IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 July 2024, accepted 23 July 2024, date of publication 1 August 2024, date of current version 14 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3437191

RESEARCH ARTICLE

Assessing Machine Learning Approaches for Photovoltaic Energy Prediction in Sustainable Energy Systems

MONTASER ABDELSATTAR[1], MOHAMED A. ISMEIL[2], (Senior Member, IEEE), MOHAMED M. A. AZIM ZAYED[3,4], (Senior Member, IEEE), AHMED ABDELMOETY[1], AND AHMED EMAD-ELDEEN[5]

[1]Electrical Engineering Department, Faculty of Engineering, South Valley University, Qena 83523, Egypt
[2]Electrical Engineering Department, Faculty of Engineering, King Khalid University, Abha 61411, Saudi Arabia
[3]Electronics Technology Department, College of Education and Technology, Beni-Suef University, Beni Suef 2722165, Egypt
[4]Department of Computer Science, College of Computer and Cyber Sciences, University of Prince Mugrin, Madinah 42241, Saudi Arabia
[5]Renewable Energy Science and Engineering Department, Faculty of Postgraduate Studies for Advanced Sciences "PSAS", Beni-Suef University, Beni-Suef 62511, Egypt

Corresponding author: Montaser Abdelsattar (Montaser.A.Elsattar@eng.svu.edu.eg)

This work was supported by the Deanship of Scientific Research, King Khalid University, through the Large Group Research Project RGP2/167/45.

ABSTRACT Precise forecasting of solar power output is crucial for integrating renewable energy into power networks, improving efficiency and dependability. This study assesses the efficacy of several Machine Learning (ML) algorithms in predicting solar power generation through a detailed performance comparison. This paper analyzes six algorithms: CatBoost, Gradient Boosting Machines (GBMs), Multilayer Perceptron (MLP) regressor, Support Vector Machines (SVMs), XGBoost, and Random Forest (RF). Using a dataset of 4213 sets of solar power generation data, each model was trained and tested, with performance evaluated based on R-squared (R^2) scores for the whole dataset, training set, and test set. Also, this study examined the mean and standard deviation of test set predictions to gauge how consistent each model was. The results showed that RF had the highest overall R^2 score of 0.940 and a training set score of 0.971. XGBoost demonstrated exceptional performance on the test set, attaining a high R^2 score of 0.822. CatBoost and GBMs exhibited strong performance, albeit with slightly lower R^2 values of 0.786 and 0.829, respectively. Although the MLP regressor and SVMs exhibited high training scores, they encountered difficulties in generalizing to unfamiliar data. This paper highlights the effectiveness of combining XGBoost and RF techniques in improving the accuracy of solar power forecasts. The investigation focuses on enhancing the precision and reliability of renewable energy projections through a comprehensive comparison of various contemporary ML techniques.

INDEX TERMS Machine learning, predictive modeling, renewable energy forecasting, solar power generation.

I. INTRODUCTION

Among the main and most durable forms of renewable energy, solar power provides an abundant and long-lasting way to produce electricity [1]. With the world facing the problems of climate change, the need for renewable energy to lower Greenhouse Gas (GHG) emissions is growing [2]. Solar energy contributes to the diversification of the energy supply and reduces reliance on unpredictable fossil fuel markets [3]. In addition, solar power plays a role in promoting sustainable development by offering a source of clean and renewable energy that may foster economic expansion without causing harm to the environment [4]. Adopting renewable energy sources such as solar power is crucial for attaining worldwide energy sustainability and successfully addressing climate change [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed F. Zobaa.

Accurate predictions of solar power are important for system management, energy storage, and resource planning. In this way, we can better prepare for how unpredictable solar power production is, which helps keep the system steady and stops problems [6]. Artificial Neural Networks (ANNs) and Machine Learning (ML) are being used in predicting more and more to make solar power predictions more accurate. These models give us a lot of useful knowledge for creating power systems and making grid control work better [7]. These advanced planning methods can help solar power systems work better by taking into account problems that arise because solar energy isn't always available [8].

Accurate solar power planning is necessary for the best energy storage and grid control. It makes sure there is a steady flow of power even when solar production is low by making it easier to plan and use stored resources [9]. By correctly guessing how much solar power will be produced, energy companies may be able to lower the costs of running energy storage systems and depend less on Backup Power Sources (BPS) [10]. Forecasts that are more accurate also help utilities and grid operators make better use of their resources, which in turn helps them make better choices about how to produce and distribute energy [11].

Correct solar power forecasts can help cut down on wasted energy and make systems work better. To get the most use out of energy and lose as little as possible, utilities should time the output of solar panels with demand predictions [12]. This alignment is beneficial for businesses, everyone in the energy supply chain, and consumers [13]. It also helps lessen the damage that energy creation does to the earth. Accomplishing a smoother change in the future with sustainable energy sources is also possible, with accurate predictions that help add Green Energy Sources (GES) to the power grid [14].

Solar power output has been predicted using conventional forecasting methods, including statistical models, to great extent. But usually they find it difficult to explain the complex and non-linear features of solar power generation [15]. The classic approaches mentioned, such as Autoregressive Integrated Moving Average (ARIMA) and linear regression, have limitations due to their dependence on linear assumptions and the requirement of a large amount of historical data [16]. Because ML methods can handle large datasets and non-linear patterns well, they are ideal for precise solar power predictions [17].

ML models, including Support Vector Machines (SVM), Random Forest (RF), and ANN, have demonstrated exceptional efficacy in capturing the complex interconnections among diverse meteorological variables and solar power output [18]. The ability of these models to dynamically adjust to fluctuating weather conditions enhances the precision of solar power forecasts, a critical factor in the administration and stability of power grids [19]. For example, the incorporation of domain expertise into physical model-ML algorithm integrations, such as hybrid models that combine deep learning and statistical methods, has further improved prediction accuracy [20].

There are more advantages to ML in solar power forecasting than just non-linearity management. It makes processing and analysis of massive volumes of data possible as well. Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are two examples of very efficient time-series data processing techniques that are essential for anticipating solar power generation based on historical weather trends [21]. Multiple studies have shown that these approaches are more effective than standard statistical models, resulting in more accurate predictions for grid operators and energy planners [22].

Furthermore, the use of ensemble approaches, which combine numerous ML algorithms, has demonstrated substantial improvements in forecasting precision by capitalizing on the advantages of individual models and reducing their limitations [23]. Ensemble approaches, such as bagging and boosting, combine the predictions of many models to create a forecast that is both more accurate and resilient [24]. This technique enhances both the precision of solar power generation estimates and the understanding of the factors that impact them.

ML methods offer significant improvements over traditional statistical models when it comes to predicting solar power. ML improves the accuracy and dependability of predictions by effectively dealing with non-linear patterns, evaluating large datasets, and employing ensemble methods. These advancements are essential for improving grid management, maximizing energy storage technologies, and enabling the incorporation of renewable energy sources into the power system [14].

The growing global demand for sustainable, environmentally friendly energy sources emphasizes the importance of solar power in the transition to renewable energy [25]. Solar power generation is, by nature, uncertain because of its reliance on the weather and other environmental conditions [26]. The predictability of this element makes it very difficult to keep the grid stable, create energy strategies, and maximize operational efficiency [27]. Therefore, effective integration of solar energy into the electrical system and maximization of its possible benefits depend on a precise forecast of solar power output. Conventional approaches to predicting solar power, which rely on statistical and physical models, have drawbacks in accurately representing the intricate, non-linear connections seen in solar energy data [26]. These approaches may not effectively utilize the given data or rapidly adjust to changing conditions, resulting in less than ideal forecasts. The development of powerful ML algorithms has a chance to greatly improve the precision and dependability of solar power projections [28]. ML models have the ability to analyze vast datasets and detect complex patterns that conventional models may fail to recognize. Utilizing ML techniques can enhance the precision of forecasts, optimize the distribution and storage of energy, save operational expenses, and improve the overall reliability of the power system [29]. The objective of this study is to examine and assess the effectiveness of several cutting-edge ML algorithms in predicting solar electricity generation.

This study seeks to employ state-of-the-art ML techniques to address the pressing need for better and more accurate solar power predictions. The selected algorithms for evaluation are CatBoost, XGBoost, Multi-Layer Perceptron (MLP) regressor, Support Vector Machine (SVM), Gradient Boosting Machines (GBMs), and RF. The selection of these algorithms is based on their proven efficacy in managing intricate, multi-dimensional data and their capacity to accurately represent non-linear associations. The study's findings will aid not just the area of renewable energy forecasting, but also policymakers, grid operators, and energy planners who need direction. This paper highlights the need to use ML algorithms to improve solar power estimations. It emphasizes the need of use innovative ways to overcome the problems of integrating renewable energy and establishing a stronger, more sustainable energy system.

As environmental awareness and demand for sustainable energy alternatives have grown, solar power has emerged as a crucial and substantial source of renewable energy [30], [31], [32]. To mitigate the effects of climate change and reduce our reliance on fossil fuels, we must optimize the use of solar energy. This is because solar power systems provide clean energy without emitting pollutants [33], [34], [35]. Nevertheless, a major obstacle in the field of solar energy is the precise forecasting of the electricity generated by solar systems [36]. Precise solar power generation forecasting requires consideration of several factors, such as meteorological conditions, solar irradiation, temperature, and the physical properties of solar panels [36], [37]. Conventional statistical approaches frequently fail to accurately represent the intricate, non-linear connections between these factors and power output [36]. Therefore, the requirement for sophisticated predictive modeling approaches is of utmost importance. ML techniques provide a hopeful resolution to this obstacle [38]. ML algorithms can depict complicated patterns and correlations in data, resulting in more accurate and consistent projections. This study attempts to analyze the performance of multiple ML algorithms in forecasting solar power generation.

This work aims to assess how well a number of ML systems estimate solar electricity generation. This is necessary to effectively include solar power into the electrical system, optimize energy distribution and storage, reduce running expenses, and improve overall system reliability. The study's main objective is to assess and compare the predictive performance of various algorithms, including CatBoost, GBM, MLP regressor, Support Vector Machine (SVM), XGBoost, and RF. This assessment will be based on metrics such as Root Mean Squared Error (RMSE) and R-squared (R^2) scores, which will be calculated using both training and test datasets. The research aims to determine the optimal algorithm by examining the strengths and drawbacks of each algorithm in dealing with the unpredictability and non-linear correlations present in solar power data. Furthermore, it investigates the significance of different variables in predicting solar power by employing techniques such as Lasso regression to select and prioritize the most important features. The study aims to graphically represent and evaluate the prediction results by comparing the expected and observed values for both the training and test datasets. Furthermore, its objective is to illustrate the relationship between significant factors, such as sun azimuth, and the expected power generation. The study seeks to provide suggestions for enhancing solar power forecasts, optimizing the integration of solar energy into the power grid, and promoting the advancement of accurate and reliable prediction models.

This research is unique because it conducts a thorough examination and comparison of powerful ML algorithms that are specifically designed for predicting solar energy in sustainable energy systems. This study differs from previous research that primarily examined individual models. Instead, it comprehensively evaluates six cutting-edge algorithms—CatBoost, GBMs, Multilayer Perceptron (MLP) regressor, SVM, XGBoost, and RF—by utilizing a real-world dataset consisting of 4213 instances of solar power generation. The research stands out by using a multi-metric approach to evaluate performance. Our methodology encompasses not only conventional R^2 values, but also the mean and standard deviation of predictions on the test set. This offers a more thorough evaluation of the prediction precision and reliability of each model. Furthermore, we utilize Lasso Regression to evaluate the significance of different variables, identifying key factors that influence solar power generation and suggesting approaches to improve our models. The research demonstrates that ensemble approaches such as RF and XGBoost exhibit superior generalization capabilities, rendering them more efficacious for real-world applications. Furthermore, our research highlights the potential of hybrid models and emphasizes our focus on future research areas, demonstrating the novel contributions of this work to the field of renewable energy forecasting.

Incorporating renewable energy sources into the current power grid is crucial for attaining sustainable energy systems. Photovoltaic (PV) solar energy has the capacity to greatly decrease GHG emissions and reduce reliance on fossil fuels. Accurate prediction of solar power generation is essential for optimizing the efficiency and dependability of power grids. The integration of solar energy into the grid has been made easier by recent developments in inverter technology. One example is the advancement of transformer-less grid-tied inverters, which has played a crucial role in enhancing the efficiency and safety of PV systems. An important technological development is the five-level switched-capacitor-based inverter, which has demonstrated considerable potential in lowering leakage currents, thus improving the safety and efficiency of the solar power integration process. In a similar vein, the six-level transformer-less inverter has been created to reduce leakage currents and enhance the overall performance of the system, hence increasing the durability and dependability of PV systems. In addition, advancements such as the implementation of single-source multilevel inverters that rely on flyback DC-DC converters have improved the efficiency and expandability of PV systems. This technology enables enhanced power conversion and enhanced voltage regulation, which are crucial for the reliable integration of solar energy into the grid. The significance of precise solar power prediction models in maximizing the utilization of renewable energy sources is highlighted by these technological breakthroughs. Enhancing the predictability of solar power output allows grid operators to more effectively control energy distribution and storage, resulting in improved dependability and efficiency of the power network. This work expands upon this existing knowledge by evaluating the effectiveness of various sophisticated ML algorithms in forecasting solar power generation. By conducting meticulous analyses of performance indicators, our objective is to provide vital insights that facilitate the seamless integration of renewable energy into current power systems.

Although hybrid models were not created in this work, the results establish a strong basis for future research in this field. The thorough examination of individual algorithms yields useful insights that can guide the creation of hybrid models. Further investigation could be conducted to integrate algorithms such as RF and XGBoost in order to capitalize on their individual strengths, thus enhancing overall prediction accuracy and robustness. Hybrid models, which merge the strong resilience and precise predictions of RF with the effectiveness and regularization features of XGBoost, have the potential to outperform other models in solar power prediction tasks. In addition, the development of adaptive models that can dynamically adapt to changing environmental conditions has the potential to greatly improve their utility and reliability. These models will guarantee strong and constant performance in several conditions, including the natural oscillations in the output of solar electricity. Using ensemble approaches, choosing pertinent features, and applying adaptive learning strategies could be one possible approach to generating these hybrid models. Combining RF and XGBoost models, coupled with a feature selection technique using Lasso Regression helps to create a strong and flexible prediction system. By choosing the most important components, this mix would guarantee stability and consistency in projections, as well as increase their accuracy. As a result, this would lead to more accurate and dependable solar power forecasts. Therefore, the foundation of the work sets the stage for future progress in creating hybrid models with the goal of improving the precision, durability, and dependability of solar power forecast models.

II. LITERATURE REVIEW

Solar power generation is a crucial component of the renewable energy industry. The process entails harnessing solar energy with the use of PV cells in order to produce electricity [18]. The use of this power generation method is very sustainable and has been widely accepted owing to its minimal impact on the environment and the abundant availability of solar energy [39]. Precise solar power generation prediction is necessary to maximize the integration of solar energy into the system, provide a steady energy supply, and maintain grid stability [40]. Higher network efficiency in the energy distribution and reduced operational costs result from more effective strategic planning and management of energy resources made feasible by forecasting [22]. Still, there are other challenges that make solar power forecasts less accurate. The unpredictable nature of the weather is a major challenge as it may cause sharp variations in solar irradiance, which affects electricity generation. Data quality is a serious issue, as errors in earlier data and differences in data collection techniques may jeopardize the trustworthiness of projections [41]. In order to address these issues, it is imperative to utilize sophisticated modeling techniques and robust ML algorithms to enhance the precision of solar power forecasts. Ultimately, this will contribute to the development of a more robust and effective energy infrastructure [42]. Through several studies, the above described issues are further examined and contrasted, as **Table 1** succinctly summarizes. The table offers an extensive comparison of many ML methods used for solar power forecasting. It emphasizes their advantages, drawbacks, and particular problems they tackle. Through the analysis of this data, we may learn important information about the efficacy of different approaches and pinpoint the best approaches to raise the precision of solar power estimates.

III. METHODOLOGY

A. DATA PRESENTATION

Table 1 gives an overview of the main descriptive statistics of the dataset and provides a detailed look at the important aspects evaluated in the research on solar power production. Including 4213 sets, this dataset provides a strong foundation for analysis and predictive modeling.

The dataset includes the lowest and highest values for each variable, which cover the whole range of meteorological circumstances and power output levels that have been recorded. The temperature varies between −5.35 °C and 34.9 °C, while the relative humidity ranges from 7% to 100%. The mean values represent the average measurement of each variable, with an average temperature of 15.068 °C and a mean relative humidity of 51.361%. These mean values provide a reference point for comprehending the usual circumstances seen throughout the time of data collection.

The standard deviation numbers show how different the values of each variable are from each other. For instance, the standard variation of temperature is 8.854 °C, which means that there is a modest amount of change around the mean. Additionally, the produced power fluctuates significantly, with a standard variation of 937.957 kW. This is due to the diverse applications of solar power in various scenarios. This wide range in produced power shows how different external factors can affect solar power output.

TABLE 1. Comparative analysis of solar power forecasting studies.

Study	Year	Algorithms Used	Dataset Characteristics	Performance Metrics	Key Findings	Limitations
[43]	2023	GBM, RF, k-NN, SVM.	1.2 MW grid-integrated solar PV, 3 years data.	RMSE, Mean Absolute Error (MAE)	GBM and RF showed high accuracy; SVM outperformed k-NN.	Accuracy varied by PV system type; limited algorithm comparison.
[44]	2023	LASSO, RF, MLP, SVR, XGB.	Real-time series data from AIT, Thailand.	Prediction accuracy.	Bayesian stacking ensemble achieved best accuracy.	Complexity of ensemble methods.
[45]	2023	RF, Decision Tree.	Historical PV system data.	Prediction accuracy.	RF outperformed Decision Tree.	Dependency on historical data quality.
[22]	2022	Ensemble ML models.	Weather data, solar irradiance.	Accuracy, placement cost.	Hybrid model outperformed individual ML models.	performance may vary with different datasets and locations.
[46]	2022	RNN, SVM, ARX, FFNN-gdx, LASSO.	PV solar cells data.	Prediction accuracy.	RNN and SVM outperformed linear models.	Non-linear models complexity.
[47]	2022	MLR, PCC, XGBoost, PCA, Ridge Regression, Autoencoder, LSTM, ARIMA	Real-world data from solar power plant sites in Germany (100-8500 kW).	RMSE, MAE, R^2.	XGBoost with feature engineering improved accuracy.	High complexity of feature engineering.
[18]	2022	ANN, RF, DT, XGB, LSTM.	NREL data, Cocoa, Florida.	RMSE, MAE, R^2.	ANN produced best forecasting results.	Limited to specific geographic data.
[48]	2021	WTP, GAN, DA.	Datasets from two regions (100–8500 kW).	MAPE, RMSE	WTP-GAN hybrid model showed high accuracy.	High complexity of solar irradiance data.
[49]	2021	SVM, ANN, ELM.	São Paulo meteorological data, 1933-2014	RMSE	SVM produced lowest RMSE, ELM had fastest training rate	Location-specific parameters affect model performance.
[38]	2020	SVM, LR, NNM, RF.	Aguascalientes, Mexico, 6 months data.	MSE, MAE.	RF showed best prediction accuracy.	Limited to six months of data.

Also, things like wind speed and direction at different heights, cloud cover at different layers, and angles of impact and zenith all have large ranges and changes. For instance, the wind speed 10 meters above ground runs from 0 to 61.18 m/s, with a mean of 16.229 m/s. This means that there are times when the wind is very strong, which could affect how well solar panels work. The screens can receive any amount of sun energy based on the angle of incidence, which is between 3.755 degrees and 121.636 degrees, with a mean of 50.837 degrees.

Table 2 is a complete statistical overview to fully understand how the dataset's main traits spread out and change over time. These new ideas will assist with later steps of analysis and predictive modeling, allowing for a more complete look at the factors that affect solar power output. We can build and test forecasting models using the 4213 observations, enhancing the accuracy of solar power output predictions.

The histogram in **Figure 1** displays the frequency distribution of the ambient temperature measured at a vertical distance of two meters. The dataset has 4213 observations, which is a sufficiently large sample size for doing statistical analysis. The temperature data is normally distributed, with a conspicuous peak at the mean and a roughly symmetrical dispersion. This pattern indicates that the majority of temperature measurements exhibit a concentration around the mean value, with only a limited number of instances recording exceptionally high or low temperatures.

Figure 1 provides valuable insights into the potential influence of temperature as a variable on solar power systems' output and efficiency. It provides details regarding the typical temperature conditions encountered during the process of gathering data on solar power generation.

Figure 2 depicts the frequency distribution of relative humidity at a height of two meters above the ground. Similar to the presentation of temperature, **Figure 2** displays the relative humidity data for a total of 4213 observations. The distribution has a minor skewness, indicating a higher occurrence of lower relative humidity values, followed by a progressive decrease as humidity levels rise. This skewness indicates that the data was taken in a location that typically experiences dry weather more often than not. This distribution provides a framework for understanding the environmental factors that influence solar power production. Relative humidity may affect the performance of solar panels

TABLE 2. Descriptive statistics of key variables in the solar power generation dataset.

Variable	Min	Max	Mean	SD
Temp 2m Above Ground (°C)	-5.35	34.9	15.068	8.854
RH 2m Above Ground (%)	7	100	51.361	23.526
MSL Pressure (hPa)	997.5	1046.8	1019.34	7.023
Total Precipitation (mm)	0	3.2	0.032	0.17
Snowfall (mm)	0	1.68	0.003	0.038
Total Cloud Cover (%)	0	100	34.057	42.844
High Cloud Cover (%)	0	100	14.459	30.712
Medium Cloud Cover (%)	0	100	20.023	36.388
Low Cloud Cover (%)	0	100	21.373	38.014
Shortwave Radiation (W/m²)	0	952.3	387.759	278.459
Wind Speed 10m (m/s)	0	61.18	16.229	9.877
Wind Dir 10m (°)	0.54	360	195.078	106.627
Wind Speed 80m (m/s)	0	66.88	18.978	12
Wind Dir 80m (°)	1.12	360	191.167	108.76
Wind Speed 900mb (m/s)	0	61.11	16.363	9.885
Wind Dir 900mb (°)	1.12	360	192.448	106.516
Wind Gust 10m (m/s)	0.72	84.96	20.583	12.649

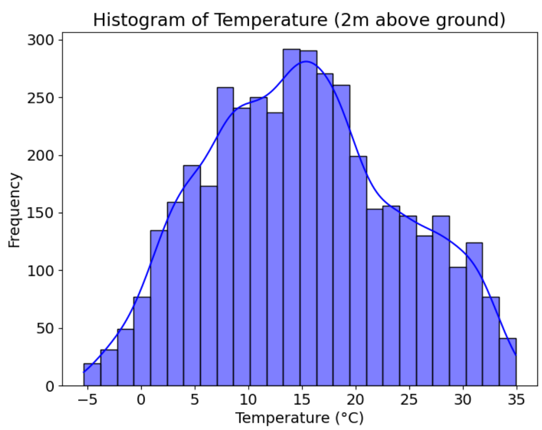

FIGURE 1. Histogram of ambient temperature at two meters above ground level.

and the total energy output by influencing the quantity of solar radiation that reaches the panels.

Figure 3 illustrates the dispersion of solar power production, quantified in kilowatts. Higher power outputs are less common, and the majority of the produced power values are focused in the lower range, as the histogram shows. Unequal distribution of solar power output is a blatant sign of significant irregularity, possibly caused by seasonal changes, weather patterns, and other environmental variables. Understanding this distribution is crucial for assessing the efficiency and dependability of solar power systems, as it highlights the difficulties in attaining steady power production. Furthermore, the fluctuation emphasizes the need for reliable prediction models to precisely estimate solar power output under different circumstances.

The scatter plot shown in **Figure 4** demonstrates the correlation between ambient temperature and solar power generation. Every data point in the dataset corresponds to each observation, and the plot demonstrates a positive correlation between temperature and power output. This indicates that increased temperatures are often associated with increased solar power output. This animation highlights the influence of temperature on the efficiency of solar electricity, offering first-hand insights into possible predictors for ML algorithms. The positive association suggests that temperature plays a crucial role in predicting solar power output. Incorporating temperature into the models may improve their accuracy and dependability.

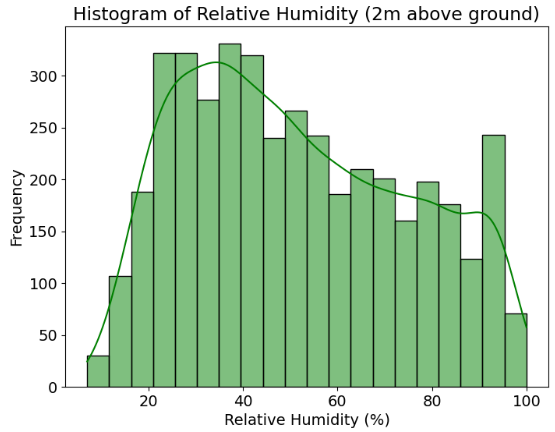

FIGURE 2. Frequency distribution of relative humidity at two meters above ground.

The heatmap in **Figure 5** depicts the correlation coefficients between various variables related to solar power generation. The findings of this comprehensive analysis reveal significant correlations between power generation, temperature, relative humidity, and total cloud cover. There is a strong negative relationship between cloud cover and power generation, as well as a positive relationship between temperature and power generation. By utilizing a heatmap, we can identify the key variables that exert the most significant influence on our predictive models for solar power generation. In order to optimize the generation of solar power, researchers must improve the accuracy of their models by gaining a deeper understanding of the correlations between different features.

B. ML ALGORITHMS

This section provides a concise summary of the ML algorithms utilized in our study for the purpose of forecasting solar power generation. We chose six different algorithms based on their demonstrated effectiveness in managing intricate datasets and their diverse methods of handling features and interpreting models. The algorithms listed are CatBoost, GBM, MLP regressor, Support Vector Machine (SVM), XGBoost, and RF. Each algorithm was extensively assessed to determine its efficacy and suitability for the assigned task. The selection of these algorithms is designed to utilize their unique benefits and address the various characteristics of our dataset, which includes a variety of environmental and

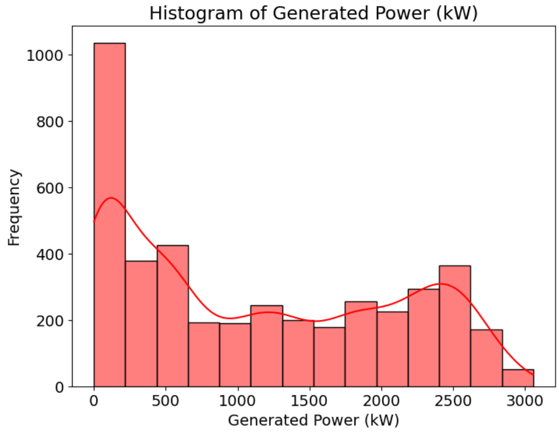

FIGURE 3. Distribution of generated solar power output in kilowatts.

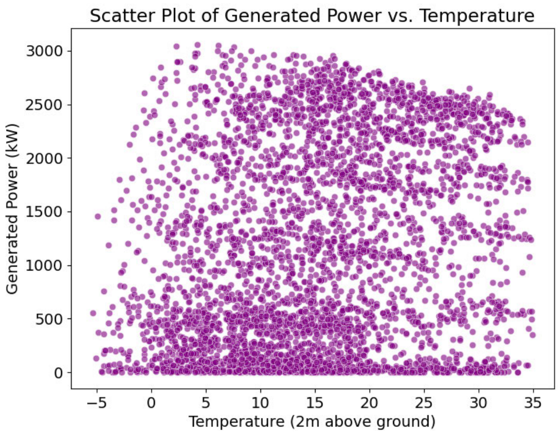

FIGURE 4. Correlation between ambient temperature and solar power output.

meteorological variables. The classification, computational complexity, feature handling capabilities, and interpretability of these algorithms are thoroughly examined in **Table 3**. This analysis attempts to provide a brief synopsis of their particular advantages and disadvantages.

Figure 6 depicts the step-by-step process employed in our methodology for forecasting solar power generation. The process begins with data preparation, where the raw data is loaded and undergoes preprocessing. This entails overseeing the handling of missing values and performing feature engineering to enhance the dataset's quality. Subsequently, the data is partitioned into distinct training and testing sets. Afterwards, the characteristics and desired outcomes are standardized in order to ensure uniformity and improve the effectiveness of the model.

During the training and prediction phase, we develop the ML model and then proceed to train it using the prepared training data. The trained model is subsequently assessed on both the training and test datasets to evaluate its performance and ability to generalize.

The evaluation and visualization phase entails the computation of essential metrics to quantify the performance of the model, which is then followed by the creation of various visual representations. Some of these methods include

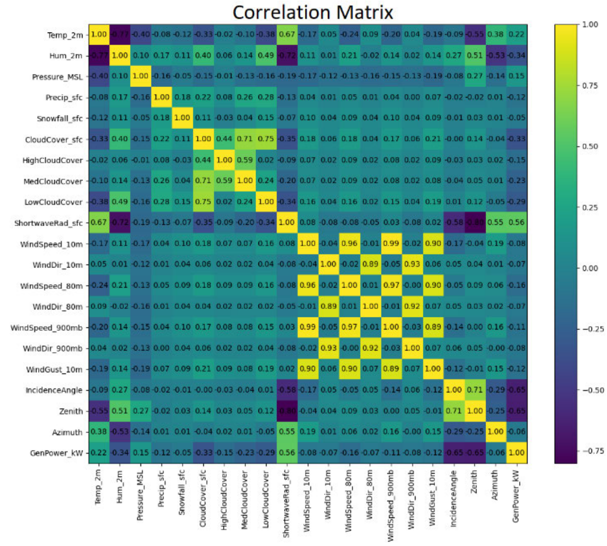

FIGURE 5. Heatmap of correlation coefficients among solar power generation variables.

creating graphs to compare predicted values with actual data, analyzing predictions in relation to the sun's azimuth, and creating a heatmap to illustrate the interconnectedness of different traits. To determine the relative importance of each attribute and the factors with the greatest influence, Lasso regression is employed.

Eventually, the comprehensive results are merged and saved to a CSV file during the result export stage. Subsequently, this file can be utilized for further examination or documentation. This systematic evaluation of ML models and their prediction capabilities provides valuable insights into the factors that influence solar power generation.

C. EVALUATION METRICS

This section provides a detailed explanation of the evaluation criteria used to determine the efficiency of ML models used to forecast solar power output. These metrics offer a comprehensive understanding of the models' accuracy and reliability. The main metrics used are the R^2 scores of each data frame, training set, and test set. Furthermore, it encompasses the average, deviation, and relative deviation of the forecasts generated on the test dataset.

The R^2 score measures the extent to which the independent variables can account for the variability in the dependent variable [50], [51]. The R^2 scores were calculated for the entire data frame using **(1)**, for the training set using **(2)**, and for the test set using **(3)**.

To understand the central tendency of the predicted values, the mean of the test set predictions is calculated [52], [53], as demonstrated in **(4)**. The dispersion of the predicted values around this mean is measured by the standard deviation [54], as shown in **(5)**. The Relative Standard Deviation (RSD) [55], depicted in **(6)**, provides a normalized measure of the variability relative to the mean. Collectively, these metrics offer a thorough understanding of the model's performance, demonstrating its precision and resilience in forecasting solar

TABLE 3. Comparative characteristics of ML algorithms used in solar power generation prediction

Algorithm	Type	Complexity	Feature Handling	Interpretability
CatBoost	Gradient Boosting Decision Trees	High	Handles categorical features natively	Moderate
GBM	Ensemble of Decision Trees	High	Requires feature preprocessing	Moderate
MLP regressor	Neural Network	High	Requires feature scaling and preprocessing	Low
SVM	Support Vector Machine	High	Requires feature scaling	Low
XGBoost	Gradient Boosting Decision Trees	High	Requires feature preprocessing	Moderate
RF	Ensemble of Decision Trees	Moderate	Handles feature interactions automatically	High

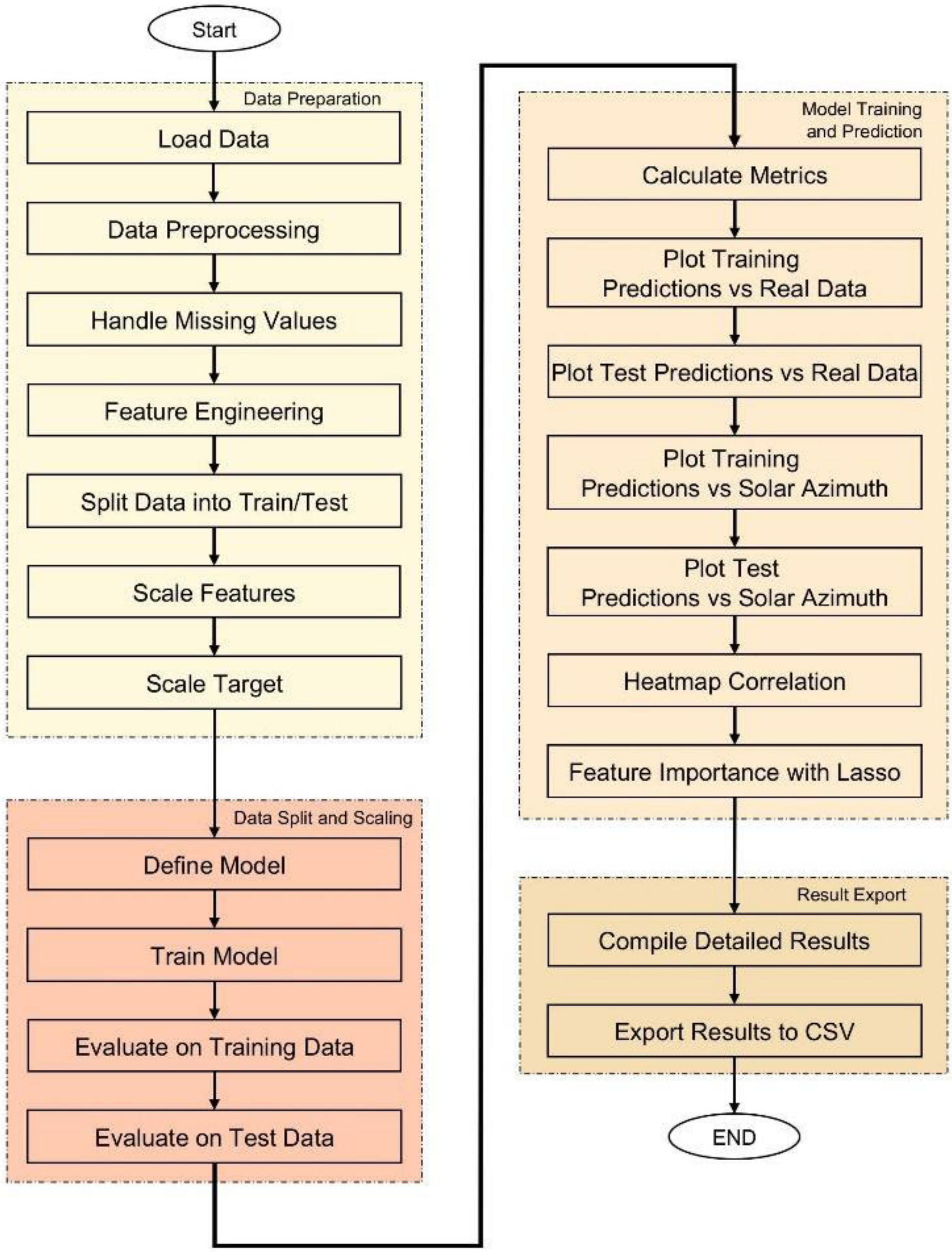

FIGURE 6. Flowchart of the predictive modeling process for solar power generation.

power generation. Assessing these metrics guarantees the reliability and interpretability of the models, thereby facilitating improved decision-making and optimization for solar power forecasting.

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

$$R^2_{\mathrm{train}} = 1 - \frac{\sum_{i=1}^{n_{train}} (y_{train,i} - \hat{y}_{train,i})^2}{\sum_{i=1}^{n_{train}} (y_{train,i} - \bar{y}_{train})^2} \tag{2}$$

$$R^2_{\mathrm{test}} = 1 - \frac{\sum_{i=1}^{n_{test}} (y_{test,i} - \hat{y}_{test,i})^2}{\sum_{i=1}^{n_{test}} (y_{test,i} - \bar{y}_{test})^2} \tag{3}$$

$$\text{Mean of Test Set} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \hat{y}_{test,i} \tag{4}$$

$$\mathrm{SD}_{\mathrm{test}} = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_{test,i} - \bar{\hat{y}}_{test})^2} \tag{5}$$

$$\mathrm{RSD} = \frac{\text{Standard Deviation of Test Set}}{\text{Mean of Test Set}} \tag{6}$$

IV. RESULTS AND DISCUSSION

In this section, we showcase the outcomes of our predictive modeling for solar power generation, utilizing a range of ML algorithms. The analysis entails a thorough assessment of the models' performance on both the training and test datasets. We use multiple metrics and visualizations to assess each algorithm's precision and ability to generalize. This section's data offers a comprehensive understanding of the models' efficacy by demonstrating how well they can use the features provided to predict power output.

Figure 7 depicts the significance of several attributes in forecasting solar power generation, as estimated by the Lasso Regression model. This study aids in determining the factors that exert the most substantial influence on the performance of the predictive model. The Lasso Regression technique uses regularization to improve the model's capacity to generalize and mitigate overfitting by punishing less significant features. This information is useful for comprehending the fundamental factors that influence solar power generation and enhancing the effectiveness of prediction models.

Figure 8 displays scatter plots comparing the actual power output to the predicted power output in the training set. Six different algorithms are used to create the plots: CatBoost, GBM, MLP regressor, SVM, XGBoost, and RF. **Figure 8(a)** through **Figure 8(f)** illustrate each subplot, which shows how well an algorithm performs. Understanding the degree of expertise each model has acquired from the training data depends on the scatter plots. A high correlation between observed and predicted values indicates excellent model performance. The majority of algorithms demonstrate a strong concentration of data points near the line of perfect prediction, indicating their successful capture of the underlying patterns in the training data. Data points dispersed and clustered along the diagonal line reveal each model's precision and dependability.

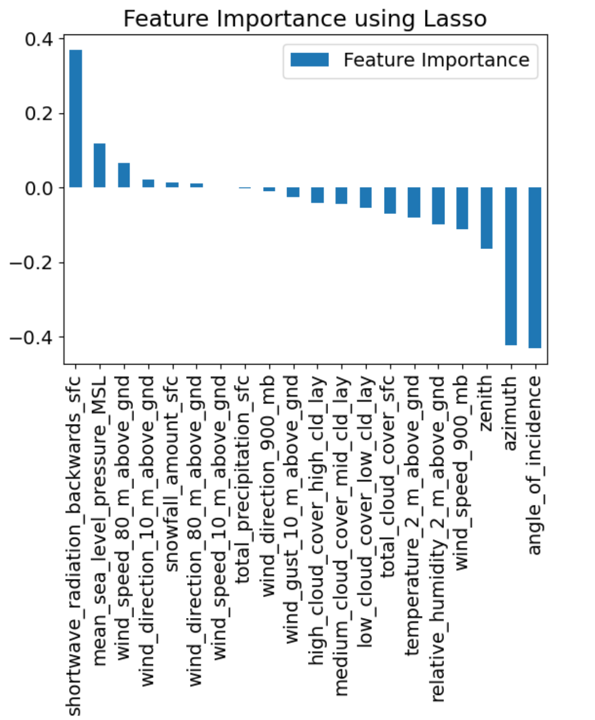

FIGURE 7. Feature importance analysis using lasso regression for solar power generation forecasting.

Figure 9 displays scatter plots comparing the actual power output to the predicted power output in the test set. The plots are generated using six algorithms: CatBoost, GBM, MLP regressor, SVM, XGBoost, and RF. Subplots in particular, **Figure 9(a)** through **Figure 9(f)** show how well the model performs on data that it has never seen before. These scatter plots are essential for evaluating each algorithm's capacity for generalization. When points on the diagonal line strongly correlate, the predictive accuracy is high. The deviations in the concentration and distribution of data points from the regression line show the extent of prediction errors. By comparing these plots with those shown in **Figure 8**, we can determine whether the models retain their predictive accuracy when applied to new data. This is crucial for real-world applications.

Figure 10 depicts the correlation between solar azimuth and power output for the training set predictions made by six algorithms: CatBoost, GBM, MLP regressor, SVM, XGBoost, and RF. The subplots **Figure 10(a)** through **Figure 10(f)** illustrate the correlation between the predictions of each model and the solar azimuth angles in the training set. The capacity of these plots to faithfully depict the impact of solar azimuth on power generation is moderately compromised. The identification of recurring trends or patterns within the data suggests that the models have adequately accounted for the azimuth angle of the sun. Any inconsistencies or abnormalities detected may suggest areas that might be improved or require new features in the model.

Figure 11 depicts the relationship between solar azimuth and power output for test set predictions using the same

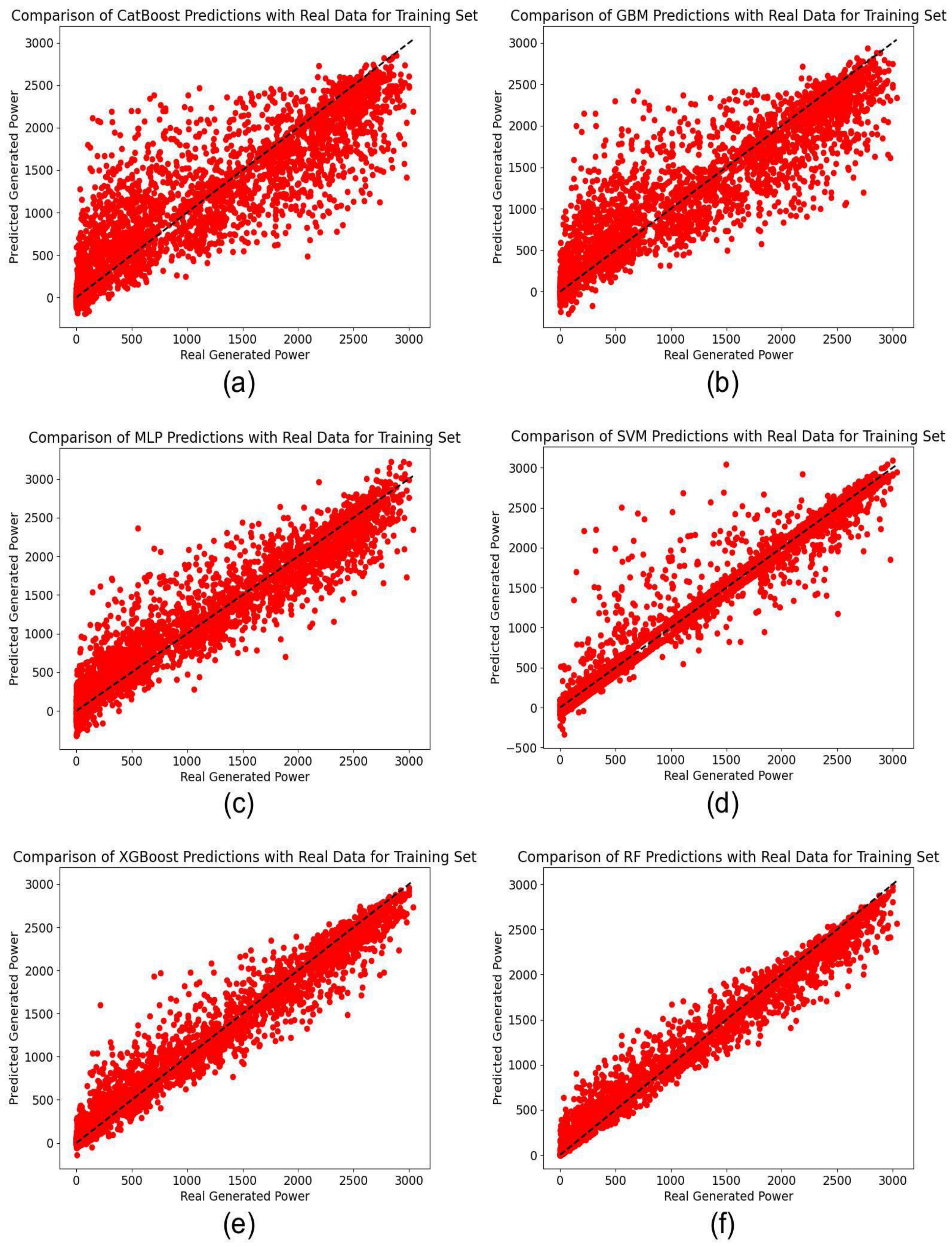

FIGURE 8. **Scatter plots of actual vs. predicted power output in the training set using different algorithms (a) CatBoost, (b) GBM, (c) MLP regressor, (d) SVM, (e) XGBoost, (f) RF.**

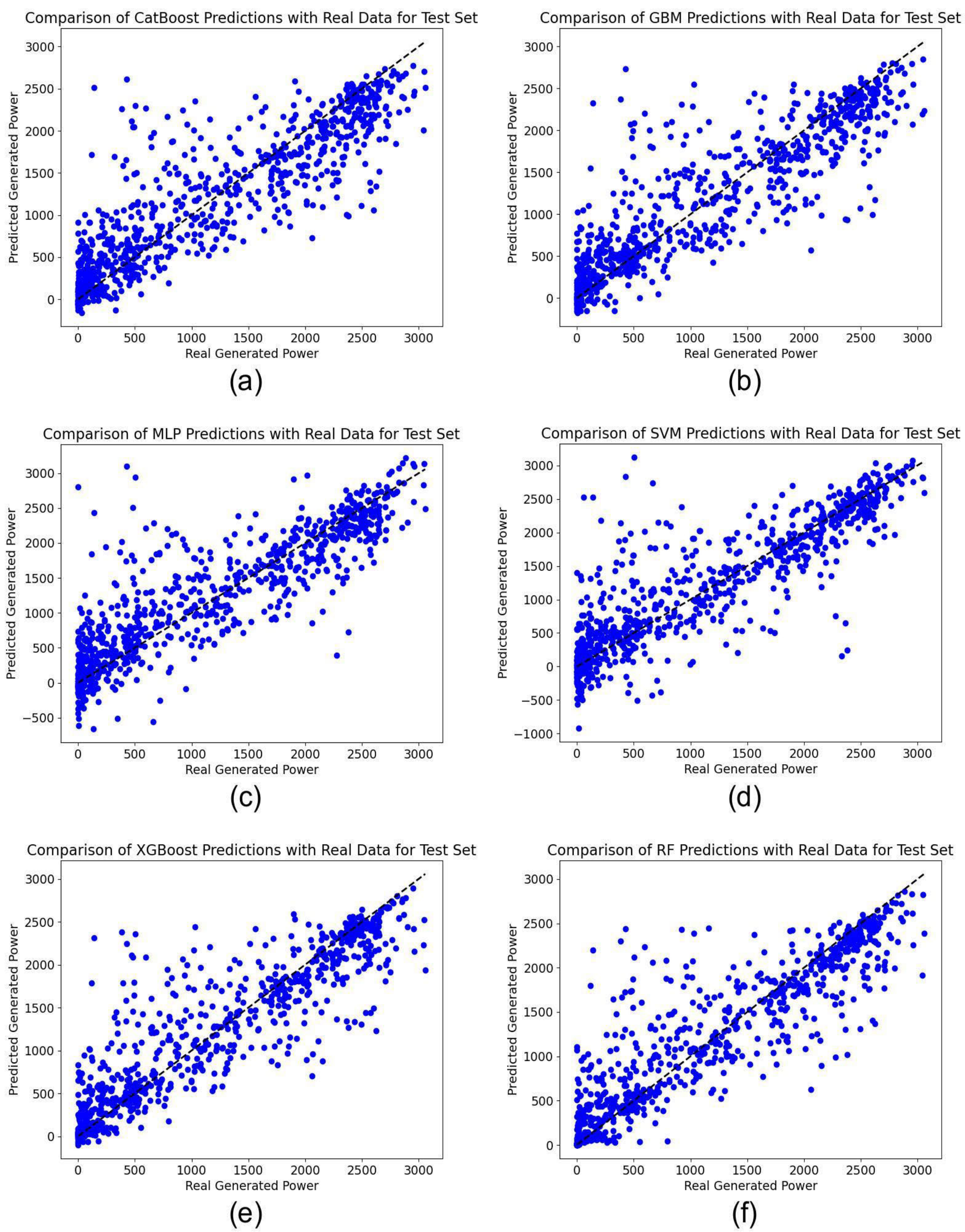

FIGURE 9. Scatter plots of actual vs. predicted power output in the test set using different algorithms (a) CatBoost, (b) GBM, (c) MLP regressor, (d) SVM, (e) XGBoost, (f) RF.

six algorithms: CatBoost, GBM, MLP regressor, SVM, XGBoost, and RF. Each subplot **Figure 11(a)** through **Figure 11(f)** shows the model's ability to predict power output relative to solar azimuth angles in the test set. These

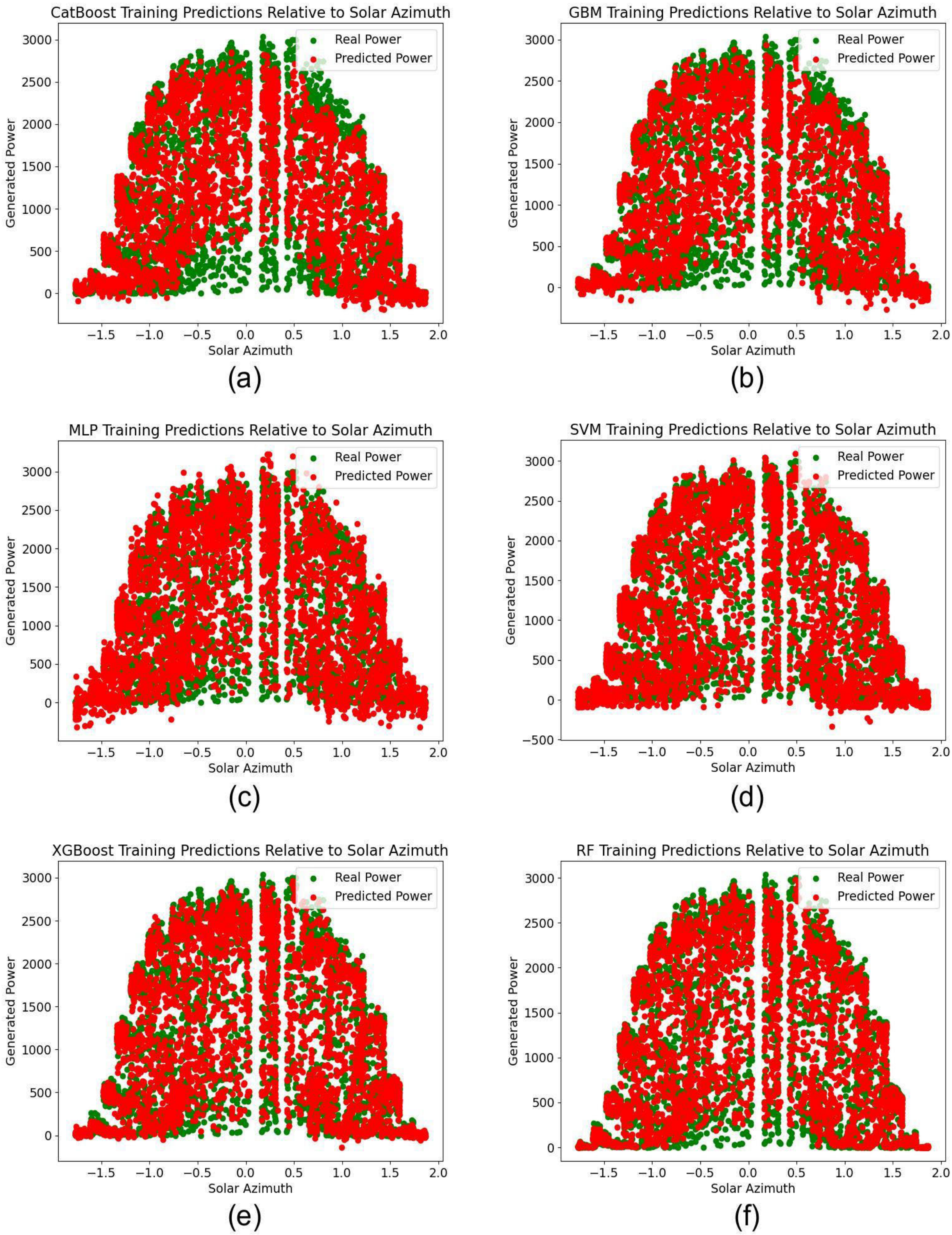

FIGURE 10. Relationship between solar azimuth and power output for training set predictions using different algorithms (a) CatBoost, (b) GBM, (c) MLP regressor, (d) SVM, (e) XGBoost, (f) RF.

plots play a critical role in evaluating the models' ability to extrapolate the discovered correlation between solar azimuth and power output to novel data. A model's performance is considered robust when the projected values and actual

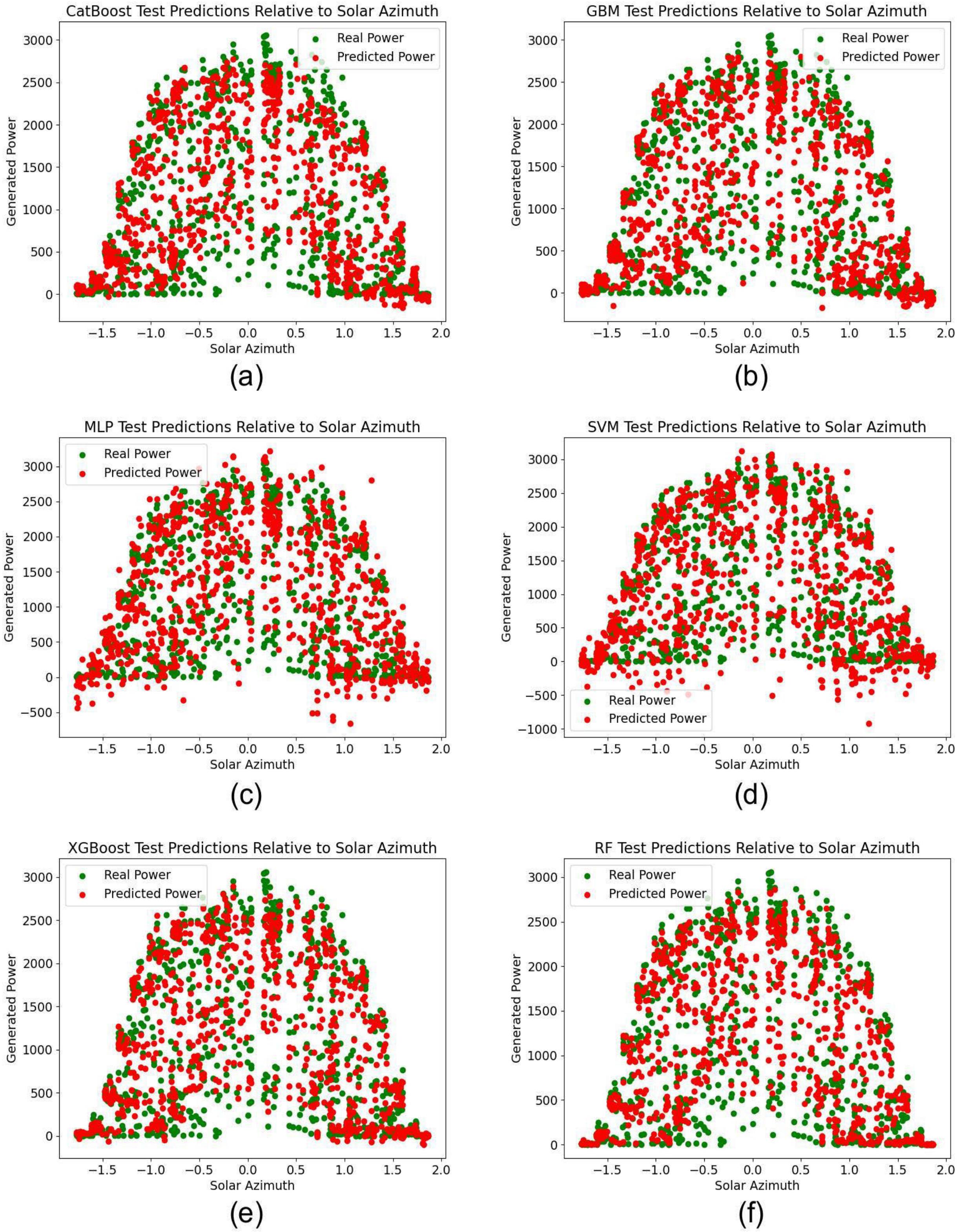

FIGURE 11. Relationship between solar azimuth and power output for test set predictions using different algorithms (a) CatBoost, (b) GBM, (c) MLP regressor, (d) SVM, (e) XGBoost, (f) RF.

solar azimuth trends show constant alignment. Substantial deviations may indicate issues with overfitting or indicate the necessity of including more variables to enhance the model's accuracy across different contexts.

TABLE 4. Descriptive statistics of key variables in the solar power generation dataset.

Model	R^2_{whole}	R^2_{train}	R^2_{test}	Mean of Test Set	SD_{test}	RSD
CatBoost	0.786	0.786	0.786	1131.562	833.260	0.736
GBM	0.829	0.836	0.802	1137.566	854.565	0.751
MLP	0.876	0.905	0.766	1183.541	908.547	0.768
SVM	0.911	0.948	0.768	1160.487	951.396	0.820
XGBoost	0.928	0.956	0.822	1146.255	869.111	0.758
RF	0.940	0.971	0.818	1147.735	854.705	0.745

Figure 7 displays the qualities on the horizontal axis and their associated significance ratings on the vertical axis. The qualities encompass climatic factors such as shortwave radiation, mean sea level pressure, wind speed and direction at different altitudes, precipitation, cloud cover at various levels, temperature, and relative humidity. In addition, the solar power generation process takes into account particular features like azimuth and angle of incidence, which have a significant impact. The research reveals that some parameters, such as shortwave radiation and mean sea level pressure, exhibit significant positive significance, demonstrating their substantial positive impact on the prediction model. On the other hand, variables such as angle of incidence demonstrate a negative significance, indicating a reverse correlation with the model's forecasts. Additional characteristics exhibit different levels of significance, underscoring their varied influence on the accuracy of predictions. This research assists in improving the predictive models for solar power generation by finding and comprehending the most significant aspects. This information is crucial for maximizing model performance, strengthening forecasting precision, and eventually improving the efficiency of solar power systems. The knowledge acquired from this study can provide direction for future research and development endeavors in the realm of renewable energy prediction.

To evaluate the ML algorithms used in this study in detail, we analyzed a large number of variables that provide insightful information about their effectiveness in solar power generation predictions. The metrics include the mean, standard deviation, and RSD of the test set predictions coupled with the R^2 scores for the whole data frame, training set, and test set. These metrics taken together show that the models can correctly forecast new, untested data and reflect the training data appropriately.

Table 4 presents the performance characteristics of the six ML algorithms employed in this study: CatBoost, GBM, Multi-Layer Perceptron MLP regressor, Support Vector Machine (SVM), XGBoost, and RF. The R^2 ratings for the entire data frame, training set, and test set offer valuable insights into the extent to which each model can account for the variability in solar power generation. The RF algorithm earned the maximum R^2 score of 0.940 for the entire data frame and 0.971 for the training set, suggesting a strong fit of the model. On the test set, XGBoost exhibited superior generalization performance, achieving a R^2 score of (0.822). This demonstrates that XGBoost achieved a favorable equilibrium between precisely fitting the training data and making accurate predictions on fresh, unknown data.

MLP regressor had the greatest mean value (1183.541 kW) in terms of the test set predictions, indicating that it consistently projected greater power outputs on average. Contrarily, CatBoost had the smallest standard deviation (833.260 kW) among the models tested, suggesting that its predictions were more uniform and less dispersed in comparison to the other models. The RSD, which normalizes the standard deviation by the mean, measures the level of forecast stability in relation to the average output. The CatBoost model had the lowest RSD (0.736), indicating a high level of stability in its predictions compared to its mean output. On the other hand, the SVM model had the greatest RSD (0.820), suggesting a greater degree of unpredictability in its predictions. Collectively, these measures effectively demonstrate the advantages and disadvantages of each method for forecasting solar power production. RF and XGBoost demonstrated robust fitting and generalization capabilities, respectively, while CatBoost exhibited the highest level of prediction consistency. These insights are critical for selecting the best model for precise and dependable solar power prediction.

V. CONCLUSION

This study assessed the efficacy of six ML algorithms, namely CatBoost, GBM, MLP regressor, SVM, XGBoost, and RF, in forecasting solar power generation. The dataset used for this analysis consisted of 4213 sets. The RF method did the best, with a R^2 score of 0.940 for the whole dataset, 0.971 for the training set, and 0.818 for the test set. This showed how accurate and useful it is in real life. Most algorithms did great on the training set, but not so well on the test set. This suggests that the models might change as they see more data. Comparing XGBoost and RF to models like MLP regressor and SVM, their better balance in performance across the training and test sets suggests that they are less prone to overfit. We found inconsistent test set forecasts based on the relative and standard deviation metrics of the models' findings. CatBoost and RF exhibited greater prediction stability compared to SVM, which displayed the highest degree of fluctuation. The application of Lasso regression identified the salient features. This improved the model and provided more insight into the factors influencing solar power generation. The model's precision is highly dependent on these novel concepts. We highly recommend the RF and XGBoost models for real-world solar power forecasting due to their exceptional accuracy, resilience, and efficacy on both training and test datasets. These findings establish a solid basis for maximizing the incorporation of solar electricity into the energy system. Potential future research can prioritize several feasible routes. An important focus is the creation of hybrid models that integrate the advantages of many methodologies, perhaps resulting in improved accuracy and resilience. Integrating domain-specific information into ML frameworks has the capacity to improve the performance of the models. Expanding the scope of datasets to include regularly updated ones has the potential to enhance the usefulness and reliability of models in different environmental conditions. Additional study might investigate the influence of seasonal fluctuations and meteorological conditions on the production of solar energy, perhaps resulting in more precise and adaptable prediction models. By focusing on these specific areas, future studies can further enhance the field of renewable energy forecasting, thereby promoting the development of more efficient and sustainable energy systems.

VI. FUTURE WORK

To enhance the precision and effectiveness of our solar power prediction models, it is advisable for future research to focus on developing hybrid models that combine the benefits of many ML methods. An effective strategy is to merge the robustness of RF with the strong ability to generalize of XGBoost. Using ensemble methods—that is, aggregating forecasts from several models—is one approach to do this. One can effectively achieve this integration using stacking, mixing, and voting, among other techniques. Stacking is an approach whereby one trains many models and then uses another model to find the best way of aggregating their forecasts, hence improving general accuracy. Blending, a less complex variant of stacking, uses a distinct dataset to train the ultimate model, hence mitigating the likelihood of overfitting. Voting methods aggregate model predictions by either averaging them (for regression tasks) or accepting the majority vote (for classification tasks). This can be accomplished through hard voting, which follows a majority rule, or soft voting, which considers weighted probabilities. Moreover, using sophisticated feature engineering approaches might enhance hybrid models even more. Lasso Regression has already demonstrated its utility in detecting relevant characteristics. Future studies might use methods to improve the feature set, such Recursive Feature Elimination (RFE) or Principal Component Analysis (PCA). Furthermore, looking at feature interactions could help produce more all-encompassing input for the models. Time-series modeling techniques help to improve the accuracy of solar power generation projections by considering the temporal elements influencing them. Better results come from combining normal regression models with techniques designed especially to control temporal dependencies—such as LSTM networks. Hybrid models gain from adaptive learning, a technique whereby a model changes constantly as fresh data becomes accessible. Combining online learning algorithms—which continuously learn from fresh data points—with conventional batch learning techniques helps one create a more agile and current model. Comprehensive evaluation systems are needed if we are to fairly gauge the performance of these hybrid models. By lowering overfitting and guaranteeing the generalizability of the model, cross-valuation methods include k-fold cross-valuation help to improve the evaluation process. Still another inspiring direction is looking at unique hybrid techniques. LSTMs for temporal dynamics and CNNs for spatial data extraction could be used, respectively. Moreover, using evolutionary algorithms and other optimization techniques helps one find the weighting of several models and the most suitable combination. Future studies employing these approaches can provide hybrid models that surpass individual algorithms in terms of expected accuracy, therefore producing more accurate and reliable solar power estimates. Maximizing the inclusion of renewable energy sources into power systems depends on these innovations, which also help to increase their efficiency and sustainability.

REFERENCES

[1] A. Shahsavari and M. Akbari, "Potential of solar energy in developing countries for reducing energy-related emissions," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 275–291, Jul. 2018, doi: 10.1016/j.rser.2018.03.065.

[2] M. Izanloo, Y. Noorollahi, and A. Aslani, "Future energy planning to maximize renewable energy share for the south Caspian sea climate," *Renew. Energy*, vol. 175, pp. 660–675, Sep. 2021, doi: 10.1016/j.renene.2021.05.008.

[3] A. Androniceanu and O. M. Sabie, "Overview of green energy as a real strategic option for sustainable development," *Energies*, vol. 15, no. 22, p. 8573, Nov. 2022, doi: 10.3390/en15228573.

[4] B. E. Logan, R. Rossi, G. Baek, L. Shi, J. O'Connor, and W. Peng, "Energy use for electricity generation requires an assessment more directly relevant to climate change," *ACS Energy Lett.*, vol. 5, no. 11, pp. 3514–3517, Nov. 2020, doi: 10.1021/acsenergylett.0c02093.

[5] M. J. B. Kabeyi and O. A. Olanrewaju, "Sustainable energy transition for renewable and low carbon grid electricity generation and supply," *Frontiers Energy Res.*, vol. 9, Mar. 2022, Art. no. 743114, doi: 10.3389/fenrg.2021.743114.

[6] H. Sharadga, S. Hajimirza, and R. S. Balog, "Time series forecasting of solar power generation for large-scale photovoltaic plants," *Renew. Energy*, vol. 150, pp. 797–807, May 2020, doi: 10.1016/j.renene.2019.12.131.

[7] A. R. Pazikadin, D. Rifai, K. Ali, M. Z. Malik, A. N. Abdalla, and M. A. Faraj, "Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ANN): A review of five years research trend," *Sci. Total Environ.*, vol. 715, May 2020, Art. no. 136848, doi: 10.1016/j.scitotenv.2020.136848.

[8] S. Ghosh and P. K. Gupta, "Forecasting of solar power volatility using GJR-GARCH method," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2021, pp. 261–266, doi: 10.1109/EPEC52095.2021.9621364.

[9] G. Moreno, C. Santos, P. Martín, F. J. Rodríguez, R. Peña, and B. Vuksanovic, "Intra-day solar power forecasting strategy for managing virtual power plants," *Sensors*, vol. 21, no. 16, p. 5648, Aug. 2021, doi: 10.3390/s21165648.

[10] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871–175880, 2020, doi: 10.1109/ACCESS.2020.3025860.

[11] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, "Solar irradiance resource and forecasting: A comprehensive review," *IET Renew. Power Gener.*, vol. 14, no. 10, pp. 1641–1656, Jul. 2020, doi: 10.1049/iet-rpg.2019.1227.

[12] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, "A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework," *Energy Convers. Manage.*, vol. 212, May 2020, Art. no. 112766, doi: 10.1016/j.enconman.2020.112766.

[13] M. Jasiński, Z. Leonowicz, J. Jasinski, L. Martirano, and R. Gono, "PV advancements & challenges: Forecasting techniques, real applications, and grid integration for a sustainable energy future," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur.*, Jun. 2023, pp. 1–5, doi: 10.1109/eeeic/icpseurope57605.2023.10194796.

[14] M. Karunanithi, A. A. S. Braitea, A. A. Rizvi, and T. A. Khan, "Forecasting solar irradiance using machine learning methods," in *Proc. IEEE 64th Int. Scientific Conf. Inf. Technol. Manage. Sci. Riga Tech. Univ. (ITMS)*, Oct. 2023, pp. 1–4, doi: 10.1109/itms59786.2023.10317687.

[15] N. C. Nath, W. Sae-Tang, and C. Pirak, "Machine learning-based solar power energy forecasting," *J. Soc. Automot. Engineers Malaysia*, vol. 4, no. 3, pp. 307–322, Sep. 2020, doi: 10.56381/jsaem.v4i3.25.

[16] A. M. Alam, I. A. Razee, and M. Zunaed, "Solar PV power forecasting using traditional methods and machine learning techniques," in *Proc. IEEE Kansas Power Energy Conf. (KPEC)*, Apr. 2021, pp. 1–5, doi: 10.1109/KPEC51835.2021.9446199.

[17] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, Jun. 2021, Art. no. 120240, doi: 10.1016/j.energy.2021.120240.

[18] Y. Essam, A. N. Ahmed, R. Ramli, K.-W. Chau, M. S. I. Ibrahim, M. Sherif, A. Sefelnasr, and A. El-Shafie, "Investigating photovoltaic solar power output forecasting using machine learning algorithms," *Eng. Appl. Comput. Fluid Mech.*, vol. 16, no. 1, pp. 2002–2034, Dec. 2022, doi: 10.1080/19942060.2022.2126528.

[19] U. Munawar and Z. Wang, "A framework of using machine learning approaches for short-term solar power forecasting," *J. Electr. Eng. Technol.*, vol. 15, no. 2, pp. 561–569, Mar. 2020, doi: 10.1007/s42835-020-00346-4.

[20] D. V. Pombo, H. W. Bindner, S. V. Spataru, P. E. Sørensen, and P. Bacher, "Increasing the accuracy of hourly multi-output solar power forecast with physics-informed machine learning," *Sensors*, vol. 22, no. 3, p. 749, Jan. 2022, doi: 10.3390/s22030749.

[21] H. Sheng, B. Ray, K. Chen, and Y. Cheng, "Solar power forecasting based on domain adaptive learning," *IEEE Access*, vol. 8, pp. 198580–198590, 2020, doi: 10.1109/ACCESS.2020.3034100.

[22] C. Vennila, A. Titus, T. S. Sudha, U. Sreenivasulu, N. P. R. Reddy, K. Jamal, D. Lakshmaiah, P. Jagadeesh, and A. Belay, "Forecasting solar energy production using machine learning," *Int. J. Photoenergy*, vol. 2022, pp. 1–7, Apr. 2022, doi: 10.1155/2022/7797488.

[23] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," *Appl. Comput. Informat.*, vol. 20, no. 3, pp. 231–250, Jun. 2024, doi: 10.1016/j.aci.2019.11.002.

[24] R. D. Kumar, K. Prakash, P. A. Sundari, and S. Sathya, "A hybrid machine learning model for solar power forecasting," in *Proc. E3S Web Conf.*, 2023, p. 04003, doi: 10.1051/e3sconf/202338704003.

[25] A. Nikas, V. Stavrakas, A. Arsenopoulos, H. Doukas, M. Antosiewicz, J. Witajewski-Baltvilks, and A. Flamos, "Barriers to and consequences of a solar-based energy transition in Greece," *Environ. Innov. Societal Transitions*, vol. 35, pp. 383–399, Jun. 2020, doi: 10.1016/j.eist.2018.12.004.

[26] O. Elalani, H. Ghennioui, and A. Ghennioui, "Intra-day variability quantification from ground-based measurements of global solar irradiance," *Int. J. Renew. Energy Res.*, vol. 10, no. 4, pp. 1576–1587, 2020, doi: 10.20508/ijrer.v10i4.11308.g8042.

[27] L. Meegahapola, A. Sguarezi, J. S. Bryant, M. Gu, E. R. Conde, and R. B. A. Cunha, "Power system stability with power-electronic converter interfaced renewable power generation: Present issues and future trends," *Energies*, vol. 13, no. 13, p. 3441, Jul. 2020, doi: 10.3390/en13133441.

[28] D. Korkmaz, "SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting," *Appl. Energy*, vol. 300, Oct. 2021, Art. no. 117410, doi: 10.1016/j.apenergy.2021.117410.

[29] Y.-J. Zhong and Y.-K. Wu, "Short-term solar power forecasts considering various weather variables," in *Proc. Int. Symp. Comput., Consum. Control (ISC)*, Nov. 2020, pp. 432–435, doi: 10.1109/IS3C50286.2020.00117.

[30] J. I. Kwak, S.-H. Nam, L. Kim, and Y.-J. An, "Potential environmental risk of solar cells: Current knowledge and future challenges," *J. Hazardous Mater.*, vol. 392, Jun. 2020, Art. no. 122297, doi: 10.1016/j.jhazmat.2020.122297.

[31] N. Novas, R. M. Garcia, J. M. Camacho, and A. Alcayde, "Advances in solar energy towards efficient and sustainable energy," *Sustainability*, vol. 13, no. 11, p. 6295, Jun. 2021, doi: 10.3390/su13116295.

[32] S. Srivastava, A. Behera, and R. Biswal, "Assessing the grip of solar energy systems on environmental sustainability—A review," *Micro Nanosystems*, vol. 14, no. 2, pp. 133–143, Jun. 2022, doi: 10.2174/1876402913666210908122052.

[33] S. Gulati and R. Jain, "Solar cells for ecological sustainable development: A review," *J. Adv. Zoology*, vol. 44, no. 6, pp. 1109–1121, Nov. 2023, doi: 10.17762/jaz.v44is6.2366.

[34] T. Tran, H. Ho, and T. Do, "Assessment the impact of solar panels on the environment," *Tra Vinh Univ. J. Sci.*, vol. 13, no. 6, pp. 1–9, 2023, doi: 10.35382/tvujs.13.6.2023.2125.

[35] A. A. Jan, F.-W. Lai, M. K. Shad, S. E. A. Ali, and S. Hamad, "Solar photovoltaic technology and its impact on environmental, social and governance (ESG) performance: A review," *KnE Social Sci.*, vol. 8, no. 20, pp. 759–774, Dec. 2023.

[36] T.-J. Ren and J.-Y. Wu, "Power forecasting of photovoltaic system using environment data and weather information," in *Proc. IET Int. Conf. Eng. Technol. Appl. (IET-ICETA)*, Oct. 2022, pp. 1–2, doi: 10.1109/IET-ICETA56553.2022.9971686.

[37] J.-H. Syu, M.-E. Wu, G. Srivastava, C.-F. Chao, and J. C. Lin, "An IoT-based Hedge system for solar power generation," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10347–10355, Jul. 2021, doi: 10.1109/JIOT.2021.3064384.

[38] L. F. J. Alvarez, S. R. González, A. D. López, D. A. H. Delgado, R. Espinosa, and S. Gutiérrez, "Renewable energy prediction through machine learning algorithms," in *Proc. IEEE ANDESCON*, Oct. 2020, pp. 1–6, doi: 10.1109/ANDESCON50619.2020.9272029.

[39] S. P. S. Ibrahim, I. Rakshitha, T. Vasisri, R. H. Aswitha, M. R. Rao, and D. V. Krishna, "Revolutionizing solar generation data mining through advanced machine learning algorithms: Novel insights and results," in *Proc. 7th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solutions (CSITSS)*, Nov. 2023, pp. 1–8, doi: 10.1109/csitss60515.2023.10334112.

[40] E. Subramanian, M. M. Karthik, G. P. Krishna, D. V. Prasath, and V. S. Kumar, "Solar power prediction using machine learning," 2023, *arXiv:2303.07875*.

[41] S. Theocharides, G. Tziolis, J. Lopez-Lorente, G. Makrides, and G. E. Georghiou, "Impact of data quality on day-ahead photovoltaic power production forecasting," in *Proc. IEEE 48th Photovoltaic Specialists Conf. (PVSC)*, Jun. 2021, pp. 0918–0922, doi: 10.1109/PVSC43889.2021.9518471.

[42] Q.-T. Phan, Y.-K. Wu, Q.-D. Phan, and H.-Y. Lo, "A novel forecasting model for solar power generation by a deep learning framework with data preprocessing and postprocessing," *IEEE Trans. Ind. Appl.*, vol. 59, no. 1, pp. 220–231, Jan. 2023, doi: 10.1109/TIA.2022.3212999.

[43] V. Raj, S.-Q. Dotse, M. Sathyajith, M. I. Petra, and H. Yassin, "Ensemble machine learning for predicting the power output from different solar photovoltaic systems," *Energies*, vol. 16, no. 2, p. 671, Jan. 2023, doi: 10.3390/en16020671.

[44] K. P. Natarajan and J. G. Singh, "Solar power forecasting using stacking ensemble models with Bayesian meta-learning," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2023, pp. 1–6, doi: 10.1109/icccnt56998.2023.10308202.

[45] R. P. Ponraj, V. Ravindran, K. Chittibabu, T. Akash, A. Chandru, and E. Louis, "Machine learning based solar PV power prediction," in *Proc. 2nd Int. Conf. Augmented Intell. Sustain. Syst. (ICAISS)*, Aug. 2023, pp. 1646–1652, doi: 10.1109/icaiss58487.2023.10250567.

[46] S. M. Babbar and L. C. Yong, "Solar power prediction using machine learning algorithms: A comparative study," in *Proc. 3rd Int. Conf. Smart Electron. Commun. (ICOSEC)*, Oct. 2022, pp. 1313–1319, doi: 10.1109/ICOSEC54921.2022.9951972.

[47] M. Mohamed, F. E. Mahmood, M. A. Abd, A. Chandra, and B. Singh, "Dynamic forecasting of solar energy microgrid systems using feature engineering," *IEEE Trans. Ind. Appl.*, vol. 58, no. 6, pp. 7857–7869, Nov. 2022, doi: 10.1109/TIA.2022.3199182.

[48] F. Meng, Q. Zou, Z. Zhang, B. Wang, H. Ma, H. M. Abdullah, A. Almalaq, and M. A. Mohamed, "An intelligent hybrid wavelet-adversarial deep model for accurate prediction of solar power generation," *Energy Rep.*, vol. 7, pp. 2155–2164, Nov. 2021, doi: 10.1016/j.egyr.2021.04.019.

[49] G. de Freitas Viscondi and S. N. Alves-Souza, "Solar irradiance prediction with machine learning algorithms: A Brazilian case study on photovoltaic electricity generation," *Energies*, vol. 14, no. 18, p. 5657, Sep. 2021, doi: 10.3390/en14185657.

[50] C. Onyutha, "A hydrological model skill score and revised R-squared," *Hydrol. Res.*, vol. 53, no. 1, pp. 51–64, Jan. 2022, doi: 10.2166/nh.2021.071.

[51] M. M. Momin, S. Lee, N. R. Wray, and S. H. Lee, "Significance tests for R2 of out-of-sample prediction using polygenic scores," *Amer. J. Human Genet.*, vol. 110, no. 2, pp. 349–358, Feb. 2023, doi: 10.1016/j.ajhg.2023.01.004.

[52] S. Weirich, M. Hecht, B. Becker, and S. Zitzmann, "Comparing group means with the total mean in random samples, surveys, and large-scale assessments: A tutorial and software illustration," *Behav. Res. Methods*, vol. 54, no. 3, pp. 1051–1062, Sep. 2021, doi: 10.3758/s13428-021-01553-1.

[53] D. Tasche, "Proving prediction prudence," *Data Sci. Finance Econ.*, vol. 2, no. 4, pp. 359–379, Sep. 2022, doi: 10.2139/ssrn.3595625. [Online]. Available: https://ssrn.com/abstract=3595625

[54] C. Andrade, "Understanding the difference between standard deviation and standard error of the mean, and knowing when to use which," *Indian J. Psychol. Med.*, vol. 42, no. 4, pp. 409–410, Jul. 2020, doi: 10.1177/0253717620933419.

[55] S. Ghosh, D. N. Ayyala, and R. Hellebuyck, "Two-sample high dimensional mean test based on prepivots," *Comput. Statist. Data Anal.*, vol. 163, Nov. 2021, Art. no. 107284, doi: 10.1016/j.csda.2021.107284.

MONTASER ABDELSATTAR was born in Abu Tisht, Qena, Egypt, in August 1983. He received the B.Sc. degree from the Electrical Engineering Department, Faculty of Engineering, Al-Azhar University, Egypt, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from the Faculty of Engineering, Minia University, Egypt, in 2011 and 2015, respectively. He was a Lecturer with the Department of Electrical and Computer Engineering, El-Minia High Institute of Engineering and Technology, El-Minia, Egypt, from October 2016 to October 2020. He was also a Lecturer with the Electrical Engineering Department, Faculty of Engineering, South Valley University, Qena, from November 2020 to September 2021, where he has been an Associate Professor and the Head of the Electrical Engineering Department, Faculty of Engineering, since October 2021. He has published over 60 technical papers in international journals and conferences and has supervised and examined more than ten M.Sc. and Ph.D. theses with South Valley University, Minia University, Sohag University, Aswan University, and other Egyptian universities. He has published three international books in the field of renewable energy. His current research interests include renewable energy sources, energy management, artificial intelligence, machine learning techniques, smart grids, micro-grid, power quality, power systems, control, electric vehicles, and optimization algorithms.

MOHAMED A. ISMEIL (Senior Member, IEEE) was born in Qena, Egypt, in October 1977. He received the B.Sc. and M.Sc. degrees in electrical engineering from South Valley University, in 2002 and 2008, respectively, and the joint Ph.D. degree from Aswan University and Technical University of Munich (TUM), Germany, in April 2014. From October 2010 to January 2013, he was a Ph.D. Student with the Department of Electrical Drive Systems and Power Electronics, TUM, as a Channel System Program. From April 2014 to September 2018, he was an Assistant Professor with the Aswan Faculty of Engineering, Aswan University. Since October 2018, he has been an Associate Professor with Qena Faculty of Engineering, South Valley University. From March 2020 to November 2022, he was the Head of the Electrical Department, Qena Faculty of Engineering. Since November 2002, he has been an Associate Professor with the College of Engineering, King Khalid University, Saudi Arabia. He has published more than 78 papers in international conferences and journals. His current research interests include power electronics applications in wind energy conversion systems, PV interface with the utility, smart grid technologies, and digital control applications (PIC, FPGA, and DSP). His main research interest includes power inverter design for renewable applications.

MOHAMED M. A. AZIM ZAYED (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Cairo University, Egypt, in 1994, the M.Sc. degree jointly from the Fontys University and Technical University Eindhoven, The Netherlands, in 1997, and the Ph.D. degree in computer sciences from Tohoku University, Japan, in 2006. He is a Professor of electronics technology with Beni-Suef University, Egypt. In 2008, he joined the College of Computer Science and Engineering, Taibah University, KSA. In 2012, he became the Chair of Networks and Communications Systems Department. In 2020, he joined the University of Prince Mugrin. He is the author of several research papers published in the most reputable journals and conferences. He is the author of the book titled *Optical Networks: A Restoration Perspective with Active Restoration*. He is also an Editor of the book titled *Wireless Sensor Multimedia Networks: Architectures, Protocols, and Applications* (CRC Press, USA). In 2021, he received the Distinguished Researcher Award at UPM. He is a member of the editorial boards of the *International Journal of Sensor and Related Networks* and *International Journal of Communication Networks and Information Security*. He is also the Chair of the Membership Development Committee of IEEE in Western KSA of R8.

AHMED ABDELMOETY was born in Amman, Jordan, in February 1999. He received the Bachelor of Science degree in electrical power and machinery engineering from South Valley University, Qena, Egypt, in 2022, where he is currently pursuing the Master of Science degree in electrical engineering. His research interests include renewable energy sources, AI, machine learning, and deep learning, aiming to advance smart energy solutions.

AHMED EMAD-ELDEEN was born in Egypt, in September 1982. He received the bachelor's and master's degrees in electronics technology and the Ph.D. degree in renewable energy science and engineering from Beni-Suef University, Beni-Suef, Egypt, in 2007, 2014, and 2020, respectively. He has specialized in the field of renewable energy science and engineering. Beginning his career with Beni-Suef University, in 2009, he initially was a Demonstrator in electronics technology. His commitment and passion for teaching and research led to his subsequent roles as an Assistant Lecturer and then as a Lecturer in renewable energy science and engineering. Currently, he holds the position of a Lecturer with the Faculty of Postgraduate Studies for Advanced Sciences, Beni-Suef University. His work focuses on renewable energy education and pioneering research in the field, contributing to the advancement of sustainable and eco-friendly energy solutions.