Intelligent and Converged Networks ISSN 2708-6240
2022, 3(4): 311−324 DOI: 10.23919/ICN.2022.0025

Multi-features fusion for short-term photovoltaic power prediction

Ming Ma, Xiaorun Tang, Qingquan Lv, Jun Shen, Baixue Zhu, Jinqiang Wang, and Binbin Yong*

Abstract: In recent years, in order to achieve the goal of "carbon peaking and carbon neutralization", many countries have focused on the development of clean energy, and the prediction of photovoltaic power generation has become a hot research topic. However, many traditional methods only use meteorological factors such as temperature and irradiance as the features of photovoltaic power generation, and they rarely consider the multi-features fusion methods for power prediction. This paper first preprocesses abnormal data points and missing values in the data from 18 power stations in Northwest China, and then carries out correlation analysis to screen out 8 meteorological features as the most relevant to power generation. Next, the historical generating power and 8 meteorological features are fused in different ways to construct three types of experimental datasets. Finally, traditional time series prediction methods, such as Recurrent Neural Network (RNN), Convolution Neural Network (CNN) combined with eXtreme Gradient Boosting (XGBoost), are applied to study the impact of different feature fusion methods on power prediction. The results show that the prediction accuracy of Long Short-Term Memory (LSTM), stacked Long Short-Term Memory (stacked LSTM), Bi-directional LSTM (Bi-LSTM), Temporal Convolutional Network (TCN), and XGBoost algorithms can be greatly improved by the method of integrating historical generation power and meteorological features. Therefore, the feature fusion based photovoltaic power prediction method proposed in this paper is of great significance to the development of the photovoltaic power generation industry.

Key words: meteorological factors; multi-features fusion; time series prediction; photovoltaic power prediction

1 Introduction

Since the beginning of the 21st century, fossil energy has gradually dried up, and environmental protection issues have become increasingly prominent. Therefore, green energy has become the mainstream direction of the world's energy development. Electricity power generation methods typically include thermal power generation, wind power generation, photovoltaic power generation, hydraulic power generation, and also nuclear power generation. In photovoltaic power generation, with the increasing proportion of installed photovoltaic capacity, the uncertainty of photovoltaic power output poses great challenges to the planning and operation of the power system. The high rejection-to-adopt rate is also one of the difficulties in the development of the photovoltaic industry. Therefore, accurate photovoltaic power prediction is of great significance for new energy grid connection and dispatching.

Generally, the photovoltaic power prediction technologies can be classified from three aspects: time scale, prediction process, and prediction form.

The time scale can be divided into medium and long-

• Ming Ma is with the State Grid Gansu Electric Power Research Institute, Lanzhou 730000, China. E-mail: 283257683@qq.com.
• Xiaorun Tang, Qingquan Lv, Baixue Zhu, Jinqiang Wang, and Binbin Yong are with the School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China. E-mail: tangxr21@lzu.edu.cn; lvqq18@lzu.edu.cn; zhubx19@lzu.edu.cn; jqwang16@lzu.edu.cn; yongbb@lzu.edu.cn.
• Jun Shen is with the School of Computing and Information Technology, University of Wollongong, Wollongong 2522, Australia. E-mail: jshen@uow.edu.au.
* To whom correspondence should be addressed.
Manuscript received: 2022-11-08; revised: 2022-12-12; accepted: 2022-12-14

term prediction, short-term power prediction, ultra-short-term power prediction, and minute-level prediction. The length of the medium and long-term forecast is one month to one year[1]. Short-term power prediction can generally predict the longest generation power of 72 hours and the shortest generation power of 15 minutes. The prediction time of ultra-short-term power prediction is much shorter, and the generated power in the next 4 hours is generally predicted. The time length of minute-level prediction is about 2 hours, and the time resolution is generally 5 minutes.

The prediction process can be generally divided into direct prediction and indirect prediction. Direct prediction establishes a photoelectric conversion model based on the irradiance data collected by the sensor, including installation angle, altitude, longitude, and latitude of the photovoltaic system, so as to directly output the photovoltaic power. Indirect prediction utilizes meteorological data, including solar irradiance, temperature, humidity, wind speed, and historical photovoltaic power generation data, to predict photovoltaic power generation[2].

In terms of the form of prediction, it can be divided into three types: point prediction, interval prediction, and probability prediction. In point prediction, the model gives the determined power generation prediction value, the interval prediction gives the upper and lower bounds of output under the confidence level, and the probability prediction gives the expected output value and probability distribution information in the future[3].

Based on the meteorological features and historical power generation data, this paper makes efforts toward a short-term power prediction for the power generation of photovoltaic power station, and uses the indirect prediction method to make a point prediction for a single photovoltaic power station.

The rest of the paper is organized as follows. Section 2 introduces the background of photovoltaic power generation prediction technology. Section 3 preprocesses and analyzes the data of photovoltaic stations, and Section 4 analyzes the impact of different feature fusion methods on the prediction of photovoltaic power generation after processing the three datasets obtained in Section 3, and this section discusses all the results. Section 5 summarizes the whole paper.

2 Related work

In literature, many scholars have conducted extensive research on photovoltaic power prediction. The following related work will be summarized according to the research methods, technical characteristics, and prediction results.

Firstly, regarding the photovoltaic power time series method, the most representative time series prediction method is the AutoregRessive Integrated Moving Average model (ARIMA), which combines the AutoRegressive (AR) model and Moving Average (MA) model and takes into account the difference between these models[4]. Therefore, for photovoltaic data with small-scale volatility, the model can effectively mine the internal patterns of the series. However, the ARIMA model cannot satisfactorily mine the fluctuation pattern of data in the case of high volatility.

Statistical methods train statistical models and output prediction results based on historical meteorological data and historical power generation data, which mainly include Artificial Neural Networks (ANN), deep learning neural networks, Support Vector Machines (SVM), K-Nearest Neighbor (KNN), etc. In Ref. [5], a novel Residuals-Based Deep Least Squares Support Vector Machine (RBD-LSSVM) is proposed. Analysis of the prediction results and comparisons with recent and past studies demonstrate the promising performance of the proposed RBD-LSSVM approach with redundancy test-based model selection method for modeling and predicting nonlinear time series. Soft Tissue Tumors (STTs) are a form of sarcoma found in tissues that connect, support, and surround body structures. Alaoui et al.[6] proposed a machine learning based approach which combines a new technique of preprocessing the data for features transformation, resampling techniques to eliminate the bias and the deviation of instability, and performing classifier tests based on the SVM and Decision Tree (DT) algorithms. The tests carried out on the dataset collected in Nur Hidayah Hospital of Yogyakarta in Indonesia showed a

great improvement compared to previous studies. These results confirm that machine learning methods could provide efficient and effective tools to reinforce the automatic decision-making processes of STT. Chandra et al.[7] proposed an SVM model based on supervised learning, which opens up a variety of ways and industrial communication technologies for 5G networks (such as factory automation). In Ref. [8], the KNN algorithm is applied to the clustering result to perform short-term predictions of the traffic flow vectors. Analysis of real world traffic data shows the effectiveness of these methods for traffic flow predictions, for they can capture the nonlinear information of traffic flows data and predict traffic flows on multiple links simultaneously.

By mixing different models and using the method of majority voting or average prediction probability, the final prediction results are often better than using only one model. Reference [9] introduced a general multi-model mixing method based on machine learning, and the typical machine learning models include Random Forests (RF)[9−11], Gradient Boosting Regression tree (GBR)[12], Recursive Partitioning And Regression Tree (RPART)[13], SVM[14−16], and Multi-Layer Perception (MLP)[17]. Experiments show that compared with the single optimal model, the model mixing effect is better, and the prediction accuracy of photovoltaic power generation has been improved by more than 30%[18].

A single model has certain limitations, as it can only capture partial features of the data, hence the prediction results cannot meet the demand for complex dynamics. Ensemble learning combines different features learned by each weak supervision model, so as to obtain a more comprehensive and stronger supervision model[19]. Ground-based cloud images and satellite cloud images are also used in photovoltaic power forecasts. The thickness of the cloud layer is calculated according to the texture and gray level of the cloud image, so as to calculate the attenuation of cloud layers to irradiance and build the attenuation model. Combined with the cloud movement model and irradiance attenuation model, the minute power prediction of photovoltaic power station is realized.

In the research of generation power prediction based on meteorological features, current existing works only use meteorological features or only use historical generation power as the features of the input model, but the fusion method of different characteristics will directly affect the accuracy of photovoltaic generation power prediction. In this paper, the representative model of photovoltaic power prediction is selected to study the impact of different feature fusion methods on the generation power prediction of photovoltaic power stations.

Next, the Recurrent Neural Network (RNN) models in the field of time series prediction are introduced, and then the theoretical basis of the machine learning models involved in the research of multi-features fusion for short-term photovoltaic power prediction is introduced in this paper, including Long Short-Term Memory (LSTM), Temporal Convolutional Network (TCN), eXtreme Gradient Boosting (XGBoost), etc.

2.1 RNN model

The traditional neural network has some inherent problems for photovoltaic power prediction. Firstly, it can not reflect the correlation within time series data; secondly, the input and output length should be fixed; thirdly, the features learned from different locations are not shared, resulting in a huge number of network training parameters. RNN partly solves the above problems of traditional neural networks.

RNN is a type of neural network used to process sequence data. It traverses all sequence elements and saves a state, which contains the information of the previous content. The output of the model depends on the information saved in the previous state. Theoretically, the RNN model can remember all the time step information. When the model is applied to tasks with a long sequence, it needs to comprehensively consider the information of the previous multiple time steps. However, the structure of the RNN is prone to causing gradient disappearance, which leads to poor accuracy of long-term prediction.

2.2 LSTM model

A memory cell is designed in the LSTM network structure. The LSTM model does not remember all the information, and it only needs to selectively remember important information and forget unimportant

information, so as to reduce the burden of memory and effectively alleviate the problem of long-term dependence. The LSTM controls the long-term and short-term memory through gating, which avoids memorizing all information in the RNN, so it is suitable for long sequence problems.

In traditional RNN, the disappearance of gradients leads to short-term memory. In LSTM, the model introduces memory cells whose values are dependent on the values of the previous memory cell and the update term weighted by the input gate. In addition, the model also introduces a forget gate to control the infinite growth of memory cells, so as to facilitate the network to learn to remember or forget, and avoid the problem of gradient disappearance. In Ref. [20], a spatial-temporal Convolutional Long Short-Term Memory (ConvLSTM) model is proposed to predict the vehicle's lateral and longitudinal driving intentions simultaneously. This network includes two modules: the first module mines the information of the target vehicle using the LSTM network and the second module uses ConvLSTM to capture the spatial interactions and temporal evolution of surrounding vehicles simultaneously when modeling the influence of surrounding vehicles. The model is trained and verified on a real road dataset, and the results show that the spatial-temporal ConvLSTM model is superior to the traditional LSTM in terms of accuracy, precision, and recall, which helps improve the prediction accuracy at different time horizons. Li and Ye[21] proposed a wireless network traffic prediction model based on long-term and short-term memory cyclic neural networks. Through simulation experiments, the throughput prediction of 5G wireless networks using different scheduling algorithms for many different types of services is studied. Simon et al.[22] verified that the long short-term memory prediction model has acceptable prediction accuracy and training speed, meets the needs of wireless network traffic prediction, and has a good application prospect.

2.3 TCN model

TCN is a special Convolution Neural Network (CNN), which uses one-dimensional convolution in combination with the one-dimensional features of series data. In addition, TCN applies causal convolution, residual connection, and expansion convolution in model design, in a creative way that effectively avoids the problem that traditional convolution neural networks cannot capture information for a long time. In a specific task, the effect of TCN can often reach or even exceed that of RNN[23, 24]. In Section 4, we compare TCN with other neural network models according to different feature fusion methods.

2.4 XGBoost model

The XGBoost model is widely used as a very effective machine learning method. It is an additive model consisting of the base learners, each of which is based on the Classification And Regression Tree (CART). Each base learner can be optimized step by step using the forward distribution algorithm. The forward distribution algorithm adopts a greedy strategy to optimize tree by tree. In the process of finding the optimal tree, although the greedy algorithm can achieve good results, in the case of a relatively large amount of data, the time complexity is too high and it is very resource intensive. Therefore, the XGBoost algorithm adopts an approximate algorithm to avoid traversing all the points and effectively reduce the time complexity.

3 Data preprocessing and analysis of photovoltaic station

This section first introduces the parameters and their meanings of the data used in experiments, then detects and identifies the abnormal points before using the data interpolation method to repair the abnormal and missing points. Next, we analyze the correlation of the repaired data, and relevant features with a high correlation with the power generation. Finally, according to the selected meteorological features and historical power generation, three feature fusion methods are designed to organize the data into three datasets. The data preprocessing structure is shown in Fig. 1.

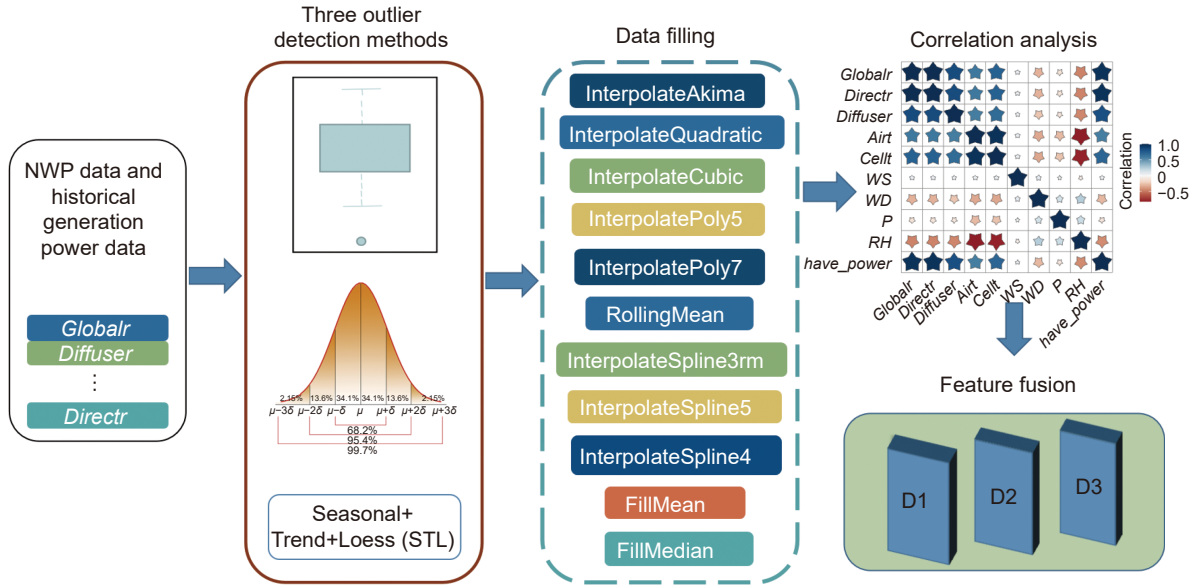

Fig. 1 Overall diagram of data preprocessing.

3.1 Photovoltaic data preparation

This paper uses the photovoltaic data of 18 power stations in Northwest China, including the actual power generation data of photovoltaic power stations and the Numerical Weather Prediction (NWP) data of the respective photovoltaic power stations. The data are collected from May 25, 2020 starting at 9:00:00 to June 30, 2021 ending at 20:55:00, with an interval of 5 minutes. NWP data include global radiation (*Globalr*), direct radiation (*Directr*), diffuse radiation (*Diffuser*), air temperature (*Airt*), ambient temperature (*Cellt*), wind speed (*WS*), wind direction (*WD*), pressure (*P*), relative humidity (*RH*), and *have_power* indicates whether there is output power. The photovoltaic power data (*power*) are discontinuous data, and the length of missing data between continuous data segments varies, some as long as five days, and some less than or equal to 2 hours. The longest continuous period length, secondary length, total length, and missing rate of data of each power station are summarised in Table 1.

- The original data collected by the photovoltaic power stations can be directly used for a short period of time, mostly 1800 continuous data points, equivalent to about continuous data points of 6.25 days, with a small amount of data. Photovoltaic power prediction requires a lot of training data to learn a generalized model. Through observation, the missing rate of data is about 5.00%, which needs to be repaired and filled to expand the amount of continuous data.
- The data of meteorological features *Globalr*, *Directr*, *Diffuser*, *WD,* and *Power* data fluctuate greatly, mainly because the intermittence and fluctuation of solar resources and cloud motion shielding affect the solar irradiation. The fluctuation of data brings some difficulties to power prediction.
- The overall change trend of meteorological features *Globalr*, *Directr*, *Diffuser,* and *Power* data is similar and has a high correlation, whereas the correlation of other meteorological features needs to be calculated.

Table 1 Statistics of continuous time periods and missing rate of some power station data.

Power station number	Length of the longest continuous time period (point)	Length of the second longest continuous time period (point)	Total length of data points	Missing rate (%)
cxjtgf	1836	1139	50 841	3.37
dhmsgf	1836	1139	48 637	4.30
dhztg2	1835	1139	50 797	5.62
dhztgf	1836	1139	50 800	5.57

Noto: 5 minutes interval between two adjencent points.

● There are many abnormal points in the dataset. For example, the *Directr* is 0 for a long time at night, which is not consistent with the trend of the data. In addition, some data exceed the historical maximum or minimum values. The abnormal data will affect the training of the subsequent model, and directly affect the accuracy, robustness, and generalization ability of the models.

3.2 Data repair

3.2.1 Outlier detection

During the data collection process, due to factors, such as equipment failure, human error, environmental influence, etc, abnormal points appear in the data. These abnormal points will affect the prediction accuracy of the follow-up models. In order to reduce the impact of outliers on model training and prediction, outliers are detected and repaired in the original data. The results of exemplar outliers detection are shown in Fig. 2.

In order to accurately and comprehensively detect data outliers, this paper uses three methods to detect outliers.

(1) Box diagram detection

The outliers detected by the box diagram are extreme values in the whole dataset. We need to find outliers by choosing a benchmark, including mild outliers and extreme outliers. Benchmark selection is based on sample maximum, sample minimum, sample median, lower quartile, upper quartile, and InterQuartile Range (IQR). Mild outliers are any value 1.5 times greater than the IQR based on all remaining values and limit outliers are any value 3 times greater than the IQR based on all remaining values

(2) 3-sigma criteria

Assuming that the data obey the normal distribution, we can calculate the mean μ and standard deviation δ of the data, and take $x=\mu$ as the symmetry axis of the image. The probability of data distribution in $(\mu-\delta, \mu+\delta)$ is 0.6828, that in $(\mu-2\delta, \mu+2\delta)$ is 0.9545, and that in $(\mu-3\delta, \mu+3\delta)$ is 0.9973. Almost all the data are concentrated in the $(\mu-3\delta, \mu+3\delta)$ interval, and the data beyond this range are regarded as outliers.

(3) STL

STL method can fully consider the seasonality, trend change, and residual error of data. Firstly, the data period is calculated by Fourier transform[25], in which there are two splitting principles, addition and multiplication. Then according to the length of the time period as a time window, a moving average is performed, and the trend data of the original data can be obtained by decomposing the data. Finally, we can analyze the decomposed residual data and use 3-sigma criteria method to find abnormal points.

3.2.2 Data interpolation

In this paper, 13 rule-based methods are used to fill in abnormal data and missing values. Taking the data of

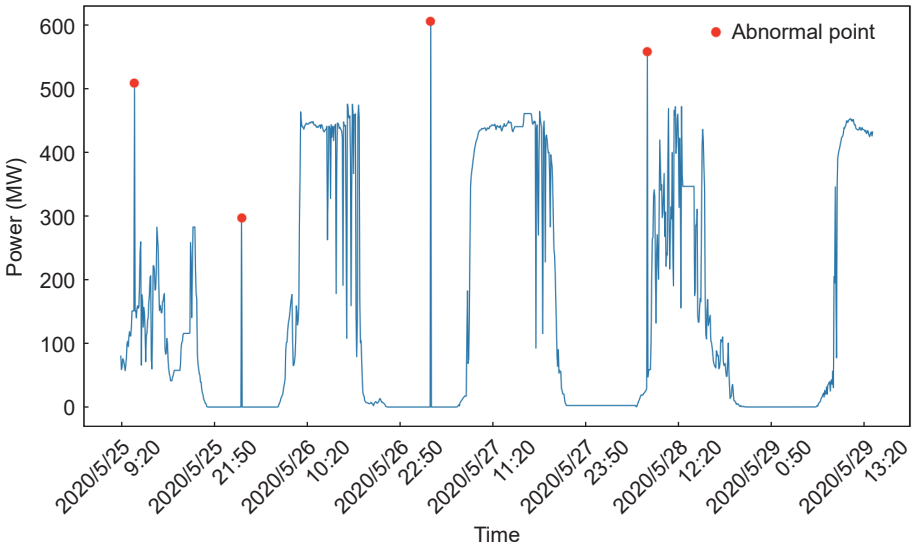

Fig. 2 Example diagram of detected abnormal points.

power station "qnxgh" for 18 consecutive days as an example, the data missing rates are set to 25%, 35%, and 40% seperately, and R2, Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are taken as the measurement indicators,

$$\mathrm{R2} = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \tag{1}$$

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{3}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \hat{y}_i)| \tag{4}$$

where m represents the number of samples, y_i represents the real value, and $\hat{y}_i$ represents the predicted value. The evaluation results are shown in Table 2. It can be seen from Table 2 that Interpolatetime method has the best effect on goodness of fit, and the Akima filling method also performs well. With the increase of the deletion rate, the goodness of fit of each method decreases, indicating that the data with a high deletion rate are not credible after data filling. Therefore, in this paper, the missing time periods are divided into two cases: more than 2 hours, and less than or equal to 2 hours.

Only the missing time periods of less than 2 hours are repaired by Interpolatetime method. The available data length after data filling is shown in Table 3.

3.3 Correlation analysis

In this paper, the Pearson correlation coefficient is used to observe the correlation between variables.

Pearson correlation coefficient is a linear correlation coefficient. It assumes that the data belong to the normal distribution to calculate the similarity between vectors. The output range is [−1, 1]. Positive values are positive correlations and negative values are negative correlations[26]. The results of the Pearson correlation coefficient between meteorological features and power generation are shown in Fig. 3.

Through analysis, the correlation between *Globalr*, *Directr*, *Diffuser,* and *power* is 0.98, which is highly correlated. There is a significant correlation between *Cellt* and *power* , and there is a relatively low correlation between *Airt*, *WD*, *RH*, and *power*. Among them, there is a positive correlation between *Airt* and *power*, *RH*, *WS*, and *power* are negatively correlated, while the correlation value between *WS* and *P* is too small, which is not correlated. To sum up, *Globalr*, *Directr*, *Diffuser*, *Cellt*, *Airt*, *WD*, and *RH* are used as the inputs of the model for the training of subsequent models.

3.4 Different feature fusion methods

The specific data format of the preprocessed data is shown in the following:

$$\begin{pmatrix} x_{0,0} & x_{0,1} & \cdots & x_{0,n} & y_{0,n} \\ x_{1,0} & x_{1,1} & \cdots & x_{1,n} & y_{1,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{m,0} & x_{m,1} & \cdots & x_{m,n} & y_{m,n} \end{pmatrix} \tag{5}$$

Table 2 Evaluation of different deletion rates for different data filling methods.

Missing data rate of power station qnxgh (%)	Method of filling missing values	Indicator			
		R2	MSE	RMSE	MAE
25	**InterpolateTime**	**0.997 310 496**	**0.011 807 172**	**0.108 660 813**	**0.022 813**
	InterpolateAkima	0.996 864 355	0.013 765 773	0.117 327 630	0.024 253
	InterpolateQuadratic	0.994 325 933	0.024 909 678	0.157 828 001	0.029 977
35	**InterpolateTime**	**0.994 298 772**	**0.025 028 917**	**0.158 205 300**	**0.036 102**
	InterpolateAkima	0.993 923 386	0.026 676 899	0.163 330 600	0.036 948
	InterpolateQuadratic	0.992 357 461	0.033 551 451	0.183 170 551	0.043 192
40	**InterpolateTime**	**0.994 113 770**	**0.025 841 094**	**0.160 751 654**	**0.040 968**
	InterpolateAkima	0.993 928 212	0.026 655 709	0.163 265 762	0.041 695
	InterpolateQuadratic	0.991 905 638	0.035 534 995	0.188 507 281	0.049 247

Table 3 Available data length after partial filling.

Power station name	Total length of data points	Deletion rate (%)	Available data length after filling
cxjtgf	50 841	3.37	42 567
dhmsgf	48 637	4.30	56 895
dhztg2	50 797	5.62	49 930
dhztgf	50 800	5.57	50 792
hdjygg	42 579	4.22	42 572
hhsdgf	39 344	4.63	39 335
hhsdww	41 461	4.69	41 453
hrwwgf	53 981	4.80	53 971

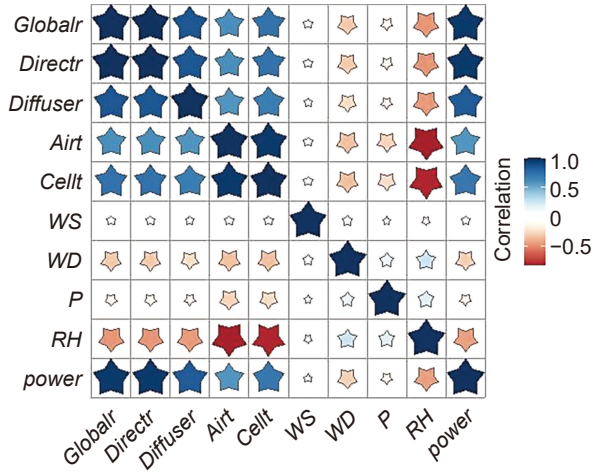

Fig. 3 Pearson correlation coefficient between meteorological features and power generation.

where each row represents the data value of a time point, which is 5 minutes different from the time point of the previous row. There are m rows and n columns in total. It represents the meteorological features value of the first row and the generating power value of the second row. This paper considers three data fusion methods to form the different datasets:

● The meteorological features of the previous day are used as features to predict the power generation of the next day. The dataset obtained according to this feature fusion method is defined as dataset D1.

● It is considered to input the meteorological features and power generation of the previous day into the model as features to predict the power generation of the next day. The dataset based on this feature fusion method is defined as D2.

● The generation power of the previous day predicts the generation power of the next day, and the dataset is defined as D3 according to this feature fusion method.

According to different feature fusion methods, the data are sorted into three datasets, namely D1, D2, and D3. In this paper, TCN, LSTM, stacked LSTM, Bi-LSTM, and XGBoost models are used to test the impacts of different feature fusion methods on model prediction.

4 Multi-features fusion photovoltaic power prediction and result analysis

In this section, Gaussian filter is used for data stationarity processing for the three datasets obtained in Section 3. The processed data are divided into training dataset and test dataset with a ratio of 9:1 for model training, and the impact of different feature fusion methods on photovoltaic power generation prediction is analyzed. All experiments in this paper use RMSE, MAE, MSE, R2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) to evaluate the model. The formulas for AIC and BIC are detailed below:

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{6}$$

$$\mathrm{BIC} = k\ln(n) - 2\ln(L) \tag{7}$$

where k is the number of model parameters, n is the number of samples, and L is the likelihood function. The experimental environment is set as Nvidia Tesla V100 GPU and TensorFlow 2.7.0.

4.1 Experiments of LSTM

In order to measure the overall performance of LSTM, we compare and analyze the evaluation indexes of each power station in each dataset. The results are shown in Table 4.

From Table 4 we can conclude that:

● The values of RMSE, MAE, and MSE of the model in D2 are significantly lower than that in D1. The R2 value in D2 dataset is larger, indicating that the model fits better, and the AIC and BIC are smaller than those on D1.

● From the evaluation indicators, the model can also achieve a good prediction effect on D3. If there is a lack of equipment deployment to collect data ending with insufficient data, it also has a certain reference value to use only the historical power generation to

Table 4 Experimental results of LSTM for three datasets.

Dataset name	RMSE	MAE	MSE	R2	AIC	BIC
D1	24.4939	9.8197	818.9468	0.9492	12658.42	12668.76
D2	**0.5632**	**0.2415**	**0.8158**	**0.9999**	**284.05**	**275.54**
D3	2.1785	1.3901	5.6741	0.9996	2098.22	2106.69

predict the power generation in the future.

● On the D1 dataset, there is a large difference between the predicted values of the model and the real values, which shows that there is a lack of reliability to predict the power generation at the future time if only using the characteristics of meteorological features.

4.2 Experiments of stacked LSTM

The results after taking the average value of each evaluation index of 18 power stations are shown in Table 5. The specific analysis is as follows:

● In D2 dataset, the values of RMSE, MAE, and MSE of the stacked LSTM model are generally lower than those of the models in D1 and D3 datasets. Compared with the performance on D1 dataset, the values of RMSE, MAE, and MSE are decreased by 97.46%, 97.35%, and 99.87%, respectively, and the R2 value is increased by 5.31%.

● In the D1 dataset fused with the characteristics of various meteorological features, the values of RMSE, MAE, and MSE of the stacked LSTM model are significantly higher than those on the D3 dataset, which indicates that there are certain limitations in using only meteorological features as feature input, and the historical power generation plays a certain role in the later stage prediction of the model, and it also indicates that the historical power generation has a strong correlation with the power generation at the current time. Its correlation is higher than the features of meteorological features.

4.3 Experiments of Bi-LSTM

Bi-LSTM is composed of LSTM layers with opposite information transmission directions. The first layer transmits information in chronological order, and the second layer transmits information in reverse chronological order. The output of Bi-LSTM is determined by the hidden layer state of the two LSTMs. The structure of Bi-LSTM is shown in Fig. 4.

In order to compare the prediction effects of Bi-LSTM in D1, D2, and D3 datasets, the experimental evaluation results obtained after averaging the prediction and evaluation indexes of 18 power stations in this section are shown in Table 6.

● From Table 6 we can observe that Bi-LSTM has the best prediction effect on D2 dataset, and the values of RMSE, MAE, and MSE are significantly lower than those on D1 and D3 datasets, indicating that the model effectively captures the correlation between the characteristics of meteorological features, historical power generation, and the current power generation.

● The R2 value of the Bi-LSTM model on D2 dataset is significantly higher than those on D1 and D3 datasets, indicating that the model has a high degree of fitting in the test set.

● The values of AIC and BIC of Bi-LSTM model on D2 dataset are lower than those on D1 and D3 datasets, which shows that the model effectively punishes the parameters that will increase the complexity of the model when training D2 dataset, thus making the model more concise and less complex.

● Bi-LSTM obtained the best evaluation index in D2 dataset. The model obtains the best evaluation value in 18 power stations in D2 dataset. The above experimental results confirm the important role of historical generation power in Bi-LSTM prediction of generation power.

● Compared with the LSTM model, the values of

Table 5 Experimental results of stacked LSTM for three datasets.

Dataset name	RMSE	MAE	MSE	R2	AIC	BIC
D1	24.5239	9.6926	817.7764	0.9495	12 653.60	12 663.93
D2	**0.6236**	**0.2570**	**1.0529**	**0.9999**	**−232.39**	**−223.88**
D3	2.3538	1.6207	7.2895	0.9995	2056.12	2064.58

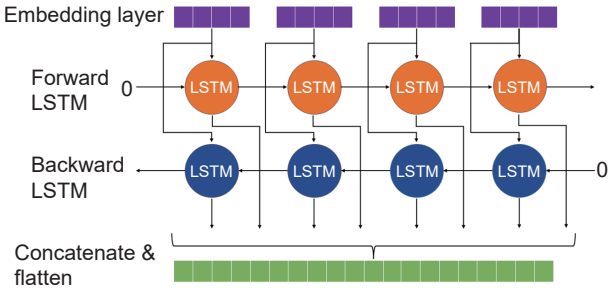

Fig. 4 Structure of Bi-LSTM.

MSE, AIC, and BIC of Bi-LSTM are lower, indicating that the model tends to choose the model with lower complexity. In addition, the overall evaluation indexes of Bi-LSTM are better than those of stacked LSTM, indicating that the characteristics of two-way learning in 18 power stations are more conducive to the prediction of power generation of photovoltaic power stations than those learned by model stacking.

4.4 Experiments of TCN

TCN network uses one-dimensional convolution and expansion convolution to realize the application of convolution kernel in the temporal problems. This section uses TCN to verify the impact of different feature fusion methods on generation power prediction. 70% of the data in datasets D1, D2, and D3 are used as training dataset, and 30% remainders are used as test dataset for experiments. The experimental results of TCN in each dataset are shown in Fig. 5 and Table 7, and we can draw the following conclusions:

- TCN performs best on D2 dataset, and the values of RMSE, MAE, and MSE are significantly lower than those on D1 and D3 datasets. In addition, R2 value of the model on the dataset is also the highest, indicating that the error between the predicted value and the real value is small and the fitting degree of the model is high. The values of AIC and BIC of TCN on D2 dataset are also the lowest, indicating that there are fewer explanatory variables and better fitting degree of the model on this dataset.
- In the experiment of TCN on D3 dataset, the values of RMSE, MAE, and MSE are lower than those on D1 dataset, and the difference values of RMSE, MAE, and MSE between D3 and D1 are small.
- TCN uses one-dimensional convolution and expansion convolution to realize the application of convolution neural network in time series prediction. However, looking at the experimental results between TCN and LSTM, actually TCN has not made a great breakthrough in the prediction of power generation of photovoltaic power plants. Compared with LSTM, the model does not respond well to the internal regular pattern of data.

4.5 Experiments of XGBoost

XGBoost algorithm filters the features according to the importance of the features, and selects the features with high importance as the input of the model. In this paper, the maximum number of trees generated is set to 5000, the learning rate is set to 0.01, and the value is set to 0. For each tree, the proportion of random sampling is set to 70%, which is 30% lower than the default value of the system, and this treatment effectively avoids the overfitting of the model. For each tree, the column sampling ratio is set to 60%, and the maximum depth of the tree is set to 4. In addition, XGBoost uses cross-validation in each iteration to obtain the optimal number of iterations.

Next, we take the prediction results that perform best among the experimental results of different feature fusion methods of each model to compare them with XGBoost algorithm. XGBoost algorithm is used to conduct the experiments in two scenarios: including the historical power generation characteristics of the power station and excluding the historical power generation characteristics of photovoltaic power station. The specific comparison results are shown in Table 8, XGBoost-P is XGBoost algorithm including historical generation power.

Table 6 Experimental results of Bi-LSTM for three datasets.

Dataset name	RMSE	MAE	MSE	R2	AIC	BIC
D1	24.4194	9.8253	813.5776	0.9496	12 651.06	12 661.40
D2	**0.5648**	**0.2826**	**0.6136**	**0.9999**	**273.89**	**265.38**
D3	2.5159	1.6084	8.4706	0.9994	2164.13	2172.59

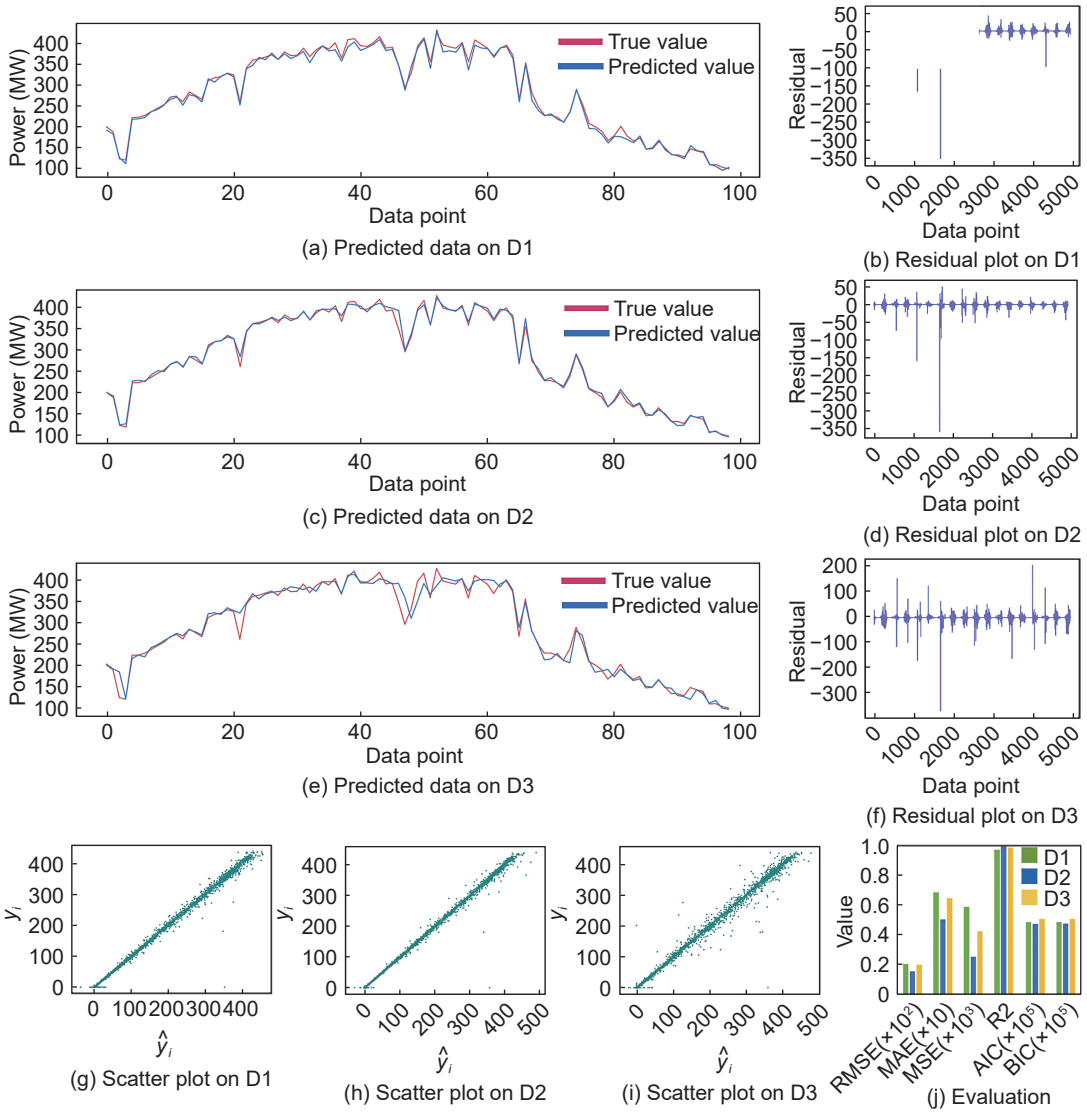

Fig. 5 Experimental results of TCN for three datasets, and the time between two adjencent points is 5 minutes.

Table 7 Experimental results of TCN for three datasets.

Dataset name	RMSE	MAE	MSE	R2	AIC	BIC
D1	19.9764	6.7851	582.1143	0.9633	48 036.92	48 050.03
D2	**15.1136**	**4.9897**	**249.3071**	**0.9854**	**46 928.96**	**46 942.06**
D3	19.4761	6.3841	419.9613	0.9766	50 036.73	50 049.83

Table 8 Experimental results of XGBoost and various models.

Model	RMSE	MAE	MSE	R2	AIC	BIC
LSTM	**0.5632**	**0.2415**	0.8158	0.9999	284.05	275.54
Stacked LSTM	0.6236	0.2570	1.0529	0.9999	**−232.39**	**−223.88**
Bi-LSTM	0.5648	0.2826	**0.6136**	0.9999	273.89	265.38
TCN	15.1136	4.9897	249.3071	**0.9854**	46 928.96	46 942.06
XGBoost	7.7436	4.0466	171.9377	0.9889	37 546.89	37 560.00
XGBoost-P	0.7704	0.4541	0.6580	0.9999	13 698.83	13 711.95

● LSTM, stacked LSTM, Bi-LSTM, and XGBoost-P algorithms have achieved good results on datasets.

● Compared with the method of meteorological features fusion only in LSTM, stacked LSTM, Bi-LSTM, and XGBoost, the error evaluation indexes (RMSE, MAE, and MSE) of the model are reduced by more than 97%, the values of AIC and BIC are increased by more than 60%, the error evaluation indexes (RMSE, MAE, and MSE) in TCN are reduced by more than 20%, and the values of AIC and BIC are increased by more than 97%.

● Compared with the method of historical generation power fusion only, the method of historical generation power and meteorological element feature fusion is also improved in the evaluation results of each model.

● The complexity of TCN model using one-dimensional convolution kernel is too high, and the fitting degree of the model to the training data is relatively high, but the generalization of new data on the test dataset is low. Therefore, the model can predict the overall change trend of the data, but the error is large for the data with large volatility.

5 Conclusion

This paper analyzes different feature fusion methods of meteorological features and historical power generation of photovoltaic power station, and explores the impact of different feature fusion methods on the power generation of photovoltaic power station.

Firstly, the box diagram method, 3-sigma criterion, and STL are used to detect the abnormal points in the photovoltaic data. Three datasets with the missing rate of 25%, 35%, and 40%, respectively, are filled. Finally, the optimal method InterpolateTime is selected to fill the missing data and abnormal points, which effectively expands the dataset.

Secondly, this paper conducts correlation analysis on the filled data and obtains the meteorological features with high correlation with power generation. Meanwhile, combined with the features of historical generation power, the data are sorted into three datasets by different feature fusion methods.

Finally, through three data fusion methods, the representative LSTM, stacked LSTM, Bi-LSTM, TCN, and XGBoost models are selected to conduct experiments on the datasets, and the experimental results are evaluated by six evaluation indexes, such as AIC for comparative analysis. Results show that the fusion method of historical power generation and meteorological elements can greatly improve the prediction accuracy of LSTM, stacked LSTM, Bi-LSTM, TCN, and XGBoost algorithms. All error evaluation indicators have decreased by more than 22.34%. The values of AIC and BIC are both increased by more than 2.30%.

In the future, we will consider combining image data to achieve multi-source heterogeneous data fusion. We can also analyze cloud shadow and cloud movement through ground-based cloud images and satellite cloud images, followed by combining meteorological features and historical power generation to perform feature fusion to predict the power generation of photovoltaic power plants in a more meaningful way.

Acknowledgment

This work was supported by the State Grid Gansu Electric Power Research Institute (Nos. SGGSKY00WYJS2100164 and 52272220002W).

References

[1] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, *Renew. Sustain. Energy Rev.*, vol. 124, p. 109792, 2020.

[2] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, Solar photovoltaic generation forecasting methods: A review, *Energy Convers. Manage.*, vol. 156, pp. 459–497, 2018.

[3] J. F. Bermejo, J. F. G. Fernández, F. O. Polo, and A. Cr. Márquez, A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources, *Appl. Sci.*, vol. 9, no. 9, p. 1844, 2019.

[4] C. H. Liu, S. C. H. Hoi, P. L. Zhao, and J. L. Sun, Online ARIMA algorithms for time series prediction, in *Proc. 30th AAAI Conf. on Artificial Intelligence*, Phoenix, AZ, USA, 2016, pp. 1867–1873.

[5] Y. H. Yu and J. Li, Residuals-based deep least square support vector machine with redundancy test based model selection to predict time series, *Tsinghua Science and*

Technology., vol. 24, no. 6, pp. 706–715, 2019.

[6] E. A. A. Alaoui, S. C. K. Tekouabou, S. Hartini, Z. Rustam, H. Silkan, and S. Agoujil, Improvement in automated diagnosis of soft tissues tumors using machine learning, *Big Data Mining and Analytics.*, vol. 4, no. 1, pp. 33–46, 2021.

[7] S. Chandra, Prateek, R. Sharma, R. Arya, and K. Cengiz, QSPCA: A two-stage efficient power control approach in D2D communication for 5G networks, *Intelligent and Converged Networks.*, vol. 2, no. 4, pp. 295–305, 2021.

[8] Y. D. Chen, Z. Yi, and J. M. Hu, Multi-dimensional traffic flow time series analysis with self-organizing maps, *Tsinghua Science and Technology.*, vol. 13, no. 2, pp. 220–228, 2008.

[9] M. Belgiu and L. Drăguţ, Random forest in remote sensing: A review of applications and future directions, *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, 2016.

[10] G. Biau and E. Scornet, A random forest guided tour, *Test*, vol. 25, no. 2, pp. 197–227, 2016.

[11] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, How many trees in a random forest? in *Proc. 8th Int. Conf. on Machine Learning and Data Mining in Pattern Recognition*, Berlin, Germany, 2012, pp. 154–168.

[12] H. W. Ma, X. Yang, J. R. Mao, and H. Zheng, The energy efficiency prediction method based on gradient boosting regression tree, in *Proc. 2nd IEEE Conf. on Energy Internet and Energy System Integration* (*EI2*), Beijing, China, 2018, pp. 1–9.

[13] C. Strobl, J. Malley, and G. Tutz, An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests, *Psychol. Methods*, vol. 14, no. 4, pp. 323–248, 2009.

[14] L. P. Wang, *Support Vector Machines*: *Theory and Applications*. Berlin, Germany: Springer, 2005.

[15] S. K. Sunori, P.B. Negi, K. Ghai, A. Mittal, M.C. Lohani, M. Manu, and P. Juneja, Prediction of air pollutant PM10 using various SVM models, in *Proc. of 2022 International Conference on Sustainable Computing and Data Communication Systems* (*ICSCDS*), doi: 10.1109/ICSCDS53736.2022.9760965.

[16] W. S. Noble, What is a support vector machine? *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.

[17] M. W. Gardner and S. R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, *Atmos. Environ.*, vol. 32, nos. 14&15, pp. 2627–2636, 1998.

[18] S. Y. Lu, Y. Hwang, I. Khabibrakhmanov, F. J. Marianno, X. Y. Shao, J. Zhang, B. M. Hodge, and H. F. Hamann, Machine learning based multi-physical-model blending for enhancing renewable energy forecast-improvement via situation dependent error correction, in *Proc. European Control Conference* (*ECC*), Linz, Austria, 2015, pp. 283–290.

[19] O. Sagi and L. Rokach, Ensemble learning: A survey, *WIREs*, *Data Mining Knowl. Discov.*, vol. 8, no. 4, p. e1249, 2018.

[20] H. Huang, Z. N. Zeng, D. Y. Yao, X. Pei, and Y. Zhang, Spatial-temporal ConvLSTM for vehicle driving intention prediction, *Tsinghua Science and Technology.*, vol. 27, no. 3, pp. 599–609, 2022.

[21] L. L. Li and T. Ye, Research on throughput prediction of 5G network based on LSTM, *Intelligent and Converged Networks.*, vol. 3, no. 2, pp. 217–227, 2022.

[22] M. Simon, S. Rikka, S. Nõmm, and V. Alari, Application of the LSTM models for baltic sea wave spectra estimation, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 83–88, 2023.

[23] S. J. Bai, J. Z. Kolter, and V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv: 1803.01271, 2018.

[24] J. Fan, K. Zhang, Y. P. Huang, Y. F. Zhu, and B. P. Chen, Parallel spatio-temporal attention-based TCN for multivariate time series prediction, *Neural Comput. Appl.*, doi: 10.1007/s00521-021-05958-z.

[25] N. Norman, The Fourier transform method for normalizing intensities, *Acta Crystallogr.*, vol. 10, no. 5, pp. 370–373, 1957.

[26] J. Benesty, J. D. Chen, and Y. T. Huang, On the importance of the Pearson correlation coefficient in noise reduction, *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 4, pp. 757–765, 2008.

Xiaorun Tang received the BEng degree from Tianjin Normal University, China in 2021. She is currently a master student at Lanzhou University. Her main research interests include machine learning and deep reinforcement learning.

Ming Ma is a senior engineer at the State Grid Gansu Electric Power Research Institute. His research interests include new energy and artificial neural network.

Qingquan Lv received the MEng degree in computer science and technology from Lanzhou University, China in 2012. Now he is a PhD candidate at Lanzhou University. His research interests include new energy, machine learning, and artificial neural network.

Jun Shen received the PhD degree from Southeast University, China in 2001. He held positions at Swinburne University of Technology in Melbourne and University of South Australia in Adelaide before 2006. He is a professor at the School of Computing and Information Technology, University of Wollongong, Australia, where he has been the head of Postgraduate Studies, and chair of School Research Committee since 2014. He is a senior member of three institutions: IEEE, ACM, and ACS. He has published more than 120 papers in journals and conferences in CS/IT areas. His expertise includes computational intelligence, web services, cloud computing, and learning technologies including MOOC. He has been an editor, PC chair, guest editor, PC member for numerous journals and conferences published by IEEE, ACM, Elsevier, and Springer. He is also a current member of ACM/AIS Task Force on Curriculum MSIS 2016.

Baixue Zhu received the BEng degree from Gansu Agricultural University, China in 2019. She is currently a master student at Lanzhou University. Her research interests include machine learning and deep learning.

Jinqiang Wang received the MEng degree from Lanzhou University, China in 2019. He is currently a PhD candidate at Lanzhou University. His main research interests include machine learning, deep reinforcement learning, and optimization theory.

Binbin Yong received the PhD degree in computer science and technology from Lanzhou University, China in 2017. Now he is an associate professor at the School of Information Science and Engineering, Lanzhou University. His research interests inclucle parallel computing of GPU, deep learning, and physics-informed neural networks.