epsr.2020.106276>
- [12] Li, Y., Basem, A., Alizadeh, A., Singh, P. K., Dixit, S., Abdulaali, H. K., Ali, R., Cajla, P., Rajab, H., & Ghachem, K. (2025). Synergizing neural networks with multi-objective thermal exchange optimization and PROMETHEE decision-making to improve PCM-based photovoltaic thermal systems. *Case Studies in Thermal Engineering*, 68, 105851. <https://doi.org/10.1016/j.csite.2025.105851>
- [13] Lin, P., Chen, H., Cheng, S., Lu, X., Lin, Y., & Sun, L. (2025). Assessment of power loss caused by soiling PV modules using a dual branch multi-modality deep learning network framework. *Renewable Energy*, 248, 122926. <https://doi.org/10.1016/j.renene.2025.122926>
- [14] Pombo, D. V., Bacher, P., Ziras, C., Bindner, H. W., Spataru, S. V., & Sorensen, P. E. (2022). Benchmarking physics-informed machine learning-based short term PV-power forecasting tools. *Energy Reports*, 8, 6512–6520. <https://doi.org/10.1016/j.egyr.2022.05.006>
- [15] Salehpour, S., Eskandari, A., Nedaei, A., & Aghaei, M. (2025). Two-stage deep Q-network reinforcement learning based ultra-efficient fault diagnosis and severity assessment scheme for photovoltaic protection. *Energy and AI*, 20, 100512. <https://doi.org/10.1016/j.egyai.2025.100512>
- [16] Saravanan, S., Kumar, R. S., Balakumar, P., & Prabakaran, N. (2025). Optimal power harvesting under partial shading: Binary Greylag Goose optimization for reconfiguration and Machine learning-Based fault diagnosis in solar PV arrays. *Energy Conversion and Management*, 333, 119808. <https://doi.org/10.1016/j.enconman.2025.119808>
- [17] Silva Santos, A., Da Silva, R. J., Montenegro, P. A., Faria, L. T., Lopes, M. L. M., & Minussi, C. R. (2025). Integrating autoencoders to improve fault classification with PV system insertion. *Electric Power Systems Research*, 242, 111426. <https://doi.org/10.1016/j.epsr.2025.111426>
- [18] Sushmi, N. B., & Subbulekshmi, D. (2025). Real-time ultra short-term irradiance forecasting using a novel R-GRU model for optimizing PV controller dynamics. *Results in Engineering*, 26, 105046. <https://doi.org/10.1016/j.rineng.2025.105046>
- [19] Zhou, N., Shang, B., Xu, M., Peng, L., & Feng, G. (2024). Enhancing photovoltaic power prediction using a CNN-LSTM-attention hybrid model with Bayesian hyperparameter optimization. *Global Energy Interconnection*, 7(5), 667–681. <https://doi.org/10.1016/j.gloi.2024.10.005>

Ph.D. Ahmed Saidie-mail: saidi.ahmed@univ-adrar.edu.dz

Ahmed Saidi received the B.Eng., M.Sc., and Ph.D. degrees in electrical engineering from the University of Béchar, Algeria, in 2013, 2015, and 2019, respectively. Currently, he is an associate professor at the Department of Electrical Engineering, University of Adrar.

His research interests include optimization, metaheuristics, renewable energy, power electronics, power generation, power grids, battery chargers, prediction, forecasting, diagnostic, fault detection, ML and artificial intelligence applied to power systems.

<https://orcid.org/0000-0003-0056-5004>**Ph.D. Abdelghani Draoui**e-mail: abdelghani.draoui@univ-tindouf.dz

Abdelghani Draoui received the B.Eng. and M.Sc. degrees in electrical engineering from the Institute of Electrical and Electronic Engineering (Ex. INELEC), University of Boumerdes, in 2013 and 2015 respectively. He received his Ph.D. degree in electric power engineering from the University of Béchar, Algeria, in 2019. Currently, he is an associate professor at the Department of Computer Science, University Center of Tindouf.

His research interests include optimization, metaheuristics, renewable energy, power electronics, power generation, power grids, battery chargers, and artificial intelligence applied to power systems.

<https://orcid.org/0000-0003-4814-898X>**Ph.D. Touhami Abdelouahed**e-mail: a.touhami@univ-boumerdes.dz

Touhami Abdelouahed was born in Bechar Algeria, he received the Master degree in 2013 from Tahri Mouhamed University, Bechar, Algeria. He obtained his Ph.D. degree in electrical engineering in 2020 from Djillali Liabes University, Sidi Bel Abbes, Algeria. He is currently a professor at M'Hamed Bougara University of Boumerdes, Algeria.

His research interests: power system operation, FACTS devices in power system, artificial intelligence and optimizations techniques.

<https://orcid.org/0000-0002-9188-8417>