IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 30, 2021, accepted May 10, 2021, date of publication May 13, 2021, date of current version May 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3080140

Ultra-Short-Term Prediction of Wind Power Based on Sample Similarity Analysis

CHANGXIN MIAO, HAO LI, XIA WANG, AND HENG LI
School of Electrical and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China
Corresponding author: Hao Li (lihao0209@cumt.edu.cn)

ABSTRACT Increasing the utilization rate of wind energy is of great significance to the improvement of energy structure, which is inseparable from the support of wind power forecasting (WPF) technology. However, it is well known that there is no certain WPF model suitable for all conditions, such as different regions or seasons. Therefore, instead of focusing on the combination of machine learning models in a specific scenario, this article proposes a two-stage modeling strategy of ''first classify and separately model, then perform pattern recognition'' from the perspective of sample similarity analysis. That is, in offline mode, the historical database is divided into multiple categories with different characteristics, and prediction models are established for each category respectively; in online mode, pattern recognition is carried out on the prediction sample to select the corresponding prediction model. In this way, the WPF problem is decomposed into two strongly related tasks: wind power mode classification and wind power numerical prediction. Furthermore, the coupling and connection between mode classification task and numerical prediction task are strengthened through the transfer learning of sample features. Around the above ideas, specific methods of how to classify, identify, and predict are proposed, which are two-level clustering, Convolutional Neural Network (CNN) classification model and Long Short-term Memory (LSTM) prediction models. Simulation results based on real-world datasets prove the effectiveness and superiority of the proposed hybrid model.

INDEX TERMS Wind power forecasting, similarity analysis, pattern recognition, mode classification, numerical prediction, CNN, LSTM.

I. INTRODUCTION

With the increasing scale of wind power integration, the randomness, volatility and intermittence of wind power have brought great challenges to the security, stability and economy of power system [1], [2]. Accurate wind power forecasting (WPF) is one of the crucial technologies to deal with this problem [3], [4].

The research of WPF is mainly divided into mechanism forecasting and data-driven forecasting, and there are essential differences between the two [5], [6]. Mechanism forecasting is based on the depiction of atmospheric motion [7]. According to the information of topography, geomorphology and meteorological environment, the model is established by using physical laws such as hydrodynamics and thermodynamics, focusing on the optimization of boundary conditions and physical solution rules. It has the characteristics of difficult modeling and large calculation load, so it has poor timeliness and is generally suitable for medium and long-term forecast [8], [9]. Data-driven WPF emphasizes finding inherent laws in historical data, using data mining methods and artificial intelligence algorithms to establish the mapping relationship between inputs and targets [10], which is generally applicable to ultra-short-term and short-term forecasting [11], [12]. The development process of data-driven WPF includes several major stages: statistical methods, traditional machine learning, and deep learning. Statistical methods represented by time series method, persistence model (PM) and Kalman filter [13], and early artificial intelligence algorithms represented by artificial neural network (ANN) [14], [15], random forest (RF) [16], and support vector machine (SVM) [17], [18], have their own application scenarios and limitations, but the common point is that most of them are shallow structures with insufficient generalization capabilities, which makes it difficult to describe complex nonlinear relationships. In recent years, deep learning has developed rapidly [19], [20], especially convolutional neural network (CNN) and recurrent neural network (RNN), which

The associate editor coordinating the review of this manuscript and approving it for publication was R.K. Saket.

have very eye-catching performance in various fields [21]. CNN has achieved great success in the field of computer vision [22], [23], and RNN has been widely used in the fields of voice recognition and power prediction due to the consideration of the sequence of time series [24], [25].

At present, the research of WPF mainly focuses on the combination of multiple forecasting models, a typical example is the combination of different neural networks, or the combination of signal decomposition technology and neural networks [26], [27]. However, in practice, it has been found that no simulation method can provide the best results for all conditions [28], [29]. Specifically, wind power in different regions, seasons, and weather conditions have different characteristics, and no certain model can achieve the best performance under all conditions. The essential reason for this problem is that the characteristics of the training set determine the training and selection of neural network parameters: when the samples of the training set have completely different and messy characteristics, it is difficult for the neural network to summarize general and adaptive law [30], [31]. Therefore, it is very necessary to analyze the similarity of samples: cluster the samples with similar characteristics into one category, and separate the samples with different characteristics [32], [33]. This is also one of the main tasks of this paper.

For WPF, the similarity of samples is reflected in the similarity of wind power characteristics or weather conditions [34], [35]. Initially, in [36], Liu *et al.* proposed to search the historical database for training samples with similar features to the prediction sample every time the prediction task was performed. However, every forecasting operation must search the entire database, which not only consumes computing resources, but also is not conducive to the timeliness of ultra-short-term prediction. The two-stage modeling method of "first classify and separately model, then perform pattern recognition" is a more efficient modeling strategy, that is, the historical database is first divided into multiple categories with different characteristics, and prediction models are established for each category respectively; then pattern recognition is carried out on the prediction sample to select the corresponding prediction model. This method can avoid repeated searches on the database and greatly improve the response speed of the model. In [37]–[39], based on weather information, historical dataset was allocated to several clusters, and then prediction models were established for different weather scenarios. Finally, the weather of the prediction sample was compared with each scenario, and the corresponding prediction model was selected. In [36], the trend of wind power was expressed as a tuple vector by piecewise linearization, and then classification and recognition were performed based on the tuple vector. In [40], [41], the historical database was divided into multiple subsets with different mathematical morphologies through data mining and K-means cluster analysis, and corresponding prediction models were established for each subset, highlighting the importance of selecting similar samples [42], [43]. Although the above studies have different standards for sample similarity, their central idea is the same, that is, to group samples with similar characteristics into one category, and then model each category separately.

Sample similarity analysis is an effective method to improve the accuracy of WPF models, but there are still gaps and deficiencies in current research:

1) Wind power mode classification is different from target recognition in the field of computer vision, such as the classification and recognition of traffic lights, the classification and recognition of cats and dogs. Green lights and red lights, cats and dogs, have clear definitions and can be accurately labeled through manually labeling samples, and then used to train the classification model. But how to meaningfully define the categories of wind power and get a sufficient number of labeled samples?

2) Not only the definition of wind power categories, but also the identification method of prediction samples is very important. The mainstream method to identify the sample category is to compare the distance or similarity between the sample and the centroid of each category, but obviously this type of method does not pay enough attention to the feature mining of wind power data.

3) Based on the sample similarity analysis, the WPF problem is decomposed into two highly related tasks: wind power mode classification and wind power numerical prediction. The current research does not pay attention to the strong correlation between the two. It is worth exploring whether mode classification task can provide valuable knowledge for numerical prediction task.

In view of the above problems, the main work of this article can be summarized as follows:

1) A two-level clustering method is proposed to divide the wind power dataset into multiple categories with different characteristics, and provide labeled samples for the training of classification model. The two-level clustering firstly divides the dataset into low-power, medium-power and high-power categories according to average power value, and then further subdivides each power level according to trend similarity.

2) In offline mode, a CNN classification model is trained with the labeled samples generated by the two-level clustering. In online mode, the CNN classification model is applied on the test set for wind power mode recognition. Compared with the shallow method of calculating the distance between the sample and the centroid of each category, CNN can perform in-depth feature analysis on the samples.

3) There is a strong coupling and correlation between the wind power mode classification and wind power numerical prediction. Based on the idea of transfer learning, the mode classification model is regarded as a pre-training model for numerical prediction model. Specifically, the sample classification features mined by the CNN are submitted to the numerical prediction task as supplementary knowledge to help the training of the long short-term memory (LSTM) prediction models. Moreover, the method of how to use CNN

features to expand the input vector information of LSTM is also discussed, including dimension direction and time direction.

4) The simulation experiment is carried out based on the real-world wind power datasets, and the results show that the method based on sample similarity analysis proposed in this paper can effectively improve the accuracy of WPF.

The paper is organized as follows. Section 2 presents the methods and related theories proposed in this article. Section 3 introduces the experimental settings and provides experimental results with detailed analysis. The conclusion is drawn in section 4.

II. METHODOLOGY

The algorithm flow of the WPF model proposed is shown in Fig.1, which includes several main steps: generating sample sets, manually labeling samples, establishing CNN classification model and LSTM prediction models, as shown in Fig.2.

A. GENERATING SAMPLE SETS

In this paper, the resolution of wind power data is set to 15min, that is, there are 4 data points per hour and 96 data points per day. The power is expressed by the standard unit value, and the normalization method is as below.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

As shown in Fig.2 Step1, a power sequence with L1 width is defined as feature window, and L2 is defined as a sliding stride to slide the historical power into segments to form sample sets. It should be noted that the width L1 of the feature window is fixed during sliding. Specifically, the feature window of the first sample selects the [1, 2, . . . , L1] data of historical data, the feature window of the second sample selects the [(1+L2), (2+L2), . . . , (L1+L2)] data of historical data, and so on. When L2 < L1, there are overlapping data between samples, and the smaller L2, the more overlapping data. When L2 = L1, there is no overlap between samples. When L2 > L1, there is no overlap between samples, and some data are discarded.

A common practice is to take every day as a sample, corresponding to L1 = 96 and L2 = 96. But the disadvantages of this approach are: 1) Wind power does not have obvious daily cycle characteristics; 2) The feature window of the samples is too wide and the data do not overlap, resulting in too few samples generated (365 samples per year), which is not conducive to the training of the prediction model, and is not applicable to new wind farms that lack historical data. Therefore, for short-term prediction, we set L1 = 32 and L2 = 8, that is, the feature window is set to 8 hours before the prediction period, and the sliding stride is set to 2 hours.

B. MANUALLY LABELING SAMPLES

1) TWO-LEVEL CLUSTERING

In order to classify samples accurately and efficiently, as shown in Fig.2 Step2, a two-level clustering method is proposed in this paper. Firstly, according to the average power in feature window of each sample, the samples are divided into three categories: low-power, medium-power and high-power. Then, each power level is finely classified according to the sample similarity. Two-level clustering follows two basic principles: (m is the number of categories)

1) m cannot be too large, otherwise it will lead to over-subdivision and too few samples in each category, making the prediction models unable to be trained effectively.

2) m cannot be too small, otherwise samples in the same category will still have large feature differences, resulting in poor classification performance.

Based on the above two principles, the selection of m is determined according to the scale of the database and the characteristics of the wind power data.

2) EVALUATION INDEX OF CLUSTERING

A distance function is defined to measure the sample similarity:

$$L_p(x, y) = \|x - y\|_p = (\sum_i |x_i - y_i|^p)^{1/p} \tag{2}$$

When $p = 1, p = 2, p = \infty$, it denotes Manhattan distance, Euclidean distance and maximum distance, respectively. In this paper, Euclidean distance is defined as the classification index:

$$L(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{3}$$

S_{ave} and S_{min} respectively refer to the average number of samples contained in each category, and the smallest number of samples in all categories. They reflect whether each category contains a sufficient number of samples.

$$S_{ave} = \text{average}\{S_1, S_2, \ldots, S_m\} \tag{4}$$

$$S_{min} = \min\{S_1, S_2, \ldots, S_m\} \tag{5}$$

where m is the number of categories, and S_i represents the number of samples in category i.

L_{max} represents the maximum distance between samples belonging to the same category, used to evaluate the similarity of samples.

$$L_{max} = \max\{L_1, L_2, \ldots, L_m\} \tag{6}$$

where L_i represents the maximum distance between samples of category i.

3) K-MEANS ALGORITHM

As a data mining method, k-means clustering algorithm can effectively identify sequences with similar features, and is widely used in WPF [44]. The main steps of the algorithm are shown in Algorithm 1.

C. CNN CLASSIFICATION MODEL

As a kind of deep networks, CNN has been successfully applied in the field of computer vision [45]. In this article,

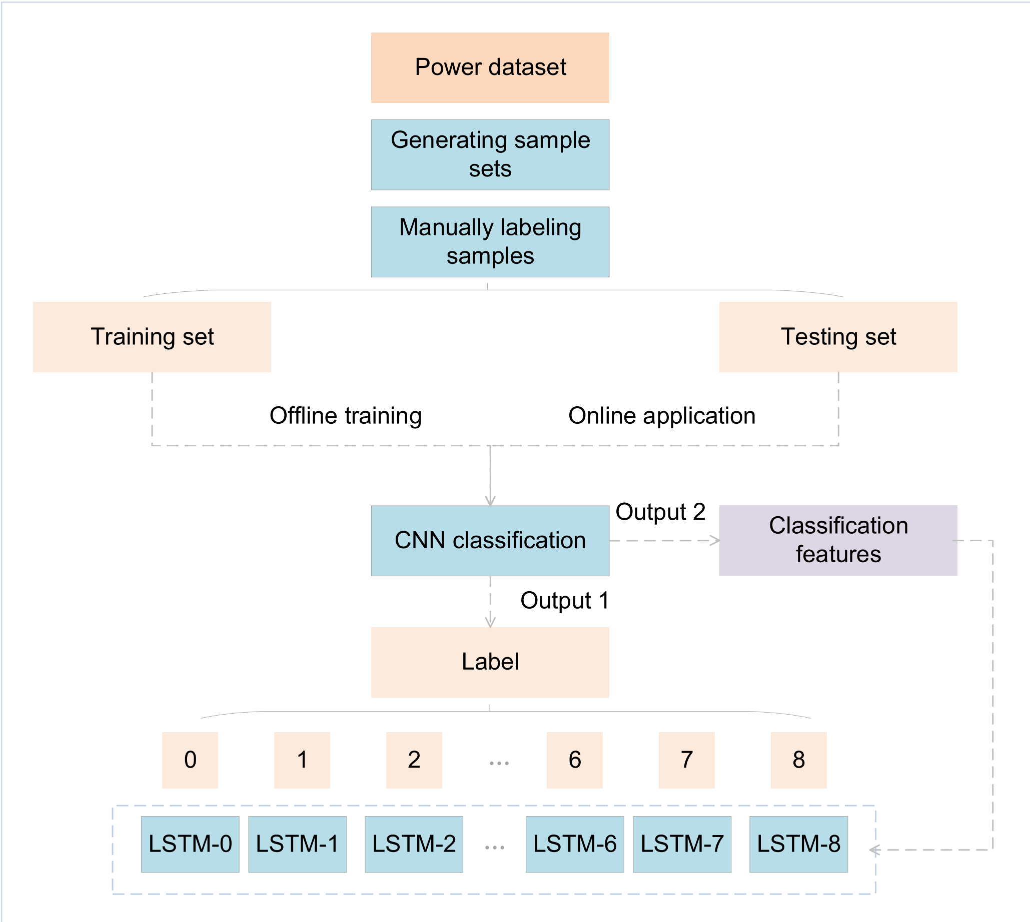


FIGURE 1. **Flow diagram of the model.**

Algorithm 1 k-means algorithm:

Input: Sample set $S = \{x_1, x_2, \ldots, x_n\}$
Randomly select k samples as initial centroid vector $C = \{c_1, c_2, \ldots, c_k\}$;
for every sample x_i
    calculate the distance between x_i and c_j;
    find the nearest centroid c_j and let $u_i = j$ where u_i represent the corresponding cluster of x_i;
end for
$U = \{u_1, u_2, \ldots, u_k\}$;
for every cluster:
    calculate the centroid of every cluster v_i;
end
$V = \{v_1, v_2, \ldots, v_k\}$, $U = \{u_1, u_2, \ldots, u_n\}$;
if $V = C$:
    return U;
else:
    $C = V$ and go to step 2;
end if
Output: vector U which shows the cluster label of every sample.

CNN is designed to perform pattern recognition and feature mining on wind power signals. Different from other time series classification methods that require artificially designed features, CNN can automatically mine and extract the internal structure of input data and generate underlying features aiming at the target. Moreover, CNN has strong robustness to translation, scaling, and rotation, because it has three important ideas different from traditional feedforward neural networks: local receptive fields, weight sharing, and pooling operations.

As shown in Fig.2 Step3, a CNN that performs time series classification tasks usually consists of two parts: in Part 1, multiple convolutional layers and pooling layers are stacked to generate underlying features of the samples; in Part 2, these features are connected to the fully connected (FC) layer to perform classification.

D. LSTM PREDICTION MODEL

As shown in Fig.2 Step 4, LSTM consists of an internal memory cell and three multiplicative gates, including the forget gate, input gate, and output gate. Each time, after receiving the input x_t, LSTM updates its internal state c_t by using

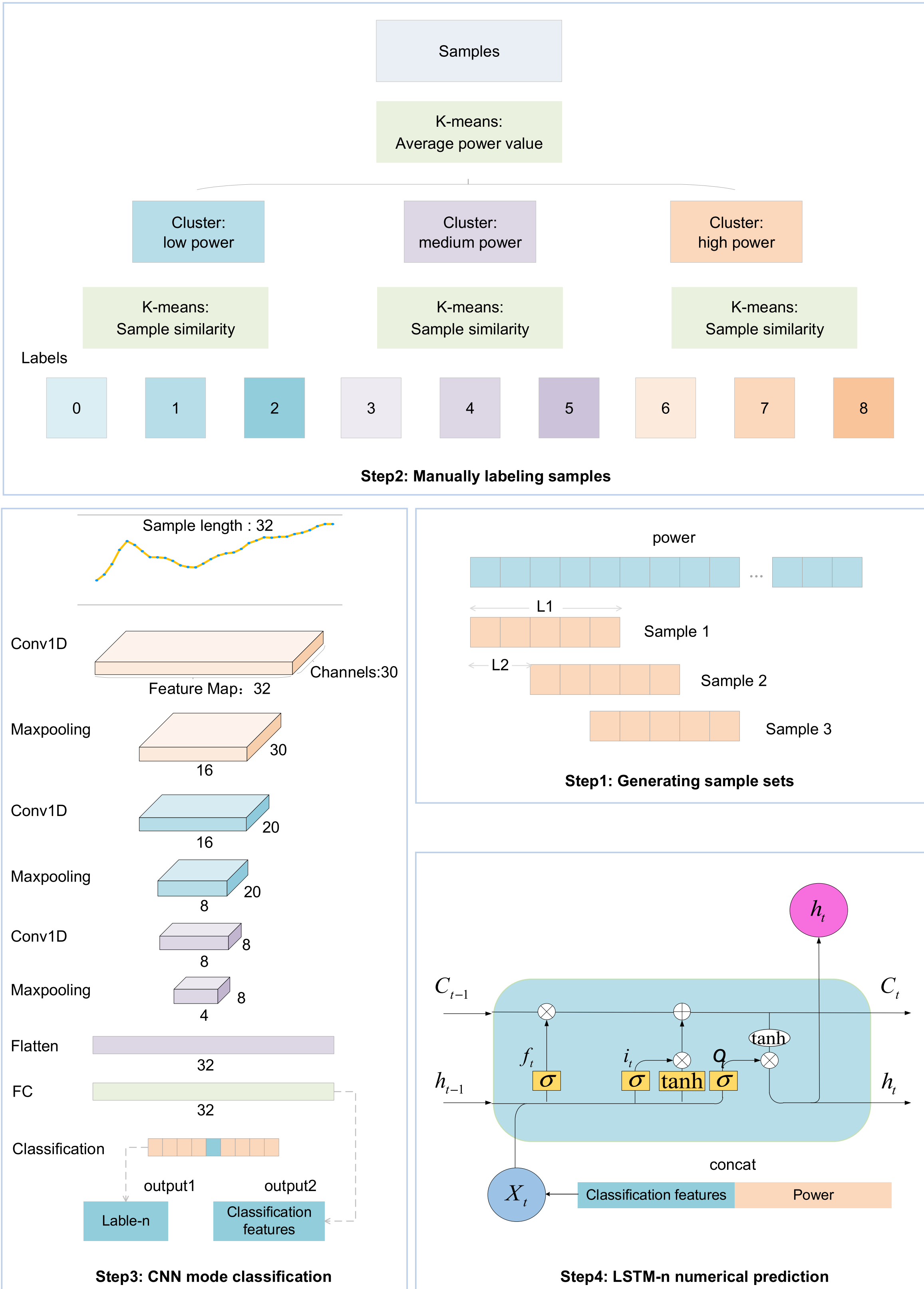


FIGURE 2. Main steps of the model.

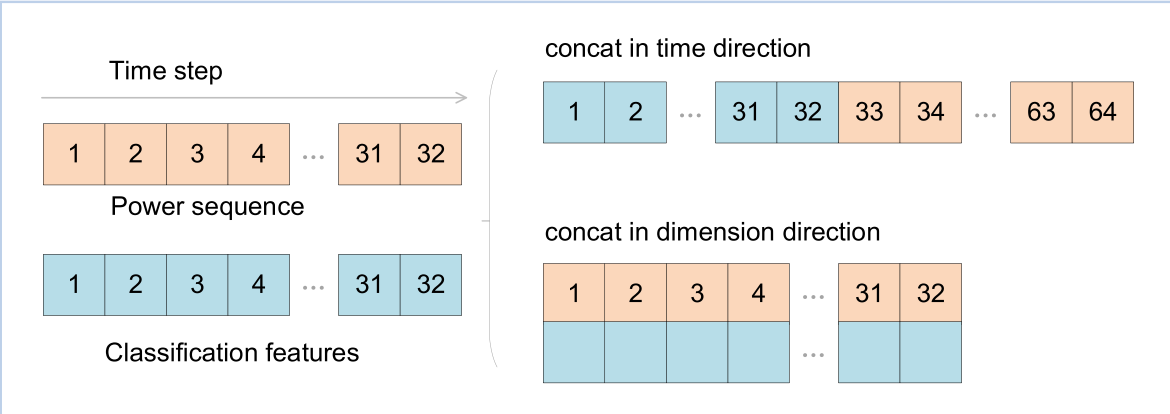


FIGURE 3. Ways to expand the input vector information of LSTM.

the current input x_t and previous internal state c_{t-1}, namely, $c_t = g(x_t, c_{t-1})$. The final state h_t will be determined by c_t and o_t. Controlled by the input and output gates, the memory cell is able to store the previous information for a long period of time. Meanwhile, the states stored in the memory cell can be cleared by the forget gate.

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \tag{7}$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \tag{8}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o) \tag{9}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes (W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{10}$$

$$h_t = o_t \otimes \tanh(c_t) \tag{11}$$

where W_{ix}, W_{ih}, W_{ic}, W_{fx}, W_{fh}, W_{fc}, W_{ox}, W_{oh}, W_{oc}, W_{cx}, W_{ch} are weight matrices for the corresponding inputs, b_i, b_f, b_o, b_c are the bias vectors, $\otimes$ denotes the Hadamard product, and σ represents the activation function.

It is worth noting that the size of LSTM input vector is (B, T, D). Where B refers to the training batch size, T refers to the time steps, and D refers to the data dimension. Therefore, there are two ways to expand LSTM input vector information using CNN classification features: time direction and dimension direction, as shown in Fig.3.

III. CASE STUDY

A. DATASET AND EVALUATION INDICATOR

In order to verify the effectiveness of the proposed method, experiments are conducted on two datasets from different regions.

The dataset-1 is opensource data provided by Elia. The data collection location is Belgium; the collection time is from January 1, 2018 to December 31, 2018; the installed wind power capacity is 806.71MW; and the data resolution is set to 15min.

The dataset-2 is opensource data provided by National Renewable Energy Laboratory (NREL). The data collection location is 78.5945° ~79.8666°W, 39.0369° ~40.3445°N; the collection time is from January 1, 2012 to December 31, 2012; the installed wind power capacity is 80MW; and the data resolution is set to 15min.

Detailed process analysis is carried out based on dataset-1. According to the rules for generating sample sets described above, a total of 4350 samples are generated for the entire year: the training set contains 3,500 samples for offline training; the test set contains 850 samples for online testing.

Mean absolute error (MAE) and root mean square error (RMSE) are the most common evaluation indicators for WPF, which can be illustrated by

$$\varepsilon_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n}|P_{Pi} - P_{Mi}| \tag{12}$$

$$\varepsilon_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{Pi} - P_{Mi})^2} \tag{13}$$

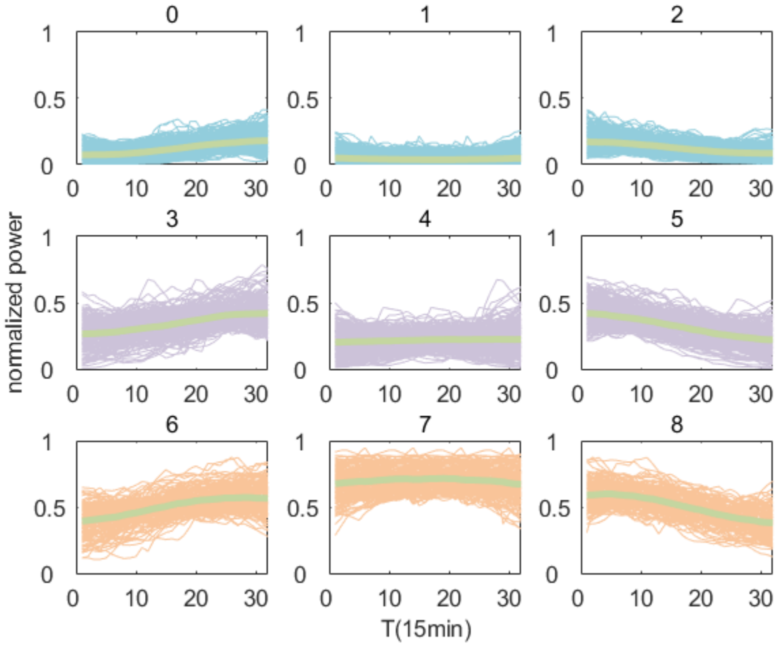
where n is the number of samples, P_{Pi} is forecasting value, P_{Mi} is measured value. The power values in this article refer to the normalized values.

B. MODEL DESCRIPTION

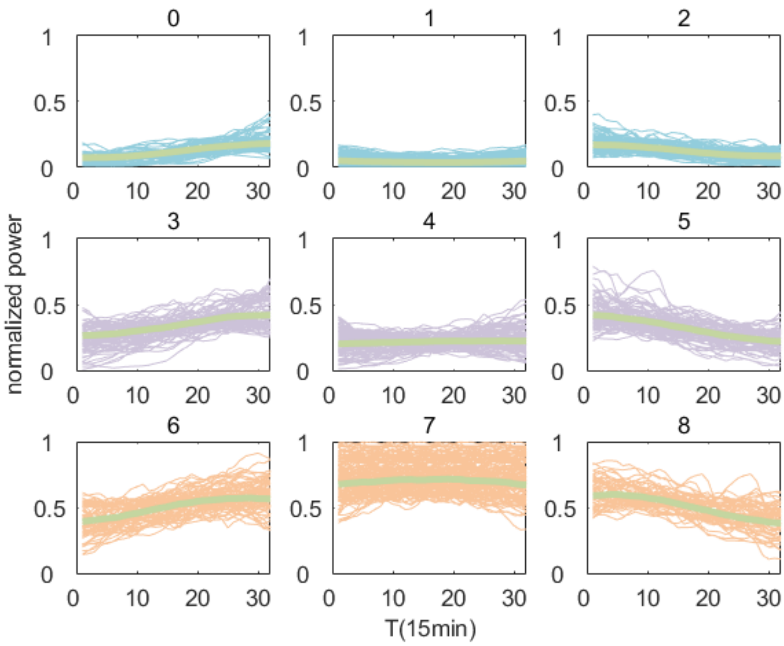
Comparative experiments are designed to verify the method proposed. The description of the baseline models is shown in Table 1.

TABLE 1. Model description.

Model	Description
PM	establish a PM for all samples.
LSTM	establish a LSTM for all samples.
MLSTMs	classification, establish LSTM for each category separately.
CNN-MLSTMs-T	classification, establish LSTM for each category separately, add CNN features in time direction.
CNN-MLSTMs-D	classification, establish LSTM for each category separately, add CNN features in dimension direction.



(a) Training set



(b) Test set

FIGURE 4. **Power curves and centroid curves of each category.**

C. CLASSIFICATION AND RECOGNITION RESULTS

Table 2 shows the different classification schemes of two-level clustering on the training set.

When the samples are divided into 6 categories, $L_{\max}$ is too large, which means that the distance between the samples of the same category is too large, and the sample similarity is poor. When the samples are divided into 12 categories, $S_{\min}$ is 84, which means that the category with the fewest samples contains only 84 samples, making it difficult to train the prediction model. Furthermore, the $L_{\max}$ of 9-categories and 12-categories are both 1.837, which shows that excessive subdivisions are meaningless. Therefore, 9-categories is the best scheme: the samples are first divided into low-power, medium-power, and high-power categories based on the

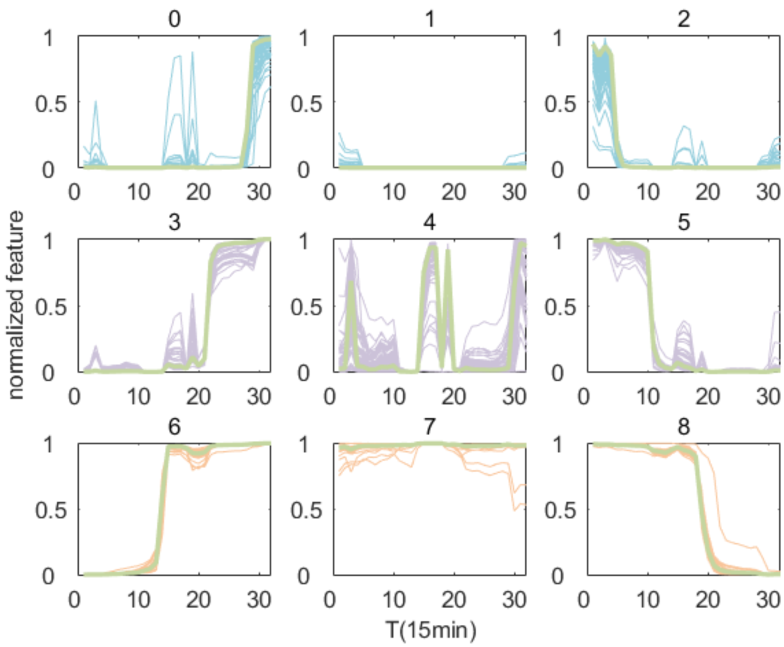


FIGURE 5. Classification feature curves and centroid curves of each category.

TABLE 2. Different schemes of two-level clustering.

First-level (average power)	Second-level (sample similarity)	categories	S_{ave}	S_{min}	L_{max}
2	3	6	583	175	2.944
3	2	6	583	177	2.906
3	3	9	389	135	1.837
3	4	12	292	84	1.837

TABLE 3. Classification and recognition results.

	0	1	2	3	4	5	6	7	8
Training set	408	1184	484	245	470	245	160	168	136
Test set	52	205	81	76	95	77	69	137	58

average power value, and then each power level is subdivided into 3 trends according to the defined distance. In this way, the samples are divided into 9 categories labeled as 0, 1, 2,... 7, 8.

Based on the labeled samples provided by two-level clustering, the CNN classification model is trained and used to perform mode recognition on the test set. The number of samples included in each category is shown in Table 3.

Fig.4(A) and Fig.4(B) respectively show the classification result of the training set and the recognition result of the test set. Taking Fig.4(A) as an example: categories with different power levels, such as category 1, category 4, and category 7, are clearly distinguished in power value; categories with the similar power level, such as category 3, category 4, and category 5, although they all belong to medium-power, their power trends are different, which are respectively rising, flat and falling. The results reflect that the method proposed has a strong ability to distinguish characteristics of samples.

Fig.5 shows the classification features, that is, the pre-training features of samples mined by CNN. Obviously, the feature vectors of samples belonging to the same category have high similarity, and the feature vectors of samples belonging to different categories have great difference. In fact, it is the feature similarity and feature difference between samples that guarantee the accuracy of classification. In more detail, the classification features of samples belonging to the same category are not exactly the same, because these features contain not only the sample similarity, but also the unique information of each sample itself.

D. PREDICTION RESULTS

For Belgium data, Table 4 and Table 5 show the prediction errors of each category from a micro perspective, and Table 6 is a summary of all categories. Fig.6 and Fig.7 are the visualization results of the prediction performance of each model. For NREL data, the performance of the models is shown in Table 7.

To sum up, it can be concluded that:

1) According to the summary results in Table 6 and Table 7, the WPF errors of the multi-category models are smaller than that of the unclassified models. Taking the Belgium 1-hour-ahead forecast results as an example: RMSE and MAE of LSTM are 0.0478 and 0.034, MAE and RMSE of MLSTMs are 0.0453 and 0.0315. Specific to a single category, such as category 1, as shown in Table 4, RMSE and MAE of MLSTMs are 0.0195 and 0.013, which are significantly

TABLE 4. 1-hour-ahead: prediction errors of each category.

		0	1	2	3	4	5	6	7	8
PM	RMSE	0.0498	0.0249	0.0316	0.0593	0.0537	0.0457	0.0719	0.0587	0.0658
	MAE	0.0383	0.0153	0.02	0.0436	0.0384	0.0354	0.0578	0.0433	0.0546
MLSTMs	RMSE	0.043	0.0195	0.0257	0.0558	0.046	0.0416	0.0708	0.0549	0.0566
	MAE	0.032	0.013	0.0183	0.0416	0.0336	0.0326	0.0568	0.0405	0.045
CNN-MLSTMs-T	RMSE	0.0392	0.0179	0.0247	0.0526	0.0431	0.0399	0.0668	0.0538	0.0578
	MAE	0.0292	0.0114	0.0171	0.041	0.0322	0.0319	0.0523	0.0399	0.0436
CNN-MLSTMs-D	RMSE	0.0408	0.0185	0.0249	0.053	0.0441	0.0429	0.0673	0.056	0.0579
	MAE	0.0312	0.0115	0.0175	0.0412	0.0339	0.0346	0.0522	0.0418	0.0466

TABLE 5. 4-hour-ahead: prediction errors of each category.

		0	1	2	3	4	5	6	7	8
PM	RMSE	0.1275	0.0791	0.0974	0.142	0.1382	0.116	0.1743	0.1445	0.1295
	MAE	0.0923	0.0489	0.0571	0.1149	0.103	0.0875	0.1369	0.1094	0.1099
MLSTMs	RMSE	0.1229	0.0652	0.0887	0.1419	0.134	0.114	0.1691	0.1369	0.1212
	MAE	0.0875	0.0453	0.057	0.1138	0.0999	0.0861	0.136	0.1102	0.1027
CNN-MLSTMs-T	RMSE	0.1178	0.0627	0.0847	0.1408	0.132	0.1126	0.1655	0.1381	0.1114
	MAE	0.0876	0.0459	0.0568	0.1141	0.0974	0.0858	0.1332	0.1112	0.0906
CNN-MLSTMs-D	RMSE	0.1182	0.0632	0.0854	0.1427	0.1323	0.1157	0.167	0.1371	0.1106
	MAE	0.0859	0.0442	0.0582	0.1131	0.1034	0.0882	0.1349	0.1114	0.092

TABLE 6. Summary results for Belgium: prediction errors of models.

Model	1h-ahead		4h-ahead	
	RMSE	MAE	RMSE	MAE
PM	0.0499	0.0347	0.1247	0.0888
LSTM	0.0478	0.034	0.1201	0.0887
MLSTMs	0.0453	0.0315	0.1184	0.0867
CNN- MLSTMs-T	0.0434	0.0300	0.1162	0.0856
CNN- MLSTMs-D	0.0445	0.0311	0.1168	0.0863

smaller than the average errors on the entire dataset. Based on these results, it is safe to say that the proposed method based on sample similarity analysis can significantly improve the accuracy of WPF.

2) According to Table 6 and Table 7, the errors of prediction models considering classification features are smaller. In both 1-hour-ahead forecast task and 4-hour-ahead forecast task, CNN-MLSTMs-T and CNN-MLSTMs-D perform better than MLSTMs, especially CNN-MLSTMs-T. In more detail, Table 4 and Table 5 show the effect of classification features on each category: for most categories, the addition of classification features enhances the feature information of samples, which is beneficial for prediction. For categories with few training samples, increasing the feature dimension of samples will lead to overfitting of the model, such as category 8 belonging to the high-power level, it contains few samples because in real-world scenarios, such strong wind energy is rarely seen. In this case, enriching historical data, considering adjacent wind farm data, and further processing features under supervision are all paths worth exploring to improve the WPF accuracy. On the whole, it is safe to say that the classification features with sample similarity are helpful

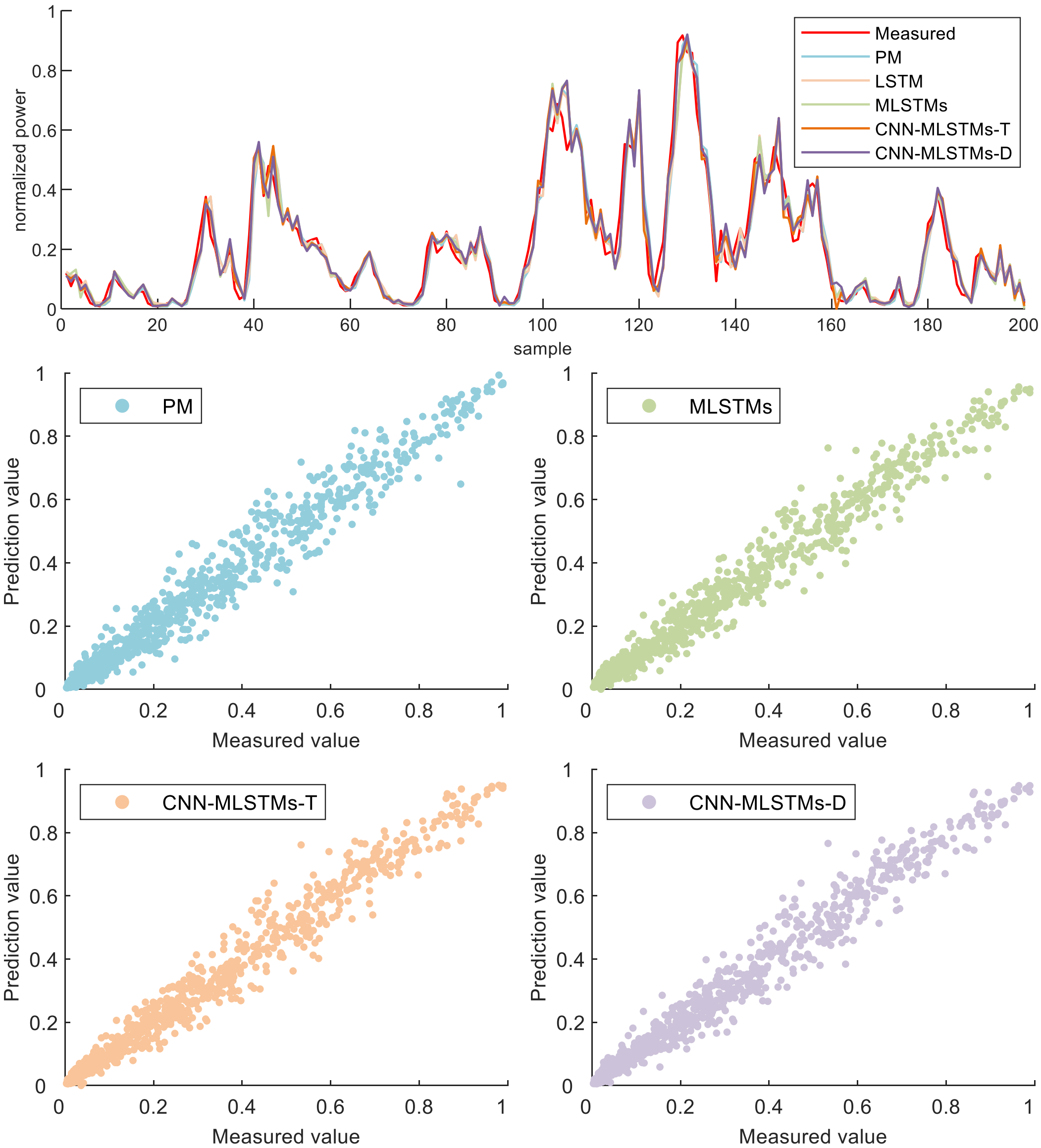


FIGURE 6. **1-hour-ahead: prediction results of models.**

TABLE 7. **Summary results for NREL: prediction errors of models.**

Model	1h-ahead		4h-ahead	
	RMSE	MAE	RMSE	MAE
PM	0.0939	0.062	0.2147	0.1506
LSTM	0.0879	0.0614	0.2051	0.1608
MLSTMs	0.0861	0.0592	0.2016	0.1545
CNN- MLSTMs-T	0.0839	0.0571	0.1998	0.1523
CNN- MLSTMs-D	0.0851	0.0582	0.2002	0.1506

to the establishment and training of prediction models, and it is meaningful to strengthen the coupling and connection between wind power mode classification and numerical prediction.

3) CNN-MLSTMs-D adds features in dimension direction, so it is necessary to align features with other input vectors, resulting in a limitation of feature size. CNN-MLSTMs-T adds features in time direction, so the calculation takes longer, which is determined by the structure of LSTM. For category 1 with 1184 training samples, the training time of CNN-MLSTMs-D and CNN-MLSTMs-T are 10.2s and 13.8s, respectively (Intel Core i5-10300H CPU/ 16.00GB RAM/ GeForce GTX 1650). Therefore, both of them can meet the time-consuming needs of ultra-short-term forecasts.

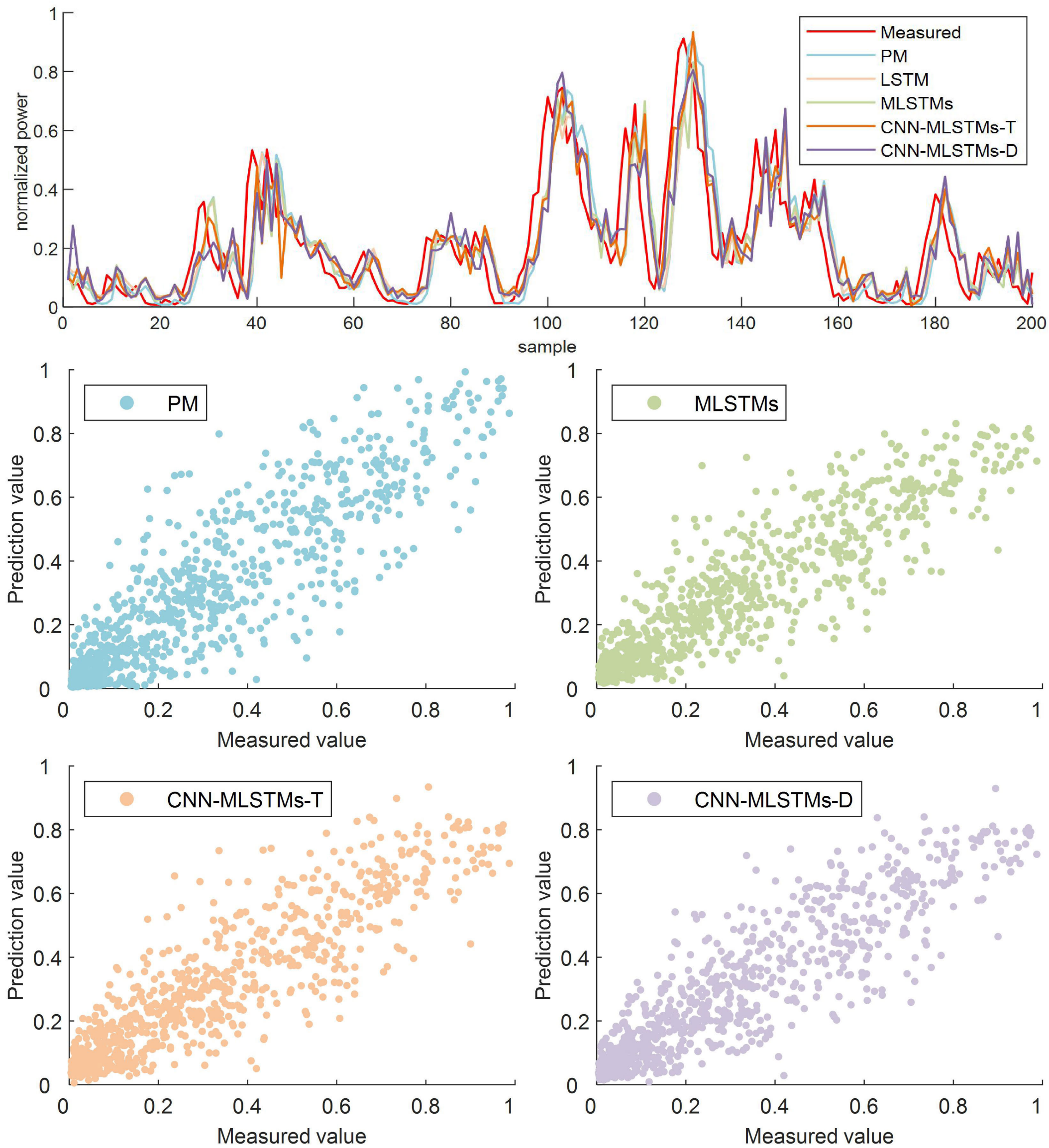


FIGURE 7. 4-hour-ahead: prediction results of models.

IV. CONCLUSION

Based on the sample similarity analysis, this paper decomposes the WPF problem into two strongly related tasks: mode classification and numerical prediction. The process of wind power mode classification is essentially the analysis of sample similarity and sample difference, which is beneficial to the selection of training samples for prediction models. In mode classification task, two-level clustering, as an accurate and efficient classification method, can divide the wind power database into multiple categories corresponding to different real-world scenarios, and provide a large number of labeled samples for CNN classification model. In numerical prediction task, the necessity of separately modeling and the role of feature transfer learning are verified. Simulation results based on real-world datasets prove the effectiveness and superiority of the above methods.

LSTM does not perform well in every category, so the combination of the sample similarity analysis idea and other hybrid models will be our future research focus. Besides, we will analyze other tasks that are strongly coupled with WPF, such as monitoring and numerical prediction of wind speed and wind direction, and working status monitoring of wind turbines, so as to realize knowledge sharing and joint optimization of multi tasks.

REFERENCES

[1] P. de Jong, A. Kiperstok, A. S. Sánchez, R. Dargaville, and E. A. Torres, "Integrating large scale wind power into the electricity grid in the Northeast of Brazil," *Energy*, vol. 100, pp. 401–415, Apr. 2016, doi: 10.1016/j.energy.2015.12.026.

[2] Y. Ren, P. N. Suganthan, and N. Srikanth, "A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 236–244, Jan. 2015, doi: 10.1109/TSTE.2014.2365580.

[3] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," *Appl. Energy*, vol. 238, pp. 368–383, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.063.

[4] I. Okumus and A. Dinler, "Current status of wind energy forecasting and a hybrid method for hourly predictions," *Energy Convers. Manage.*, vol. 123, pp. 362–371, Sep. 2016, doi: 10.1016/j.enconman.2016.06.053.

[5] A. Ahmed and M. Khalid, "A review on the selected applications of forecasting models in renewable power systems," *Renew. Sustain. Energy Rev.*, vol. 100, pp. 9–21, Feb. 2019, doi: 10.1016/j.rser.2018.09.046.

[6] E. B. Ssekulima, M. B. Anwar, A. A. Hinai, and M. S. E. Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review," *IET Renew. Power Gener.*, vol. 10, no. 7, pp. 885–989, Aug. 2016, doi: 10.1049/iet-rpg.2015.0477.

[7] L. Landberg, "Short-term prediction of the power production from wind farms," *J. Wind Eng. Ind. Aerodyn.*, vol. 80, nos. 1–2, pp. 207–220, 1999, doi: 10.1016/S0167-6105(98)00192-5.

[8] M. Lange and U. Focken, *Physical Approach to Short-Term Wind Power Prediction*. Berlin, Germany: Springer, 2006, pp. 1–208.

[9] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using Gaussian processes and numerical weather prediction," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 656–665, Mar. 2014, doi: 10.1109/TPWRS.2013.2282366.

[10] G. W. Chang, H. J. Lu, Y. R. Chang, and Y. D. Lee, "An improved neural network-based approach for short-term wind speed and power forecast," *Renew. Energy*, vol. 105, pp. 301–311, May 2017, doi: 10.1016/j.renene.2016.12.071.

[11] Y.-T. Wu and F. Porté-Agel, "Simulation of turbulent flow inside and above wind farms: Model validation and layout effects," *Boundary-Layer Meteorol.*, vol. 146, no. 2, pp. 181–205, Feb. 2013, doi: 10.1007/s10546-012-9757-y.

[12] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Convers. Manage.*, vol. 118, pp. 193–203, Jun. 2016, doi: 10.1016/j.enconman.2016.03.078.

[13] K. Chen and J. Yu, "Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach," *Appl. Energy*, vol. 113, pp. 690–705, Jan. 2014, doi: 10.1016/j.apenergy.2013.08.025.

[14] S. Robert, L. Foresti, and M. Kanevski, "Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks," *Int. J. Climatol.*, vol. 33, no. 7, pp. 1793–1804, Jun. 2013, doi: 10.1002/joc.3550.

[15] B. Saavedra-Moreno, S. Salcedo-Sanz, L. Carro-Calvo, J. Gascón-Moreno, S. Jiménez-Fernández, and L. Prieto, "Very fast training neural-computation techniques for real measure-correlate-predict wind operations in wind farms," *J. Wind Eng. Ind. Aerodyn.*, vol. 116, pp. 49–60, May 2013, doi: 10.1016/j.jweia.2013.03.005.

[16] A. Lahouar and J. B. H. Slama, "Hour-ahead wind power forecast based on random forests," *Renew. Energy*, vol. 109, pp. 529–541, Aug. 2017, doi: 10.1016/j.renene.2017.03.064.

[17] Y. Zhang, J. Le, X. Liao, F. Zheng, and Y. Li, "A novel combination forecasting model for wind power integrating least square support vector machine, deep belief network, singular spectrum analysis and locality-sensitive hashing," *Energy*, vol. 168, pp. 558–572, Feb. 2019, doi: 10.1016/j.energy.2018.11.128.

[18] S. Salcedo-Sanz, E. G. Ortiz-García, Á. M. Pérez-Bellido, A. Portilla-Figueras, and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4052–4057, Apr. 2011, doi: 10.1016/j.eswa.2010.09.067.

[19] Y.-L. Hu and L. Chen, "A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm," *Energy Convers. Manage.*, vol. 173, pp. 123–142, Oct. 2018, doi: 10.1016/j.enconman.2018.07.070.

[20] H. Liu, X.-W. Mi, and Y.-F. Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network," *Energy Convers. Manage.*, vol. 156, pp. 498–514, Jan. 2018, doi: 10.1016/j.enconman.2017.11.053.

[21] S. Liang, L. Nguyen, and F. Jin, "A multi-variable stacked long-short term memory network for wind speed forecasting," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 4561–4564, doi: 10.1109/BigData.2018.8622332.

[22] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019, doi: 10.1016/j.energy.2019.03.081.

[23] M. Morchid, "Parsimonious memory unit for recurrent neural networks with application to natural language processing," *Neurocomputing*, vol. 314, pp. 48–64, Nov. 2018, doi: 10.1016/j.neucom.2018.05.081.

[24] R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, J. Zhang, R. Zhang, and Z. Zhang, "LSTM-EFG for wind power forecasting based on sequential correlation features," *Future Gener. Comput. Syst.*, vol. 93, pp. 33–42, Apr. 2019, doi: 10.1016/j.future.2018.09.054.

[25] E. López, C. Valle, H. Allende, E. Gil, and H. Madsen, "Wind power forecasting based on echo state networks and long short-term memory," *Energies*, vol. 11, no. 3, p. 526, Feb. 2018, doi: 10.3390/en11030526.

[26] P. Jiang, X. Li, and Y. Dong, "Research and application of a new hybrid forecasting model based on genetic algorithm optimization: A case study of Shandong wind farm in China," *Math. Problems Eng.*, vol. 2015, pp. 1–14, Jan. 2015, doi: 10.1155/2015/740490.

[27] H. A. Nielsen, T. S. Nielsen, H. Madsen, M. J. S. I. Pindado, and I. Marti, "Optimal combination of wind power forecasts," *Wind Energy*, vol. 10, no. 5, pp. 471–482, Sep. 2007, doi: 10.1002/we.237.

[28] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks," *Appl. Energy*, vol. 107, pp. 191–208, Jul. 2013, doi: 10.1016/j.apenergy.2013.02.002.

[29] H. Liu, H.-Q. Tian, X.-F. Liang, and Y.-F. Li, "Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks," *Appl. Energy*, vol. 157, pp. 183–194, Nov. 2015, doi: 10.1016/j.apenergy.2015.08.014.

[30] Z. Zhang, X. Cui, X. Lin, and L. Zhang, "An algorithm of training sample selection for integrated circuit device modeling based on artificial neural networks," in *Proc. IEEE Int. Conf. Electron Devices Solid-State Circuits (EDSSC)*, Aug. 2016, pp. 314–317, doi: 10.1109/EDSSC.2016.7785271.

[31] Z. Jiang, X. Zhu, W.-T. Tan, and R. Liston, "Training sample selection for deep learning of distributed data," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2189–2193, doi: 10.1109/ICIP.2017.8296670.

[32] B. Hadjadji and Y. Chibani, "Optimized selection of training samples for one-class neural network classifier," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 345–349, doi: 10.1109/IJCNN.2014.6889429.

[33] Z. Lv, L. Wang, Z. Guan, J. Wu, X. Du, H. Zhao, and M. Guizani, "An optimizing and differentially private clustering algorithm for mixed data in SDN-based smart grid," *IEEE Access*, vol. 7, pp. 45773–45782, 2019.

[34] J. Yan, Y. Liu, S. Han, and M. Qiu, "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 613–621, Nov. 2013, doi: 10.1016/j.rser.2013.07.026.

[35] M. Barman, N. B. D. Choudhury, and S. Sutradhar, "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India," *Energy*, vol. 145, pp. 710–720, Feb. 2018, doi: 10.1016/j.energy.2017.12.156.

[36] Y. Liu, C. Liu, W. Li, and D. Zhang, "Multi-time scale power prediction of wind farm cluster based on profile pattern matching," *Zhongguo Dianji Gongcheng Xuebao/Proc. Chin. Soc. Electr. Eng.*, vol. 34, no. 25, pp. 4350–4358, Sep. 2014, doi: 10.13334/j.0258-8013.pcsee.2014.25.018.

[37] L. Qin, Y. Xiong, and K. Liu, "Weather division-based wind power forecasting model with feature selection," *IET Renew. Power Gener.*, vol. 13, no. 16, pp. 3050–3060, Dec. 2019, doi: 10.1049/iet-rpg.2019.0263.

[38] X. Zhang, R. Wang, T. Liao, T. Zhang, and Y. Zha, "Short-term forecasting of wind power generation based on the similar day and Elman neural network," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 647–650, doi: 10.1109/SSCI.2015.99.

[39] L. Dong, L. Wang, S. F. Khahro, S. Gao, and X. Liao, "Wind power day-ahead prediction with cluster analysis of NWP," *Renew. Sustain. Energy Rev.*, vol. 60, pp. 1206–1212, Jul. 2016, doi: 10.1016/j.rser.2016.01.106.

[40] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Appl. Energy*, vol. 141, pp. 190–199, Mar. 2015, doi: 10.1016/j.apenergy.2014.12.039.

[41] Y. Hao, L. Dong, X. Liao, J. Liang, L. Wang, and B. Wang, "A novel clustering algorithm based on mathematical morphology for wind power generation prediction," *Renew. Energy*, vol. 136, pp. 572–585, Jun. 2019, doi: 10.1016/j.renene.2019.01.018.

[42] G. Sun, C. Jiang, P. Cheng, Y. Liu, X. Wang, Y. Fu, and Y. He, "Short-term wind power forecasts by a synthetical similar time series data mining method," *Renew. Energy*, vol. 115, pp. 575–584, Jan. 2018, doi: 10.1016/j.renene.2017.08.071.

[43] T.-C. Fu, "A review on time series data mining," *Eng. Appl. Artif. Intell.*, vol. 24, no. 1, pp. 164–181, Feb. 2011, doi: 10.1016/j.engappai.2010.09.007.

[44] K. Wang, X. Qi, H. Liu, and J. Song, "Deep belief network based *k*-means cluster approach for short-term wind power forecasting," *Energy*, vol. 165, pp. 840–852, Dec. 2018, doi: 10.1016/j.energy.2018.09.118.

[45] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017, doi: 10.21629/JSEE.2017.01.18.

CHANGXIN MIAO received the Ph.D. degree in electrical engineering from the China University of Mining and Technology. He is currently an Associate Professor with the School of Electrical and Power Engineering, China University of Mining and Technology. His research interests include power quality management and application of artificial intelligence in power systems.

HAO LI is currently pursuing the master's degree in electrical engineering with the School of Electrical and Power Engineering, China University of Mining and Technology. His research interests include machine learning and application of artificial intelligence in power systems.

XIA WANG is currently pursuing the master's degree in electrical engineering with the School of Electrical and Power Engineering, China University of Mining and Technology. Her research interests include machine learning and application of artificial intelligence in power systems.

HENG LI is currently pursuing the master's degree in electrical engineering with the School of Electrical and Power Engineering, China University of Mining and Technology. His research interests include power quality management and power electronics technology.