cigre Turkey

TEPES
TURKISH JOURNAL OF ELECTRICAL POWER AND ENERGY SYSTEMS

TEPES, [epub ahead of print], 2024 **DOI:** 10.5152/tepes.2024.24027

RESEARCH ARTICLE

Improving Accuracy in Solar Power Plant Power Generation Prediction: A Hybrid Model Proposal

Necati Aksoy, V. M. Istemihan Genc

Department of Electrical Engineering, İstanbul Technical University, İstanbul, Türkiye

Cite this article as: N. Aksoy and V. M. I. Genc, "Improving accuracy in solar power plant power generation prediction: A hybrid model proposal," *Turk J Electr Power Energy Syst.*, Published online November 11, 2024. doi 10.5152/tepes.2024.24027.

ABSTRACT

Renewable energy sources are increasingly critical in addressing global energy needs while reducing carbon emissions and energy costs. Accurate forecasting of power generation in solar power plants is essential for efficient energy management and planning. This study introduces a novel hybrid prediction model that combines several prevalent machine learning algorithms to improve the accuracy of solar power generation forecasting. Using real meteorological and production data, the proposed model significantly outperforms individual prediction models. The hybrid model's integration of meteorological data ensures more reliable short-term and long-term power predictions, contributing to improved decision-making in solar plant operations. The results demonstrate the advantages of this approach, providing valuable insights into enhancing the predictability and operational efficiency of solar power plants.

Index Terms—Hybrid predictive model, machine learning, renewable energy ensemble learning, solar power prediction.

I. INTRODUCTION

Renewable energy sources, by reducing dependence on fossil fuels, offer a sustainable solution to future energy crises. This transition not only provides greater flexibility in energy usage but also promotes equitable global energy distribution. Among renewable energy technologies, solar power stands out due to the sun's abundant and nearly inexhaustible energy supply. While continuous advancements are being made in solar energy conversion technologies, optimizing the efficiency of conventional solar panels remains a prominent area of academic interest. Increasing the efficiency of energy generation from solar panels is closely linked to maximizing solar energy utilization, particularly solar radiation. In this context, sensor-based solar tracking systems have been developed to enhance the capture of solar radiation, while accurate prediction of power generation based on radiation levels offers an alternative approach.

Power generation prediction can be approached through either rule-based or artificial intelligence (AI) methods, with machine learning (ML), a key subset of AI, offering a range of highly effective predictive algorithms. The integration of ML techniques into solar power forecasting has gained significant attention in recent studies [1]. Furthermore, deep learning, a specialized branch of ML, has demonstrated considerable advantages in enhancing solar power predictions [2, 3]. For example, satellite imagery has been utilized as a dataset for forecasting [4], and meteorological data have been incorporated alongside satellite images to train predictive models [5, 6]. In addition to predicting the overall power generation of a solar power plant, models have been developed based on data from individual solar panels [7, 8]. Moreover, research has extended beyond plant-level forecasting to examine the effects of power generation predictions on microgrid operations [9-11]. Solar power estimation is also very important for energy management systems that use these estimates. In [12], solar power values are used in the reward function of a reinforcement learning-based energy management system for microgrids to optimize itself. The proposed hybrid solar power forecasting model will help such energy management systems to optimize themselves and improve dynamic management decisions. These studies demonstrate that predictive models can facilitate real-time operation of microgrids, enabling day-ahead planning and optimizing energy distribution.

ML models are also widely used in other renewable energy applications, such as wind power generation and wind speed estimation [13]. While predicting wind turbine output is relatively straightforward when wind speed and direction are known, forecasting solar power generation presents a greater challenge due to the variability

Corresponding author: Necati Aksoy, aksoyn18@itu.edu.tr

Received: September 5, 2024
Accepted: October 11, 2024
Publication Date: November 11, 2024

of meteorological conditions. Despite this complexity, prediction models for solar power have produced promising results.

However, the development of predictive models alone is insufficient; performance evaluations and model improvements are necessary. Even accurate models may not perform optimally in the face of sharp changes in power generation. To address these challenges, neural networks and sensor optimization solutions have been proposed [14], and combining ML models with other optimization techniques has been shown to reduce prediction errors [15]. Additionally, bootstrapping techniques have been introduced to enhance model performance [16]. An alternative approach to improving predictions is the use of hybrid models, which combine multiple forecasting models, either using the same algorithm with different parameters or using entirely different algorithms [17, 18]. While combining models based on the same algorithm has yielded some improvements, hybrid models that integrate different algorithms have shown superior results [19]. This concept of combining multiple predictive models presents challenges in determining how best to integrate them into a unified output.

Building upon these advancements in solar power generation prediction, this study proposes an innovative hybrid model that integrates several prevalent ML algorithms. By employing a specifically developed hybridization approach, the combined model's performance is evaluated against individual models to determine its overall effectiveness in improving prediction accuracy.

II. POWER PREDICTIVE METHODS

ML, encompassing the concept of deep learning, is a subfield within the broader AI paradigm. A key focus of ML is predictive modeling—anticipating future outcomes based on historical data. In this study, we introduce a novel approach to prediction using ML techniques. ML is categorized into three primary branches: supervised learning, unsupervised learning, and reinforcement learning [20]. Supervised learning, which is the focus of this study, relies on labeled datasets to predict outcomes. This branch is primarily used for tasks such as regression, where the goal is to predict continuous values, and classification, which assigns data to predefined categories. In the following section, we briefly discuss the predictive models employed in this study.

Main Points

- Introduction of a Hybrid Model: A novel hybrid model combining four Machine Learning (ML) algorithms improves solar power generation forecasting.
- Improved Accuracy: The hybrid model achieves superior accuracy (94.22%) compared to individual machine learning models.
- Use of Meteorological Data: The integration of real meteorological and production data enhances the reliability of short-term and long-term predictions.
- Operational Impact: The model supports better decision-making in solar plant operations, including battery storage and system planning.

A. Polynomial Regression

In many datasets, there can be a linear relationship between certain input variables and the corresponding output. In such cases, linear regression models often yield satisfactory predictions. However, when higher-order relationships exist—such as when second, third, or higher degrees of the input variables influence the output—linear models fall short, and polynomial regression (Poly. Reg.) becomes necessary to capture the non-linear dynamics of the data.

Poly. Reg. is an extension of linear regression, particularly useful when the relationship between the input variables and the output is non-linear. A basic linear regression model is represented by (1).

$$\mathbf{Y} \approx \beta_0 + \beta_1 \mathbf{X} + \in \tag{1}$$

where Y is the target variable, X is the input vector, β_0 and β_1 are coefficients, and $\in$ represents the error term. This model can be generalized to multiple inputs using the multiple linear regression (2):

$$\mathbf{Y} \approx \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \ldots + \beta_n \mathbf{X}_n + \in \tag{2}$$

Although this form of regression works well for many problems, it may be insufficient when the relationship between the input and output is non-linear. In such cases, Poly. Reg., which incorporates higher-order terms, is applied. The general form of a Poly. Reg. model is given by (3):

$$\mathbf{Y} \approx \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2^2 + \beta_3 \mathbf{X}_3^3 + \ldots + \beta_n \mathbf{X}_n'' + \in \tag{3}$$

This approach allows the model to capture more complex relationships within the data. In the present study, a Poly. Reg. model was utilized to predict solar power generation based on meteorological data, with a polynomial degree of 4 to better account for the nonlinear interactions in the data. This enhanced the accuracy of the predictions, providing a more robust model for forecasting solar power output.

B. Support Vector Machines

Support Vector Machine (SVM) is a supervised ML algorithm widely used for both classification and regression tasks, though it is primarily employed in classification. In this study, the regression variant, known as Support Vector Regression (SVR), is used to develop the predictive model. SVM operates on the same fundamental principles as SVM. The key objective of SVR is to identify the best-fit line (or kernel function) that optimally represents the data. In SVR, this best-fit line corresponds to a hyperplane that maximizes the number of data points within the margin boundaries.

SVR relies on selecting and configuring a kernel function based on the problem at hand. For classification tasks, the kernel function is chosen to effectively separate the data, while for regression tasks, the kernel is selected to best fit the data. The most commonly used kernel functions include the Gaussian kernel, the Gaussian Radial Basis Function (RBF), the polynomial kernel, and the linear kernel, with RBF often delivering robust results across various problem domains. The goal is to position as many data points as possible

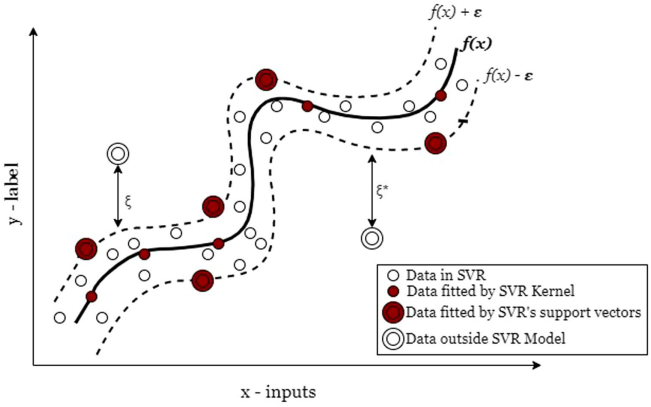

Fig. 1. Overall framework of Support Vector Machine.

within the margins of the support vectors and minimize errors for data points that lie outside these margins. A detailed framework of SVR can be seen in Fig. 1.

In addition to the brief overview provided here, several review articles further elaborate on the use of SVM in both classification and regression contexts, demonstrating its effectiveness as a powerful tool for diverse prediction tasks [22].

C. Decision Tree

Decision Tree (DT) is a ML algorithm that partitions a dataset into regions based on feature values. Although predominantly used for classification, it also serves as an effective regression tool. For regression tasks, the algorithm creates a tree structure by dividing the dataset into distinct regions. Formally, the dataset's features X_1, X_2, X_3, … X_p are mapped to regions R_1, R_2, R_3, … R_j.

$$X_1, X_2, \ldots X_p \rightarrow R_1, R_2, R_3, \ldots R_j$$

where *Xs* represent features of the dataset and *Rs* are regions that the DT algorithms create. In each region R_j, predictions are made based on the average output values ($\hat{y}_i$.) within that region. The selection of regions is based on minimizing the Residual Squares Error function, calculated as:

$$\sum_{j=1}^{J}\sum_{j\in 1}\left(y_i - \hat{y}_{R_j}\right)^2 \tag{4}$$

where $\hat{y}_{R_j}$ the average of outputs in R_j region. Region selection in DTs is accomplished through recursive binary splitting, as described by the formula given in (5).

$$R_1(j,s) = \{X \mid X_j < s\} \qquad R_2(j,s) = \{X \mid X_j \geq s\} \tag{5}$$

where *j* and *s* represent the feature and split point that minimize the following expression in (6).

$$\sum_{i:x_i \in R_1(j,s)} \left(y_i - \hat{y}_{R_1}\right)^2 + \sum_{i:x_i \in R_2(j,s)} \left(y_i - \hat{y}_{R_2}\right)^2 \tag{6}$$

Once the dataset is divided into regions, the process halts if any region contains too few data points, ensuring that the model does not overfit. A general framework for DT demonstrating regions and data analysis is given in Fig. 2.

In DT, features such as X_1 and X_2 are used to identify distinct regions in the dataset. The surfaces that correspond to these regions are depicted by the tree structure, as shown in Fig. 2. The upward arrows in the figure represent the average values of the points within each region, which are used for making predictions. However, if the data is divided into too many regions, the model risks overfitting—memorizing the data rather than generalizing from it.

To mitigate overfitting, tree pruning is applied. This technique adjusts the size of the tree by either reducing the number of terminal nodes or minimizing the squared error function. The pruning process is based on the following formula given in (7).

$$\sum_{m=1}^{|T|}\sum_{x_i \in R_m}\left(y_i - \hat{y}_{Rj}\right)^2 + \alpha|T| \tag{7}$$

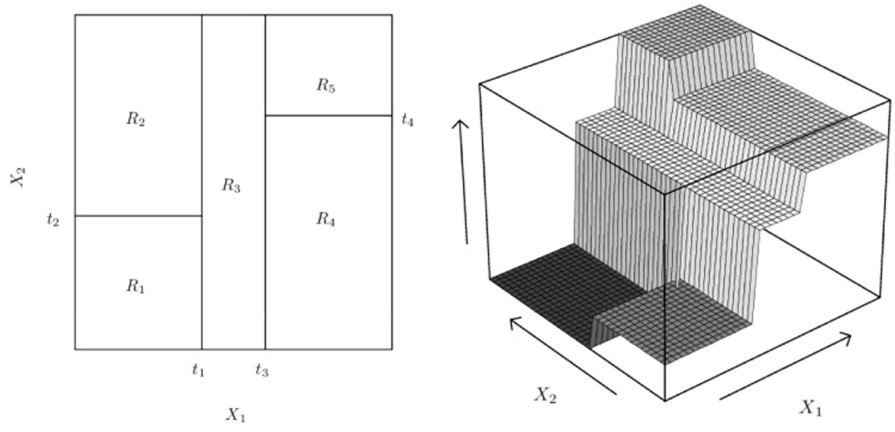

Fig. 2. Determining the regions on features of the dataset in Decision Tree.

where, T is a subtree of T_0 ($T \subset T_0$). $|T|$ is the number of terminal nodes in tree T. R_m represents the region associated with the terminal node m and α is a hyperparameter controlling the complexity of the model. To reduce the output of this formula, the model either decreases the number of nodes or lowers the squared error. The optimal tree size is determined through k-fold cross-validation, which ensures that the tree is neither too large (overfitting) nor too small (underfitting), leading to the best predictive performance. Ultimately, DTs construct a multidimensional structure over the dataset to generate predictions. This approach has proven to be one of the most effective and widely used methods for both classification and regression tasks, with its adequacy and efficiency well-established in various applications [23].

D. Random Forrest

The Random Forest (RF) algorithm is fundamentally built on DTs and belongs to the class of Ensemble Learning (EL) methods. It specifically utilizes bagging, a technique in which multiple DTs are constructed on different subsets of the same dataset, and the final prediction is the average of the individual tree outputs. The key objective of RF is to reduce the correlation between trees by constructing smaller, randomized trees on distinct sub-datasets, thereby enhancing prediction accuracy.

Several DT algorithms are used as the foundation of RF, such as ID3 [24], C4.5 [25], and CART [26]. While CART relies on the principles explained in the previous section, algorithms like ID3 and C4.5 utilize entropy as an information measure to evaluate the quality of a split at a particular node. This measure, called information gain, determines the feature with which the tree will begin. The entropy for a dataset is calculated as follows:

$$info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{8}$$

where, p_i represents the probability of selecting a feature. After calculating the entropy for the entire dataset, the entropy for individual features is determined:

$$info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} . I(D_j) \tag{9}$$

The *information gain* is then calculated as the difference between the overall entropy and the feature-specific entropy:

$$Gain(A) = info(D) - info_A(D) \tag{10}$$

The feature with the highest information gain is chosen as the root node of the tree, and subsequent branches are formed by repeating this process for the remaining features.

Once all the trees in the RF have been constructed, the algorithm aggregates the predictions from each tree, and the average of these predictions—after reducing their correlation—provides the final output. One of the main strengths of the RF algorithm is its applicability to both classification and regression tasks, yielding high-quality results in both domains [27]. Additionally, RF is highly resilient to overfitting; if a sufficient number of trees is included, the likelihood of overfitting decreases.

Several hyperparameters need to be fine-tuned to optimize the performance of RF, such as the number of trees, the maximum number of features used at each split, and the minimum number of leaves required to split a node. Proper tuning of these parameters is essential for achieving high predictive accuracy. The number of trees and the randomness introduced in generating subsets are especially critical to ensuring robust performance. Although RFs can be computationally intensive due to their parallel structure, they consistently deliver excellent predictive results, making them a cornerstone of EL approaches [28].

III. IMPLEMENTATION AND PROPOSED METHOD

This section provides a detailed description of the dataset used for training the proposed algorithms and outlines the key implementation parameters. Additionally, the characteristics of the proposed method are explained.

A. Dataset

The solar power generation primarily depends on two factors: the type of solar panels used (such as crystalline or semiconductor-based panels) and the prevailing meteorological conditions at the time of generation. While the structure of the panels impacts the power plant's performance in a stable manner, real-time weather conditions dynamically influence the amount of energy generated, either positively or negatively. Given these considerations, the models developed in this study rely on meteorological data. Specifically, weather data collected at intervals of less than 5 minutes over the past 5 years, along with total energy output from solar power plants in the region, are used to train the models.

The dataset comprises eight meteorological features, including ambient temperature, direct radiation, diffuse radiation, ultraviolet (UV) radiation, wind speed, wind direction, precipitation, and atmospheric pressure. Given that this data spans 5 years, using all features for training would significantly increase the computation time. To address this, the *P*-value for each feature was calculated, allowing the selection of the most impactful variables for model training. Fig. 3 presents a bar graph depicting the importance of each feature in relation to solar power generation.

Before training, data preprocessing was performed to handle missing values, and the dataset was adjusted to an hourly frequency. The preprocessed dataset was saved for further use. Rather than utilizing all eight meteorological variables, the analysis focused on four features considered most directly related to power generation: ambient temperature, global radiation, diffuse radiation, and ultraviolet radiation.

Fig. 4 shows the variation in ambient temperature over 140 randomly selected hours. Changes in ambient temperature affect the temperature of the solar panels, which in turn influences energy production. Among the selected features, radiation data is considered the most significant factor affecting power generation. Fig. 5 visualizes three types of radiation—global, diffuse, and ultraviolet—over the same 140-hour period. As shown, global radiation is the dominant factor, followed by diffuse radiation and ultraviolet radiation.

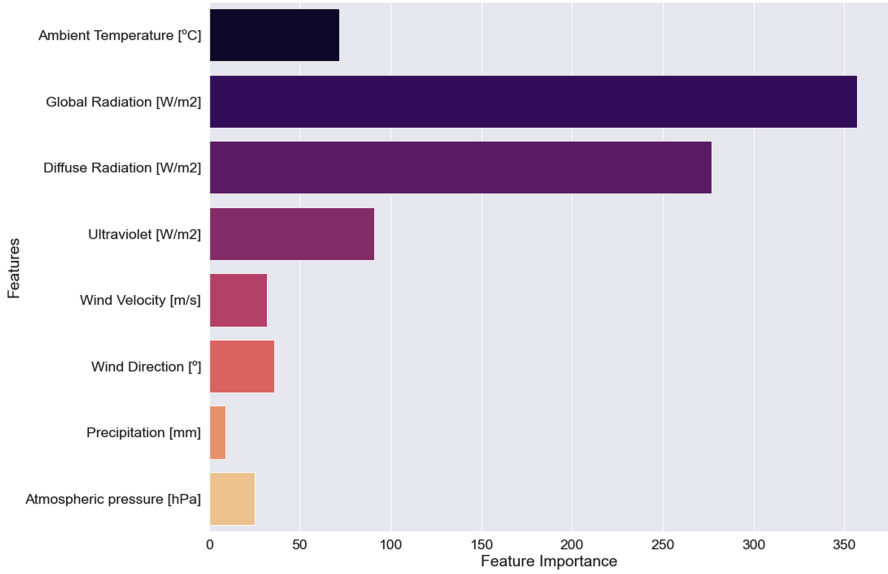

Fig. 3. Data analysis: feature importance levels of the dataset.

To further assess the relationship between radiation data and energy production, Fig. 6 visualizes the power generation data from a solar power plant in the same region over the same time period.

A comparison between Figs. 5 and 6 reveals a clear relationship between the variations in radiation and power generation. To further illustrate this correlation, a heatmap was created to show the relationships between the various meteorological features and power generation.

The heatmap confirms a strong correlation between power generation and different types of radiation, with global radiation exhibiting the highest correlation with energy output. Additionally, diffuse and ultraviolet radiation also demonstrate notable effects on power generation. Based on these findings, ML models are trained using four key input variables—ambient temperature, global radiation, diffuse radiation, and ultraviolet radiation—along with the power generation data as the output label.

B. Evaluation Methods for Model Performances

Loss (error) functions are essential tools widely used in various areas of AI to evaluate the performance of trained models. In deep learning, they are employed to adjust the weights and biases of neural networks via backpropagation. In ML, they measure the discrepancy between the model's predicted outputs and the actual values. One of the most commonly used loss functions is the Mean Square Error (MSE), defined in (11).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{11}$$

Where y_i represents the actual data and $\hat{y}_i$ denotes the predicted values for index *i*. Mean square error is a robust metric for evaluating the performance of models by quantifying the average squared difference between predicted and actual values. Another closely related metric is the Root Mean Square Error (RMSE):

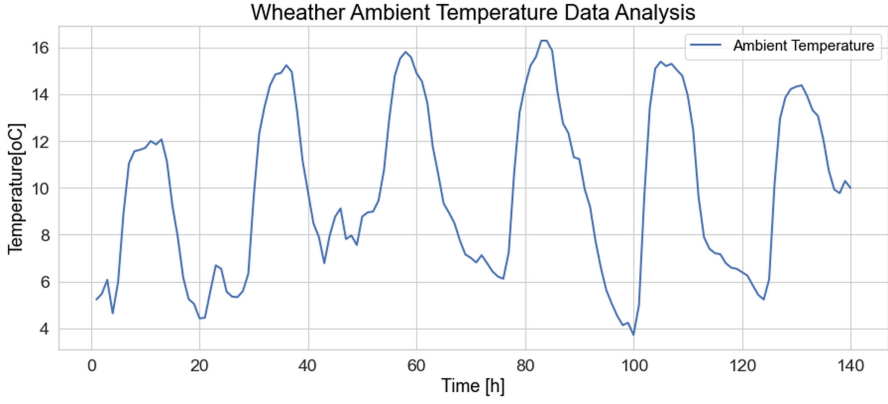

Fig. 4. Ambient temperature over 140 randomly selected hours.

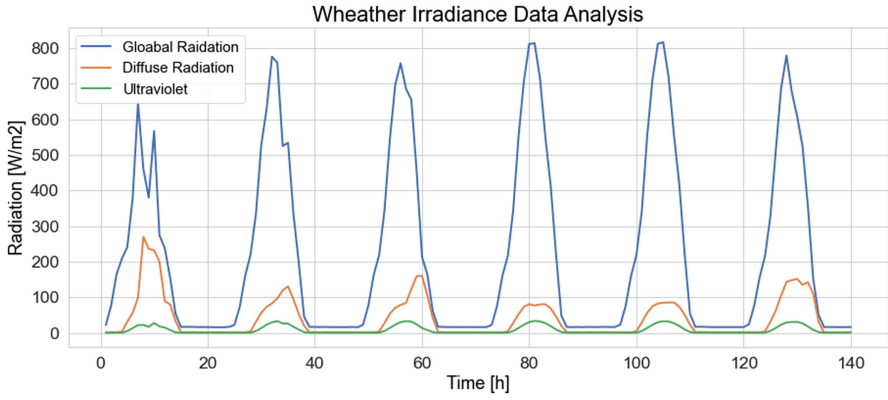

Fig. 5. Three types of radiation over 140 randomly selected hours.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{12}$$

While MSE provides the average squared deviation, RMSE gives the average magnitude of the prediction error. Both metrics are useful for understanding different aspects of model performance. Additionally, the Mean Absolute Error (MAE) is used in this study, defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{13}$$

These three error metrics—MSE, RMSE, and MAE—are utilized to evaluate and enhance the performance of the models and inform the unification strategy in the proposed hybrid approach.

C. Proposed Hybridization Method

Developing a forecasting model, training it, and deploying it after achieving satisfactory results has been extensively explored in the literature. An additional approach involves training a single model multiple times with different kernel functions and hyperparameters, yielding positive outcomes. Another prominent method is hybridization, where multiple models are combined to improve prediction accuracy. This study focuses on the latter approach, emphasizing the creation and integration of multiple ML models.

The models are trained using the algorithms outlined in Section II and the dataset described in the subsection *A. Dataset.* After training, each model is evaluated using the test dataset to generate prediction values. These predictions are then combined using an innovative voting method, formalized in (14):

$$\hat{y} = \frac{R_1^2.y_1 + R_2^2.y_2 + R_3^2.y_3 + R_4^2.y_4}{4}\ldots. \\ \ldots+\left(y^{\alpha} - y^{\beta}\right)\left(c_0\left(R^{2^{\alpha}} - R^{2^{\beta}}\right)\right) \tag{14}$$

where

- $\hat{y}$ is the final prediction value.
- y_1, y_2, y_3, and y_4, are predicted values of four ML models, respectively.
- $R_1^2, R_2^2, R_3^2, R_4^2$: R-square scores of four ML models, respectively.
- y^{α} is the predicted value of the model that has the highest accuracy score.

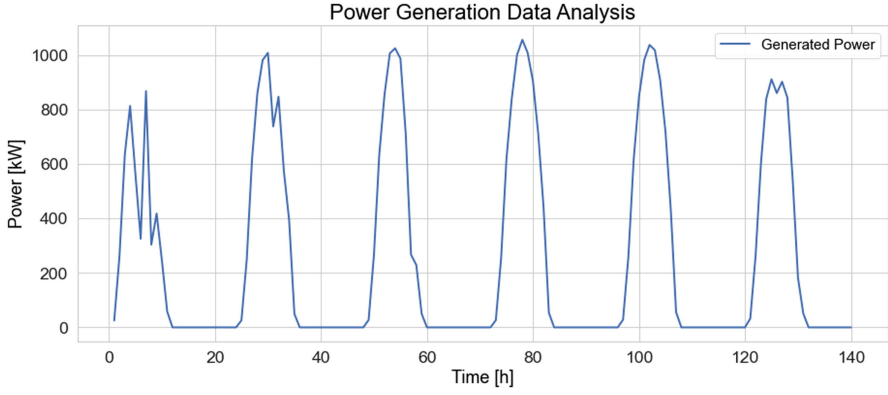

Fig. 6. Power generation data over the same 140 hours.

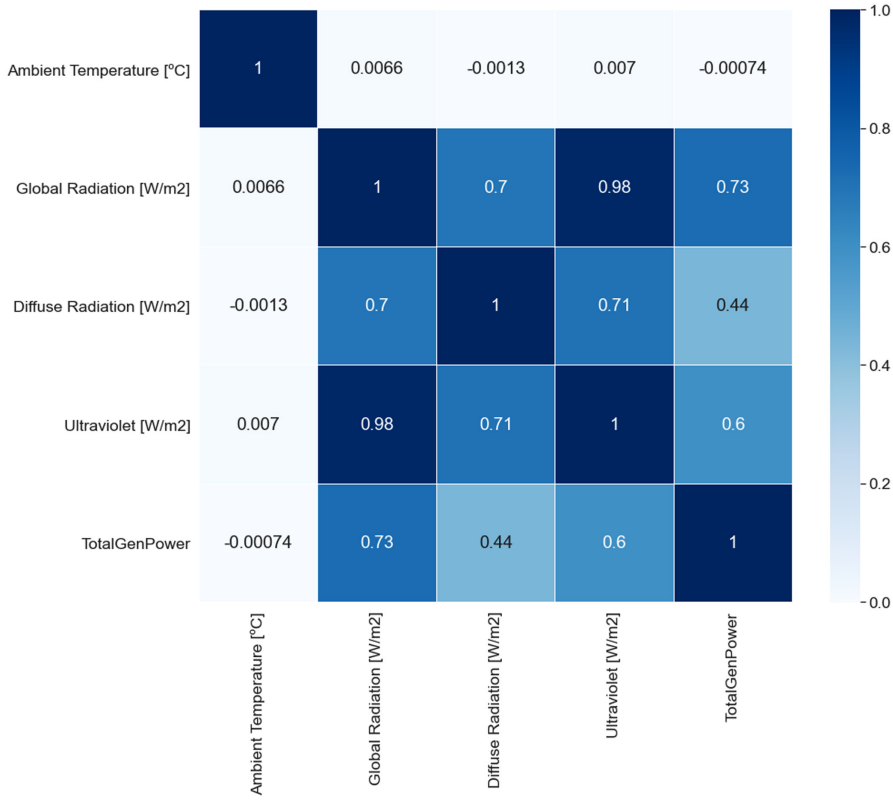

Fig. 7. Heatmap of correlations between features and power generation.

- y^{β} is the predicted value of the model that has the lowest accuracy score.
- c_0 consolidation and improvement constant, which is selected as 1.5 for four ML models.
- $R^{2^{\alpha}} - R^{2^{\beta}}$: difference between the best and worst accuracy score.

This hybrid approach ensures that the model with the highest accuracy contributes the most to the final prediction, while the model with the lowest accuracy has the least influence. Moreover, the difference between the best and worst models is factored into the final output, further enhancing the overall prediction. Fig. 8 illustrates the general framework of the proposed hybrid prediction model.

IV. PERFORMANCE ANALYSIS

The predictive models, developed using four distinct ML algorithms, were trained on meteorological data and power generation records from the past 5 years. This section examines the performance of these models. In addition to the error metrics outlined in Section III, model accuracy is also considered a key performance measure. The *R*-squared (R^2) value is used to calculate accuracy, defined as [29]:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i-1}^{n} \left(y_i - \bar{y}_i\right)^2} \tag{15}$$

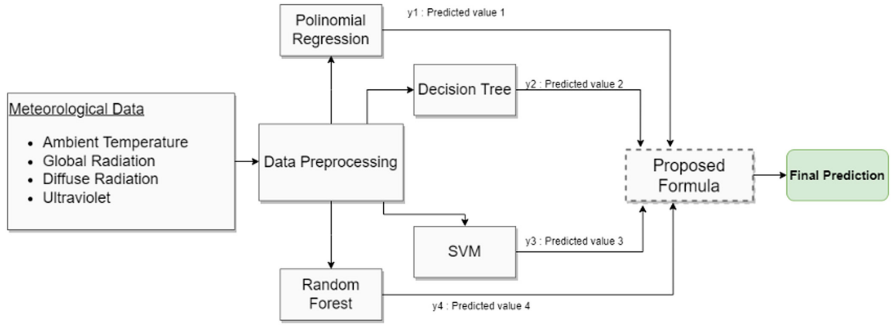

Fig. 8. General framework of the proposed hybrid prediction model.

TABLE I.
PERFORMANCE RESULTS AND COMPARISON

	Performance Matrices				
Method	**Acc. (%)**	**MSE**	**RMSE**	**MAE**	**R^2**
Poly.Reg.	89.83%	10387	101.9	64.53	0.8983
SVM	81.92%	14893	133.0	72.45	0.8192
DT	87.47%	12992	113.9	42.45	0.8747
RF	93.01%	6933	83.27	35.82	0.9301
Hybrid	94.22%	5899	76.81	32.21	0.9422

Poly. Reg.: Polynomial Regression, SVM: Support Vector Machine, DT: Decision Tree, RF: Random Forest, MSE: Mean Square Error, RMSE: Root Mean Square Error, MAE: Mean Absolute Error, R2: R-squared Value.

where SS_{RES} represents the sum of squared residuals from the regression model, and SS_{TOT} is the total sum of squares from the mean model. The R^2 value, ranging between 0 and 1, reflects model performance, where values closer to 1 indicate better accuracy. Multiplying R^2 by 100 yields the accuracy percentage.

For a comprehensive performance analysis, the accuracy percentage, R^2, MSE, RMSE, and MAE values were calculated for all models. Additionally, the proposed hybrid model's prediction results were evaluated using the same metrics. Table I summarizes the performance and error metrics for all models.

Evaluating the error metrics, the RF algorithm demonstrated the smallest deviation between actual and predicted values, while the SVM algorithm exhibited the largest deviation. Interestingly, the model produced a lower MAE than the Poly. Reg., despite its lower accuracy, indicating a smaller average error between the actual and predicted values.

Table I also shows that the proposed hybrid model, using the unification method, achieved the best overall performance, both in terms of accuracy and error metrics. The hybrid model had the lowest RMSE, signifying minimal deviation, and the lowest MAE, indicating the smallest difference between actual values and predictions. While the error metrics—MSE, RMSE, and MAE—are often considered similar, they highlight distinct aspects of model performance.

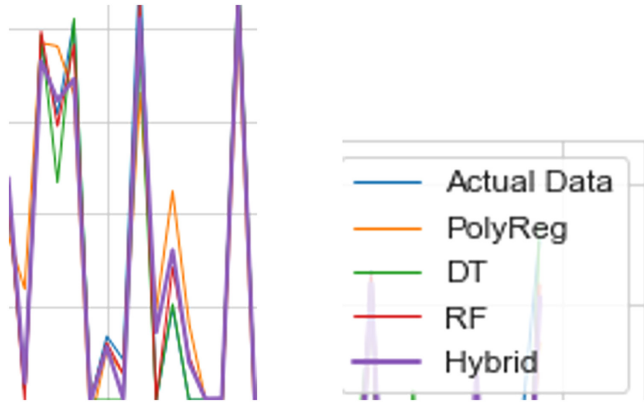

Fig. 10. Closer examination of model performance in sharp transition regions.

Figure 9 provides a visual comparison of the proposed hybrid prediction model with the other ML models. In this figure, the predicted values from the hybrid model are shown in purple, while actual values are plotted in blue. Additionally, the Poly. Reg. results are shown in orange, the DT model in green, and the RF model in red. As evident from the graph, the Poly. Reg. model tends to overestimate at sharp transition points, and the DT model often produces predictions significantly higher than the actual values in certain cases.

A closer view of the sharp transition regions is presented in Fig. 10, where the proposed hybrid model (in purple) demonstrated the closest match to actual values in these regions. Accurate predictions in areas with rapid fluctuations in power generation are crucial for effective decision-making in power plant operations. The zoomed-in graph emphasizes the success of the proposed method in predicting sharp changes with high accuracy.

While the RF model delivered relatively good prediction results, the hybrid model produced predictions almost identical to actual values. To further assess the hybrid model's performance, Fig. 11 visualizes both the actual and predicted data, with the forecasted results shown in orange and actual values in blue. Apart from a few

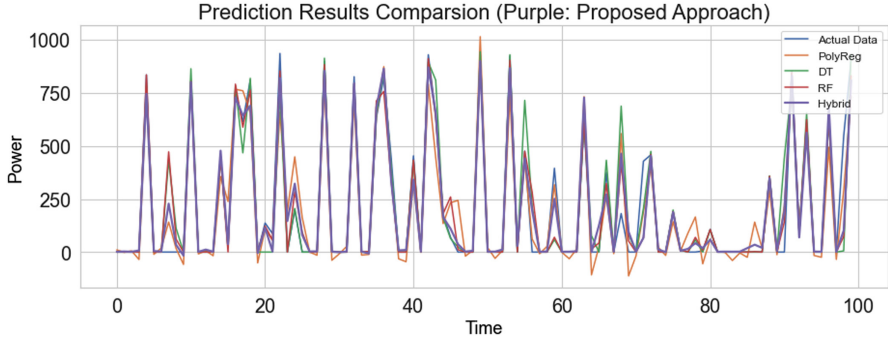

Fig. 9. Comparison of the proposed method with the prevalent machine learning models.

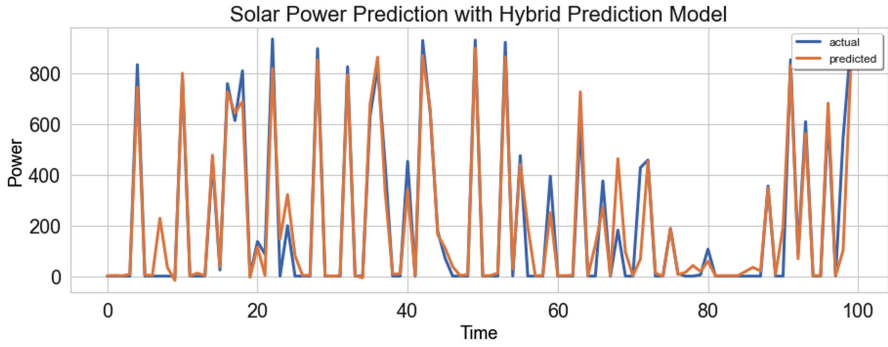

Fig. 11. Solar power prediction using proposed hybrid model (actual vs predicted values).

sharp transition points, the two sets of data overlap nearly perfectly. Notably, the hybrid model tends to slightly underestimate power generation at sharp transitions, a more cautious approach that is advantageous for avoiding operational errors.

V. CONCLUSION

The dynamic nature of meteorological conditions significantly affects the power generation of solar power plants, making accurate forecasting essential for operational decision-making. This study addresses the need for more reliable predictions by developing a hybrid prediction model that combines multiple ML algorithms. By training these models on real meteorological and production data, the hybrid model achieved superior accuracy compared to individual models, reducing prediction errors and deviations.

Briefly, our research aimed to advance predictive modeling by proposing a hybrid prediction model that combines four prevalent ML methods. The individual methods, including Poly. Reg. (89.83% accuracy), SVM (81.92% accuracy), DT (87.47% accuracy), and RF (93.01% accuracy), each demonstrated notable performance. However, our proposed hybrid prediction method surpassed them all, achieving an impressive accuracy of 94.22%. Not only did our hybrid model exhibit superior accuracy, but it also outperformed the prevalent methods in terms of error metrics. Comparative results of MSE, RMSE, and MAE consistently favored our hybrid prediction method, highlighting its efficacy in minimizing prediction errors.

These quantitative outcomes emphasize the practical relevance and potential impact of our hybrid model in solar power prediction. The robust performance across various metrics positions our approach as a promising solution for predictive tasks. Moreover, the proposed model enhances decision-making processes in solar plant operations, such as managing battery storage systems and planning for system decommissioning or expansion. The study highlights the critical role of accurate forecasting in maximizing the efficiency of solar power plants and the broader potential for ML to contribute to renewable energy optimization. The hybrid model's performance underscores its value in guiding future strategies for sustainable energy production. In summary, our work not only contributes to the ongoing discourse on improving prediction accuracy but also establishes a foundation for the adoption of hybrid models in solar power prediction. The superior results in both accuracy and error metrics underscore the significance of our proposed approach, marking a noteworthy advancement in predictive modeling.

Availability of Data and Materials: The data that support the findings of this study are available on request from the corresponding author.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – N.A., I.G.; Design – N.A.; Supervision – I.G.; Resources – N.A.; Materials – N.A., I.G.; Data Collection and/or Processing – N.A., I.G.; Analysis and/or Interpretation – N.A.; Literature Search – N.A., I.G.; Writing – N.A.; Critical Review – N.A., I.G.

Declaration of Interests: The authors have no conflicts of interest to declare.

Funding: This study received no funding.

REFERENCES

1. E. D. Obando, S. X. Carvajal, and J. Pineda Agudelo, "Solar radiation prediction using machine learning techniques: A review," *IEEE Lat. Am. Trans.*, vol. 17, no. 4, pp. 684–697, 2019. [CrossRef]
2. M. Elsaraiti, and A. Merabet, "Solar power forecasting using deep learning techniques," *IEEE Access*, vol. 10, pp. 31692–31698, 2022. [CrossRef]
3. R. Chang, L. Bai, and C.-H. Hsu, "Solar power generation prediction based on deep learning," *Sustain. Energy Technol. Assess.*, vol. 47, p. 101354, 2021. [CrossRef]
4. H. S. Jang, K. Y. Bae, H.-S. Park, and D. K. Sung, "Solar power prediction based on satellite images and support vector machine," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1255–1263, 2016. [CrossRef]
5. J. Zheng *et al.*, "Time series prediction for output of multi-region solar power plants," *Appl. Energy*, vol. 257, p. 114001, 2020. [CrossRef]
6. M. H. Shams, H. Niaz, B. Hashemi, J. J. Liu, P. Siano, and A. Anvari-Moghaddam, "Artificial intelligence-based prediction and analysis of the oversupply of wind and solar energy in power systems," *Energy Convers. Manag.*, vol. 250, p. 114892, 2021. [CrossRef]
7. Q. Liu, and Q.-J. Zhang, "Accuracy improvement of energy prediction for solar-energy-powered embedded systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 24, no. 6, pp. 1–13, 2016. [CrossRef]
8. F. Serttas, F. O. Hocaoglu, and E. Akarslan, "Short term solar power generation forecasting: A novel approach," in International Conference on Photovoltaic Science and Technologies (PVCon), vol. 2018, 2018, pp. 1–4. [CrossRef]
9. R. Sabzehgar, D. Z. Amirhosseini, and M. Rasouli, "Solar power forecast for a residential smart microgrid based on numerical weather

predictions using artificial intelligence methods," *J. Build. Eng.*, vol. 32, p. 101629, 2020. [CrossRef]

10. R. H. M. Zargar, and M. H. Yaghmaee Moghaddam, "Development of a markov-chain-based solar generation model for smart microgrid energy management system," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 736–745, 2020. [CrossRef]
11. F. Ramahatana, J. Le Gal La Salle, P. Lauret, and M. David, "A more efficient microgrid operation through the integration of probabilistic solar forecasts," *Sustain. Energy Grids Netw.*, vol. 31, p. 100783, 2022. [CrossRef]
12. N. Aksoy, and I. Genc, "Energy storage management for microgrids using n-step bootstrapping," in *IEEE Can. J. Electr. Comput. Eng.* IEEE Publications, vol. 46, no. 2, pp. 107–116, 2023. [CrossRef]
13. A. T. Eseye, J. Zhang, D. Zheng, H. Ma, and G. Jingfu, "A double-stage hierarchical ANFIS model for short-term wind power prediction," in 2nd International Conference on Big Data Analysis (ICBDA), vol. 2017. IEEE Publications, 2017, pp. 546–551. [CrossRef]
14. K. Li, R. Wang, H. Lei, T. Zhang, Y. Liu, and X. Zheng, "Interval prediction of solar power using an Improved bootstrap method," *Sol. Energy*, vol. 159, pp. 97–112, 2018. [CrossRef]
15. A. Asrari, T. X. Wu, and B. Ramos, "A hybrid algorithm for short-term solar power prediction—Sunshine state case study," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 582–591, 2017. [CrossRef]
16. B. Zazoum, "Solar photovoltaic power prediction using different machine learning methods," *Energy Rep.* The 8th International Conference on Power and Energy Systems Engineering, vol. 8, pp. 19–25, 2022. [CrossRef]
17. L. A. Dao, L. Ferrarini, and D. La Carrubba, *"Improving Solar and PV Power Prediction with Ensemble Methods", arXiv*, 2020. [CrossRef]
18. M. Nejati, and N. Amjady, "A new solar power prediction method based on feature clustering and hybrid-classification-regression forecasting," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1188–1198, 2022. [CrossRef]
19. S. Wen, C. Zhang, H. Lan, Y. Xu, Y. Tang, and Y. Huang, "A hybrid ensemble model for interval prediction of solar power output in ship onboard power systems," *IEEE Trans. Sustain. Energy*, vol. 12, no. 1, pp. 14–24, 2021. [CrossRef]
20. E. Alpaydin, *Introduction to Machine Learning*. Boston: The MIT Press, 2014.
21. P. Rivas-Perea, J. Cota-Ruiz, D. G. Chaparro, J. A. P. Venzor, A. Q. Carreón, and J. G. Rosiles, "Support vector machines for regression: A succinct review of large-scale and linear programming formulations," *Int. J. Intell. Sci.*, vol. 03, no. 1, pp. 5–14, 2013. [CrossRef]
22. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020. [CrossRef]
23. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statical Learning*. New York: Springer, 2021.
24. J. R. Quinlan, "Introduction to decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 2013. [CrossRef]
25. J. R. Quinlan, *Programs for Machine Learning*. Amsterdam: Elsevier Press, 1993, p. C4.5.
26. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York, United States of America: Chapman Hall/London, New York: CRS, 1984.
27. K. Gao, Y. Wang, and L. Ma, "Belief entropy tree and random forest: Learning from data with continuous attributes and evidential labels," *Entropy*, vol. 24, no. 5, 2022. [CrossRef]
28. N. Aksoy, *A Novel Artificial Intelligence Based Energy Management System for Microgrids* PhD thesis. Istanbul: Istanbul Technical University, TR, 2023. Available: http://hdl.handle.net/11527/24381.
29. M. Purlu, and B. E. Turkay, "Estimating the distributed generation unit sizing and its effects on the distribution system by using machine learning methods," *Elektronika ir Elektrotechnika*, vol. 27, no. 4, pp. 24–32, 2021.