



Article XGBoost–SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions

Bowen Zhou ^{1,2,*}, Xinyu Chen ^{1,2,*}, Guangdi Li ^{1,2}, Peng Gu ^{1,2}, Jing Huang ³ and Bo Yang ^{1,2}

- ¹ College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; liguangdi@mail.neu.edu.cn (G.L.); gupeng@mail.neu.edu.cn (P.G.); yangbo@ise.neu.edu.cn (B.Y.)
- ² Key Laboratory of Integrated Energy Optimization and Secure Operation of Liaoning Province, Northeastern University, Shenyang 110819, China
- ³ State Grid Electric Power Research Institute Wuhan Efficiency Evaluation Company Limited, Wuhan 430072, China; huangjing1@sgepri.sgcc.com.cn
- * Correspondence: zhoubowen@ise.neu.edu.cn (B.Z.); 2170937@stu.neu.edu.cn (X.C.)

Abstract: Sustainability can achieve a balance among economic prosperity, social equity, and environmental protection to ensure the sustainable development and happiness of current and future generations; photovoltaic (PV) power, as a clean, renewable energy, is closely related to sustainability providing a reliable energy supply for sustainable development. To solve the problem with the difficulty of PV power forecasting due to its strong intermittency and volatility, which is influenced by complex and ever-changing natural environmental factors, this paper proposes a PV power forecasting method based on eXtreme gradient boosting (XGBoost)-sequential forward selection (SFS) and a double nested stacking (DNS) ensemble model to improve the stability and accuracy of forecasts. First, this paper analyzes a variety of relevant features affecting PV power forecasting and the correlation between these features and then constructs two features: global horizontal irradiance (GHI) and similar day power. Next, a total of 16 types of PV feature data, such as temperature, azimuth, ground pressure, and PV power data, are preprocessed and the optimal combination of features is selected by establishing an XGBoost-SFS to build a multidimensional climate feature dataset. Then, this paper proposes a DNS ensemble model to improve the stacking forecasting model. Based on the gradient boosting decision tree (GBDT), XGBoost, and support vector regression (SVR), a base stacking ensemble model is set, and a new stacking ensemble model is constructed again with the metamodel of the already constructed stacking ensemble model in order to make the model more robust and reliable. Finally, PV power station data from 2019 are used as an example for validation, and the results show that the forecasting method proposed in this paper can effectively integrate multiple environmental factors affecting PV power forecasting and better model the nonlinear relationships between PV power forecasting and relevant features. This is more applicable in the case of complex and variable environmental climates that have higher forecasting accuracy requirements.

Keywords: PV power forecasting; natural environmental factors; XGBoost–SFS; double nested stacking; optimal combination of features

1. Introduction

The goal of sustainable development is to protect and enhance the quality of life of future generations while meeting the needs of the present. PV power generation is one of the sustainable and clean energy sources, which converts solar energy into electricity. PV power prediction refers to the prediction and estimation of the power that is generated from a PV power generation system for a future period of time. This prediction is of great significance for power grid dispatching, energy planning, and market operation. By accurately predicting the power from a PV power generator, we can better regulate the operation of the power system, improve energy utilization efficiency, and reduce the risk of imbalance between power supply and demand. Therefore, the prediction of PV



Citation: Zhou, B.; Chen, X.; Li, G.; Gu, P.; Huang, J.; Yang, B. XGBoost–SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions. *Sustainability* **2023**, *15*, 13146. https://doi.org/10.3390/ su151713146

Academic Editor: Valeria Palomba

Received: 10 July 2023 Revised: 23 August 2023 Accepted: 29 August 2023 Published: 1 September 2023



Copyright: © 2023 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/). power generation has become an important part of sustainable development strategies. With the policy of "promoting the digital and intelligent development of energy" and the advantages of the green and clean industry chain, coupled with strong support from relevant national policies, China's PV industry is currently experiencing unprecedented growth [1]. However, the environment and climate will always and significantly impact the forecasting of PV power generation. The randomness of variables and time fluctuations pose a threat to grid security and power supply reliability when PV systems are connected to the grid [2]. For the safety and stability of the power grid and the economic operation of PV power stations, it is crucial to research effective methods that improve the forecasting accuracy of PV power generation [3,4].

Extensive research has been conducted by various scholars on methods for PV power forecasting [5]. In References [6–9], Mikkel L. Sørensen et al. introduced and researched a series of new methods for multivariate prediction of solar power generation. They proposed new prediction methods such as point forecasting, forecast reconciliation, and a new hybrid framework to increase the processing efficiency and accuracy of the model. In Reference [10], Yuan-Kang Wu et al. summarized and compared various new PV power forecasting methods and discussed the input selection of PV power forecasting models that reduced prediction uncertainty and maintained system security. Honglu Zhu et al. described one notable method in Reference [11] that involved the combination of wavelet decomposition (WD) and artificial neural networks (ANNs). This hybrid model utilizes theoretical solar irradiance and meteorological variables as inputs and employs WD to extract valuable information while filtering out disturbances. The results demonstrate faster calculation speeds and improved prediction accuracy. Xwégnon Ghislain Agoua et al. proposed another approach in Reference [12]; the approach presents a very short-term PV power forecasting model that leverages distributed power stations as sensors and exploits a spatiotemporal dependence to enhance forecasting accuracy. This method boasts low computational requirements and is well-suited for large-scale applications. Pengtao Li et al. introduced a PV power generation forecasting method combining wavelet packet decomposition (WPD) and long short-term memory (LSTM) network [13]. The original PV power generation series is divided into sub-series using WPD, assigning each sub-series to a separate LSTM network. The forecasting results of each LSTM network are then reconstructed, and a linear weighting methodology is applied to improve the final forecasting results. In Reference [14], Xing Luo et al. have taken into account the domain knowledge specific to PVs and introduced a physical LSTM (PC-LSTM) constraint model for hourly PV power generation prediction. This model addresses the limitations of recent machine learning algorithms that rely heavily on extensive data applications by incorporating physical constraints in the prediction process. In Reference [15], VanDeventer et al. proposed a genetic-algorithm support-vector machine model. This model utilizes SVM classifiers to analyze historical weather data and employs ensemble technology optimized using genetic algorithms to enhance model accuracy. In Reference [16], Mingzhang Pan et al. developed a support vector machine (SVM) model for ultra-short-term PV power forecasting. The model incorporates data preprocessing techniques and optimizes parameters using ant colony optimization (ACO). The results demonstrate a significant improvement in peak power and nighttime forecasting accuracy. Additionally, in Reference [17], Fei Wang et al. presented a day-ahead PV power generation forecasting model based on the partial daily pattern forecasting (PDPP) framework. By accurately predicting daily patterns within this framework, the performance of the forecasting model is further enhanced. In Reference [18], Ajith Gopi et al. used three data-based artificial intelligence (AI) technologies, namely adaptive neural fuzzy inference system (ANFIS), response surface method (RSM), and artificial neural network (ANN), to develop a prediction model to predict the annual power generation and performance ratio (PR) of installed PV systems. The results indicate that ANFIS is the most accurate performance ratio prediction model and will become a valuable tool for policy makers, solar researchers, and solar farm developers. In Reference [19], M. Talaat et al. used a hybrid model of an artificial neural network (ANN) and multiverse optimization (MVO)/genetic algorithm (GA) to predict PV output power, efficiency, and battery temperature. In addition, the relationship between the efficiency of PV panels and battery temperature was also studied, and the results showed that the efficiency and accuracy of PV prediction were significantly improved. To summarize, the utilization of machine learning algorithms has become a prominent research focus in the domain of PV power prediction. Researchers are actively exploring various methodologies to improve accuracy and reliability in this field. However, the feature data that is currently used in research often have problems, such as large data volume and invalid data; high levels of data noise due to measurement errors or sensor drift; nonlinear relationships between PV power and reliability is thallenging to establish forecasting models; and unclear feature relationships between data [20]. In the face of complex influencing factors and numerous training samples, the forecasting model's robustness, stability, and forecasting accuracy are poor, and there are problems such as overfitting and difficulty in determining the structure, which significantly affects the accuracy of PV power generation forecasting [21].

The field of PV power prediction has witnessed the emergence of numerous forecasting algorithms thanks to the rapid development of artificial intelligence. Machine learning algorithms, in particular, offer advantages such as high forecasting accuracy, algorithm adaptability, and scalability. However, they also have shortcomings, including poor antiinterference and sensitivity to algorithm and parameter selections [22]. Deep learning algorithms, on the other hand, possess excellent accuracy and generalization abilities. However, they require better interpretability, longer training times, and actual data and computational resources [23]. In recent years, the application of the stacking algorithm for forecasting has received significant attention. The stacking algorithm leverages the strengths of multiple models, enhancing their generalization abilities. It offers several advantages, such as strong interpretability, high algorithm stability, and accuracy [24]. As a result, local and international experts have extensively researched and applied the stacking algorithm in the context of PV power prediction [25–27]. In Reference [28], Hongchao Zhang et al. proposed multiple stacking models to predict PV power generation using two datasets. The results demonstrate that the stacking models outperform single models in terms of forecasting accuracy. In Reference [29], Elizabeth Michael et al. introduced a hybrid short-term solar irradiance forecasting model that combines a convolutional neural network (CNN) with stacked LSTM. This model significantly enhances the accuracy of solar irradiance forecasting. In Reference [30], Xifeng Guo et al. proposed a stacking ensemble learning method for PV power generation forecasting. The model is trained iteratively using data from the data collection system (DCS). The results indicate that the model achieves high forecasting accuracy and contributes to power grid stability. In Reference [31], Abdallah Abdellatif et al. employed three machine learning models, namely random forest regression (RFR), XGBoost, and adaptive boost (AdaBoost), to construct a stacking ensemble model. The findings indicate that this approach enhances the accuracy of PV power generation prediction. In Reference [32], Waqas Khan et al. presented a stacking model that combines artificial neural networks (ANNs), LSTM, and XGBoost. This model aims to mitigate risks associated with uncertainty in individual models and contributes to the stability of PV power generation predictions. In summary, the majority of research conducted by both domestic and foreign scholars has primarily focused on the composition and parameter adjustment of stacking ensemble algorithms. While this approach has yielded improvements in algorithm performance, it may limit the algorithm's scalability and hinder its optimization. However, there is a need for further research on the principles and structure of the stacking algorithm. With the ongoing development of machine learning and other related technologies, it is crucial to prioritize the innovation and optimization of the stacking algorithm's structure. This can facilitate better performance optimization, enhanced reliability, interpretability, and scalability of the model [33,34]. From the above literature, we can put forward the following research hypotheses that optimizing PV feature data can improve the quality and correlation of input data to the forecasting model, that ensemble models can increase the stability and robustness of the PV power forecasting

model, and that improving models can effectively handle the time-varying volatility and randomness of PV power forecasting caused by complex and variable environmental and other factors.

To address the time-varying volatility and randomness of PV power forecasting due to complex and variable environmental factors, this paper proposes a new method to optimize PV feature data to ensure high-quality and correlated input data for the model. This paper also adopts integrated models to improve the prediction of PV power generation and improve existing fusion models, which can further improve the stability and robustness of the prediction model. This paper proposes a PV power forecasting method based on XGBoost–SFS and the double nested stacking ensemble model, and the proposed method is validated using a PV power station dataset from 2019. The results show that the proposed method can make important features play a more significant role in complex and everchanging environments and has better vital generalization ability and better forecasting accuracy than the comparative model. Therefore, the innovative contributions of this paper can be summarized as follows:

- Two robust features, GHI and similar day power, are constructed based on 14 types of features, such as temperature, azimuth, and ground pressure. This enriches feature samples and enhances the expressiveness of the data, facilitating a more accurate and reliable PV power forecast;
- XGBoost–SFS is constructed to filter out influential features in a complex and variable environment, reduce the impacts of redundant features on the forecasting accuracy and the model's computation, and improve computational efficiency;
- A DNS ensemble model is proposed; the metamodel of the basic stacking model is used to build another stacking model; the actual PV power station data verifies the high forecasting accuracy and stability of the model.

This paper is organized as follows: Section 1 presents the introduction and emphasizes the importance of accurate forecasting of power generation; Section 2 performs feature construction using a dataset of known data features and provides a detailed overview of XGBoost–SFS; Section 3 describes the DNS ensemble model proposed in this paper and the forecasting process based on XGBoost–SFS and DNS ensemble models; Section 4 compares and analyzes the experimental data and the research results are demonstrated; Section 5 compares and analyzes the results of the research methods proposed in different literature; Section 6 summarizes this paper's main conclusions and discusses possible future research; Section 7 gives some recommendations based on the research in this paper.

2. XGBoost-SFS Combined Feature Search Model

2.1. Feature Construction

In this paper, data were collected from a PV power station in 2019, and 14 features, such as temperature, azimuth, relative humidity, and ground pressure, were included. The period of meteorological data is 15 min, and the time resolution of historical PV power is 15 min. The collected features affecting PV power generation are shown in Table 1.

Table 1. Features affecting PV power generation.

No.	Name	No.	Name
1	Temperature	8	Snowfall depth
2	Azimuth	9	Ground pressure
3	Dew point temperature	10	Zenith angle
4	Diffuse horizontal irradiance (DHI)	11	Fixed tilt irradiance (FTI)
5	Direct normal irradiance (DNI)	12	Tracking tilt irradiance (TTI)
6	Atmospheric precipitable water	13	Height 10m wind direction
7	Relative humidity	14	Height 10m wind speed

GHI is a crucial factor in PV (PV) power generation as it directly impacts the overall power output of solar panels. Generally, a higher GHI leads to increased power generation, while a lower GHI results in a lower power generation capacity. Moreover, fluctuations in GHI can significantly influence the performance of a PV power system. As GHI decreases, the temperature of the PV panel rises, subsequently reducing its efficiency and resulting in a decline in power generation [35]. Consequently, GHI serves as a critical indicator for assessing the power generation capabilities of PV power stations. GHI can be defined as follows:

$$GHI = DHI + (DNI \times COS\theta), \tag{1}$$

where *GHI* is the global horizontal irradiance; *DHI* is the diffuse horizontal irradiance; *DNI* is the direct normal irradiance; and θ is the zenith angle. After multiple data analyses, it was finally observed that there is a strong positive correlation between GHI and PV power, as shown in Figure 1. In the context of dynamic solar radiation conditions, incorporating historical data of past radiation and power generation can enhance the accuracy of models and provide more reliable prediction results.



Figure 1. Relationship between GHI and PV power.

2. Similar Day Power

Indeed, it is challenging to eliminate the influence of various variables when relying solely on historical data for making accurate forecasts for PV power generation, as weather conditions, seasons, and time of day all play significant roles. However, incorporating the concept of similar day power, derived from matching historical data with weather data, can significantly enhance the accuracy of PV power forecasts [36]. Figure 2 illustrates the cyclical and strong correlation between similar day power and PV power. By constructing similar day power and mining the power variation patterns associated with similar weather conditions, seasons, and times from historical data, forecasting accuracy can be significantly improved.

2.2. XGBoost

XGBoost is a high-performance machine learning model based on gradient-boosting decision trees. The main idea is to train a series of individual decision tree models using multiple iterations and then combine the results of these models to carry out forecasting. In each iteration round, the model prioritizes samples with high error rates and adjusts the contribution of each underlying decision tree model by weighting it to gradually improve

its performance at each step of the model [37]. For the PV power forecasting problem, the XGBoost base regression tree model can be formulated as follows [38]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k(x_i) \in F,$$
(2)

where x_i is the *i*th sample input value; \hat{y}_i is the *i*th sample predicted value; *K* is the number of trees; f_k is a function in the set *F* of functions; *F* is the set space of all trees; and *k* represents the *k*th tree [39].



Figure 2. Relationship between similar day power and PV power.

The expression of the objective function of XGBoost can be defined as follows:

$$X = \sum_{i=1}^{n} l(y, \hat{y}) + \sum_{k=1}^{K} \Omega(f_k),$$
(3)

where $l(y, \hat{y})$ is the error between the model forecasting result and the actual value; and $\Omega(f_k)$ is the regularization term that controls the complexity of the model.

The overfitting phenomenon during model training is reduced by adding a penalty term to the regularization function as follows:

$$\Omega(f_k) = \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T \omega_j^2, \tag{4}$$

where λ is the fraction of control leaf nodes; *T* is the number of leaf nodes; γ is the number of control leaf nodes; and ω_i is the fraction of the *j*th leaf node.

2.3. SFS

Feature selection plays a crucial role in the process of feature engineering. Its objective is to identify the most relevant subset of features for a given problem. By removing irrelevant or redundant features, feature selection can reduce the cost of model training, improve model performance, enhance model accuracy, and reduce runtime [40].

In generating feature subsets, the search's starting point can be divided into three categories—forward search, backward search, and a random selection—while according to the search strategy, it can be divided into three forms: complete search, sequence search, and random search [41]. According to the time complexity and practicality of the model, this paper chooses to use the SFS model, which is a feature search model based on the

greedy model. Its basic idea is to start from an empty feature subset and find and add one optimal feature at a time until the SFS feature subset is a feature search model with low time complexity and high interpretability [42]. The specific implementation of SFS is shown in Figure 3:

- 1. The search's starting point is feature set Y's empty set;
- 2. The importance of the original features is sorted in a non-increasing order with a specific rule;
- 3. The *i*th feature is added to Y to form a new feature subset Y_i ;
- 4. Feature subset Y_i is evaluated, and the feature subset is determined according to the stop criterion; if it is not determined, features are added to update the feature subset Y_i ; if it is determined, the process of adding features is stopped, and the optimal feature subset is outputted.



Figure 3. SFS implementation flowchart.

2.4. XGBoost-SFS

In machine learning, the frequently utilized data encounter issues such as highdimensional features and redundant attributes, which subsequently impact the effectiveness of model training. To address this concern, feature selection is employed with the goal of identifying the most pertinent and valuable features from the initial dataset. By carrying this step out, the performance of machine learning or statistical models can be enhanced. Consequently, feature selection plays a pivotal role in the research process of feature engineering. Its primary objective is to pinpoint features with significant predictive power for the target variable, allowing the model to achieve superior generalization and more accurate predictions [43].

Traditional feature selection methods focus on analyzing the correlation between individual features and the target variables. However, these methods often need to pay more attention to the interaction between different features during the model's training. As a result, in certain cases, these methods may yield unreliable feature selection outcomes. Furthermore, they fail to capture the effects of feature combinations on log loss, resulting in information loss and a subsequent decline in model performance. Consequently, there is room for improvement in achieving more effective dimensionality reduction results [44].

XGBoost–SFS consists of two modules, XGBoost and SFS, and this paper utilizes the XGBoost–SFS combination method to select the features from the original dataset to obtain the optimal feature subset. The XGBoost–SFS combination model's flowchart is shown in Figure 4, and the specific process is described as follows:



Figure 4. XGBoost–SFS flowchart.

- 1. The original dataset often contains duplicate, missing, irrelevant, or abnormal data, which can significantly impact the outcomes of model training. Therefore, performing data preprocessing operations on the model dataset is essential to address these issues effectively;
- 2. The feature selection process begins by training an XGBoost model using the preprocessed dataset. The importance of each feature is then determined based on the gain of

the model's structural score. Subsequently, the features are sorted in a non-increasing order with respect to their importance. The SFS method is utilized to select features until the optimal feature subset is obtained iteratively;

3. The feature search process implemented by SFS is as follows: Initially, the null set of features from the original data is used as the input for the model. The SFS Model is invoked to generate a new feature subset at each iteration. With each newly generated feature subset, an evaluation criterion, such as XGBoost's root-mean-square error, is utilized to assess its performance. If the generated feature subset satisfies the minimum root-mean-square error stop criterion, the optimal feature subset is outputted. If the criterion is not met, the above feature search process is repeated until a feature subset satisfying the stop criterion is attained [45].

3. DNS Ensemble Model

Stacking is an excellent ensemble model that combines multiple machine models to capture complex semantic information in data more comprehensively, significantly reducing model bias and improving model performance and generalization by leveraging the strengths of different models. The overall framework of stacking is shown in Figure 5 and consists of two layers of models. The first layer is the base model, which extracts features from the training data and constructs new datasets that are strongly correlated with the original inputs. The second layer is the metamodel, which is used to integrate the output of the base model [46,47].



Figure 5. Stacking ensemble model framework.

Nevertheless, the traditional stacking approach can encounter challenges due to variations in assumptions regarding the input data distribution caused by the utilization of different underlying models. This inconsistency in data distribution can result in the subpar performance of the metamodel and the inability to capture important interaction features. Additionally, the simplistic structure of traditional stacking limits its applicability range and flexibility. To address these limitations, this paper introduces an improved DNS ensemble model. The DNS ensemble model aims to enhance the performance and flexibility of the traditional stacking model. The model process of the DNS ensemble model using 5-fold cross-validation is shown in Figure 6. The DNS ensemble model constructs a metamodel of the traditional stacking model into a new stacking model, which extracts deeper and more relevant features with respect to the results and can more comprehensively explore higher-order interaction relations in the feature space, thus improving the discriminative power and robustness of the model; in addition, the flexibility and applicability of the model can be increased by appropriately adjusting the structure.



Figure 6. DNS ensemble model flowchart.

In this paper, we selected the XGBoost–SFS and DNS ensemble model to establish the PV power forecasting method, and the method's flowchart is shown in Figure 7. The main steps of the forecasting method are as follows:

- 1. Original data are collected, including historical power generation data and feature data;
- 2. Relevant feature data are analyzed, and PV-power-related features are constructed;
- 3. The feature dataset is preprocessed, and this includes outlier detection, missing value filling, and data normalization;
- 4. The optimal subset of features affecting PV power generation is filtered based on XGBoost–SFS;
- 5. The optimal subset of features is fed into the DNS ensemble model for forecasting;



6. Forecasting results are outputted, and the results are compared and analyzed under multiple conditions.

Figure 7. Flowchart of forecasting based on XGBoost–SFS and DNS ensemble model.

4. Case Results and Analysis

4.1. Feature Engineering

In this paper, a PV power forecasting model based on XGBoost–SFS and DNS ensemble model is established for the aforementioned historical PV power data collected at a PV power station from 1 August to 31 August 2019, as well as 14 features, such as temperature, azimuth, relative humidity, ground pressure, and 2 features constructed for a total of 16 features. The mean absolute error (MAE) and root-mean-square error (RMSE) are selected as the performance evaluation indexes, as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|,$$
(5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2},$$
(6)

where *n* is the total number of data; and \hat{y}_i and y_i denote the predicted and actual values of the *i*th PV power generation, respectively.

4.1.1. Data Preprocessing

The original feature dataset has noisy data, such as outliers and missing values, so it is necessary to perform data preprocessing operations on the original dataset, including three parts: outlier detection, missing value filling, and data normalization:

- Outlier detection: Outliers refer to data points that deviate significantly from the typical sample, and this deviation may be due to measurement errors, input errors, or other reasons. The 3σ principle is considered a common method for determining outliers, which is based on the assumption of normal distribution and statistically treats data points beyond the three times the standard deviation of the average as outliers; by this method, potential outliers can be screened and eliminated;
- 2. Missing value filling: Missing value filling refers to the filling of missing values in the dataset and completing these missing data points by using certain processing

methods without affecting the overall distribution and accuracy of the data. The K nearest neighbors (KNN) model can be applied to the missing value filling task, and the basic idea is to predict the value of a new sample by finding the values of existing k samples that are most similar to the new sample;

3. Data normalization: Data normalization is a data processing method used to reduce different data scales to the same range so that different variables can represent the optimal data variables to better fit the requirements of most machine models, making the model more accurate, exhibiting rapid convergence, and improving the predictive capability of the model. The data normalization formula can be defined as follows:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}},\tag{7}$$

where y_i is the normalized data; x_i is the original data; x_{min} is the original data minimum; and x_{max} is the original data maximum.

4.1.2. Feature Optimization

According to the XGBoost–SFS method described earlier, the features are optimized. Initially, an XGBoost forecasting model is established, which calculates the importance of each feature and ranks them accordingly. Figure 8 depicts the scores of each feature, reflecting their respective impacts on PV power forecasting.



Figure 8. The score of each PV power features.

Features like GHI and similar daily power demonstrate significant influence, while relative humidity, dew point temperature, and snowfall depth have relatively smaller impacts, with the latter features being negligible. Next, the SFS algorithm is employed to search for different combinations of features by sorting them in a non-increasing order. The optimal feature subset is determined based on the root-mean-square error (RMSE) of the XGBoost model. A smaller RMSE indicates a higher forecasting accuracy for the corresponding feature combination. Figure 9 showcases the relationship between the number of features and the RMSE. It is observed that when six features are selected (GHI, similar day power, FTI, TTI, DHI, and zenith angle), the RMSE is minimized, indicating that this combination of features represents the optimal feature subset for the forecasting model.



Figure 9. Relationship between the number of features and RMSE.

4.2. PV Power Forecasting

4.2.1. Model Design

The data are used as the training set with historical PV power data and 16 related features obtained from a PV power station from 1 August to 28 August 2019, for a total of 2688 datapoints, and a test set with data related to PV power forecasting for the next three days from 29 August to 31 August. In the stacking model, the selection of suitable ensemble models can effectively improve robustness and forecasting accuracy. The forecasting results of SVR, KNN, random forest (RF), GBDT, and XGBoost are compared and analyzed below, and shown in Figure 10 and Table 2.



Figure 10. PV power forecasting results of each model.

Stacking can combine the forecasting results of different models, thus using each model to observe data from different data spaces and structures in order to improve accuracy and stability. Based on the above forecasting results comparison, this paper selects XGBoost, GBDT, and SVR with higher forecasting accuracy as the base models of each layer of the DNS ensemble model to achieve multifaceted feature extraction, and it selects the SVR, a model with stronger robustness and generalization abilities, as the final metamodel for the DNS ensemble model to further avoid overfitting.

Model	Hyperparameters	RMSE	MAE
SVR	kernel: RBF, c: 100, gamma: 10^{-4}	0.298	0.174
KNN	neighbors: 4	1.341	0.649
RF	max_depth: 3, min_samples_leaf: 2, n_estimators: 150	0.505	0.277
GBDT	learning_rate: 0.05, max_depth: 3, min_samples_split: 2, n_estimators: 300	0.343	0.199
XGBoost	gamma: 0.01, learning_rate: 0.02, max_depth: 5, min_child_weight: 2, n_estimators: 500, subsample: 0.5	0.266	0.138

Table 2. Hyperparameters and PV power forecasting errors of each model.

4.2.2. Feature Optimization Results Analysis

Figure 11 compares the forecasting results of the DNS ensemble model and the actual PV power generation data with feature optimization and without feature optimization. The model without feature optimization inputs all features, and the model with feature optimization inputs six features with the highest feature contributions in Figure 8. Table 3 and Figure 11 show that the feature-optimized model has fewer forecasting errors and higher accuracies than the feature-unoptimized model with all features considered. This indicates that in machine learning, too many features may lead to the problem of overfitting, with excellent performance for training data but poor performance for new datasets. This is because the model focuses excessively on noise and random errors in the training data while ignoring the trends and patterns in the actual data.



Figure 11. Comparison of feature optimization results.

 Table 3. Feature optimization results.

Model	RMSE	MAE
Feature Optimization Model	0.212	0.109
Feature Unoptimized Model	0.248	0.134

To rectify this issue and enhance the performance of forecasting models, it is essential to prioritize feature optimization and to carefully select a subset of features that yields the most significant contribution.

4.2.3. Model Forecasting Results Analysis

To comprehensively evaluate the forecasting performance and rationality of the PV power forecasting model based on the XGBoost–SFS and DNS ensemble model, the model is compared with both single forecasting models and the traditional stacking forecasting model. Given the distinct patterns of PV power generation throughout various seasons and months, representative monthly PV power generation data is specifically selected to validate the model's feasibility. Figure 12 visually presents the PV power forecasting outcomes for the respective test sets in March, June, September, and December. Additionally, Table 4 provides a comparison of the forecasting errors associated with each model.



Figure 12. Comparison of forecasting results in different months: (**a**) PV power forecasting results for each model in March; (**b**) PV power forecasting results of each model in June; (**c**) PV power forecasting results of each model in December.

Figure 12 illustrates the distinctive peak-shaped power forecasting curves, indicating the different levels of power generation for each month. Higher power generation levels were observed in June and September compared to March and December. The fluctuations in September were more pronounced due to the varying natural influences and unstable light radiation. Importantly, the DNS ensemble model consistently outperformed the traditional stacking and single models in monthly load forecasting. The results demonstrated that the proposed model achieved increased stability in volatile points, such as peaks and troughs, and was closely aligned with the changing trends in the actual curve, indicating a higher degree of fit.

Month	Model	RMSE	MAE
	DNS ensemble model	0.274	0.173
	Traditional Stacking model	0.287	0.179
March	SVR	0.294	0.189
	GBDT	0.349	0.196
	XGBoost	0.298	0.181
	DNS ensemble model	0.267	0.153
	Traditional Stacking Model	0.286	0.158
June	SVR	0.304	0.171
	GBDT	0.337	0.176
	XGBoost	0.296	0.161
	DNS ensemble model	0.223	0.134
	Traditional Stacking model	0.253	0.151
September	SVR	0.288	0.162
	GBDT	0.314	0.158
	XGBoost	0.266	0.153
	DNS ensemble model	0.361	0.191
	Traditional Stacking model	0.395	0.204
December	SVR	0.438	0.244
	GBDT	0.419	0.217
	XGBoost	0.401	0.214

 Table 4. Comparison of forecasting errors in different months.

Table 4 quantitatively confirms the superior performance of the DNS ensemble model compared to other comparable models. The model achieved lower MAE and RMSE values, indicating higher forecasting accuracy. These findings underscore the effectiveness and robustness of the DNS ensemble model in PV power generation forecasting.

Overall, the PV power forecasting model based on the XGBoost–SFS and DNS ensemble model demonstrated its potential for reliable and accurate future predictions. The model exhibited superior performance compared to other models, achieving higher accuracy and stability in capturing the distinct patterns of PV power generation throughout different months and seasons. These results contribute to the overall understanding and evaluation of the proposed forecasting approach and highlight its significance in supporting efficient and effective PV power management.

4.2.4. Special Weather Forecasting Results Analysis

To assess the forecasting performance and stability of the PV power forecasting model, which is based on the XGBoost–SFS and DNS ensemble model, under complex and variable weather conditions, the PV power data for rainfall and snowfall are specifically considered. These data are then compared to those obtained from single forecasting models and the traditional stacking forecasting model, aiming to provide a comprehensive evaluation. Figure 13 presents the PV power forecasting results in the presence of rainfall and snowfall. These results serve to highlight the model's ability to handle and adapt to adverse weather conditions. Moreover, Table 5 provides a quantitative comparison of the forecasting errors for each model, offering insights into their respective accuracies.



Figure 13. Special weather forecasting results of different models: (**a**) rainfall weather forecasting results of different models; (**b**) snowfall weather forecasting results of different models.

Weather	Model	RMSE	MAE
	DNS ensemble model	0.276	0.166
	Traditional Stacking model	0.311	0.185
Rainfall	SVR	0.324	0.192
	GBDT	0.359	0.196
	XGBoost	0.318	0.189
	DNS ensemble model	0.316	0.195
	Traditional Stacking model	0.326	0.202
Snowfall	SVR	0.338	0.208
	GBDT	0.364	0.221
	XGBoost	0.334	0.203

Table 5. Comparison of forecasting errors of special weather.

Figure 13a compares the forecasting results and real power of each model under rainfall weather conditions with the corresponding forecasting errors presented in Table 5. During rainfall, the initial day experiences a decrease in PV power generation due to high rainfall and thick cloud cover, resulting in decreased light intensity. In subsequent days, as rainfall decreases, PV power increases. However, the instability of natural factors in the rainfall weather contributes to large fluctuations in PV power generation.

The forecasting results of each model under rainfall conditions can be seen in Figure 13a. It is evident that the DNS ensemble model demonstrates more stable and accurate forecasting results compared to other comparable models. The DNS ensemble model closely aligns with the trend of the real power curve, especially at points displaying significant fluctuations or at peak and trough turning points. This highlights the model's ability to capture the complex and variable nature of rainfall conditions and provide reliable predictions.

Table 5 presents the forecasting errors for each model under rainfall conditions. The DNS ensemble model exhibits lower MAE and RMSE values compared to other comparative models. This indicates that the DNS ensemble model has superior prediction accuracy and a more stable prediction ability. The lower forecasting errors further demonstrate the effectiveness of the model for accurately forecasting PV power generation during rainfall weather conditions.

Moving on to Figure 13b, it compares the forecasting results and the real power of each model under snowfall weather conditions, with the associated forecasting error displayed

in Table 5. Snowfall weather introduces complex and variable natural factors that weaken light intensity. Additionally, snow accumulation on PV modules hinders their normal functioning, resulting in lower and more volatile PV power generation.

Similar to rainfall conditions, the DNS ensemble model outperforms other models in accurately predicting PV power generation under snowfall conditions. The DNS ensemble model closely follows the trend of the real power curve, especially during significant fluctuations or at peak and trough turning points. This indicates the model's ability to adapt to the complexities of snowfall weather and deliver reliable forecasting outcomes.

Table 5 also demonstrates that the DNS ensemble model has lower MAE and RMSE values compared to other models under snowfall conditions. This further emphasizes the model's superior prediction accuracy and stability. The lower forecasting errors suggest that the DNS ensemble model is well-suited for complex and variable weather conditions, including both rainfall and snowfall.

In conclusion, Figure 13 and Table 5 provide an insightful analysis of the forecasting results and errors for each model under rainfall and snowfall weather conditions. The DNS ensemble model consistently exhibits more stable and accurate forecasting results compared to other models. Its ability to closely align with the real power curve, especially during significant fluctuations or at peak and trough turning points, demonstrates its effectiveness in capturing the complexities of rainfall and snowfall weather. With lower MAE and RMSE values, the DNS ensemble model showcases superior prediction accuracy and stability, making it a robust choice for forecasting PV power generation in various weather conditions.

5. Discussions

In this section, the method proposed in this paper is first analyzed according to the above study with the research methods from the other literature. In addition, the results of the proposed model are compared with those from the other literature to highlight the superiority of the proposed model.

5.1. Research Methods Analysis of PV Power Literature

A summary of the existing methods for PV power forecasting is shown in Table 6. In References [48,49], Dazhi Yang et al. used the time series method to master the trend and rule of historical PV data through statistical analysis, fully consider the randomness of accidental factors, and then process the data appropriately, to predict the power generation at a certain time in the future. Common time series forecasting methods include the trend forecasting method, moving average method, exponential smoothing method, etc. The time series method requires less historical data and faster prediction speed, but because it fails to take into account external factors such as weather, economy, and society the prediction accuracy is not high.

Ref. No	Methods	Advantages	Disadvantages
[48,49]	Time series method	less historical data and faster prediction speed	less external factors and low accuracy
[50-52]	Regression analysis method	simple to model and fast to calculate	not fit nonlinear data well
[15,16,26,28]	Traditional machine learning methods	flexible and can fully consider external factors	dependence on feature and parameter
[11,17,29]	Neural network method	handle large amounts of data and complex patterns	a lot of data and weak interpretation

Table 6. Comparison of PV power forecasting research methods.

Ref. No	Methods	Advantages	Disadvantages
[30,31,33,34]	Combined models method	avoid limitations and high performance	overfitting and parameter selection
[53–56]	Transfer learning, big data analysis and others	improve the forecasting effect and mine valuable information and patterns	data quality, overfitting and underfitting risks, data security
proposed method	Double Nested Stacking Ensemble Model	adapt changing factors, stronger stability, and higher forecasting accuracy	high model complexity and high calculation cost

Table 6. Cont.

In References [50–52], Mohamed Abuella et al. used regression analysis to analyze the causal relationships between the predicted variables and the results through mathematical statistics. The model has the advantages of simple structure, easy implementation, and fast calculation speeds, so it is widely used in the field of PV power forecasting. Common methods include multiple linear regression, least square methods, and so on. The prediction method of regression analysis is simple to model and fast to calculate, but it can not fit nonlinear data well; especially when there is a strong correlation between variables, the prediction effect will be worse.

In References [15,16,26,28], VanDeventer et al. used SVM, RF, and other traditional machine learning methods to obtain the patterns and rules of PV power generation from historical data. Compared with the above two methods, traditional machine learning methods are more flexible and can fully consider the influence of multiple external factors on PV power changes. However, it also has some limitations such as the dependence on feature selection and model parameter tuning.

In References [11,17,29], Honglu Zhu et al. used deep learning neural network methods to automatically learn and extract features from historical PV data for higher prediction accuracy. Compared to traditional machine learning methods, deep learning neural network methods can handle large amounts of data and complex patterns. However, this method requires a lot of data and computational resources to train and adjust the model, and the interpretation of the model is relatively weak. Therefore, when applying deep learning neural network methods, trade-offs and choices need to be made according to the actual needs of the problem and the available resources.

In References [30,31,33,34], Xifeng Guo et al. used a combined model to flexibly utilize the advantages of multiple algorithms and avoid the limitations of a single model, thus achieving high prediction performance. However, this may lead to a series of risks, such as overfitting and difficult parameter selection.

In References [53–56], Yugui Tang et al. used transfer learning, big data analysis, and other methods to forecast the power generation combined with the model. They can use the knowledge and features learned from the source domain to improve the forecasting effects of the target domain and can mine valuable information and patterns from large-scale data to provide support for decision-making. However, these methods still have some problems in practice, such as data quality, overfitting and underfitting risks, data privacy, and security.

In this paper, we proposed a new PV power forecasting method based on eXtreme gradient boosting–sequential forward selection and a double nested stacking ensemble model by combining data processing, feature engineering, and model fusion optimization. The proposed method can effectively adapt to changing environmental factors, with stronger stability, generalization ability, and higher forecasting accuracy. However, the method faces the disadvantages of high model complexity and high calculation costs. These shortcomings can be solved by simplifying the model and developing distributed computing methods.

5.2. Comparative Studies

Table 7 shows the performance comparison between the proposed model and the existing PV power forecasting models.

Ref. No	Methods	RMSE	MAE
[49]	Time series method	1.781	-
[52]	Regression analysis method	8.345	9.230
[15]	Traditional machine learning methods	11.226	-
[11]		7.193	3.639
[57]	 Neural network method	19.78	-
[58]	_	45.11	-
[28]		47.78	-
[31]	Combined models method	13.95	8.79
[59]	_	51.35	-
[54]	Transfer learning, big data analysis and others	18.04	-
proposed method	Double Nested Stacking Ensemble Model	0.212	0.109

Table 7. Comparison of different PV power forecasting models.

The model proposed in this paper has the lowest RMSE (0.212). In addition, the MAE value of the proposed model is also the smallest, 0.109. Compared with the model proposed in this paper, the RMSE and MAE of other models are higher. The RMSE and MAE of the method used in Reference [11] are 7.193 and 3.639, while the RMSE and MAE of the model used in Reference [31] are 13.95 and 8.79, respectively. In addition, Reference [52] has poor performance compared with the proposed model, with RMSE values of 8.345 and 9.230, respectively. According to the results in Table 7, the MAE and RMSE of the proposed model are both the best, and the proposed model has the best forecasting effect compared with other comparison models and shows smaller errors than other models in the literature. Therefore, according to the comparison results, the model proposed in this study is recommended for the prediction of PV power forecasting.

6. Conclusions

In recent years, the PV power generation industry has achieved leapfrog developments by virtue of its own clean, environmental protection, and pollution-free advantages, as well as the strong support of relevant national policies. However, a variety of complex environmental factors such as sunlight intensity and ambient temperature will affect the output power of solar power generation, which makes PV power generation output show strong indirectness and time volatility. Therefore, it is of great significance to study how to effectively improve the forecasting accuracy of PV power output for component scheduling and power management of PV power grid systems. Therefore, this paper proposes a PV power forecasting method that leverages XGBoost–SFS and the DNS ensemble model, and its effectiveness is demonstrated using practical examples. The main conclusions can be summarized as follows:

- The XGBoost–SFS method successfully establishes a feature determination approach for PV power forecasting by considering how multiple features impact accuracy and efficiency. It can accurately evaluate the contribution and relevance of each feature variable during the forecasting process;
- 2. The DNS ensemble model is employed to extract more representative features of PV power generation and to explore higher-order feature interactions. By observing

the data space from different perspectives, the model enhances forecasting accuracy and stability;

3. The analysis indicates that the RMSE of the PV power forecasting method based on XGBoost–SFS and the DNS ensemble model is 0.274, 0.267, 0.223, 0.361, 0.276, and 0.316 in different seasons and weather, and the MAE is 0.173, 0.153, 0.134, 0.191, 0.166, and 0.195 in different seasons and weather. All evaluation indicators are superior to the comparative model. The model outperforms traditional stacking, XGBoost, SVR, and other methods. It adapts to complex and variable forecasting environments and tackles issues like overfitting and poor interpretability. Consequently, it significantly improves forecasting accuracy and generalization capabilities.

PV power forecasting research plays an important role in the field of future sustainability. It can provide accurate predictions of the capacity, energy output, and efficiency of photovoltaic power generation systems to help optimize the operation and planning of photovoltaic power plants, thereby improving energy efficiency and economic efficiency. This will contribute to the achievement of the sustainable development goals, promote the increase in clean energy, reduce the dependence on traditional energy sources, and achieve a low-carbon and environmentally friendly energy transition. In the future development of PV power forecasting research, interdisciplinary cooperation will be very important. Combining knowledge and technology in fields such as meteorology, power systems, and data science, collaborative research and integrated innovation will further promote the development and application of PV power forecasting, contributing to the popularization and sustainable development of clean energy. In future studies, exploring combinations of transfer learning or establishing a distributed computing environment is suggested. This can effectively reduce the time complexity of the algorithm, make the training data requirements more manageable, and enhance the model's robustness.

7. Recommendations

PV power generation has many advantages, such as environmental protection, renewable, economically viable, distributed, long life and multi-function, which make PV power generation an important part of the global energy transition. With the advancement of technology, the development prospects of PV power generation are brighter and expected to become the dominant energy supply mode in the future. With large-scale PV access to the power grid, the instability brought by it will have a huge impact on the safe and stable operation of the power grid. PV power forecasting technology has become the key to improving the quality of PV grid connection, optimizing the dispatch to the power grid, and ensuring the safe and stable operation of the power grid. Therefore, it is very important to establish an accurate PV power forecasting system. This paper proposes the XGBoost–SFS and double nested stacking ensemble model to improve the stability and accuracy of the forecasting. The following recommendations can be drawn from this study:

- 1. The proposed model can analyze the uncertainty of the randomness and forecasting error from meteorological data, which is helpful to evaluate the robustness of the model and provide a basis for decision-makers to manage risks in practical applications;
- The proposed model can be applied to other systems, such as a cyber-physical-social system, building energy management system, etc. Through the accurate forecasting of power generation, the scheduling and operation of the system are optimized, and the planning and sustainable development of the system are enhanced;
- 3. The proposed model can adapt to complex environmental factors and predict power generation under various weather conditions, so the model can be extended to other places;
- 4. The proposed model can promote a reliable supply of clean energy for sustainability, enhance the efficiency of grid operations, promote the development of renewable energy, optimize energy use and energy management, and support clean energy finance and investment to further promote economic, social, and environmental sustainability.

Author Contributions: Conceptualization, X.C. and B.Z.; methodology, X.C. and B.Z.; software, X.C. and G.L.; validation, G.L. and P.G.; formal analysis, B.Z. and P.G.; data curation, B.Z.; writing—original draft preparation, X.C.; writing—review and editing, X.C. and J.H.; visualization, B.Y.; supervision, J.H. and B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China, grant number U22B20115, in part by the Applied Fundamental Research Program of Liaoning Province, grant number 2023JH2/101600036, in part by the Science and Technology Projects in Liaoning Province, grant number 2022-MS-110, and in part by the Guangdong Basic and Applied Basic Research Foundation, grant number 2021A1515110778.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Special thanks to the Intelligent Electrical Science and Technology Research Institute, Northeastern University (China), for providing technical support for this research.

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

References

- 1. Xin, B.; Shan, B.; Li, Q.; Yan, H.; Wang, C. Rethinking of the "three elements of energy" toward carbon peak and carbon neutrality. *Proc. CSEE* **2022**, *42*, 3117–3126.
- Fouad, M.; Shihata, L.A.; Morgan, E.I. An integrated review of factors influencing the perfomance of PV panels. *Renew. Sustain.* Energy Rev. 2017, 80, 1499–1511. [CrossRef]
- Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of PV power forecasting. Sol. Energy 2016, 136, 78–111.
- Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Horan, B.; Stojcevski, A. Forecasting of PV power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* 2018, *81*, 912–928. [CrossRef]
- Singla, P.; Duhan, M.; Saroha, S. A comprehensive review and analysis of solar forecasting techniques. *Front. Energy* 2021, 16, 187–223. [CrossRef]
- 6. Sørensen, M.L.; Nystrup, P.; Bjerregård, M.B.; Møller, J.K.; Bacher, P.; Madsen, H. Recent developments in multivariate wind and solar power forecasting. *Wiley Interdiscip. Rev. Energy Environ.* **2023**, *12*, e465. [CrossRef]
- 7. Zheng, J.; Du, J.; Wang, B.; Klemeš, J.J.; Liao, Q.; Liang, Y. A hybrid framework for forecasting power generation of multiple renewable energy sources. *Renew. Sustain. Energy Rev.* 2023, 172, 113046. [CrossRef]
- Tawn, R.; Browell, J. A review of very short-term wind and solar power forecasting. *Renew. Sustain. Energy Rev.* 2022, 153, 111758. [CrossRef]
- 9. Krishnan, N.; Kumar, K.R.; Inda, C.S. How solar radiation forecasting impacts the utilization of solar energy: A critical review. J. Clean. Prod. 2023, 388, 135860. [CrossRef]
- 10. Wu, Y.-K.; Huang, C.-L.; Phan, Q.-T.; Li, Y.-Y. Completed review of various solar power forecasting techniques considering different viewpoints. *Energies* 2022, *15*, 3320. [CrossRef]
- Zhu, H.; Li, X.; Sun, Q.; Nie, L.; Yao, J.; Zhao, G. A power prediction method for PV power plant based on wavelet decomposition and artificial neural networks. *Energies* 2015, 9, 11. [CrossRef]
- 12. Agoua, X.G.; Girard, R.; Kariniotakis, G. Short-term spatio-temporal forecasting of PV power production. *IEEE Trans. Sustain. Energy* **2017**, *9*, 538–546. [CrossRef]
- 13. Li, P.; Zhou, K.; Lu, X.; Yang, S. A hybrid deep learning model for short-term PV power forecasting. *Appl. Energy* **2020**, 259, 114216. [CrossRef]
- 14. Luo, X.; Zhang, D.; Zhu, X. Deep learning based forecasting of PV power generation by incorporating domain knowledge. *Energy* **2021**, 225, 120240. [CrossRef]
- 15. VanDeventer, W.; Jamei, E.; Thirunavukkarasu, G.S.; Seyedmahmoudian, M.; Soon, T.K.; Horan, B.; Mekhilef, S.; Stojcevski, A. Short-term PV power forecasting using hybrid GASVM technique. *Renew. Energy* **2019**, *140*, 367–379. [CrossRef]
- 16. Pan, M.; Li, C.; Gao, R.; Huang, Y.; You, H.; Gu, T.; Qin, F. PV power forecasting based on a support vector machine with improved ant colony optimization. *J. Clean. Prod.* **2020**, *277*, 123948. [CrossRef]
- Wang, F.; Xuan, Z.; Zhen, Z.; Li, K.; Wang, T.; Shi, M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Convers. Manag.* 2020, 212, 112766. [CrossRef]
- 18. Gopi, A.; Sharma, P.; Sudhakar, K.; Ngui, W.K.; Kirpichnikova, I.; Cuce, E. Weather impact on solar farm performance: A comparative analysis of machine learning techniques. *Sustainability* **2022**, *15*, 439. [CrossRef]
- 19. Talaat, M.; Said, T.; Essa, M.A.; Hatata, A. Integrated MFFNN-MVO approach for PV solar power forecasting considering thermal effects and environmental conditions. *Int. J. Electr. Power Energy Syst.* **2022**, *135*, 107570. [CrossRef]

- 20. Abdelmoula, I.A.; Elhamaoui, S.; Elalani, O.; Ghennioui, A.; El Aroussi, M. A PV power prediction approach enhanced by feature engineering and stacked machine learning model. *Energy Rep.* **2022**, *8*, 1288–1300. [CrossRef]
- Başaran, K.; Bozyiğit, F.; Siano, P.; Yıldırım Taşer, P.; Kılınç, D. Systematic literature review of PV output power forecasting. *IET Renew. Power Gener.* 2020, 14, 3961–3973. [CrossRef]
- Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.-L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* 2017, 105, 569–582. [CrossRef]
- Wang, K.; Qi, X.; Liu, H. A comparison of day-ahead PV power forecasting models based on deep learning neural network. *Appl. Energy* 2019, 251, 113315. [CrossRef]
- 24. Ofori-Ntow, E., Jr.; Ziggah, Y.Y.; Rodrigues, M.J.; Relvas, S. A New Long-Term PV Power Forecasting Model Based on Stacking Generalization Methodology. *Nat. Resour. Res.* 2022, *31*, 1265–1287. [CrossRef]
- 25. Huang, H.; Zhu, Q.; Zhu, X.; Zhang, J. An Adaptive, Data-Driven Stacking Ensemble Learning Framework for the Short-Term Forecasting of Renewable Energy Generation. *Energies* **2023**, *16*, 1963. [CrossRef]
- 26. Tan, Z.; Zhang, J.; He, Y.; Zhang, Y.; Xiong, G.; Liu, Y. Short-term load forecasting based on integration of SVR and stacking. *IEEE Access* 2020, *8*, 227719–227728. [CrossRef]
- 27. Zhang, Q.; Wu, J.; Ma, Y.; Li, G.; Ma, J.; Wang, C. Short-term load forecasting method with variational mode decomposition and stacking model fusion. *Sustain. Energy Grids Netw.* **2022**, *30*, 100622. [CrossRef]
- 28. Zhang, H.; Zhu, T. Stacking model for PV-power-generation prediction. Sustainability 2022, 14, 5669. [CrossRef]
- Elizabeth Michael, N.; Mishra, M.; Hasan, S.; Al-Durra, A. Short-term solar power predicting model based on multi-step CNN stacked LSTM technique. *Energies* 2022, 15, 2150. [CrossRef]
- 30. Guo, X.; Gao, Y.; Zheng, D.; Ning, Y.; Zhao, Q. Study on short-term PV power prediction model based on the Stacking ensemble learning. *Energy Rep.* 2020, *6*, 1424–1431. [CrossRef]
- 31. Abdellatif, A.; Mubarak, H.; Ahmad, S.; Ahmed, T.; Shafiullah, G.; Hammoudeh, A.; Abdellatef, H.; Rahman, M.; Gheni, H.M. Forecasting PV power generation with a stacking ensemble model. *Sustainability* **2022**, *14*, 11083. [CrossRef]
- 32. Khan, W.; Walker, S.; Zeiler, W. Improved solar PV energy generation forecast using deep learning-based ensemble stacking approach. *Energy* **2022**, 240, 122812. [CrossRef]
- 33. Shi, J.; Zhang, J. Load forecasting based on multi-model by stacking ensemble learning. Proc. CSEE 2019, 39, 4032–4042.
- 34. Dong, Y.; Zhang, H.; Wang, C.; Zhou, X. Wind power forecasting based on stacking ensemble model, decomposition and intelligent optimization algorithm. *Neurocomputing* **2021**, *462*, 169–184. [CrossRef]
- 35. Mondol, J.D.; Yohanis, Y.G.; Norton, B. Solar radiation modelling for the simulation of PV systems. *Renew. Energy* **2008**, *33*, 1109–1120. [CrossRef]
- 36. Zhou, Y.; Zhou, N.; Gong, L.; Jiang, M. Prediction of PV power output based on similar day analysis, genetic algorithm and extreme learning machine. *Energy* **2020**, *204*, 117894. [CrossRef]
- Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* 2021, 54, 1937–1967. [CrossRef]
- Li, X.; Ma, L.; Chen, P.; Xu, H.; Xing, Q.; Yan, J.; Lu, S.; Fan, H.; Yang, L.; Cheng, Y. Probabilistic solar irradiance forecasting based on XGBoost. *Energy Rep.* 2022, *8*, 1087–1095. [CrossRef]
- Cao, H.; Yang, L.; Li, H.; Wang, K. Net Power Prediction for High Permeability Distributed PV Integration System. J. Phys. Conf. Ser. 2023, 2418, 012069. [CrossRef]
- 40. Venkatesh, B.; Anuradha, J. A review of feature selection and its methods. Cybern. Inf. Technol. 2019, 19, 3–26. [CrossRef]
- Chen, R.; Sun, N.; Chen, X.; Yang, M.; Wu, Q. Supervised feature selection with a stratified feature weighting method. *IEEE Access* 2018, 6, 15087–15098.
- 42. Vandana, C.; Chikkamannur, A.A. Feature selection: An empirical study. Int. J. Eng. Trends Technol. 2021, 69, 165–170.
- 43. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. Comput. Electr. Eng. 2014, 40, 16–28.
- Eseye, A.T.; Lehtonen, M.; Tukia, T.; Uimonen, S.; Millar, R.J. Machine Learning Based Integrated Feature Selection Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems. *IEEE Access* 2019, 7, 91463–91475. [CrossRef]
- 45. Ding, J.; Fu, L. A Hybrid Feature Selection Algorithm Based on Information Gain and Sequential Forward Floating Search. J. Intell. Comput. 2018, 9, 93. [CrossRef]
- Divina, F.; Gilson, A.; Goméz-Vela, F.; García Torres, M.; Torres, J.F. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* 2018, 11, 949. [CrossRef]
- 47. Ribeiro, M.H.D.M.; da Silva, R.G.; Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int. J. Electr. Power Energy Syst.* 2022, *136*, 107712. [CrossRef]
- Yang, D.; Dong, Z. Operational PVs power forecasting using seasonal time series ensemble. Sol. Energy 2018, 166, 529–541. [CrossRef]
- Sharadga, H.; Hajimirza, S.; Balog, R.S. Time series forecasting of solar power generation for large-scale PV plants. *Renew. Energy* 2020, 150, 797–807. [CrossRef]
- Abuella, M.; Chowdhury, B. Solar power probabilistic forecasting by using multiple linear regression analysis. In Proceedings of the SoutheastCon 2015 Conference, Fort Lauderdale, FL, USA, 9–12 April 2015; pp. 1–5.
- 51. Abuella, M.; Chowdhury, B. Solar power forecasting using support vector regression. arXiv 2017, arXiv:1703.09851.

- 52. Sheng, H.; Xiao, J.; Cheng, Y.; Ni, Q.; Wang, S. Short-term solar power forecasting based on weighted Gaussian process regression. *IEEE Trans. Ind. Electron.* 2017, *65*, 300–308. [CrossRef]
- Tang, Y.; Yang, K.; Zhang, S.; Zhang, Z. PV power forecasting: A hybrid deep learning model incorporating transfer learning strategy. *Renew. Sustain. Energy Rev.* 2022, 162, 112473. [CrossRef]
- Zhou, S.; Zhou, L.; Mao, M.; Xi, X. Transfer learning for PV power forecasting with long short-term memory neural network. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 125–132.
- 55. Devaraj, J.; Madurai Elavarasan, R.; Shafiullah, G.; Jamal, T.; Khan, I. A holistic review on energy forecasting using big data and deep learning models. *Int. J. Energy Res.* 2021, *45*, 13489–13530. [CrossRef]
- 56. Galicia, A.; Talavera-Llames, R.; Troncoso, A.; Koprinska, I.; Martínez-Álvarez, F. Multi-step forecasting for big data time series based on ensemble learning. *Knowl. Based Syst.* **2019**, *163*, 830–841. [CrossRef]
- Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Almohaimeed, Z.M.; Muhammad, M.A.; Khairuddin, A.S.M.; Akram, R.; Hussain, M.M. An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants. *Energies* 2022, 15, 2243. [CrossRef]
- 58. Zhen, Z.; Liu, J.; Zhang, Z.; Wang, F.; Chai, H.; Yu, Y.; Lu, X.; Wang, T.; Lin, Y. Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. *IEEE Trans. Ind. Appl.* **2020**, *56*, 3385–3396. [CrossRef]
- 59. Kumari, P.; Toshniwal, D. Extreme gradient boosting and deep neural network based ensemble learning approach to forecast hourly solar irradiance. *J. Clean. Prod.* 2021, 279, 123285.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.