

Article

Enhanced Random Forest Model for Robust Short-Term Photovoltaic Power Forecasting Using Weather Measurements

Mohamed Massaoudi ^{1,2,*} , Ines Chihi ^{3,4,5}, Lilia Sidhom ^{4,5}, Mohamed Trabelsi ⁶, Shady S. Refaat ¹ 
and Fakhreddine S. Oueslati ²

- ¹ Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha 3263, Qatar; Shady.khalil@qatar.tamu.edu
- ² Laboratoire Matériaux Molécules et Applications (LMMA) à l'IPEST, Carthage University, Tunis 1054, Tunisia; fakhreddine.oueslati@enicarthage.rnu.tn
- ³ Département Ingénierie, Faculté des Sciences, des Technologies et de Médecine, Campus Kirchberg, Université du Luxembourg, 1359 Luxembourg, Luxembourg; ines.chihi@uni.lu
- ⁴ Laboratory of Energy Applications and Renewable Energy Efficiency (LAPER), El Manar University, Tunis 1068, Tunisia; Lilia.Sidhom@enib.rnu.tn
- ⁵ National Engineering School of Bizerta, Carthage University, Tunis 7080, Tunisia
- ⁶ Department of Electronic and Communications Engineering, Kuwait College of Science and Technology, Doha District, Block 4, Doha P.O. Box 27235, Kuwait; m.trabelsi@kcst.edu.kw
- * Correspondence: mohamed.massaoudi@qatar.tamu.edu; Tel.: +974-70-062-097

Abstract: Short-term Photovoltaic (PV) Power Forecasting (STPF) is considered a topic of utmost importance in smart grids. The deployment of STPF techniques provides fast dispatching in the case of sudden variations due to stochastic weather conditions. This paper presents an efficient data-driven method based on enhanced Random Forest (RF) model. The proposed method employs an ensemble of attribute selection techniques to manage bias/variance optimization for STPF application and enhance the forecasting quality results. The overall architecture strategy gathers the relevant information to constitute a voted feature-weighting vector of weather inputs. The main emphasis in this paper is laid on the knowledge expertise obtained from weather measurements. The feature selection techniques are based on local Interpretable Model-Agnostic Explanations, Extreme Boosting Model, and Elastic Net. A comparative performance investigation using an actual database, collected from the weather sensors, demonstrates the superiority of the proposed technique versus several data-driven machine learning models when applied to a typical distributed PV system.

Keywords: smart grid; Photovoltaic (PV) Power Forecasting; weather sensors; random decision forest; feature importance; energy management



Citation: Massaoudi, M.; Chihi, I.; Sidhom, L.; Trabelsi, M.; Refaat, S.S.; Oueslati, F.S. Enhanced Random Forest Model for Robust Short-Term Photovoltaic Power Forecasting Using Weather Measurements. *Energies* **2021**, *14*, 3992. <https://doi.org/10.3390/en14133992>

Academic Editor: Alon Kuperman

Received: 27 May 2021

Accepted: 28 June 2021

Published: 2 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the years, the exponential increase in global energy demand has become the leading cause of the rapid depletion of fossil fuels and increased Greenhouse Gas (GHG) emissions of conventional generators [1]. To effectively satisfy the meteoric growth in energy consumption, the world has taken serious initiatives to deploy RES on a larger scale. [2]. Solar Energy (SE) hold out the greatest promise for modern humankind among all RES, being free, clean, and abundantly available [3]. For these reasons, it keeps increasing its share in the energy-mix in the face of diminishing conventional fossil fuel energy sources and rising environmental protection concerns [3]. However, the discontinuity of PV power flow brings into question the reliability of the high penetration of PV systems, which affect the dispatch accuracy greatly. Moreover, the negative effects of the sudden weather change on the PV farms threatens the grid stability and rises the cumbersome costs of the allocation of the spinning reserve [3]. Therefore, PV Power Forecasting (PPF) is a pivotal element for reliable power supply as it significantly reduces the sensitivity of energy systems to weather intermittency. PPF is mandatory for PV generators as it has a direct impact on the

stability and reliability of the grid. Achieving accurate forecasting for PV power generation will facilitate the SE integration to the power system.

In this context, the research community has been focusing on the development of effective forecasting techniques to handle pattern dependencies [4,5]. With computer hardware and software development, forecasting models take advantage of High-Performance Computing (HPC) to achieve higher effectiveness. The energy forecasting methods provided by the PV generators can be generally classified into two categories: traditional methods and Artificial Neural Networks (ANN)-based methods [6]. Traditional methods mostly include statistical methods. Statistical methods include regression techniques, Exponential Smoothing (ES) [7], Autoregressive (AR), Moving Average (MA), and their generalizations such as Autoregressive Integrated Moving Average with exogenous inputs (ARIMAX) methods also known as Box-Jenkins models [8,9]. These models include a few model parameters leading to higher simplicity and interpretability. In [10], a PVPF method-based Autoregressive Integrated Moving Average (ARIMA) has been adopted for the design of an energy management system. Paper [7] exploits ES State Space (ESSS) for short-term solar irradiance model. Nevertheless, the direct PPF is not considered in this study [7]. In summary, the traditional former approaches do not make use of the historical data generated by weather stations leading to poor forecastability potential. Reciprocally, ANN becomes one of the most commonly used approaches for PVPF [11]. It has been reported in [12] that ANNs are easy to use for RES designs, especially for solar irradiance with related PV power [13].

Pioneering work is presented in [14], where it is shown that ANN can generate deterministic and probabilistic PV power for three days ahead. An Analog Ensemble (AnEN) model has boosted the ANN accuracy using computed astronomical variables and past predictions of a deterministic Numerical Weather Prediction (NWP) model. However, the computational requirements for the model training are cumbersome. The finding is consistent with the results of recent studies by [15], which employs Multi-Layer Perceptron (MLP)-based ANN. These contributions are completed in [16], where a comparison of different models is performed. Although the MLP method enhances the prediction performance, the uncertainty resulting from the assumptions of the pre-processed features could form a barrier to its practical implementation [16]. In [17], a radial basis function (RBF)-based ANN has been proposed for an online PVPF. Such observations are confirmed in [18] where Feed-Forward Neural Networks (FFNNs) and RBF, two variants of ANN, are also used for solar PV power production predictions. The proposed model provides a Root Mean Square Error RMSE = 10.59% in a typical Autumn day. Ultimately, in [19], an optimized ANN for one day ahead PVPF has been presented. The proposed model makes use of dust and temperature to follow a PV plant, yielding a coefficient of determination $R^2 = 91.4\%$.

To sum up, the ANN-based models are perfectly tailored for PVPF. The concern, however, has been raised that the above methods lack an end-to-end process for selecting essential features among the provided inputs, leading to a tedious manual preprocessing phase. Furthermore, these models sacrifice interpretability over high forecasting accuracy. In other words, the recently cited forecasting methods failed to provide an empirical ability to develop insights on how they use different inputs, resuming in burdensome problem for the industrial acceptability of these methods.

However, it is evident that the input parameters do not have an equal contribution to the domain knowledge. Therefore, several methods were proposed to benefit from this inequality. In earlier work, the authors employed an enhanced RF for classification purposes [20]. Their proposed approach lies in the integration of a slowness index with a feature ranking and selection process. According to the simulation results, a classification technique of static and dynamic nodes was adopted to mitigate the overlap between classes. Next, the Relative Mean Decrease Gini (RMDG) method is employed to determine the significance of the feature inputs to the domain knowledge and rank them accordingly. Although the proposed method outperforms a variety of prediction techniques, the compu-

tational effort is high. This burden is due to the large number of possibilities investigated, especially with a larger dimensional space.

This paper deploys a state-of-the-art method to improve the performance potential of RF model. The system employs three techniques to rank the weather and power inputs through a novel feature selection procedure. This ranking leads to a Weighted Feature Vector Importance (FVI). The forecasting model relies on FVI coefficient to generate multiple forecast outputs for each feature input. The final output result is concluded from a summation of the single forecasts. The main contributions are outlined as follows:

1. An effective feature engineering technique is deployed based on six input parameters using three different approaches. The features are classified according to their weights. FVI is calculated using the input parameters relevant to the PPF model.
2. A new approach-based-multimodal prediction system is comprehensively investigated.
3. The performance superiority of the proposed approach versus Decision Trees (DT), K-Nearest Neighbors (KNN), and Random decision forest (RF) is demonstrated using a real data set.

The remainder of the paper is organized as follows: Section 2 comprehensively defines the problem statement and the main contributions. Then, Section 3 presents the related works for PV power forecasting and the common taxonomies and methodologies. Section 4 introduces the proposed methodology and investigates the FVI formulation. Afterward, Section 5 illustrates the implementation results, interpretations, and provides a comprehensive comparison with the state-of-the-art models. Finally, Section 6 discusses the presented results and concludes the study.

2. Literature Review

In [21,22], the PPF was classified into four distinguished classes according to the forecasting horizon, the forecasting method, and forecasting output, as shown in Figure 1.

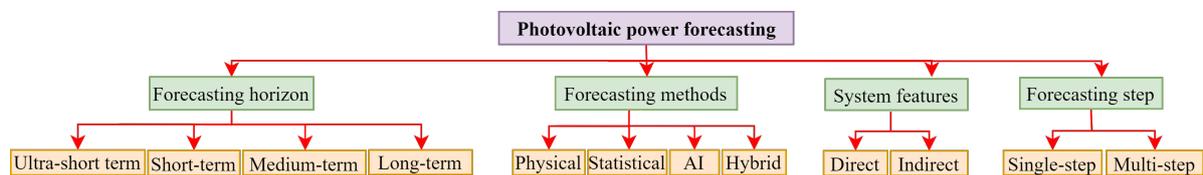


Figure 1. General taxonomy of PV power forecasting.

Conceptually, the determination of a specific parameter variation essentially lies in Physical Models (PMs), statistical techniques, and Artificial intelligence (AI) techniques [23]. The PMs consist of real-world natural conversion formulas to conduct a deterministic closed-form solution for future behavior. PMs are commonly deployed with low-complexity systems and target short-term dependencies [23]. On the other hand, statistical forecasting is carried out through extensive numerical patterns analysis based on statistical theory. Statistical algorithms require a dataset acquisition to build their domain knowledge since they neglect the investigated physical process [24]. Moreover, statistical and physical models are found not great enough to be effective with unsatisfactory accuracy in numerous complex problems such as renewable energy forecasting and weather forecasting. AI techniques have been achieving worldwide acceptance for their accurate results and excellent generalization capabilities [25]. Although AI is very promising for power systems due to the abundance of computational resources and high-resolution databases, ML techniques have only been accorded to a few considerations compared to statistical and physical methods. For PV power forecasting, ML and statistical techniques are greatly influenced by the horizon time series prediction [26].

According to the time domain, there are four distinguished forecasting horizons, specifically ultra-short (USTF), short-term forecasting (STF) required to be valid for seconds to one day, medium-term forecasting for one day to weeks, and Long-Term Forecasting (LTF) that may be valid for years [27]. For example, the USTF forecasting was comprehen-

sively investigated by the authors in [28]. The authors' work consists of the implementation of the underlying Local Sensitive Hash algorithm (LSH). The used taxonomy takes into account four weather conditions, specifically clear, cloudy, rainy, and snowy weather. LSH profoundly investigates the coupling correlated weather features. The methodology adopted for LSH system classifies the PV power segments and generates a PPF output. In [28], the authors proposed a hybrid method for an accurate hourly PV power prediction based on a gradient-descent backpropagation method (BP), Schema Frog Leaping Algorithm (SFLA), and Artificial Neural Network (ANN) named BP-SFLA-ANN model. Their proposed BP-SFLA-ANN model consists of using SFLA as a mediator between BP and ANN models. BP model provides the values of the primary hyperparameters of ANN to let the SFLA start from this initial selection to further search for more suitable parameters of a typical ANN. The interaction between SFLA and the BP led to a superior ANN accuracy and less computational burden compared to an SFLA-ANN without the initial tuning of BP.

So far, the forecasting methodologies can be classified as physical methods, statistical methods [29–31], AI methods or a mix of them (hybrid models) [32,33]. The physical models use NWP models or satellite imagery alongside physical considerations such as meteorological or topological data. However, physical models are restricted to tedious mathematical approaches for specific PV plants, leading to poor generalization potential and complicated modeling [32,34]. Statistical models employ prediction models such as Moving Average (MA) and Autoregressive (AR) [35]. AI methods employ computational intelligence to predict the PV output accurately, taking advantage of the evolved enhancement in hardware and software [36]. The optimal models are often a combination of physical and statistical models [37]. According to the literature, it has been found that the combination of different forecasting models could enhance the performance and efficiency of the overall prediction paradigm [38,39].

Moreover, PV generation forecast methodologies can be taxonomized, taking into account the relationship between inputs and estimated outputs: direct and indirect. The indirect PV power forecasting approach is the estimation of a key relevant element, such as the irradiance and the temperature, leading to an accurate PV power prediction [40], while the direct forecasting approach only considers the PV power as the output to be predicted from weather conditions. Some of the PV power forecasting methods are provided in Table 1.

Table 1. A summary of the literature studies.

Methods	Ref.	Class	Error Metrics	Lowest Error	Time Step	Data Set Location
PEEC	[41]	Physical	NMAE, WMAE	NMAE = 0.5%	1-h	Politecnico di Milano
PMCV	[42]	Physical	RMSE, MAE, MBE	NMAE = 13%	24-h/48-h	Hungaria
LSSVR–NARX	[43]	Statistical	MAE, MBE, MSE, RMSE, R^2	$R^2 = 92.03\%$	2-h	Casablanca, Morocco
ARMAX	[8]	Statistical	RMSE, MAD, and MAPE	MAPE = 38.88%	24-h	Coloane island of Macau
ARMA	[44]	Statistical	MAE, MRE	MAE = 1.16 MW	15 min	IEEE14 bus system
SVM	[45]	AI	MRE, RMSE	RMSE = 1.57 MW	24-h	PV station in China
CNN-LSTM	[31]	AI	MAE, RMSE, R^2	$R^2 = 99.93\%$	15/45 min	Limberg, Belgium
RNN	[46]	AI	R^2	$R^2 = 99.94\%$	15–90 min	Flanders, Belgium
ANN	[47]	AI	RMSE	RMSE = 0.07 KW	24-h	Amman, Jordan

In particular, an ANN-based Statistical Feature Parameters (ANN-SFP) has been implemented for solar irradiance prediction [48]. The proposed model provides a 24-h weather forecast on the hourly level for all the daylight hours of the next day. However, the proposed model is incapable of following the PV generation on an overcast and cloudy days. A three-stage prediction approach named optimized multi-layer backpropagation neural network has demonstrated better system performance than the state-of-art for ultra-short-term PPF [49]. This approach relies on the seasonal division of weather data

set to guarantee the adequate repartition of sample features for different stages. Nevertheless, the splitting process for the meteorological database can threaten the inherent consistency of the overall data set. Additionally, forecasting PV power generation located in the north of Italy for the next day could be conducted using a Physical Hybrid ANN (PHANN) [50]. By fusing the ANN with the physical model of the clear sky solar irradiance method, the proposed model improved prediction accuracy in some of the selected days but unable to provide stable and improved forecast results in peculiar weather conditions. Seasonal Autoregressive Integrated Moving Average (SARIMA)-Random Vector Functional Link (RVFL) model is employed for a short-term PPF [51]. A maximum Overlap Discrete Wavelet Transform technique is implemented to assist a hybrid model for better generalization potential. The produced power profile of a Silicon-crystalline PV module yielded an $R^2 = 92.4\%$ for a single-step-ahead prediction. However, the accuracy drops as the prediction time window become wider (Three steps ahead). In [22], an ensemble of ANN has been proposed to conduct short-term solar forecasting using day-ahead weather forecasts. Despite outperforming several benchmarks, the proposed model cannot capture the fast variation of the weather conditions. A high-precision Convolutional Neural Network (CNN)-based PPF named PVPNet is depicted to predict the PV generation for one day ahead [52]. The proposed deep learning model has been found highly sensible to representation learning and the quality of data [25]. In [53], an LSTM-based attention mechanism has been proposed for STPF. Nevertheless, the prediction system is limited to a single-step forecasting strategy. A Random Forest solar power forecast based on classification optimization was presented and analyzed for PPF [54]. Despite the high system complexity, the proposed model makes it possible to forecast solar irradiance on a 24-h basis with high precision.

3. Problem Statement and Contributions

For ML techniques, the forecasting accuracy is essentially related to three factors, namely bias, variance, and noise. Inappropriate tuning of the aforementioned factors leads to overfitting or underfitting. A comprehensive adjustment of these factors is a core solution in improving prediction results. Ideally, the bias describes the mismatched samples values between measurements and forecasts taken during the learning process. Although the variance is the quantification of the squared deviation of a random feature from its mean [55]. Erroneous predictions are due to high variance or high bias. Mathematically, let y be a variable output generated by a function f with a set of variable vector X , we assume that \hat{f} is a forecast of $f(x)$. Then, the computation of error $Err(x)$ is given by [56]:

$$Err(x) = E[(y - \hat{f}(x))^2] \quad (1)$$

where an output y can be calculated as follows [56]:

$$y = f(X) + \varepsilon \quad (2)$$

where ε denotes the error term. Thus, the prediction error is given by:

$$Err(x) = (E[\hat{f}(x)] - f(x))^2 + E[\hat{f}(x) - E[\hat{f}(x)]]^2 + \sigma_\varepsilon^2 \quad (3)$$

where σ_ε^2 denotes the irreducible error. The optimal goal is to minimize both the variance and bias at the same time to reduce the errors. However, the bias-variance tradeoff is inversely proportional, as presented in Figure 2.

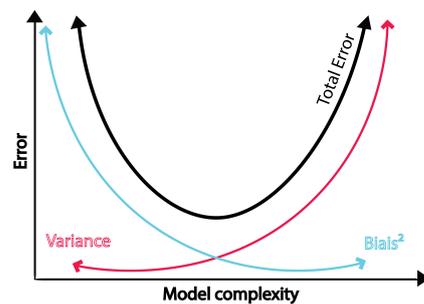


Figure 2. Bias-variance Tradeoff curve.

Therefore, the optimization task is required to obtain the desired outputs. In order to avoid high variance, we employed RF Ensemble presented in this paper. This variance optimization is achieved using randomized replication of the original dataset to construct sub-modules. The prediction output is presented by averaging these models' outputs. Although RF reduces the variance of one predictor, the system remains biased. The bias taken from the original model before the subdivision stays unchanged [57]. The bias optimization is carried out in this study using the interpretability of Feature Importance (FI).

Recently, FI has been commonly deployed, especially for high-dimensional data. It consists of the evaluation of the inputs' sensitivity to the output. The probability value (p -value) consists of the measurement of the evidence occurrence by calculating the probability of action when the null hypothesis is correct. The p -value of the distributed variables importance allows the system to determine the features' contribution as indicators of a future target behavior. The features with statistical significance are given higher importance from the p -value coefficient and vice versa. Additional information from p -values that improves the model accuracy is given in [58]. The Feature Attribute Coefficient (FAC) deploys this knowledge to optimize the variance/bias of the ANN. Feature relative importance (FRI) introduces metrics weights to emphasize the significance of variables to the model. The feature weights are combined in a vector named the null importance. In the case of small datasets, instead of putting a threshold for weights values and the removal of every feature that has a lower weight, our proposed technique takes into account all the features that have a physical interaction with the output. In this paper, a novel Voted Feature Weighting (VFW) is introduced and deeply investigated. This procedure reduces system complexity and computational burden. The weights are fed in an ensemble learning system for the aim of achieving further accuracy. In the proposed model, Feature Importance (FI) is considered a crucial part of the decision-making in the PPF system. This is related to the role of FI ranking in avoiding multicollinearity and low accuracy caused by the arbitrary variable selection. To the best of the authors' knowledge, the architecture of the proposed machine learning model has not been reported in the literature. The implementation results have been verified and validate the effectiveness of the proposed model for bias correctness.

4. Proposed Methodology

The importance of an input depends on whether the forecasting performance varies dramatically when such input is replaced with random noise [59]. Thus, the selection of the best features or the best combination of features has an utmost importance on the prediction model performance. The proposed methodology consists of introducing a preprocessing approach based on the p -value information associated with the RF model. For variable importance quantification, the p -value is designed to measure the feature relevance using Gini index (impurity). For every input parameter, the p -value is measured according to three feature ranking techniques: Elastic Net, Local Interpretable Model-agnostic Explanations (LIME), and Extreme Gradient Boosting (XGBoost). The Feature Vector Importance (FVI) is unified for each method to fairly grasp the non-linear relationships among candidate attributes and assess interactions to showcase the most effective combination of features. The global FVI is concluded using the average FVI methods output. Afterward, with every elimination of one feature, RF model generates an output result using the rest of the data.

Then, a voted ensemble technique makes a multiscale prediction for input vectors and multiply the probability distribution by the FVI. This perfectly tailored forecasting system assumes that the RF acquires n features. The prediction system is divided into n subsystems. For each subsystem, a k feature parameter is eliminated from the database. With the Bagging model, every subsystem gives a prediction output \bar{y}_i . Let us make $w_i \in [0, 1]^d$ be the importance rate of each feature. The final output is concluded by summing the weighted subsystems products by an importance factor explained in:

$$Y_i = \sum_{i=1}^n w_i \bar{y}_i \quad (4)$$

where the weight values w_i denote adjusted using three potential FRI methods. The usefulness of using multiple techniques simultaneously lies in the variant architectures of these tools. Regarding the fact that the selection of the most suitable FRI method is confusing, these three methods are taken into consideration to give more integrity to the domain knowledge. Assuming N is the number of feature weighted methods. The correlation could be averaging or voting as follows:

$$w_{feature} = \frac{1}{N} \sum_{j=1}^N w_j \quad (5)$$

In the study, averaging is the primary case deployed. Then, a voting output result is comprehensively analyzed. Assuming w is the reweighting feature coefficient. \bar{x}_{ij} is the feature weighted of x_i , which is computed as:

$$\bar{x}_{ij} = \frac{x_{ij}}{s(w_j)}; \quad j = \{1, \dots, d\} \quad (6)$$

where s denotes a positive coefficient, w is the weight vector, and x_{ij} is the j th feature of x_i . The proposed method allows RF to overcome overfitting by an additional correctness vector. By using a distinguished feature importance, this paper verifies the contribution of multiple FRI techniques to the model accuracy as shown in Figure 3.

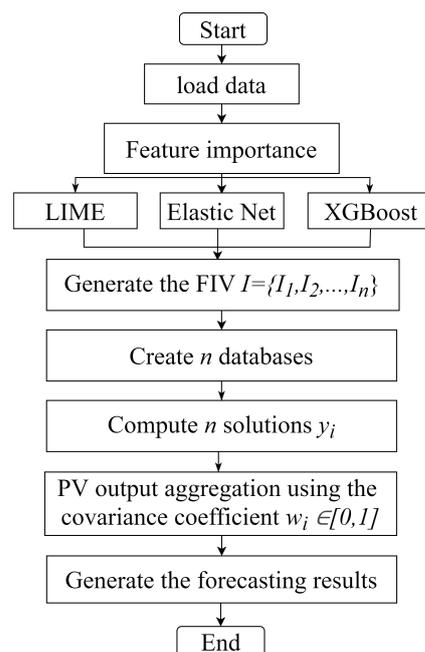


Figure 3. Flow diagram for proposed model implementation.

5. Case Study

Typically, forecasting techniques applied for smart grid operations are validated through a meteorological database and a real power system to verify the efficiency and feasibility of the proposed model.

5.1. PV System Description

In this study, the data used for the numerical validation of the proposed model is obtained from an open source from a large-scale PV plant in the Desert Knowledge Australia Solar Center (DKASC), Alice Springs (AS), Australia, at a latitude $2376'$ S and a longitude $133^{\circ}87'$ E [60]. AS has a desert climate with scarce rainfall and frequent clear skies during the dry days and, therefore, comparatively rare output volatility in the PV generation due to sky cover during that period. Rainy days are frequently registered between November and February during the wet season leading to high PV uncertainty. For showcasing the PV systems' repartition in DKASC, Figure 4, is represented.

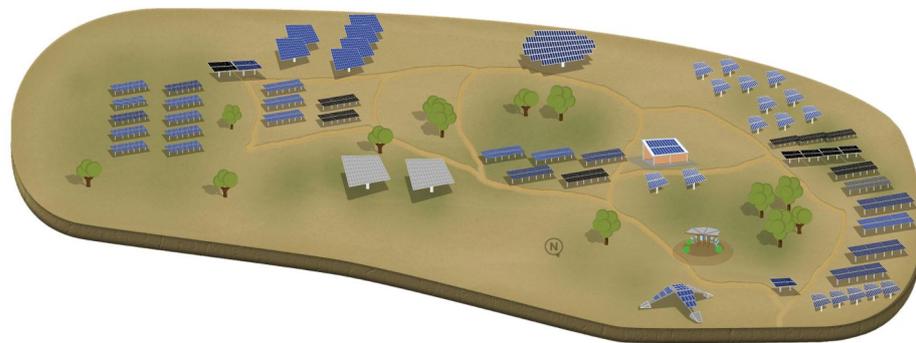


Figure 4. Representative chart of the DKASC PV plant [60].

This PV plant relies on high-resolution sensors for PV systems of different technologies and configurations to record data every five minutes. The DKASC consists of a demonstration facility of 38 sites to build a high confidence level of PV technologies with different manufactures and stakeholders. The detailed characteristics of the used PV plant are summarized in Table 2. The explanatory labels consist of time indicator, relative humidity, wind speed, and its direction, horizontal irradiation, relative horizontal irradiation, temperature, and PV power output. The above-mentioned parameters allow the PPF techniques to tackle every slight change that could affect the PV power generation. The measurements were taken from 1 April 2016 to 1 August 2019, which provides sufficient information for training and validation. 248,503 samples were pre-processed and split into 3 phases, namely training, validation, and testing tiers. Regarding the validation process, 17,280 samples are devoted to the analysis of the prediction quality.

Table 2. The related characteristics of the PV plant.

System Specification	Characteristics
Array rating	191.74 kW
Average of Powering	141 house
Location	Alice Springs, Australia
PV technology	Crystalline Silicon, CdTe/CIGS
First operating installation	Since 2008
Array area	$4 \times 38.37 \text{ m}^2$
Type of tracker	Fixed: Ground Mount, Single Axis, Dual Axis
Inverter size/type	46 kW, SMA/Sunny Mini Central 6000A

5.2. Feature Engineering

In real-world problems, several preprocessing steps were taken into account for better interpretability of the obtained data. The complete steps of the adopted data preprocessing strategy are given in Figure 5.

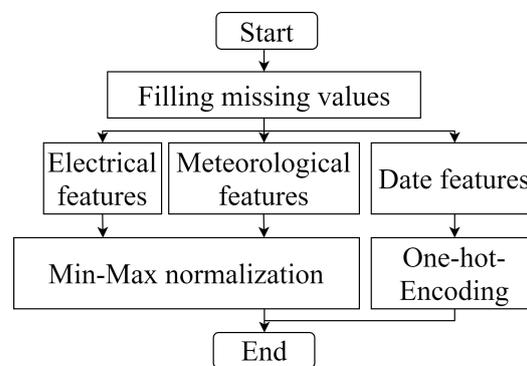
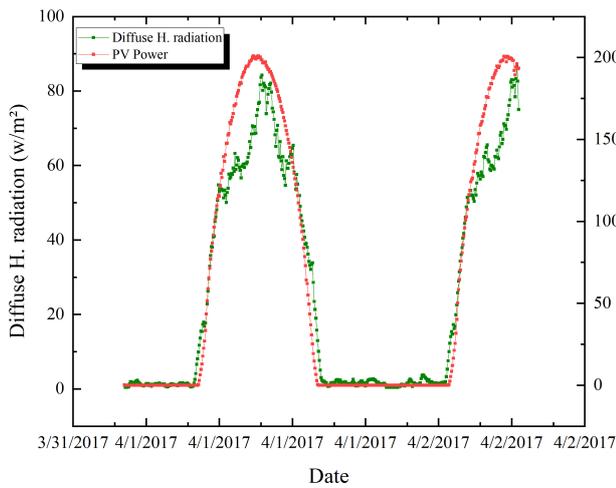


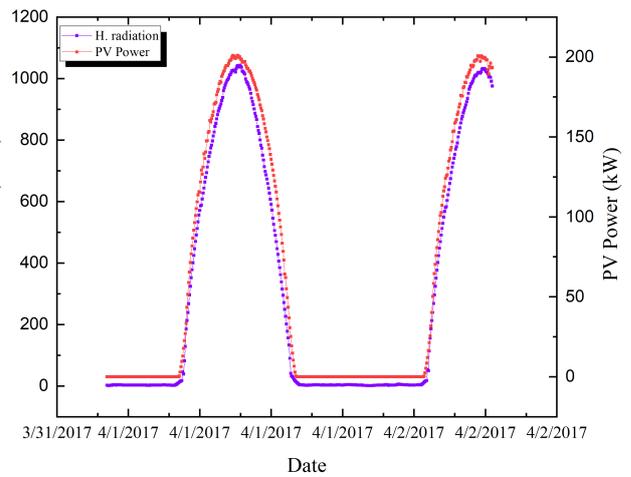
Figure 5. Flowchart of the adopted data preprocessing strategy.

Specifically, the acquired data have been cleaned from outliers, missing, and redundant data, which require a huge effort to be smoothed accordingly. For instance, for special cases where it can be found some samples were missing the values for the same times from the previous or following day were inserted. To deal with this issue, an output of zero was given to larger missing boxes. These samples are later excluded from the database since they do not give any significance to the variability analysis. As a result of the data cleaning process, the wind speed has been removed from the system inputs since it includes many apparent wrong measurements (negative values) and missing values. The generated data contains electrical features such as the PV power generation, meteorological features such as the wind orientation, temperature, and horizontal radiation, and date features. Using one-hot encoding method, the date features are transformed into numerical values to be used in the forecasting system for all data to be time-synchronized. However, the date features are excluded from the feature inputs since the irradiation features already embody time and seasonal variation tendencies. Finally, the resulting samples are standardized by the Min–Max normalization method to the range of $[0, 1]$ to prevent the model saturation during the learning process and promote the efficiency of the forecasting system [33]. The original PV power and its related features are shown in Figure 6.

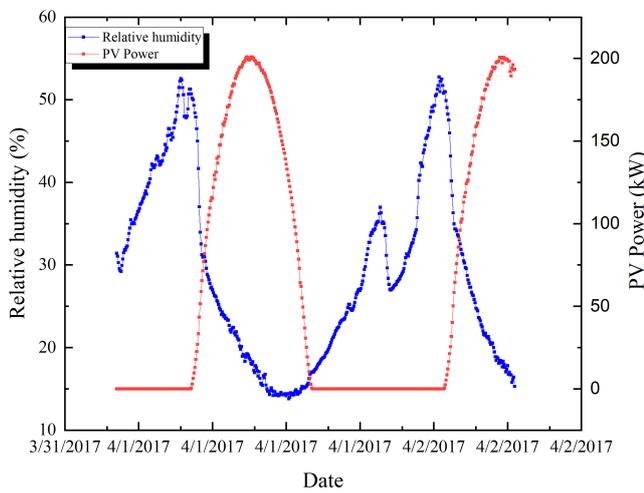
As can be seen from Figure 6, the related indicators have a direct relation with the PV power output. However, these correlations differ from one input to another. For example, the previous PV power and the horizontal radiation are perfectly tailored for the PPF contrary to the wind direction, which shows less variation with the PV power. Regardless of the weather indicators, the accumulation of PV power records over the years may be taken into consideration as a reliable measure of future PV power predictions. In the simulation part, one year (2017) data set is used for training prior to the beginning of the yearly test period. Subsequently, the testing data measurements during 2018 were used for the evaluation process. As shown in Figure 7, the historical PV power curves are illustrated for the 1st of August and the 1st of April of four consecutive years and the monthly PV power during three successive years.



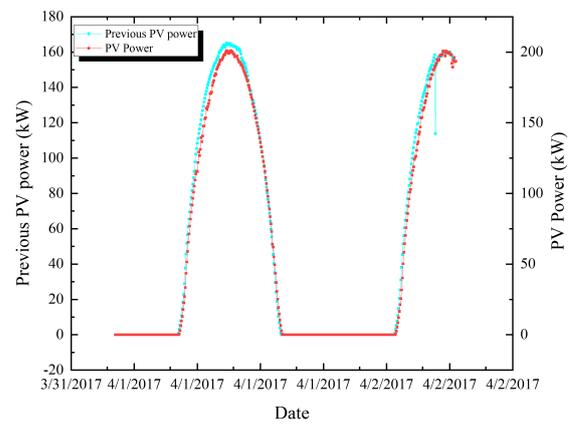
(a) PV power vs diffuse horizontal radiation



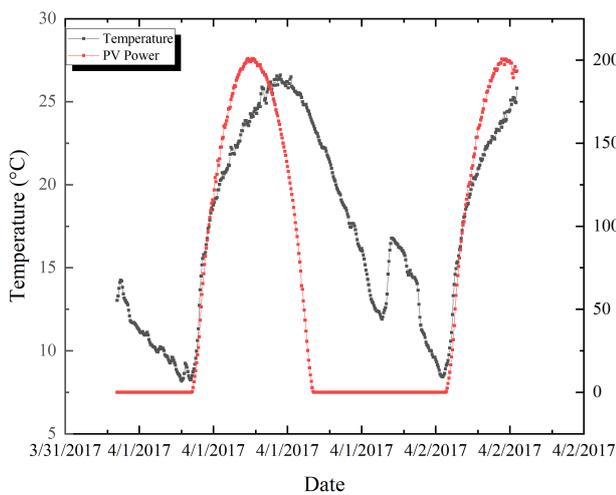
(b) PV power vs horizontal radiation



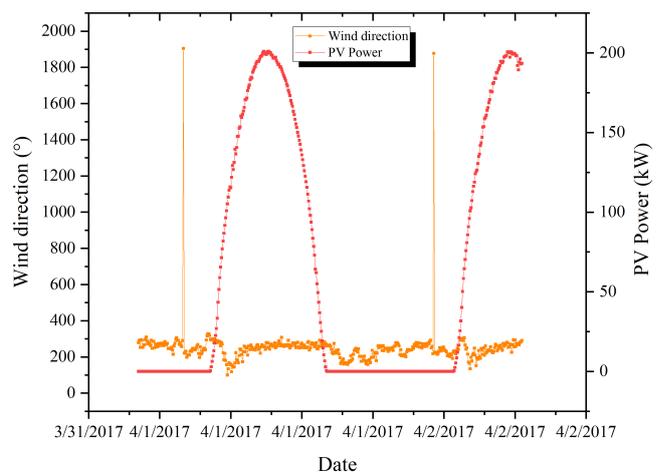
(c) PV power vs relative humidity



(d) PV power vs previous PV power



(e) PV power vs temperature



(f) PV power vs wind direction

Figure 6. PV power time series with the related feature inputs: (a) diffuse horizontal radiation (b) horizontal radiation (c) relative humidity (d) previous PV power (e) temperature (f) wind direction.

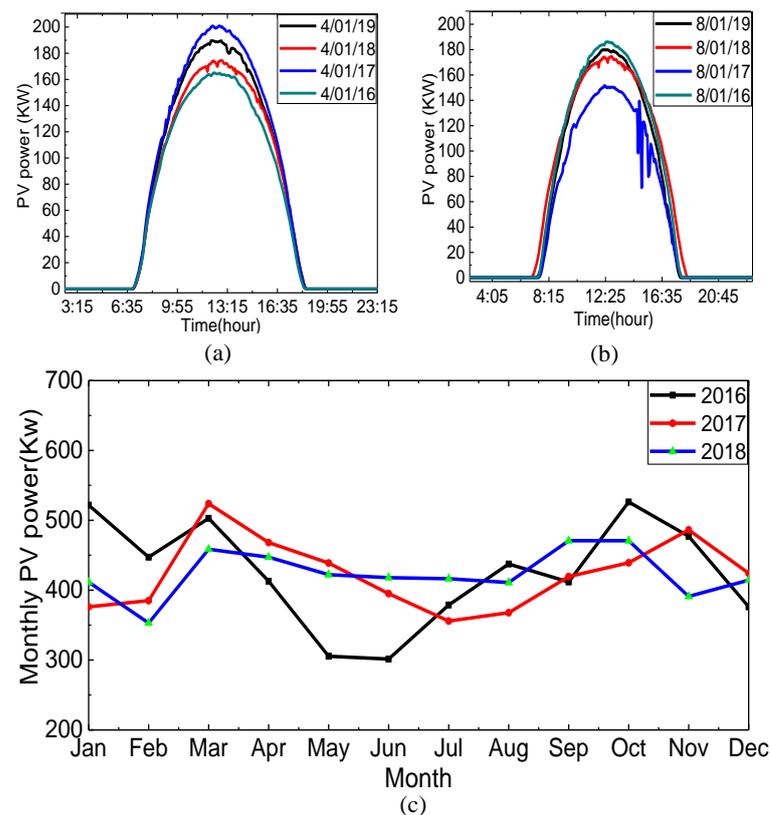


Figure 7. PV power variation (a) on April 1st in the past 4 years (b) on August 1st in the past 4 years (c) between 2016–2018.

It can be noticed that the real generated PV power in 2018 for the spring season (Figure 7a) and the winter season (Figure 7b) is remarkably close to the first previous last year accordingly. This impressive behavior leads to conclude that the lagged yearly PV output presents an important feature indicator for the estimation of the PV power. It can be noticed from Figure 7c) that the measured power values during 2017 and 2018 have a close variation, while the 2016 values are less correlated with the following years. This difference is noticeable during January, April, May, and June months. The historical PV power series at the same instant from the previous year is associated with the weather parameters and the hourly time indicator. These inputs face many processing stages. The first step is composed of feature engineering and data cleaning. The missing and odd data are removed from the dataset. Next, the extracted inputs samples pass by a feature selection stage to evaluate their importance. The P -value of each input indicates its relevance to the PV power output. For a given temporal resolution of 5 min, a total of 288 samples are collected per day.

5.3. Feature Vector Construction

The weather dataset consists of the sensor measurements, including the temperature ($^{\circ}\text{C}$), relative humidity (%), wind direction ($^{\circ}$), PV power (kW), and horizontal and vertical solar radiation (W/m^2). It may be intuitively understood that the chosen features are relevant to PV generation. Feature selection methods lie in shaping the FVI. Figure 8 presents the simulation results of a set of feature importance ranking. The annotation gives LIME the yellow color, Elastic Net the grey color, and XGBoost the red color.

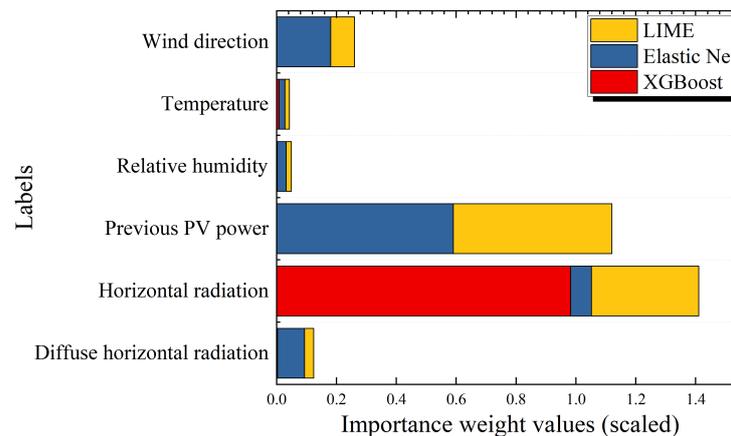


Figure 8. Relative importance of the candidate variables, using LIME, Elastic Net, and XGBoost.

The proportional importance distribution is not uniformly partitioned, as shown in Figure 8. For instance, XGBoost only considers the horizontal radiation as an informative feature while the Elastic Net does not attribute high p -value to Horizontal radiation. On the contrary, LIME gives the most physically comprehensive results. The yearly lagged PV power and horizontal radiation are the most correlated features with the current PV power. Each feature is given a relevant weight value w_i with $\sum_{i=1}^n w_i = 1$. The weights calculation takes into consideration the inputs permutation, and the percentage of the error caused by the exclusion of the corresponding feature. The higher p -value reflects, the closer the behavior of a feature inputs to the output predicted.

This diversity contributes to the system accuracy from the FVI coefficients. The horizontal irradiation followed by the previous PV power from the same instance in the neighboring year gains more importance. Next, the wind direction comes third, followed by the diffuse horizontal irradiation. Finally, the relative humidity takes place to finish with the temperature parameter in the last position with a lower relevant information according to the accumulation of the three methods p -values.

5.4. Simulation Results and Comparison with Benchmark Models

The proposed paradigm passes by four stages, specifically, data processing and feature engineering, object determination, model constriction, and evaluation as shown in Figure 9.

The data are normalized using Min–Max normalization in the data preprocessing stage. The Min–Max normalization is defined as follows:

$$x_n = \frac{x_r - x_{min}}{x_{max} - x_{min}} \quad (7)$$

where x_n denotes the normalized weather variable, x_r is the real value. Here, x_{min} and x_{max} are the minimum and maximum values. The hybrid model employs a Randomized Search tool for hyperparameter optimization. The outputs for this tool assign to the modified RF a minimum sample leaf of 20, maximum leaf nodes of 100 and maximum depth equal to 8. For reference models, Table 3 illustrates the hyperparameter of benchmarks.

The trained model is verified on a testing dataset. All experimental models run in Python 3.6.7 programming environment. The hardware is a Lenovo personal computer (PC) with Intel Core i7 9th Generation and 16 GB of memory. The Windows 10 operates a graphic card of NVIDIA GeForce GTX 1650.

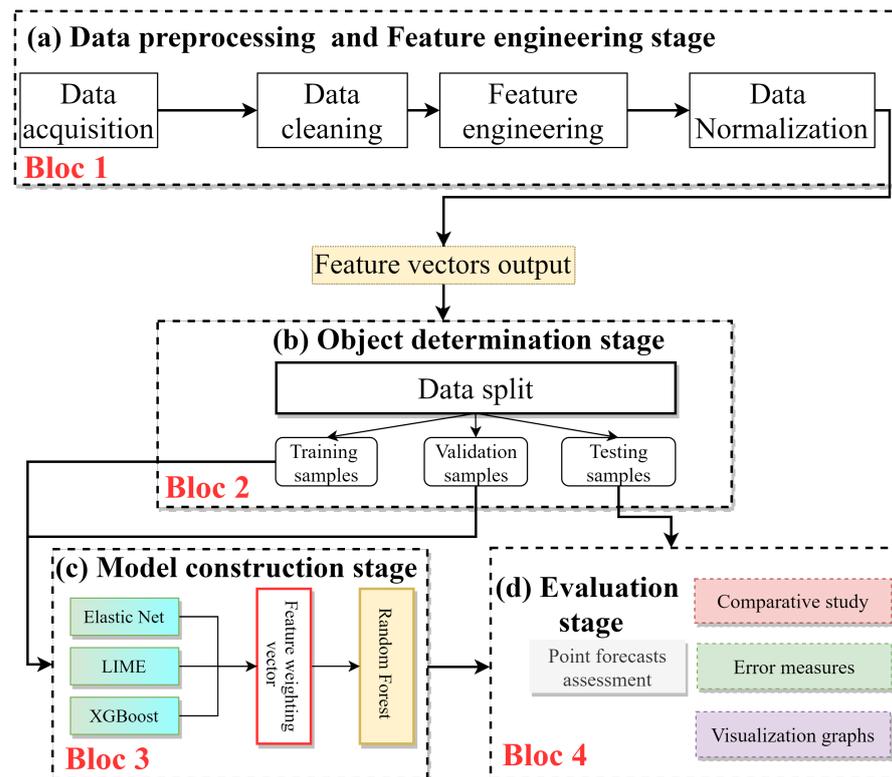


Figure 9. Structure of the proposed methodology with (a) data preprocessing and feature engineering, (b) object determination, (c) model construction (d) evaluation.

Table 3. Hyperparameters settings for reference models.

Base Models	Hyperparameter Settings
DT	maximum depth = 3; minimum samples leaf = 3; maximum leaf nodes = 5; minimum impurity decrease = 0.2
KNN	The algorithm is KDTree; the nearest neighbor number is 7; the leaf size is 90; the distance function is Minkowski distance
RF	The maximum depth is 50; the minimum samples split is 10; The number of estimators is 140

Score metrics between the actual power y_i and the forecast points \hat{y}_i were computed in terms of coefficient of determination (R^2), RMSE and Mean Absolute Error (MAE) as follows [61]:

$$MAE = \frac{1}{n} \sum_{i=0}^{n-1} |y_i - \hat{y}_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \hat{y}_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=0}^{n-1} (\hat{y}_i - y_i)^2}{\sum_{i=0}^{n-1} (\bar{y} - y_i)^2}, \bar{y} = \sum_{i=0}^{n-1} y_i \quad (10)$$

where n denotes the total number of samples. The simulation results are compared with those results of the KNN, RF, and DT models. The forecasting horizon is investigated for the short-term, specifically, for 5-min daily time interval during the year of 2018. The testing data are evaluated using RMSE and MAE, and R^2 score metrics. Figure 10 presents the PV power variations in four types of days arbitrary selected, including rainy, sunny to cloudy, sunny, foggy to cloudy days.

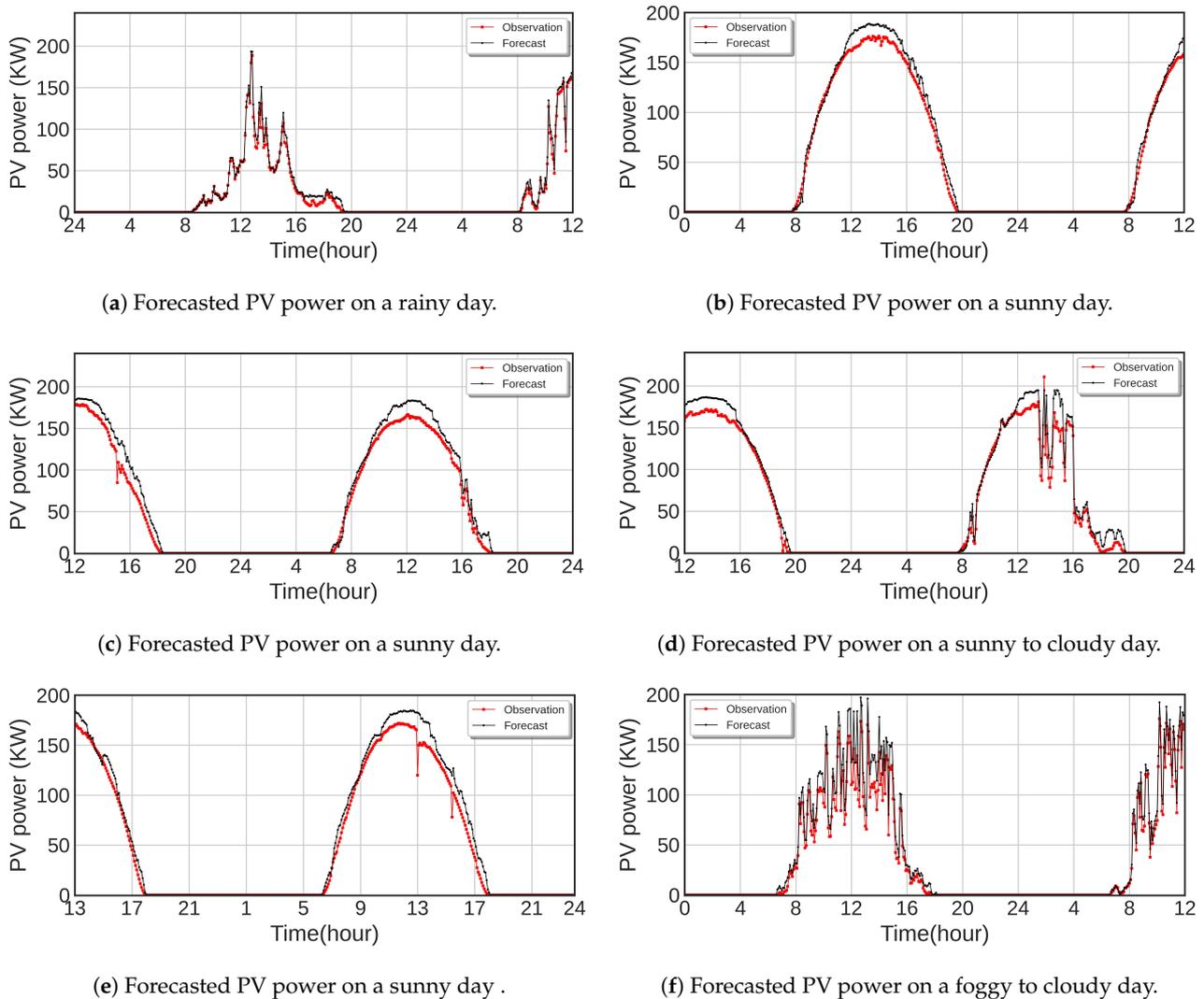


Figure 10. Actual and forecasted PV power using DKASC dataset for different weather patterns with (a) March 20, Rainy. (b) May 20, sunny. (c) June 20, sunny. (d) July 20, sunny to cloudy. (e) September 20, Sunny. (f) November 20, foggy to cloudy.

Regarding Figure 10, it can be noticed that the proposed model exhibits satisfactory forecasting performance according to the forecasting curves between the ground truth and the forecasted PV power. From Figures 10b,c,e, the the actual PV power ramps smoothly during the sunny day. With no abrupt change, the proposed model efficiently provides precise estimations of the PV power output. Even with sudden changes, especially in the middle of the day, it can be noticed from Figure 10a,d,f that the proposed hybrid machine learning model was able to follow the curve shape of the PV power generation during the rainy, foggy to cloudy weather from the close distance between the forecasted and real points. The forecast ability of the proposed method seems to be very promising in all the seasons of the year with different climatic conditions. To showcase the prediction performance of the proposed model in a more intuitive way, Figure 11 presents the scatter plot and error distributions of the proposed model.

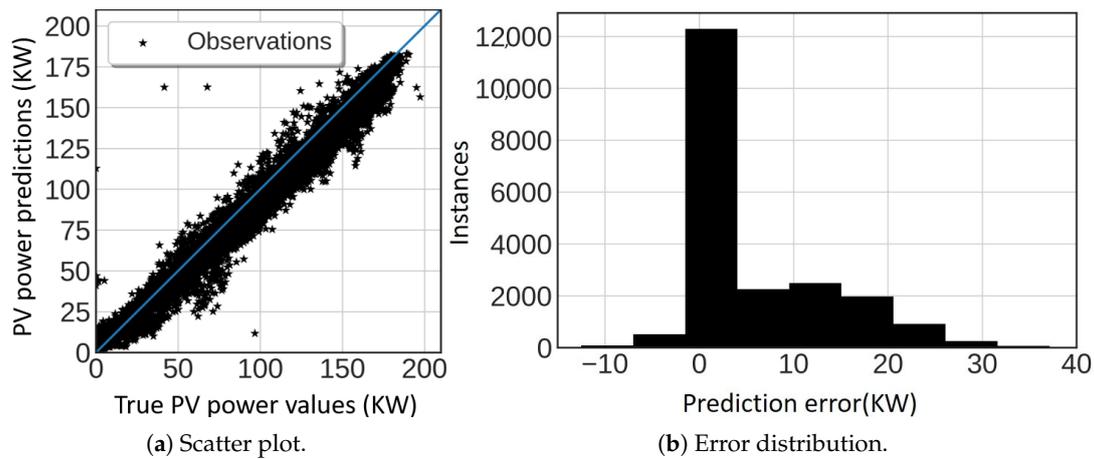


Figure 11. Scatter plots (a) and error distributions (b) of PV measured power and forecasted power.

Regarding Figure 11b, it is apparent that the forecasted points are consistent with the actual values. In Figure 11b, the majority of the instance are concentrated on the zero axes. It should be emphasized that the difference of the proposed range between -10 kW and 40 kW in the worst-case scenario. To better examine the model performance, Figure 12 illustrates a 10-fold Cross-validation (10-CV) curve.

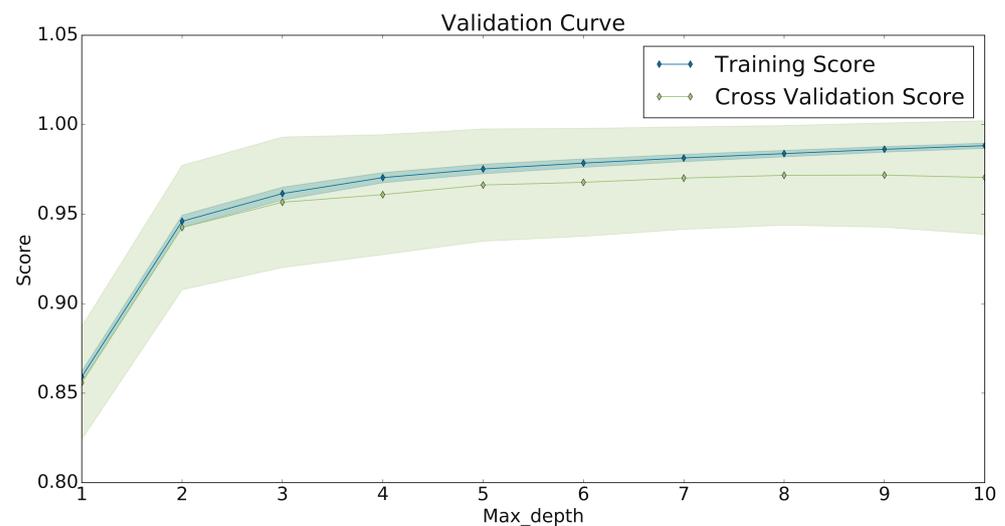


Figure 12. 10-CV flowchart using the R^2 measure.

According to Figure 12, the model conducts a coefficient of determination $R^2 = 96\%$, which reflects the high potential of the proposed approach in improving the existing RF. The simulation results confirm that the bias correctness is significantly contributing to the prediction accuracy. The proposed technique is effective in diminishing the errors coming from the misleading inputs. Therefore, the proposed architecture generalization for ML model improvement worth further investigation. The model performance assessment requires the deployment of different methodologies and comparative analysis to ML models. Table 4 includes MAE, and RMSE scores for ML models.

Table 4. Forecast error metrics of the simulated predictors for various weather conditions.

Weather Condition	Model	RMSE (kW) \pm SD	MAE (kW) \pm SD
Sunny	KNN	11.00	5.94
	RF	12.14	6.70
	DT	12.41	6.84
	Improved RF	9.60	5.23
Partially cloudy	KNN	12.74	8.08
	RF	17.17	11.26
	DT	17.49	11.40
	Improved RF	10.79	6.43
Cloudy/foggy	KNN	17.75	12.68
	RF	19.69	14.00
	DT	19.96	14.11
	Improved RF	11.65	8.51
Rainy	KNN	3.68	1.48
	RF	8.49	5.00
	DT	8.84	5.20
	Improved RF	1.41	0.65
Overall	KNN	11.29 \pm 5.05	7.04 \pm 4.03
	RF	14.37 \pm 4.35	9.24 \pm 3.58
	DT	14.68 \pm 4.33	9.39 \pm 3.55
	Improved RF	8.36 \pm 4.08	5.21 \pm 2.88

Table 4, compares different predictors in terms of RMSE and MAE errors and different weather conditions. From Table 4, the proposed method is highly effective according to the registered low error values. From Table 4, the proposed method outperforms the list of models depicted, achieving a mean RMSE = 8.36 kW and a mean MAE = 5.21 kW. The high accuracy achieved takes advantage of the bias correctness of RF model. Alternatively, the original RF and DT produces large MAE, and RMSE values than the proposed model resulting in poor performance. In particular, RF generates an RMSE = 14.37 kW and a MAE = 9.24 kW. This superiority is provided by the p -value adjustment based on multiple features selection methods. Thus, the model ensures the correct repartition of categorical feature inputs instead of selecting a particular threshold and extracting the feature corresponding to higher p -values from a single assessment method. Although the subsequent heavy computation of p -values calculations, the proposed model is high performing in PV power series forecasting.

5.5. Discussion and Analysis

As the weather indicators have a disproportionate relevance on the PV power generation, the intuition of associating an importance vector to emphasize the relevance of each variable input seems to be promising for the overall forecasting system accuracy. From the above-mentioned results (Section 5.4), it can be said that the proposed model shows excellent predictive performance for different meteorological conditions. This will permit the generalization of the proposed model. More specifically, the strong competitive advantage of using the proposed model is evident during the rainy and cloudy days since the prediction results are very close to the real values. Therefore, the proposed site-specific hybrid model can be applied to similar PV power systems with different climatic conditions and different locations. Despite the PV power output is sensitive to chaotic meteorological conditions, the proposed model has the potential to capture the trend of the PV power generation with dramatic variability. Compared to the original RF model, the proposed method significantly improves the forecast accuracy by giving more importance to the feature relevance. In fact, it generates forecasting results with the lowest RMSE and MAE

on most types of the day. The results further reveal the robustness of the proposed method. The superior accuracy of the proposed model is primarily due to attributing a coefficient that describes the importance of each feature to the PV power output, which provides an effective means to approximate inherent invariant features and structures.

6. Conclusions

For a reliable and secure operation of power systems, this paper seeks to explore the problem of predicting PV power generation for efficiently manage the capacity of the intermittent asynchronous PV generators. To overcome this challenge, an Enhanced Random Forest (ERF) model was first proposed to increase system forecasting accuracy based on the adequate understanding of the unequal influence of the input indicators on the PV power. To distinguish the relevance of the variables, a feature vector importance has been constructed based on three methods, specifically, Elastic Net, Local Interpretable Model-agnostic Explanations (LIME), and Extreme Gradient Boosting (XGBoost). A multivariate dataset from Desert Knowledge Australia Solar Center (DKASC) has been employed to validate the efficiency of the proposed method. The numerical performance investigation in sunny, rainy, and cloudy days demonstrate that the proposed model is effective, simple, explainable, and more accurate than the benchmark models with an overall RMSE = 8.36 kW and MAE = 5.21 kW. The proposed model is perfectly tailored to fulfill short-term PV power forecasting needs with high efficiency. Although the proposed model seems to be suitable for PV power systems, there are many areas that can be improved and optimized, such as the deep investigation of the weather patterns to forecasting performance substantially and empower the proposed approach stability for different climatic conditions.

Author Contributions: M.M.: Conceptualization, Methodology, Software, original draft preparation. I.C.: Validation, writing review & editing, project administration, funding acquisition. L.S.: Validation, formal analysis, Data curation, Writing review and editing. M.T.: Formal analysis, investigation, writing review & editing, resources. S.S.R.: Proofreading and editing, review and editing. F.S.O.: Supervision, project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SE	Solar Energy
LSSVR	Least squares support vector regression
MAD	Mean Absolute Deviation
RMSE	Root Mean Square Error
NMAE	Normalized Mean Square Error
WMAE	Weighted Mean Square Error
R^2	Coefficient of determination
MAPE	Mean Absolute Percent Error
CNN	Convolutional Neural Network
MRE	Mean Relative Error
RNN	Recurrent Neural Network
NARX	Non linear Auto-Regressive with Exogenous inputs
ARMAX	Autoregressive-Moving-Average model with Exogenous inputs
SD	Standard Derivation
ARMA	Autoregressive-Moving-Average

References

1. Rajagukguk, R.A.; Ramadhan, R.A.; Lee, H.J. A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power. *Energies* **2020**, *13*, 6623. [[CrossRef](#)]
2. Zervos, A.; Lins, C.; Muth, J. *RE-Thinking 2050: A 100 European Union*; Erec: Brussels, Belgium, 2010.
3. Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Shah, N.M. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renew. Power Gener.* **2019**, *13*, 1009–1023. [[CrossRef](#)]
4. Massaoudi, M.; Refaat, S.S.; Abu-Rub, H.; Chihi, I.; Oueslati, F.S. PLS-CNN-BiLSTM: An End-to-End Algorithm-Based Savitzky-Golay Smoothing and Evolution Strategy for Load Forecasting. *Energies* **2020**, *13*, 5464. [[CrossRef](#)]
5. Salamanis, A.I.; Xanthopoulou, G.; Bezas, N.; Timplalexis, C.; Bintoudi, A.D.; Zyglakis, L.; Tsolakis, A.C.; Ioannidis, D.; Kehagias, D.; Tzovaras, D. Benchmark Comparison of Analytical, Data-Based and Hybrid Models for Multi-Step Short-Term Photovoltaic Power Generation Forecasting. *Energies* **2020**, *13*, 5978. [[CrossRef](#)]
6. Alam, A.M.; Razee, I.A.; Zunaed, M. Solar PV Power Forecasting Using Traditional Methods and Machine Learning Techniques. In Proceedings of the 2021 IEEE Kansas Power and Energy Conference (KPEC), Manhattan, KS, USA, 19–20 April 2021; IEEE: New York, NY, USA, 2021; pp. 1–5.
7. Dong, Z.; Yang, D.; Reindl, T.; Walsh, W.M. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy* **2013**, *55*, 1104–1113. [[CrossRef](#)]
8. Li, Y.; Su, Y.; Shu, L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* **2014**, *66*, 78–89. [[CrossRef](#)]
9. Işığık, E.; Öz, R.; Tarkun, S. Forecasting and Technical Comparison of Inflation in Turkey With Box-Jenkins (ARIMA) Models and the Artificial Neural Network. *Int. J. Energy Optim. Eng. (IJEOE)* **2020**, *9*, 84–103. [[CrossRef](#)]
10. Van der Meer, D.; Mouli, G.R.C.; Mouli, G.M.E.; Elizondo, L.R.; Bauer, P. Energy management system with PV power forecast to optimally charge EVs at the workplace. *IEEE Trans. Ind. Inform.* **2016**, *14*, 311–320. [[CrossRef](#)]
11. Ding, M.; Wang, L.; Bi, R. An ANN-based approach for forecasting the power output of photovoltaic system. *Procedia Environ. Sci.* **2011**, *11*, 1308–1315. [[CrossRef](#)]
12. Karabacak, K.; Cetin, N. Artificial neural networks for controlling wind-PV power systems: A review. *Renew. Sustain. Energy Rev.* **2014**, *29*, 804–827. [[CrossRef](#)]
13. Al-Dahidi, S.; Ayadi, O.; Adeeb, J.; Louzazni, M. Assessment of artificial neural networks learning algorithms and training datasets for solar photovoltaic power production prediction. *Front. Energy Res.* **2019**, *7*, 130. [[CrossRef](#)]
14. Cervone, G.; Clemente-Harding, L.; Alessandrini, S.; Delle Monache, L. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renew. Energy* **2017**, *108*, 274–286. [[CrossRef](#)]
15. Shah, A.A.; Ahmed, K.; Han, X.; Saleem, A. A Novel Prediction Error Based Power Forecasting Scheme for Real PV System using PVUSA Model: A Grey Box Based Neural Network Approach. *IEEE Access* **2021**. [[CrossRef](#)]
16. Almonacid, F.; Pérez-Higueras, P.; Fernández, E.F.; Hontoria, L. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Convers. Manag.* **2014**, *85*, 389–398. [[CrossRef](#)]
17. Chen, C.; Duan, S.; Cai, T.; Liu, B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. Energy* **2011**, *85*, 2856–2870. [[CrossRef](#)]
18. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. A multivariate ensemble framework for short term solar photovoltaic output power forecast. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017; IEEE: New York, NY, USA, 2017; pp. 1–5.
19. Al-Dahidi, S.; Ayadi, O.; Alrbai, M.; Adeeb, J. Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction. *IEEE Access* **2019**, *7*, 81741–81758. [[CrossRef](#)]
20. Chai, Z.; Zhao, C. Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification. *IEEE Trans. Ind. Inform.* **2019**, *16*, 54–66. [[CrossRef](#)]
21. Mellit, A.; Massi Pavan, A.; Oglari, E.; Leva, S.; Lughi, V. Advanced methods for photovoltaic output power forecasting: A review. *Appl. Sci.* **2020**, *10*, 487. [[CrossRef](#)]
22. Massucco, S.; Mosaico, G.; Saviozzi, M.; Silvestro, F. A hybrid technique for day-ahead PV generation forecasting using clear-sky models or ensemble of artificial neural networks according to a decision tree approach. *Energies* **2019**, *12*, 1298. [[CrossRef](#)]
23. Oglari, E.; Dolara, A.; Manzolini, G.; Leva, S. Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renew. Energy* **2017**, *113*, 11–21. [[CrossRef](#)]
24. Massaoudi, M.; Chihi, I.; Sidhom, L.; Trabelsi, M.; Refaat, S.S.; Abu-Rub, H.; Oueslati, F.S. An effective hybrid NARX-LSTM model for point and interval PV power forecasting. *IEEE Access* **2021**, *9*, 36571–36588. [[CrossRef](#)]
25. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Chihi, I.; Oueslati, F.S. Deep Learning in Smart Grid Technology: A Review of Recent Advancements and Future Prospects. *IEEE Access* **2021**, *9*, 54558–54578. [[CrossRef](#)]
26. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.; Mucci, D. Day-ahead hourly forecasting of power generation from photovoltaic plants. *IEEE Trans. Sustain. Energy* **2017**, *9*, 831–842. [[CrossRef](#)]
27. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2019**, *31*, 2727–2740. [[CrossRef](#)]
28. Yang, M.; Huang, X. Ultra-short-term prediction of photovoltaic power based on periodic extraction of PV energy and LSH algorithm. *IEEE Access* **2018**, *6*, 51200–51205. [[CrossRef](#)]

29. Graupe, D.; Krause, D.; Moore, J. Identification of autoregressive moving-average parameters of time series. *IEEE Trans. Autom. Control* **1975**, *20*, 104–107. [[CrossRef](#)]
30. Da Costa Lopes, F.; Watanabe, E.H.; Rolim, L.G.B. A control-oriented model of a PEM fuel cell stack based on NARX and NOE neural networks. *IEEE Trans. Ind. Electron.* **2015**, *62*, 5155–5163. [[CrossRef](#)]
31. Li, G.; Xie, S.; Wang, B.; Xin, J.; Li, Y.; Du, S. Photovoltaic Power Forecasting With a Hybrid Deep Learning Approach. *IEEE Access* **2020**, *8*, 175871–175880. [[CrossRef](#)]
32. Tao, C.; Shanxu, D.; Changsong, C. Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement. In Proceedings of the 2nd International Symposium on Power Electronics for Distributed Generation Systems, Hefei, China, 16–18 June 2010; IEEE: New York, NY, USA, 2010; pp. 773–777.
33. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Chihi, I.; Oueslati, F.S. An Effective Ensemble Learning approach-Based Grid Stability Assessment and Classification. In Proceedings of the 2021 IEEE Kansas Power and Energy Conference (KPEC), Manhattan, KS, USA, 19–20 April 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
34. Razagui, A.; Abdeladim, K.; Semaoui, S.; Arab, A.H.; Boulahchiche, S. Modeling the forecasted power of a photovoltaic generator using numerical weather prediction and radiative transfer models coupled with a behavioral electrical model. *Energy Rep.* **2020**, *6*, 57–62. [[CrossRef](#)]
35. Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [[CrossRef](#)]
36. Massaoudi, M.; Abu-Rub, H.; Refaat, S.S.; Chihi, I.; Oueslati, F.S. Accurate Smart-Grid Stability Forecasting Based on Deep Learning: Point and Interval Estimation Method. In Proceedings of the 2021 IEEE Kansas Power and Energy Conference (KPEC), Manhattan, KS, USA, 19–20 April 2021; IEEE: New York, NY, USA, 2021; pp. 1–6.
37. Sansa, I.; Missaoui, S.; Boussada, Z.; Bellaaj, N.M.; Ahmed, E.M.; Orabi, M. PV power forecasting using different artificial neural networks strategies. In Proceedings of the 2014 First International Conference on Green Energy ICGE 2014, Sfax, Tunisia, 25–27 March 2014; IEEE: New York, NY, USA, 2014; pp. 54–59.
38. Mohanty, S.; Patra, P.K.; Sahoo, S.S.; Mohanty, A. Forecasting of solar energy with application for a growing economy like India: Survey and implication. *Renew. Sustain. Energy Rev.* **2017**, *78*, 539–553. [[CrossRef](#)]
39. Lobaccaro, G.; Carlucci, S.; Löfström, E. A review of systems and technologies for smart homes and smart grids. *Energies* **2016**, *9*, 348. [[CrossRef](#)]
40. Massaoudi, M.; Chihi, I.; Sidhom, L.; Trabelsi, M.; Oueslati, F.S. Medium and Long-Term Parametric Temperature Forecasting using Real Meteorological Data. In Proceedings of the IECON 2019–45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 14–17 October 2019; IEEE: New York, NY, USA, 2019; Volume 1, pp. 2402–2407.
41. Dolara, A.; Leva, S.; Manzolini, G. Comparison of different physical models for PV power output prediction. *Sol. Energy* **2015**, *119*, 83–99. [[CrossRef](#)]
42. Mayer, M.J.; Gróf, G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl. Energy* **2021**, *283*, 116239. [[CrossRef](#)]
43. Fentis, A.; Bahatti, L.; Tabaa, M.; Mestari, M. Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations. *Int. J. Energy Environ. Eng.* **2019**, *10*, 189–206. [[CrossRef](#)]
44. Lu, J.; Wang, B.; Ren, H.; Zhao, D.; Wang, F.; Shafie-khah, M.; Catalão, J.P. Two-tier reactive power and voltage control strategy based on ARMA renewable power forecasting models. *Energies* **2017**, *10*, 1518. [[CrossRef](#)]
45. Shi, J.; Lee, W.J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans. Ind. Appl.* **2012**, *48*, 1064–1069. [[CrossRef](#)]
46. Li, G.; Wang, H.; Zhang, S.; Xin, J.; Liu, H. Recurrent neural networks based photovoltaic power forecasting approach. *Energies* **2019**, *12*, 2538. [[CrossRef](#)]
47. Alomari, M.H.; Adeeb, J.; Younis, O. Solar photovoltaic power forecasting in Jordan using artificial neural networks. *Int. J. Electr. Comput. Eng. (IJECE)* **2018**, *8*, 497–497. [[CrossRef](#)]
48. Wang, F.; Mi, Z.; Su, S.; Zhao, H. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies* **2012**, *5*, 1355–1370. [[CrossRef](#)]
49. Hu, Y.; Lian, W.; Han, Y.; Dai, S.; Zhu, H. A seasonal model using optimized multi-layer neural networks to forecast power output of PV plants. *Energies* **2018**, *11*, 326. [[CrossRef](#)]
50. Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies* **2015**, *8*, 1138–1153. [[CrossRef](#)]
51. Kushwaha, V.; Pindoriya, N.M. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew. Energy* **2019**, *140*, 124–139. [[CrossRef](#)]
52. Huang, C.J.; Kuo, P.H. Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. *IEEE Access* **2019**, *7*, 74822–74834. [[CrossRef](#)]
53. Zhou, H.; Zhang, Y.; Yang, L.; Liu, Q.; Yan, K.; Du, Y. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. *IEEE Access* **2019**, *7*, 78063–78074. [[CrossRef](#)]
54. Liu, D.; Sun, K. Random forest solar power forecast based on classification optimization. *Energy* **2019**, *187*, 115940. [[CrossRef](#)]
55. Geman, S.; Bienenstock, E.; Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **1992**, *4*, 1–58. [[CrossRef](#)]

-
56. Breiman, L. *Bias, Variance, and Arcing Classifiers*; Technical Report, Tech. Rep. 460; Statistics Department, University of California: Berkeley, CA, USA, 1996.
 57. Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [[CrossRef](#)] [[PubMed](#)]
 58. Al Iqbal, R. Empirical learning aided by weak domain knowledge in the form of feature importance. In Proceedings of the 2011 International Conference on Multimedia and Signal Processing, Guilin, China, 14–15 May 2011; IEEE: New York, NY, USA, 2011; Volume 1, pp. 126–130.
 59. Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* **2017**, *10*, 1168. [[CrossRef](#)]
 60. DKA Solar Centre. Available online: <http://dkasolarcentre.com> (accessed on 23 September 2019).
 61. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.