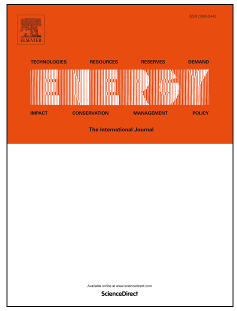
Journal Pre-proof



A Hybrid Ensemble Optimized BiGRU Method for Short-Term Photovoltaic Generation Forecasting

Yeming Dai, Weijie Yu, Mingming Leng

PII: S0360-5442(24)01231-3

DOI: https://doi.org/10.1016/j.energy.2024.131458

Reference: EGY 131458

To appear in: *Energy*

Received Date: 17 November 2023

Revised Date: 6 April 2024

Accepted Date: 26 April 2024

A Hybrid Ensemble Optimized BiGRU Method for Short-Term Photovoltaic Generation Forecasting

Yeming Dai [a*], Weijie Yu [a], Mingming Leng [b]

[a] *School of Business, Qingdao University, Qingdao 200071, China;*

[b] *Faculty of Business, Lingnan University, Hong Kong, China.*

Corresponding author: Yeming Dai, Email: yemingdai@163.com;
Weijie Yu, Email: 2865128697@qq.com;
Mingming Leng, Email: mmleng@ln.edu.hk.

ABSTRACT： In the context of prominent energy crisis, photovoltaic power (PV) generation has received increasing attention, then accurate PV generation forecasting is crucial for ensuring the smooth operation of power stations. However, existing research is insufficient in comprehensively analyzing the impact of PV generation volatility. To fill the gaps and enhance the prediction accuracy, this paper proposes a new hybrid forecasting method. We first introduce the Locally Weighted Scatterplot Smoothing (LOWESS) method to process the data and enhance the data stability, and use Pearson correlation coefficient (PCC) and Random Forests (RF) for feature selection to improve the quality of input data. Then we use Attention mechanism and Convolutional Neural Network (CNN) layer to optimize Bi-directional Gate Recurrent Unit (BiGRU) model and form a new hybrid model. Finally, based on the Bagging algorithm, we use ensemble learning to further optimize the hybrid BiGRU model to enhance the depth and performance. The proposed method is validated through case analysis results from two different locations, Xuhui District in Shanghai, China and the DKASC area in Alice Springs, Australia. The results demonstrate that, compared with other models, the developed method exhibits exceptional prediction performance and effectively enhances the accuracy of PV generation forecasting.
Keywords: photovoltaic power generation; Locally Weighted Scatterplot Smoothing; feature selection; ensemble learning; Bi-directional Gate Recurrent Unit

Nomenclature			
BiGRU	Bi-directional Gate Recurrent Unit	MSE	Mean Square Error
BiLSTM	Bidirectional Long Short-Term Memory	NN	Neural Network
BP	Back Propagation	PCC	Pearson Correlation Coefficient
CNN	Convolutional Neural Network	PV	Photovoltaic Power
DKASC	Desert Knowledge Australia Solar Centre	RF	Random Forests
ELM	Extreme Learning Machine	RMSE	Root Mean Square Error
GRU	Gate Recurrent Unit	SVM	Support Vector Machine
LOWESS	Locally Weighted Scatterplot Smoothing	TCN	Temporal Convolutional Network
LSTM	Long Short-Term Memory	VMD	Variational Mode Decomposition
MAE	Mean Absolute Error	XGBoost	eXtreme Gradient Boosting
MAPE	Mean Absolute Percentage Error		

1. Introduction

With the swift progress of renewable energy production technology and the growing awareness of environmental protection, renewable energy has been developed and utilized by more and more countries in recent years. Amongst various generating technologies and methods of renewable energy, PV stands out as it directly converts solar energy into electrical energy and numerous advantages such as huge energy storage, high flexibility, low developmental cost, and long service life, which making it the most widely used renewable energy [1]. Chinese government has promulgated specific documents such as "two-carbon policy" to promote the development of PV generation for clean and low-carbon transformation [2,3]. Similarly, the Australian government has also initiated several initiatives to promote the utilization of PV energy, such as Solar School Project, Solar Home and Community Program, and National Renewable Energy Target Plan. These initiatives have resulted in over 2 million households installing rooftop solar panels, and the number of installations continues to grow each year [4].

As PV generation is highly dependent on weather conditions, ensuring precise forecasts of PV generation is essential for optimizing the operation efficiency of PV stations. Existing studies have demonstrated that the precision and applicability of PV generation forecasting are heavily determined by the prediction period and used method. Based on the duration of the forecast period, various methods of prediction are categorized into ultra-short-term [5], short-term [6], and medium to long-term [7]. For ultra-short-term prediction, the accuracy can be within seconds to minutes, making it ideal for real-time scheduling of power grids of different scales. This allows for timely reservation of reserve capacity for power grid [5]. Short-term forecasting, on the other hand, covers a range of one hour to one day. This type of forecasting is crucial in economic dispatch and grid decision-making, balancing electricity market transactions, adjusting unit commitments, and optimizing power supply plans [6]. In addition, in terms of medium to long-term forecasting, the forecasting cycle for power grid planning ranges from daily to weekly, monthly, and yearly, so as to provide a long-term plan for equipment maintenance and new energy base station locations [7]. Since short-term load prediction is of great importance in economic dispatch and decision-making of power grid, the downstream electricity demand of PV generation can be meet in time and promote effective communication with the upstream information. Therefore, this paper focuses on the short-term forecasting of PV generation to provide accurate power supply information to the dispatching center, then enhance the security and efficiency of power grid and maintain balance and stability in the PV generation.

At present, the forecasting methods for PV generation [8] can be categorized as continuous methods [9], physical methods [10] and statistical methods. Statistical methods include prediction based on time series [11] and artificial intelligence prediction [12]. Due to low computational requirements, minimum delay, and reasonable accuracy, the continuous method [9] is usually used for ultra-short-term and short-term prediction. However, as the duration increases, it is prone to a substantial decline in the prediction precision.

On the other hand, the physical method [10] is more suitable for long-term predictions, as it takes into account factors such as pressure and topography. Although the accuracy of physical method is high when the weather variables are relatively stable, it may not be able to cope well when the meteorological variables change suddenly. Prediction based on time series [11] can effectively predict future variable changes and identify trends and development rules, but this method highlights time series without considering external factors, which leads to high prediction error. Artificial intelligence prediction [12] is dominated by machine learning algorithms, with RF [13] and Neural Networks (NN) [14] being common examples. Among these, the neural network algorithm is well-suited for different prediction cycle ranges, as it can handle massive amounts of data, remove outliers and improve model accuracy, so as to effectively solve the problems in continuous prediction, physical prediction and statistical prediction [15].

The complete forecasting procedure for PV generation comprises two components: data preprocessing and model forecasting [16], which all affect the prediction accuracy. The core of data processing involves transforming the initial data into an understandable or mineable format with the aim of enhancing the reliability of the input data. However, when collecting data, some problems such as data missing and outliers are often encountered, which will lead to low-quality mining results and reduce the precision of the later prediction model. Data processing methods such as data decomposition [17], smoothing denoising [18] and feature selection [19] can effectively solve the above problems [20]. Due to natural factors such as solar radiation and humidity, PV generation data can be highly volatile and intermittent. Therefore, it is crucial to smooth out the data in research associated with PV generation forecasting, particularly in addressing the problem of significant data fluctuations. Our previous study [13] had found that LOWESS has the best smoothing effect on photovoltaic power data, which can effectively improve the stability of PV generation and enhance predictive accuracy. Therefore, this paper chooses to use LOWESS smoothing method to process data and enhance the precision of PV generation forecasting. Though data smoothing technology can decrease data fluctuations and improve the precision of subsequent prediction models, but the authenticity of original data may be altered to some extent. Therefore, how to enhance the prediction performance under the premise of reducing the error between the smoothing and the real data becomes the key to the research.

Feature selection is of great importance in the process of PV generation forecasting. It can identify the most relevant and most influential features of PV generation, and reduce data noise and processing time, thereby improving prediction accuracy and efficiency. However, different feature selection methods and parameter settings may have different effects on prediction results. Therefore, the applicability and accuracy of feature selection methods need to be fully considered. In order to determine which feature has a significant impact on photovoltaic power generation, we need to use some common feature selection algorithms. Among them, PCC is a frequently employed method, which can measure the degree of correlation between different features and help us evaluate and select the most appropriate feature combination [21]. In addition, RF algorithm, as the basic tool of classification or regression accuracy as the criterion function and using sequence backward selection for feature selection, is also widely used. For example, [22] used RF to assess the significance of each feature category in identifying the topology of distribution network, and realized the effect of feature category screening and dimension reduction. In the field of wind speed prediction, [23] used RF to analyze the decomposed subsequence data, eliminated redundant data, and improved the utilization of data information. In forecasting PV generation, it is also essential to screen out the most effective features of PV generation from many natural factors and socio-economic factors, so as to realize feature space dimension compression, data dimension reduction and simplification. Thus, this paper chooses PCC as an index to measure the correlation degree between PV data, and uses RF to further screen data, retain effective features, and enhance the precision of PV generation forecasting.

After data processing, the subsequent stage involves selecting an appropriate model for forecasting. As one of the typical representatives of neural networks, Gate Recurrent Unit (GRU) shows good prediction performance in PV generation prediction [24], but it may cause information attenuation when transmitted forward. The Bi-directional Gate Recurrent Unit (BiGRU) model [25] effectively solves above existing problem of GRU. BiGRU is composed of two unidirectional GRUs superimposed together, taking into account the past information and future information of prediction point, which further enhances forecasting precision. For instance, in [26], a BiGRU ultra-short-term PV prediction method rooted in self-organizing map clustering and quadratic decomposition was proposed. The findings indicate that BiGRU outperforms GRU in terms of prediction accuracy. A hybrid optimized prediction model for PV generation is developed in [27] by leveraging Variational Mode Decomposition (VMD), CNN, and BiGRU. CNN is employed to discover the inherent connection between the feature matrix and the target variable through analysis. BiGRU network predicted the future values for each sub-mode. Ultimately, the predicted sub-mode value is combined to obtain the ultimate forecast for PV. [28] also proposed a new hybrid deep learning framework that includes one-dimensional convolutional layer, BiGRU, self-attention mechanism and transfer learning. The validity and applicability of BiGRU forecasting model in PV generation prediction are proved by the dataset case of California, USA.

Nevertheless, the unpredictability and sporadic nature of PV generation pose challenges for making single prediction and hybrid optimized prediction models more prone to randomness and instability [29]. Ensemble learning addresses this issue by constructing and combining multiple base learners, and effectively enhance the stability and precision of forecasting by combining a strong learner with stronger learning effect to. For example, [30] constructed a hybrid forecasting model for container throughput using selective deep integration, and verified that the proposed ensemble forecasting model suggested better forecasting precision compared to an individual forecasting model. The study presented in [31] proposed the AdaBoost-GRU ensemble prediction model and compared it with the AdaBoost-LSTM, single LSTM and GRU models. The results indicated that the AdaBoost-GRU model had significantly lower prediction errors than the other three models and yielded the highest prediction performance. In [32], the Bagging algorithm was utilized to integrate the GRU prediction model and develop an ensemble model for stock price forecasting. The conclusions illustrated the effectiveness of the ensemble prediction model in comparison to single GRU, BP and ELM models. In [33], a BiGRU ensemble prediction model based on the Bagging algorithm was proposed for short-term load prediction. The results show that the ensemble algorithm outperforms other algorithms such as BP, LSTM, GRU, and BiGRU. Another study in [34] proposed a short-term wind power forecasting model that used Attention mechanism, GRU and Stacking multi-algorithm fusion to improve the prediction accuracy. It follows that, although the ensemble forecasting model based on GRU has been developed for price forecasting, load forecasting, and wind power forecasting, it is less applied in the forecasting field of PV generation. Therefore, we introduce the ensemble learning to optimize the BiGRU model and apply it to photovoltaic power generation prediction.

In summary, we initially select LOWESS smoothing and Pearson correlation to process the data, and utilize RF for data feature selection to comprehensively enhance the quality of input data. Then, we use Attention mechanism and CNN layer to optimize BiGRU model to form a hybrid BiGRU model. Finally, we further optimize the hybrid BiGRU model by applying the Bagging algorithm and adopting the concept of ensemble learning. The paper primarily focuses on leveraging ensemble learning algorithm to construct a strong and stable prediction model. By implementing a series of optimization techniques, incorporating Attention mechanism，CNN and utilizing ensemble learning, we significantly enhance the overall predictive capability and reliability for proposed prediction method.

The main contributions of the paper can be outlined as follows:

(1) We use LOWESS smoothing technology to process the data, reduce the variability in PV generation data and the predictive mistake.

(2) We employ the PCC and RF method to extract the attributes of environmental elements influencing PV generation, reduce the feature dimension, and enhance the quality of input data.

(3) We optimize the BiGRU model with Attention mechanism and CNN layer to form a hybrid BiGRU model, and stack two BiGRU layers to improve the depth and prediction performance.

(4) We further introduce the ensemble learning to construct a hybrid ensemble optimized BiGRU prediction model, alleviate the randomness and instability of the hybrid optimized BiGRU model, and enhance the forecasting stability and precision of the model.

(5) We collect the PV generation data at the Xuhui District in Shanghai, China and the DKASC in Alice Springs, Australia, and use our proposed model and other several common single prediction models, hybrid optimization prediction models and ensemble optimized prediction models to illustrate the efficiency of our method.

This paper is structured in the following manner: Section 2 presents the basic concepts of LOWESS smoothing, PCC, RF procedure and the initial BiGRU model. Section 3 presents the structure of hybrid optimized BiGRU model with CNN and Attention mechanism that is further optimized by ensemble learning. In Section 4, we present the specific steps for proposed prediction method are presented. Section 5 present our case study, encompassing an overview of the experimental data, the criteria chosen for evaluation, the configuration of model parameters, outcomes of data processing, and an all-encompassing examination of the experimental findings. Section 6 provides an overview of the content presented in this paper.

2. Methodology

This section outlines the fundamental techniques for data processing and forecasting. The data processing techniques employed include LOWESS smoothing for data smoothing, Pearson correlation for correlation analysis, and Random Forests for feature extraction. For prediction, we utilize BiGRU prediction model as the foundation forecasting model.

2.1 LOWESS smoothing

LOWESS smoothing is a non-parametric regression technique that is useful for both prediction and smoothing. This method involves selecting a point *x*, and then fitting a weighted linear regression to the data point that lie within a certain distance of *x*. The weights used in the regression are determined by a weight function, *W*, which is typically a cubic function as shown in formula (1). This process is repeated for all *n* data points, resulting in *n* weighted regression lines [35] so as to make LOWESS adapt to various types of data and exhibits strong capabilities in handling missing values and outliers.

$$W_{(x)} = \begin{cases} (1-|x|)^3, if\,|x|<1; \\ 0, \qquad if\,|x|\geq 1. \end{cases} \tag{1}$$

2.2 Pearson Correlation Coefficient

PCC is commonly adopted to measure the degree of linear correlation between variable X and Y for correlation analysis. Formula (2) provides the calculation for obtaining the sample correlation coefficient r, which can be estimated by first determining the covariance and sample's standard deviation, which r represents the PCC, X_i and Y_i are the i-th observed values of variable X and Y, $\bar{X}$ and $\bar{Y}$ are the mean values of variable X and Y, respectively. PCC often executes promptly upon obtaining cleaned and feature-extracted data with fast computing power. It has a range of [-1, 1], where values closer to 1 or -1 indicate a stronger relationship.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{2}$$

2.3 Random Forests

RF is an ensemble machine learning approach that leverages both bootstrap resampling and random node classification techniques to build numerous independent decision trees. The ultimate classification outcomes are determined through a voting mechanism [36]. Variables' significance can serve as a valuable tool for selecting features in high-dimensional data, aiding in the examination of the significance of factors influencing PV generation. The formula for judging the importance of features using RF is (3). Among them, the significance of data is represented by VI. Error 1 refers to the inaccuracy of out-of-bag data when constructing an RF model with in-bag data, while error 2 pertains to the out-of-bag error resulting from random alterations of certain data elements within out-of-bag data samples. N denotes the total count of decision trees in the RF. If the random noise is added, the value of error 2 increases, indicating that this feature is important and has a great influence on the forecasting results. RF can obtain unbiased estimation of the true error and prevent overfitting and achieve higher accuracy compared to many individual algorithms.

$$VI = \sum(error2 - error1)\big/N \tag{3}$$

2.4 BiGRU prediction model

The hybrid optimized BiGRU model is to add a hidden layer to the unidirectional GRU model for optimization and expansion. In fact, it is composed of a forward GRU that receives forward input and a backward GRU that learns reverse input. Two hidden layers are used to extract historical and forthcoming data, and finally connected to the same output layer. At every time step, the input simultaneously supplies two GRUs in opposite directions, while the output is determined by two one-way GRUs. This arrangement effectively captures both past and future state information, enhancing the predictive model's capacity for generalization and accuracy [26]. Formula (4) is the calculation formula of BiGRU, where $\vec{h}_t$ and $\overleftarrow{h}_t$ respectively denotes the outputs of the forward and backward GRU layers at time step "t". Figure 1 is the structure diagram of BiGRU model.

$$y_t = \left[\vec{h}_t, \overleftarrow{h}_t\right] \tag{4}$$

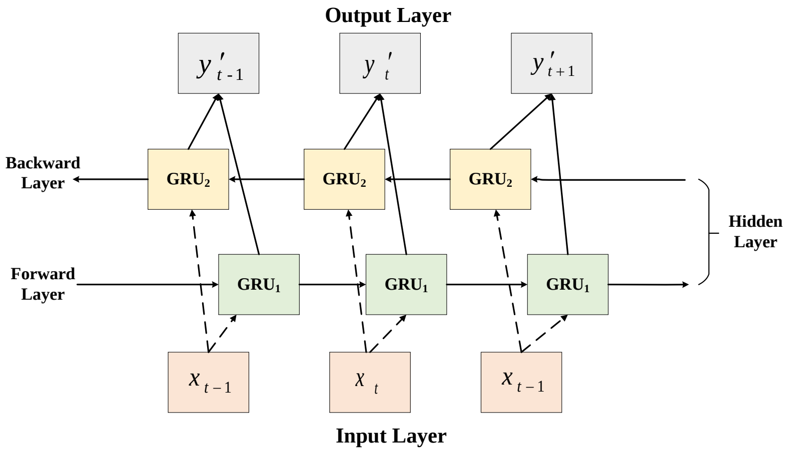


Fig. 1. Structure diagram of BiGRU model

2.5 Bagging algorithm

Bagging algorithm [37], is an ensemble learning method by using a certain combination strategy to combine multiple base learners to construct a learner with stronger learning effect, which can effectively utilize the characteristics of each base learner and enhance the model's learning performance. To obtain final strong learner, the main process involves randomly sampling m training data sets from the initial training set and using them to train the base learner [38]. The m prediction results generated by the m prediction models are then averaged and weighted [32]. The specific process is shown in Figure 2. Since ensemble learning can use multiple weak classifiers to form a strong classifier, it effectively enhances the model's ability to generalize and improves prediction performance.

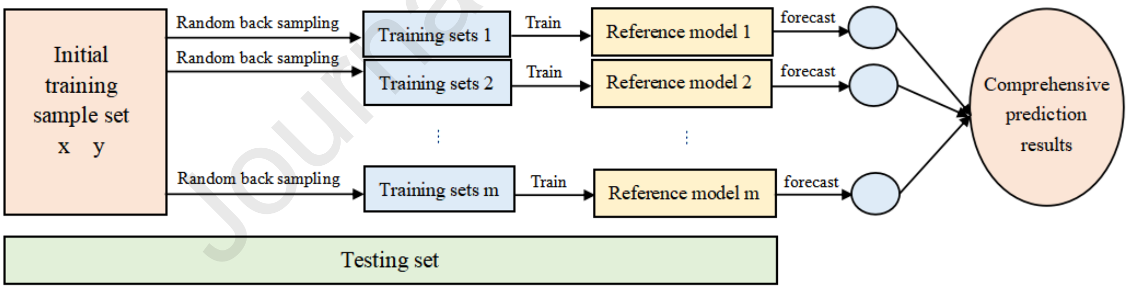


Fig. 2. Bagging algorithm Process

3. The hybrid ensemble optimized BiGRU prediction model

Above all, we proposed a hybrid optimized BiGRU prediction model, which is performed by employing one-dimensional CNN for convolution kernel operation to extract relevant features, introducing the Attention mechanism to assign weight coefficients to features, improving the attention towards important information, and stacking BiGRU to enhance the model's depth and feature extraction ability. However, both GRU model and hybrid optimized BiGRU model belong to a single model. During the model's training phase, there is a risk of randomness and instability. Therefore, we introduce the Bagging algorithm in the ensemble learning, and uses the hybrid optimized BiGRU as the base learner to construct the hybrid ensemble optimized BiGRU prediction model, which reduces the variance of prediction results, avoids the over-fitting problem, enhances the stability of model training process, and further enhances the prediction precision of the hybrid optimized BiGRU. The process of hybrid ensemble optimized BiGRU prediction model is depicted in Figure 3.

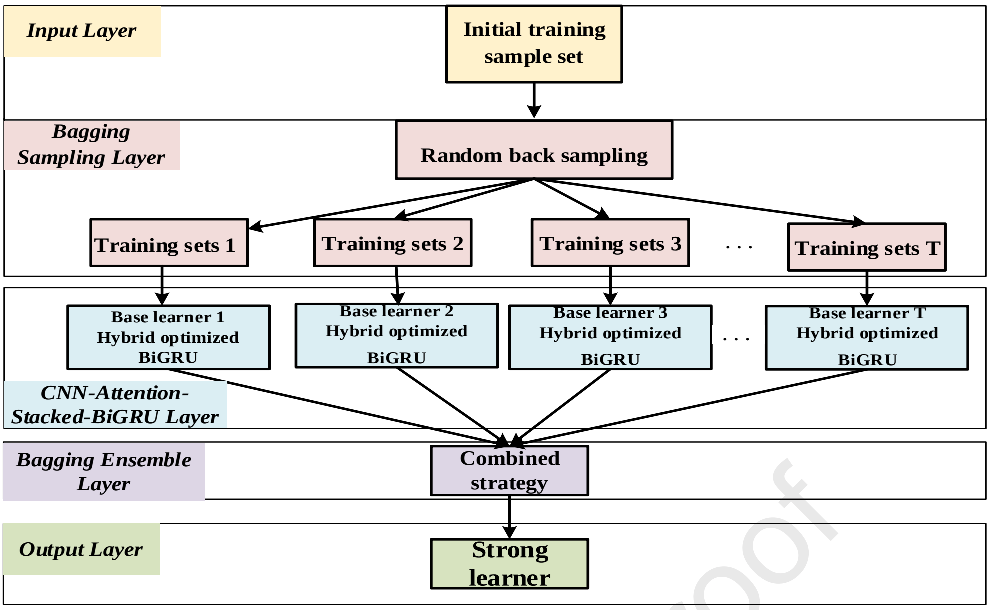


Fig. 3. The hybrid ensemble optimized BiGRU prediction model structure

The main framework of proposed hybrid ensemble optimized BiGRU prediction model includes input layer, bagging sampling layer, hybrid optimized BiGRU layer, bagging integration layer and output layer. Firstly, a sample set with m samples is randomly choosen from the initial training dataset of input layer. After collecting T times, T training data sets are obtained. Secondly, hybrid optimization of BIGRU is performed by employing one-dimensional CNN for convolution kernel operation to extract relevant features, introducing the Attention mechanism to assign weight coefficients to features, improving the attention towards important information, and stacking BiGRU to enhance the model's depth and feature extraction ability. Then, the hybrid optimized BiGRU model is used as the base learner under the Bagging algorithm, and the sampled sample set is used for training. Finally, T hybrid optimized BiGRU is predicted in parallel to obtain output values, and T predicted output values are averaged to acquire the final forecasting results.

4. The hybrid ensemble optimized BiGRU forecasting method based on LOWESS smoothing and feature selection

The proposed hybrid ensemble optimized BiGRU forecasting method mainly includes four key stages: data smoothing, feature selection, model forecasting, and evaluation and comparison of prediction results.

Stage1: Data smoothing. Based on the previous research, we choose the best performance LOWESS smoothing to analyze the actual PV generation data, so as to reduce the fluctuation range of daily power generation and enhance the data stability.

Stage2: Feature selection. Based on the previous research, we use the PCC with the highest applicability to analyze smoothed PV generation data and the influencing factors, and then use the RF to further screen the remaining features after the correlation analysis, streamline the crucial attributes and enhance the overall data input quality.

Stage3: Model forecasting. Based on the BiGRU model, we increase the Attention mechanism of assigning weight parameters, use the CNN layer to extract depth features, and superimpose BiGRU to enhance the depth of the neural network, which is called the hybrid optimized BiGRU model. Then, by introducing the idea of Bagging ensemble learning, a strong learner is built based on the hybrid optimized BiGRU model, and the ensemble optimized BiGRU prediction model is constructed to forecast the high-quality dataset after the above data processing steps.

Stage4: Evaluation and comparison of prediction results. MSE, RMSE, MAPE and MAE are used to evaluate the prediction results. For the purpose of reflect the advantages of the proposed model, the hybrid ensemble optimized BiGRU is compared and analyzed with the ensemble optimized BiLSTM, ensemble BiLSTM, ensemble BiGRU and BiLSTM and BiGRU.

The diagram illustrating the forecasting process is depicted in Figure 4.

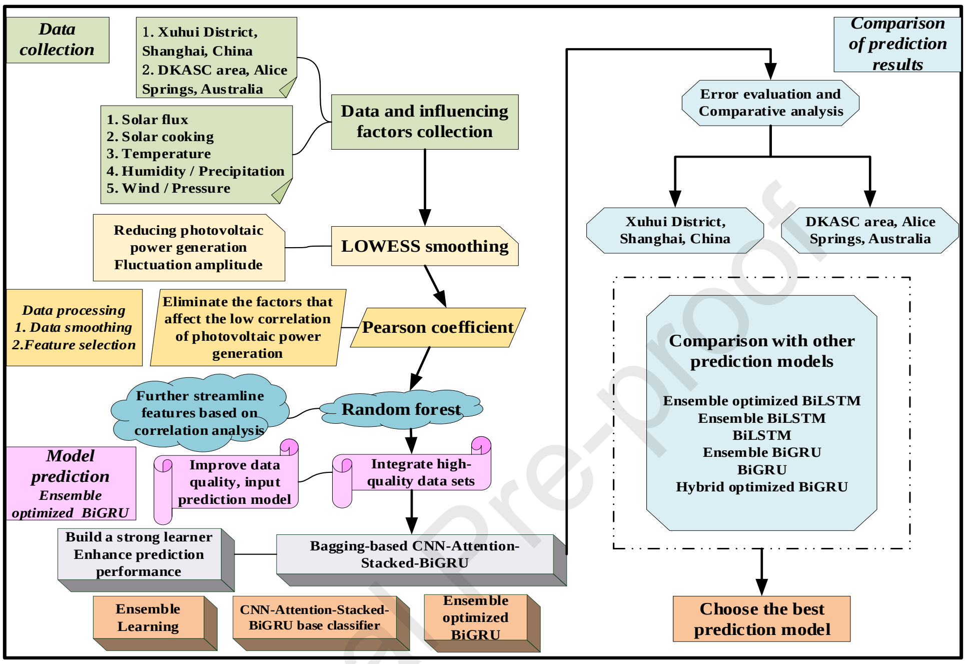


Fig. 4. Flow chart of hybrid ensemble optimized BiGRU prediction

5 Case study

5.1 The experimental data

5.1.1 Data sources

To ensure the representativeness and objectivity of the data, our experiment collects data sets from two photovoltaic power stations, i.e., Xuhui District Government [39] in Shanghai, China and DKASC [40] in Alice Springs, Australia. In fully analyze the efficacy of proposed method, the case selects May, July, November and December of 2021 as the test set. The two cities are chosen for their favorable climate and abundance of solar radiation, as well as their similar and comparable latitudes. Shanghai is located in the Northern and Eastern Hemispheres, while Alice Springs is situated in the Southern and Western Hemispheres. This combination of cities from different hemispheres makes them more representative. Figures 5 and 6 provide specific information on the locations of photovoltaic data sources, circumstances of photovoltaic plants, and the division of training and test sets in detail. The data interval is one day.

(1) Xuhui District, Shanghai, China (121.43 E, 31.18 N)

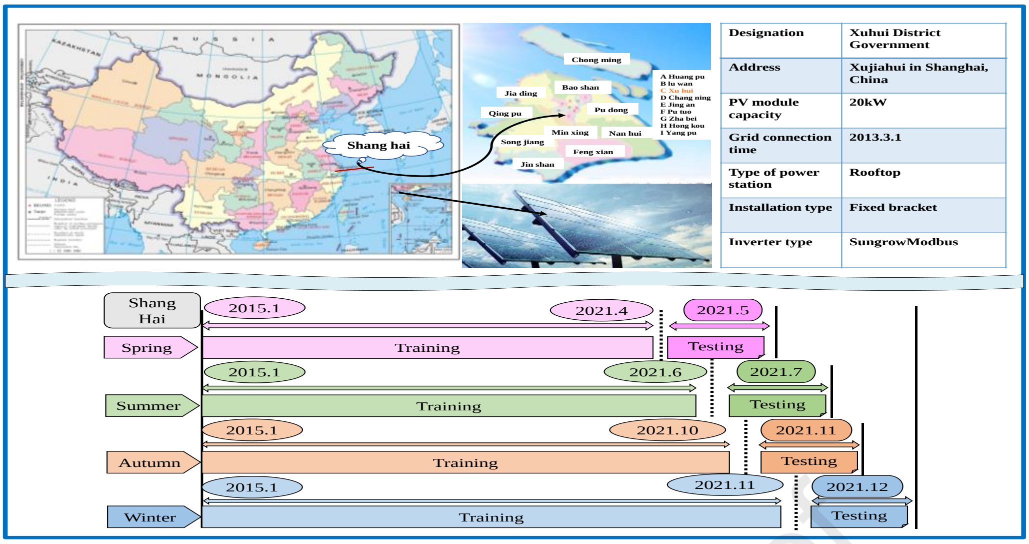


Fig. 5. Photovoltaic in Xuhui District, Shanghai, China.

(2) DKASC, Alice Springs, Australia (133.52 W, 24.17 S)

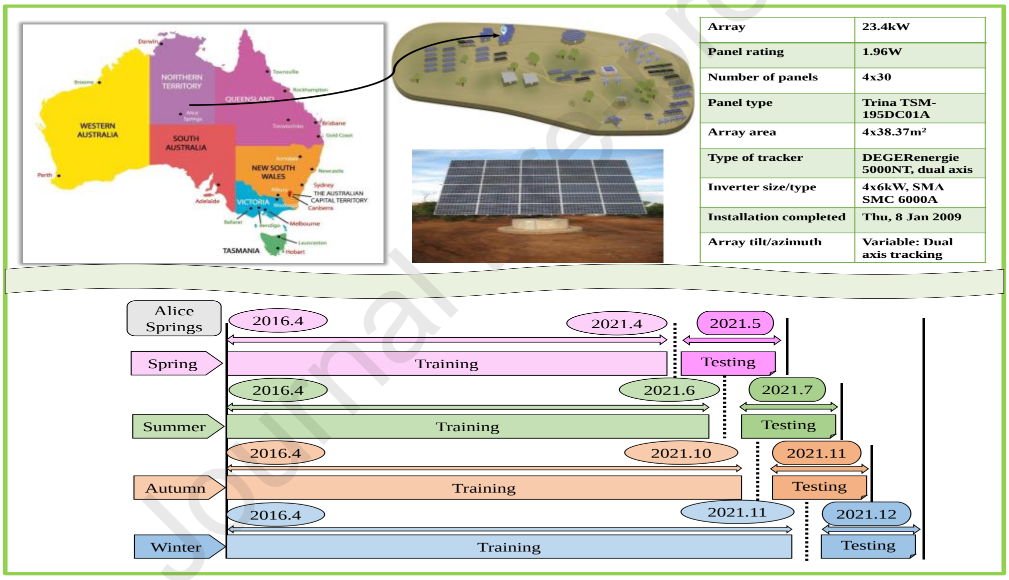


Fig. 6. Photovoltaic in DKASC, Alice Springs, Australia.

5.1.2 Influencing factors

To predict photovoltaic power generation accurately, it is crucial to consider the various factors that affect it. Our experiments have identified five main categories of factors: Part **A**: solar fluxes and related; Part **B**: parameters for solar cooking; Part **C**: Temperatures; Part **D**: Humidity/Precipitation; and Part **E**: Wind/Pressure. Table 1 shows a comprehensive array of the 31 refined influencing factors [41].

Table 1. Classification and refinement of influencing factors.

Influencing factors classification	Influencing factors refinement
	(1) All sky surface shortwave downward irradiance
	(2) Clear sky surface shortwave downward irradiance
	(3) All sky insolation clearness index
	(4) All sky surface longwave downward irradiance (thermal infrared)
	(5) All sky surface photosynthetically active radiation(PAR)total
Part **A**. Solar fluxes and related	(6) Clear sky surface photosynthetically active radiation (PAR) total

	(7) All sky surface UVA irradiance
	(8) All sky surface UVB irradiance
	(9) All sky surface UV index
Part **B**. Parameters for solar cooking	(1) Wind speed at 2 meters
Part **C**. Temperatures	(1) Temperature at 2 meters (2) Dew/Frost point at 2 meters (3) Wet bulb temperature at 2 meters (4) Earth skin temperature (5) Temperature at 2 meters range (6) Temperature at 2 meters maximum (7) Temperature at 2 meters minimum
Part **D**. Humidity/Precipitation	(1) Specific humidity at 2 meters (2) Relative humidity at 2 meters (3) Precipitation
Part **E**. Wind/Pressure	(1) Surface pressure (2) Wind speed at 10 meters (3) Wind speed at 10 meters maximum (4) Wind speed at 10 meters minimum (5) Wind speed at 10 meters range (6) Wind direction at 10 meters (7) Wind speed at 50 meters (8) Wind speed at 50 meters maximum (9) Wind speed at 50 meters minimum (10) Wind speed at 50 meters range (11) Wind direction at 50 meters

5.2 Evaluation metrics

In this paper, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are selected as the criteria for evaluating the prediction models. y_i and $y_i^{'}$ in formula (5) ~ (8) symbolize the true value and the predicted value, respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^{'}\right| \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i^{'}\right)^2} \tag{7}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i^{'}\right|}{y_i}\times 100\% \tag{8}$$

5.3 Parameter settings of experimental

(1) RF

The RF model's precision can be influenced by several parameters such as the quantity and maximum depth of decision trees, the settings for randomness, and the minimum number of samples required for leaf nodes. On the basis of reference [36], we determined the values of the relevant parameters through a large number of experiments, as shown in Table 2.

Table 2. Parameter settings of Random Forest.

Algorithm	Parameter	Parameter meaning	The parameter value
Random Forest	n_estimators	the number of decision trees	200
	max_depth	the maximum depth of decision trees	3
	random_state	the setting of random numbers	42
	min_samples_leaf	the minimum sample number of leaf nodes.	2

(2) Bagging ensemble learning algorithm

The accuracy of Bagging ensemble learning algorithm is affected by several parameters such as the number of base learners, neurons and unit layers, iteration times, hidden layer width, time step, one dimensional convolution, one-dimensional maximum pooling and fully connected. On the basis of reference [32], we determined the values of the relevant parameters through a large number of experiments, as listed in Table 2.

Table 3. Parameter setting of Bagging ensemble learning algorithm

Algorithm	Parameter	Parameter meaning	The parameter value
Bagging Ensemble Learning	n_splits	the number of base learners	6
	epochs	iteration times	40
	batch_size	hidden layer width	128
	time_steps	time step	1
	unit	the number of neurons	256
	num_layers	the number of unit layers	2
	filters		16
	kernel_size	one dimensional convolution	1
	activation		relu
	pool_size	One-dimensional maximum pooling	1
	activation	fully connected	sigmoid

(3) BiGRU

The BiGRU model's precision may be influenced by several parameters such as the number of neurons and unit layers, the settings for time step, the width of hidden layer, and the iteration times. On the basis of reference [26], we determined the values of the relevant parameters through a large number of experiments, as shown in Table 4.

Table 4. Parameter settings of BiGRU.

Algorithm	Parameter	Parameter meaning	The parameter value
BiGRU	unit	the number of neurons	256
	num_layers	the number of unit layers	2
	time_step	time step	3
	batch_size	hidden layer width	256
	epochs	iteration times	50

5.4 Data processing results

5.4.1 Data smoothing

In this experiment, we utilized LOWESS smoothing to process actual PV generation data. We mainly collect the data of Xuhui District, Shanghai, China in May 2021 and DKASC, Alice Springs, Australia in May 2021 to demonstrate the smoothing effect. Fig. 7 indicate that LOWESS smoothing can reduce abnormal data noise, make the processed data smoother, and try to ensure the authenticity of actual data.

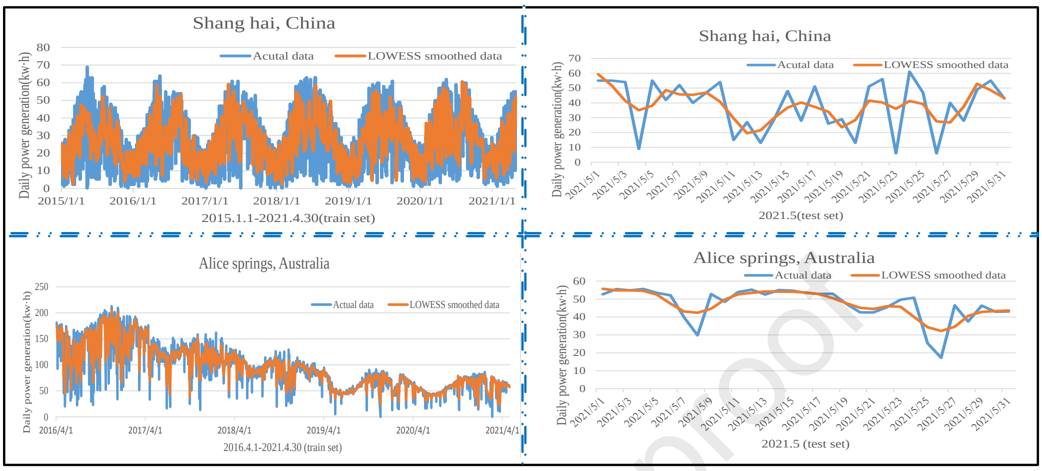


Fig. 7. Demonstration of the smoothing effect.

5.4.2 Feature selection

After smoothing the actual data, we use PCC to eliminate uncorrelated and very weakly correlated features. For the remaining features, RF are used for further refinement to enhance the quality of input data. Tables 5 and 6 display the results of correlation analysis (correlation⩾0.2) and the final features (importance ranking) after RF filtering.

Table 5. Pearson correlation analysis results.

Influencing factors	Pearson Xu hui	Pearson DKASC
A(1)	0.755	0.307
A(2)	0.716	0.225
A(3)	0.509	0.509
A(4)	0.302	0.302
A(5)	0.773	0.307
C(7)	0.473	0.473
C(6)	0.609	0.209
C(5)	0.385	0.385
B(4)	0.544	0.544
B(3)	0.489	0.489
B(2)	0.434	0.434
B(1)	0.541	0.541
A(6)	0.693	0.220
D(1)	0.455	0.455
D(2)	-0.296	-0.241
A(8)	0.761	0.302
A(7)	0.776	0.307

Table 6. Random Forest further feature selection results.

District	Pearson+ Random Forest	No correlation analysis+ Random Forest
Xu hui RF≥0.03	A(7): 0.45 A(2): 0.11 C(5): 0.08 A(1): 0.08 A(5): 0.08 D(2): 0.03 C(6): 0.03	A(1): 0.56 A(2): 0.10 C(5): 0.06 A(6): 0.03
DKASC RF≥0.1 (No correlation analysis：RF≥0.05)	C(6)：0.16 A(8)：0.15 D(2)：0.15 A(6)：0.12	A(8)：0.08 A(1)：0.05 A(5)：0.05

5.5 Experiment results

In this study, we utilize high-quality data processed by LOWESS smoothing, Pearson correlation coefficient and Random Forest to create a new dataset. This dataset is then used to verify the proposed hybrid ensemble optimized BiGRU prediction model. We compared the performance of this model with several common single models, hybrid optimized models and ensemble optimized models. Among them, the single models include BiLSTM, BiGRU, SVM, XGBoost and TCN; the hybrid optimized models include hybrid optimized BiGRU and hybrid optimized BiLSTM; the ensemble optimized models include ensemble BiGRU, ensemble BiLSTM, hybrid ensemble optimized BiGRU and hybrid ensemble optimized BiLSTM. The study is conducted in two locations: Shanghai, China and Alice Springs, Australia. The model prediction results of two cases are as follows:

5.5.1 Xuhui District, Shanghai, China

(1) The necessity of data smoothing

The prediction data results and prediction errors show that the smoothed forecast outcomes greatly surpass the unsmoothed forecast outcomes in terms of quality. As shown in Figure 8, LOWESS smoothing reduces the fluctuation and improves the quality of input data by denoising and smoothing the data. The smoothed prediction results have a higher fitting degree with the real data and a smaller forecasting error. Among the test months selected for different seasons, the prediction errors in May and July are larger than those in November and December, and the goodness of fit is not as good as that in November and December. This is because, in terms of natural factors, the climate of Shanghai belongs to subtropical monsoon climate, with abundant rainfall. 60 % of the rainfall in the whole year is concentrated in the flood season from May to September, and the rainfall in the remaining months is moderate. Therefore, the smoothing technology can effectively enhance the forecasting precision and reduce the forecasting error, which reflects the necessity of introducing LOWESS smoothing in this paper.

Fig. 8. Comparison of integrated results among actual data, smoothed, unsmoothed prediction. (Xu hui)

(2) The applicability of correlation analysis

As shown in Figure 9, the results of PCC and non-correlation analysis are basically consistent with the trend of real data, but the correlation prediction is closer to the real data. The forecasting error shows that PCC analysis has the lowest prediction error under most evaluation criteria and has a good forecasting effect. From the monthly point of view, the forecasting errors in May and July are the largest. This is consistent with the fact that the flood season in Shanghai is concentrated in May to September, during which the photovoltaic power supply fluctuates greatly. The prediction error in November and December is small, and the rainfall in autumn and winter is moderate, then the photovoltaic power supply is relatively gentle. Therefore, selecting the appropriate correlation analysis method is necessary to enhance the quality of input data.

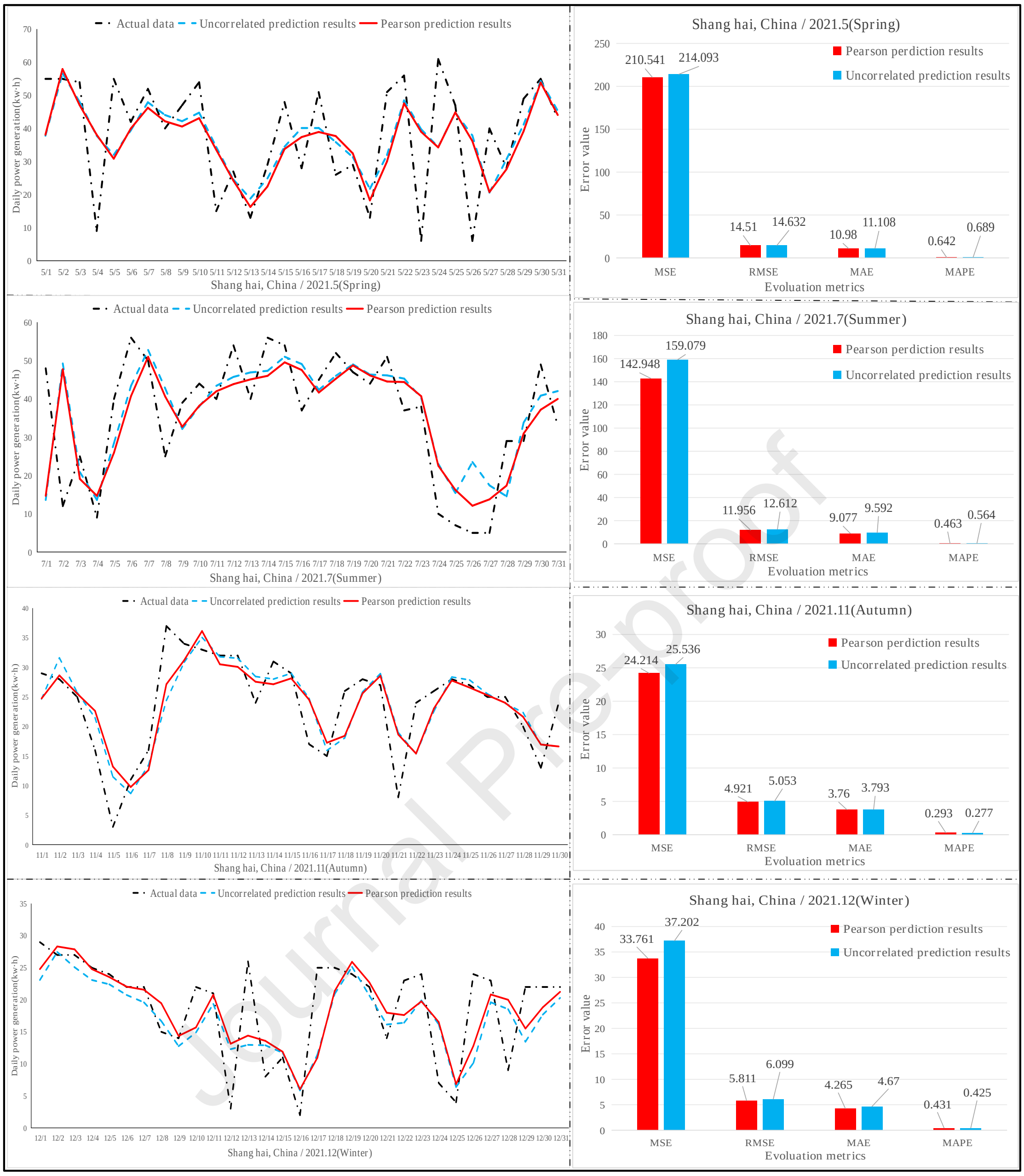


Fig. 9. Comparison of integrated results between actual data and correlation analysis prediction. (Xu hui)

(3) The precision of prediction model

To reflect the effectiveness of proposed forecasting method, we choose the single models such as SVM, XGBoost, TCN, BiLSTM, BiGRU and the hybrid optimized prediction models including hybrid optimized BiLSTM, hybrid optimized BiGRU, and the ensemble prediction models including ensemble BiLSTM, ensemble BiGRU and ensemble optimized BiLSTM for comparison. Since it has been demonstrated that the hybrid optimized BiGRU model has better forecasting results than a single BiGRU model [27], our experiments are no longer compared with single prediction models such as GRU and LSTM. The comparison between the actual data and the forecast results of different models are displayed in Figure 10. The specific analysis is as follows:

a. From the fitting degree of prediction results and actual data, the proposed hybrid ensemble optimized BiGRU prediction model shows the closest fit to actual data and the best fitting effect, both in overall trend and local amplification area. The prediction results of other models such as ensemble optimized BiLSTM, ensemble BiLSTM, ensemble BiGRU, hybrid optimized BiLSTM, hybrid optimized BiGRU and TCN models are in good agreement with actual data, but lower than the hybrid ensemble optimized BiGRU model. On the other hand, the single BiLSTM, BiGRU, XGBoost and SVM models have poor fitting with actual data and their prediction effect is not well.

b. From the perspective of prediction error, although hybrid optimized BiLSTM and hybrid optimized BiGRU has the lowest MAE value in May 2021, and BiGRU shows a lower prediction error in July 2021, however, it does not affect that the hybrid ensemble optimized BiGRU model has the lowest forecasting error in most cases.

c. From the monthly point of view, compared with November and December, the fluctuation range and prediction error of May and July are larger, and the predicted data and the actual data are less fitted, which is consistent with the climate characteristics of Shanghai flood season concentrated in May-September and less precipitation in winter.

d. Under the premise of consistent data processing results, the proposed prediction method has lower prediction error, the highest fitting degree with actual data, and is more suitable than other models for PV generation forecasting.

Combining the prediction results above, we find that the prediction process with LOWESS smoothing, PCC, RF and the hybrid ensemble optimized BiGRU model exhibits superior forecasting precision and better performance. To fully reflect the efficacy of proposed forecasting method, we choose the forecasting results of DKASC in Alice Springs, Australia as a validation below.

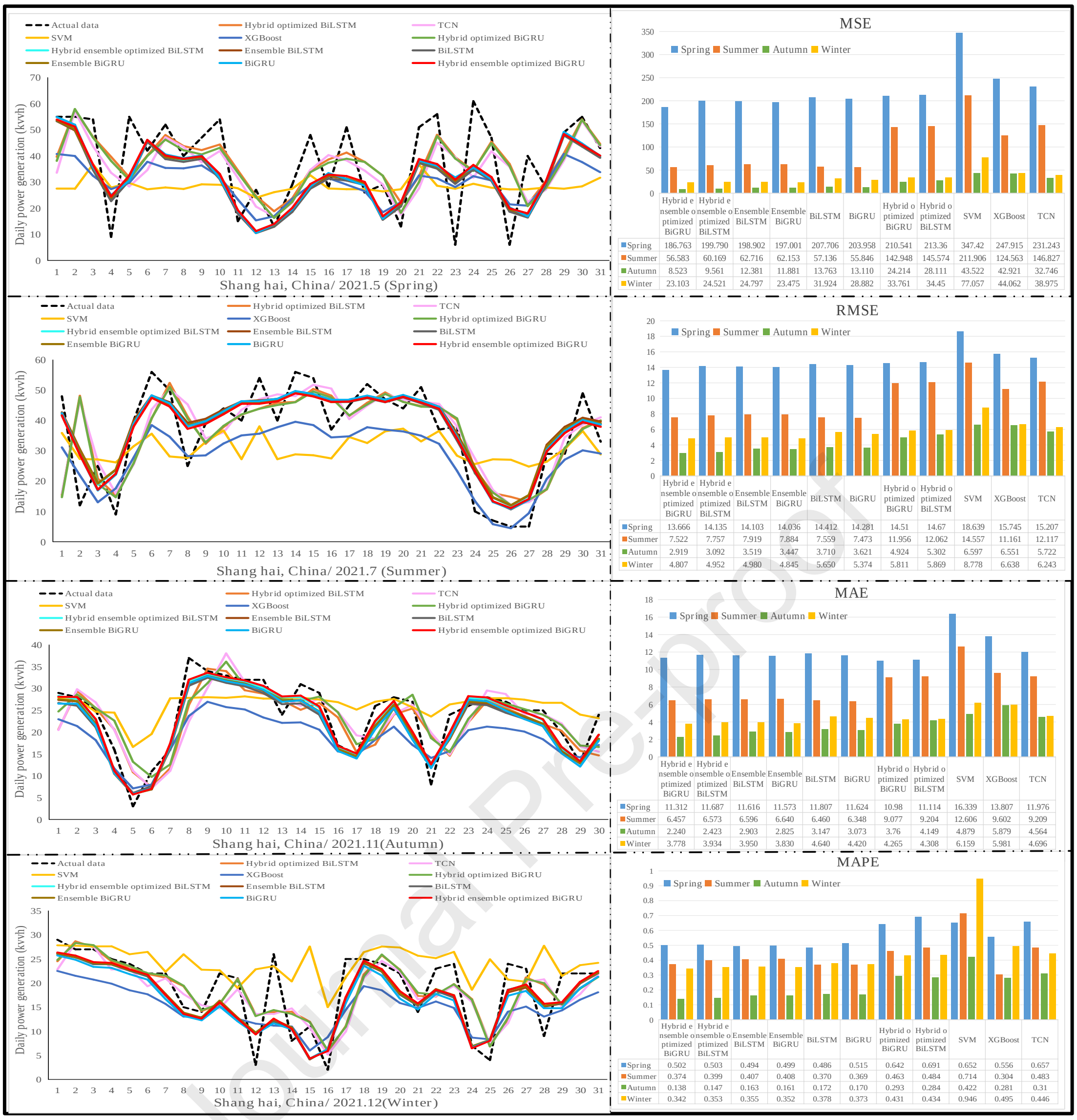


Fig. 10. Comparison of actual data and prediction results of different models (Xuhui District, Shanghai, China)

5.5.2 DKASC area, Alice Springs, Australia

(1) The necessity of data smoothing

As shown in Figure 11, the smoothed prediction results fit the actual data more precisely than the unsmoothed results, resulting in lower prediction errors. DKASC in Alice Springs boasts a tropical desert climate that is characterized by low rainfall and dry weather throughout a whole year, which making it less vulnerable to weather fluctuations. Moreover, its strategic location and high exposure to solar radiation make it an ideal site for photovoltaic power generation. The actual power generation data for the four months indicates that DKASC, despite its arid climate and limited rainfall, is still susceptible to abnormal fluctuations. The smoothed prediction results from both markets demonstrate that data smoothing still plays a key role in improving the forecasting precision. LOWESS smoothing is a typical method in reducing the magnitude of fluctuations, smoothing the data, and enhancing the forecasting accuracy.

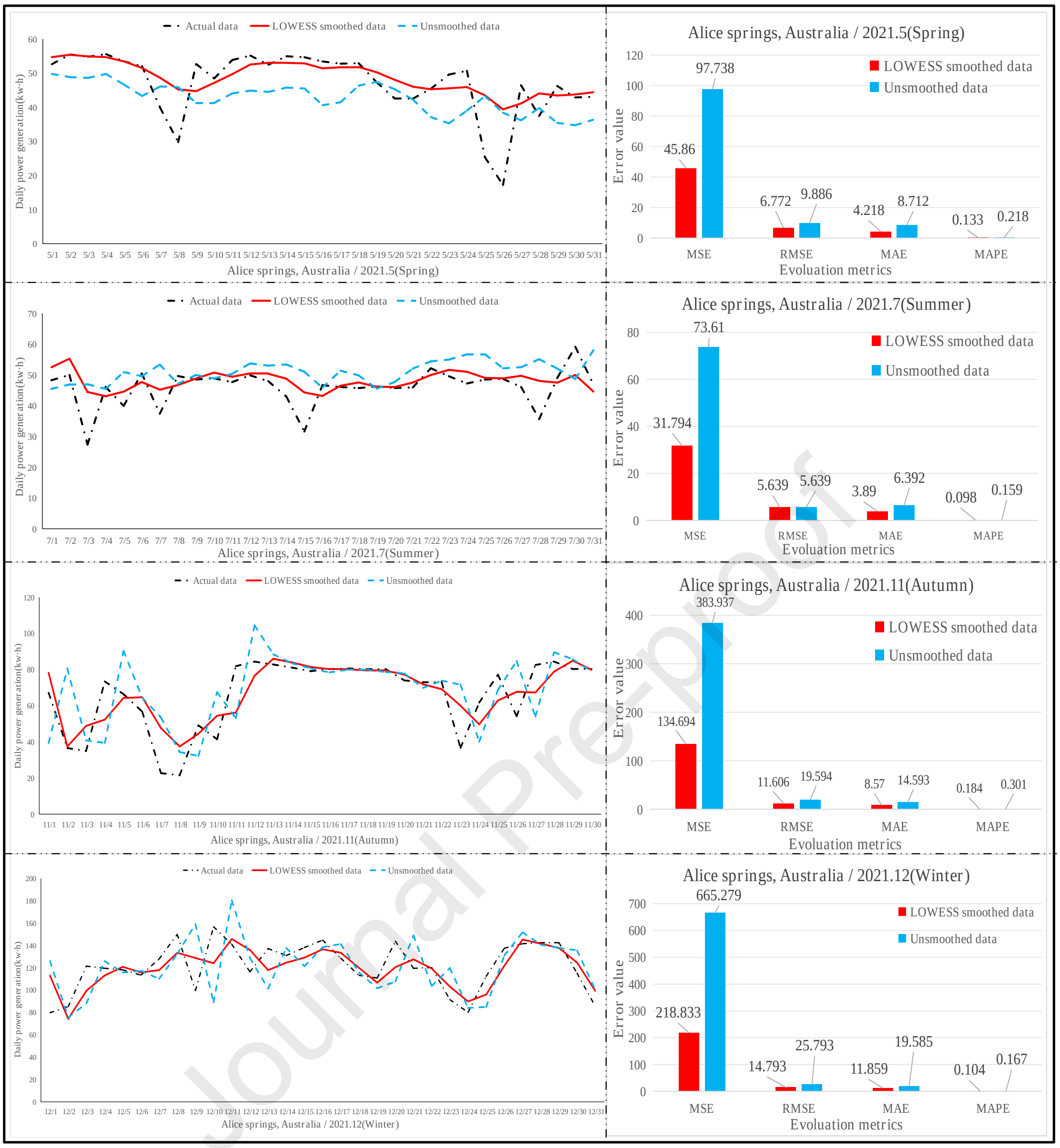


Fig. 11. Comparison of integrated results among actual data, smoothed, unsmoothed prediction. (DKASC)

(2) The applicability of correlation analysis

As shown in Figure 12, the prediction results of non-correlation analysis have a low fitting degree with the real data, while Pearson correlation analysis is basically consistent with the real data. The prediction error indicates that the forecasting error of Pearson correlation is the lowest in MSE, RMSE and MAE, and the prediction effect is good. The experimental results continue the results of correlation analysis and prediction comparison of Shanghai photovoltaic power supply, and prove the applicability of correlation analysis.

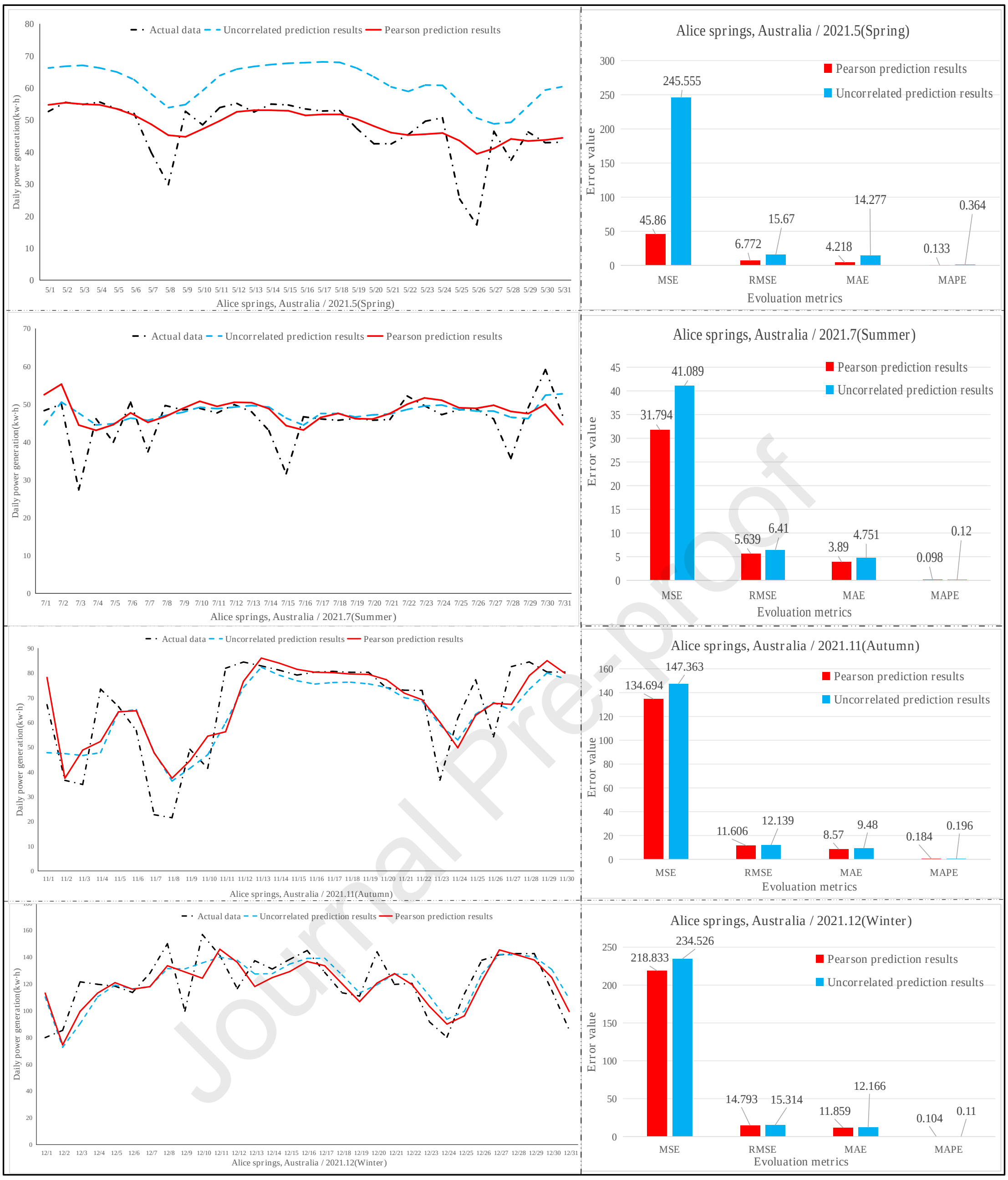


Fig. 12. Comparison of integrated results between actual data and correlation analysis prediction. (DKASC)

(3) The precision of prediction model

The comparison between the actual data and the forecast results of models is displayed in Figure 13. The specific analysis is as follows:

a. According to the fitting degree between the forecasted data by different models and the real data, the hybrid ensemble BiGRU still shows a high fitting degree and the best fitting effect, which continues the case analysis results of Shanghai. Among the prediction models, the prediction results of SVM, XGBoost, TCN, BiLSTM and BiGRU have a poor prediction performance and a low goodness of fit with actual data.

b. From the perspective of prediction error, compared with the results of Shanghai case analysis, the prediction error of Australian data set can better reflect the forecasting performance advantage of the hybrid ensemble optimized BiGRU model. Only in July 2021, the hybrid optimized BiGRU outperform the proposed model in some evaluation criteria. Nonetheless, overall, the hybrid ensemble optimized BiGRU model exhibits the lowest prediction error in most error evaluation criteria.

c. According to the month, the DKASC area belongs to the tropical desert climate, with scarce precipitation and sufficient solar radiation throughout the year. The fluctuation of power generation data in May, July and November is small, and the fluctuation in December is large. Therefore, among the four months, the prediction error in December is the largest, and the fitting degree with the real data is the lowest, which is consistent with the results of Shanghai case analysis.

d. In contrast to other models, the hybrid ensemble optimized BiGRU model exhibits the strongest fitting degree with the real data, the lowest prediction error and the best prediction performance. Through the comparison of various models, it can be found that the hybrid optimized forecasting models has a better forecasting performance than the individual forecasting models, but worse than that of ensemble prediction model. In the four ensemble models, the forecasting impact of the proposed model is significantly better than the other three.

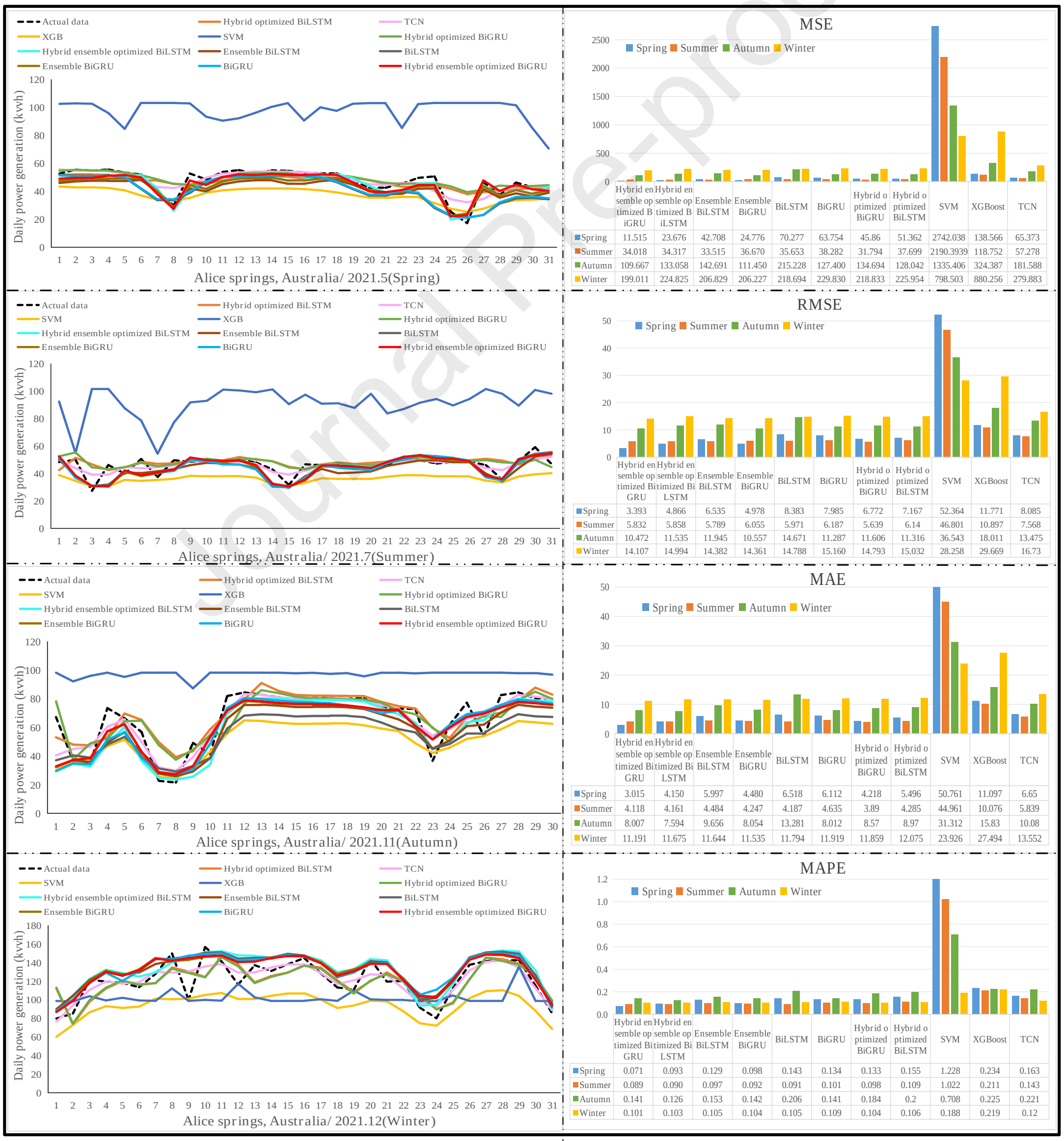


Fig. 13. Comparison of actual data and prediction results of different models (DKASC area, Alice Springs, Australia)

According to the above conclusions and the analysis results of two groups of cases, it can be seen that the proposed model is the closest to the overall trend of actual data, and the fitting effect is the best. The predicted results also conforms to the climate change characteristics of two regions. Although the prediction error is less different from other model error values, it has the lowest value in most months and evaluation criteria. From the fitting degree and prediction error value with actual data, the hybrid ensemble optimized BiGRU model has the highest forecasting performance, which is helpful to enhance the prediction accuracy of PV generation.

6 Conclusion

In order to improve the accuracy and stability of PV generation and reduce the risk of instability and volatility in the training process, this paper presents a hybrid ensemble optimized BiGRU method based on Bagging algorithm. In terms of model prediction, this paper combines the LOWESS data smoothing method in conjunction with the Pearson correlation analysis method, which is found to have more strong applicability in a previous study. Additionally, Random Forest feature selection is employed to improve the quality of data processing. In terms of model prediction, we utilize the hybrid optimized BiGRU with CNN and Attention mechanism as the base learner and employ the Bagging algorithm in ensemble learning to construct a strong learner. The case analysis results from two different areas demonstrate that the presented forecasting method significantly enhance the precision of PV generation prediction, effectively mitigate the risk of instability and randomness during the training process, and improve prediction stability and prediction performance. The main conclusions are as follows:

（1）Under the hybrid ensemble optimized BiGRU model, the data processing methods of LOWESS smoothing, Pearson correlation analysis and Random Forest feature selection can still effectively enhance the quality of input data and the forecasting precision.

（2）Under the premise of consistent data processing results, in contrast to other models, the presented hybrid ensemble optimized BiGRU model has the highest prediction precision from the fitting degree and prediction error with actual data.

（3）Under the same conditions of the case data set, the presented hybrid ensemble optimized BiGRU prediction model has significantly lower forecasting error and higher forecasting accuracy than other prediction models. The MSE value has the most significant decrease, with a maximum decrease of nearly 60.4 %.

（4）The prediction effect of hybrid optimized model is better than that of individual prediction model, and the forecasting effect of the ensemble model is better than that of hybrid optimized prediction model.

（5）Under the unified conditions of ensemble algorithm, the optimized method has better forecasting impact than the un-optimized method.

Although our prediction method improves the accuracy of photovoltaic power generation, there are still some questions worth addressing. For instance, is there a more precise prediction method to identify fluctuations when actual data encounters abnormal fluctuations? This will be the focus of our future research. Futhermore, although collecting PV data by the hour presents challenges, there are theoretical similarities between short-term photovoltaic generation forecasting and short-term hourly power forecasting. Therefore, we are optimistic about the application prospect of the hybrid ensemble optimized BiGRU model in the field of short-term hourly power forecasting.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.72371139), Humanities and Social Science Fund of Ministry of Education of China (No.20YJA630009) and Shandong Natural Science Foundation of China (No.ZR2022MG002).

References

[1] Lin G Q, Li L L, Tseng M L, et al. An improved moth-flame optimization algorithm for support vector machine prediction of photovoltaic power generation[J]. Journal of Cleaner Production, 2020, 253: 119966.

[2] Liu H, Shi Y, Guo L, et al. China's energy reform in the new era: Process, Achievements and Prospects[J]. Management World, 2022, 38(7): 6-24.

[3] Li J, Fang Z, Tan Y, et al. Application prospect analysis of electrochemical energy storage technology in county-wide rooftop photovoltaic system[J]. Acta Energiae olaris Sinica, 2022, 43(4): 1-12.

[4] Mulleriyawage U G K, Shen W X. Optimally sizing of battery energy storage capacity by operational optimization of residential photovoltaic-Battery systems: An Australian household case study[J]. Renewable Energy, 2020, 160: 852-864.

[5] Meng A, Xu X, Chen J, et al. Ultra short term photovoltaic power prediction based on reinforcement learning and combined deep learning model[J]. Power System Technology, 2021, 45(12): 4721-4728.

[6] Qiao S, Wang L, Zhang P, et al. Short-term load forecasting by long and short-term temporal networks with attention based on model decomposition[J]. Power System Technology, 2022, 46(10): 3940-3951.

[7] Ding S, Li R, Tao Z. A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting[J]. Energy Conversion and Management, 2021, 227: 113644.

[8] Ahmed R, Sreeram V, Mishra Y, et al. A review and evaluation of the state-of-the-art in photovoltaic solar power forecasting: Techniques and optimization[J]. Renewable and Sustainable Energy Reviews, 2020, 124: 109792.

[9] Kumler A, Xie Y, Zhang Y. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique[J]. Solar Energy, 2019, 177: 494-500.

[10] Gong Y, Lu Z, Qiao Y, et al. An overview of photovoltaic energy system output forecasting technology[J]. Automation of Electric Systems, 2016, 40(4): 140-151.

[11] Hassan J. ARIMA and regression models for prediction of daily and monthly clearness index[J]. Renewable Energy, 2014, 68: 421-427.

[12] Abdel B M, Hawash H, Chakrabortty R K, et al. photovoltaic-Net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production[J]. Journal of Cleaner Production, 2021, 303: 127037.

[13] Dai Y, Wang Y, Leng M, et al. LOWESS smoothing and Random Forest based GRU model: A short-term photovoltaic power generation forecasting method[J]. Energy, 2022, 256: 124661.

[14] Dai Y, Zhou Q, Leng M, et al. Improving the Bi-LSTM model with XGBoost and attention mechanism: A combined approach for short-term power load prediction[J]. Applied Soft Computing, 2022, 130: 109632.

[15] Ren X, Zhang F, Zhu H, et al. Quad-kernel deep convolutional neural network for intra-hour photovoltaic power forecasting[J]. Applied Energy, 2022, 323: 119682.

[16] Wang L, Mao M, Xie J, et al. Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model[J]. Energy, 2023, 262: 125592.

[17] Sun S, Wei Y, Wang S. Exchange rates forecasting with decomposition-clustering-ensemble learning approach[J]. Systems Engineering-Theory & Practice, 2022, 42(3): 664-677.

[18] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM[J]. Energy Conversion and Management, 2018, 159: 54-64.

[19] Xiu X, Yu L, Lin Z, et al. Combination forecasting of short-term sales for fresh products based on feature fusion[J]. Journal of Management Sciences in China, 2022, 25(12): 102-123.

[20] Wang Y, Wang H, Srinivasan D, et al. Robust functional regression for wind speed forecasting based on Sparse Bayesian learning[J]. Renewable Energy, 2019, 132: 43-60.

[21] Gallardo R P, Ríos A M, Ramírez J S. Analysis of the solar and wind energetic complementarity in Mexico[J]. Journal of Cleaner Production, 2020, 268: 122323.

[22] Yang X, Jiang J, Liu F, et al. Distribution Network Topology Identification Based on Attention Mechanism and Convolutional Neural Network[J]. Power System Technology, 2022, 46(5): 1672-1682.

[23] Zhang C, Ma H, Hua L, et al. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction[J]. Energy, 2022, 254: 124250.

[24] Phan Q T, Wu Y K, Phan Q D, et al. A novel forecasting model for solar power generation by a deep learning framework with data preprocessing and postprocessing[J]. IEEE Transactions on Industry Applications, 2023, 59(1): 220-231.

[25] He Y, Wang Y. Short-term wind power prediction based on EEMD–LASSO–QRNN model[J]. Applied Soft Computing, 2021, 105: 107288.

[26] Dong X, Zhao H, Zhao S, et al. ULTRA-SHORT-TERM forecasting method of photovoltaic power based on som clustering, secondary decomposition and BiGRU[J]. Acta Energiae olaris Sinica, 2022, 43(11): 85-93.

[27] Zhang C, Peng T, Nazir M S. A novel ensemble photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables[J]. Electric Power Systems Research, 2022, 213: 108796.

[28] Niu T, Li J, Wei W, et al. A hybrid deep learning framework integrating feature selection and transfer learning for multi-step global horizontal irradiation forecasting[J]. Applied Energy, 2022, 326: 119964.

[29] Abbasi E, Moghaddam M R A, Kowsari E. A systematic and critical review on development of machine learning based-ensemble models for prediction of adsorption process efficiency[J]. Journal of Cleaner Production, 2022, 379: 134588.

[30] Xiao J, Wen Z, Liu B, et al. A hybrid model based on selective deep-ensemble for container throughput forecasting[J]. Systems Engineering-Theory & Practice, 2022, 42(4): 1107-1128.

[31] Busari G A, Lim D H. Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance[J]. Computers & Chemical Engineering, 2021, 155: 107513.

[32] Niu H, Zhao Y. Predicting stock price index using Bagging algorithm and GRU model[J]. Computer Engineering and Applications, 2022, 58(12): 132-138.

[33] Wang K, Zhang Z, Han A, et al. Short-term load forecasting based on Bidirectional GRU neural network integrated by Bagging algorithm[J]. Proceedings of the CSU-EPSA, 2021, 33(10): 24-30.

[34] Zhou H, Deng Z, Xia Y, et al. A new sampling method in particle filter based on Pearson correlation coefficient[J]. Neurocomputing, 2016, 216: 208-215.

[35] Cleveland W S. Robust locally weighted regression and smoothing scatterplots[J]. Journal of the American statistical association, 1979, 74(368): 829-836.

[36] Huo W, Li W, Zhang Z, et al. Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and Random Forest feature selection[J]. Energy Conversion and Management, 2021, 243: 114367.

[37] Zhang C, Peng T, Nazir M S. A novel ensemble photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables[J]. Electric Power Systems Research, 2022, 213: 108796.

[38] Zhao Y, Hao J, Li J. A Trimmed Average Based Neural Network Ensemble Approach for Time Series Forecasting[J]. Chinese Journal of Management Science, 2022, 30(3): 211-220.

[39] https://www.lvsedianli.com/perHome.html.

[40] https://dkasolarcentre.com.au/download/notes-on-the-data.

[41] https://power.larc.nasa.gov/data-access-viewer/

Highlights

- LOWESS smoothing technology is used to reduce the variability of PV generation data
- Pearson correlation coefficient and Random Forest are used to for extract feature
- Attention mechanism and CNN layer are used to optimize BiGRU model
- Two BiGRU layers are stacked to improve the depth and prediction performance
- Ensemble learning is introduced to further optimize the hybrid BiGRU model

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Yeming Dai reports financial support was provided by Shandong Natural Science Foundation of China. Yeming Dai reports financial support was provided by Chinese Education Department.