

## ORIGINAL RESEARCH PAPER

# Ultra-short-term multi-step wind power forecasting based on CNN-LSTM

Qianyu Wu<sup>1</sup> | Fei Guan<sup>1</sup> | Chen Lv<sup>2</sup> | Yongzhang Huang<sup>1</sup>

<sup>1</sup> State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, People's Republic of China

<sup>2</sup> China Electric Power Research Institute, Beijing, China

## Correspondence

Qianyu Wu, State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing, People's Republic of China.  
Email: 18810663690@163.com

**Funding:** Funder: Science and Technology Project of State Grid. Grant No.: 5201011600TS.

## Abstract

The fluctuation and intermission of large-scale wind power integration is a serious threat to the stability and security of the power system. Accurate prediction of wind power is of great significance to the safety of wind power grid connection. This study proposes a novel spatio-temporal correlation model (STCM) for ultra-short-term wind power prediction based on convolutional neural networks-long short-term memory (CNN-LSTM). The original meteorological factors at multi-historical time points of different sites throughout the target wind farm can be reconstructed into the input window of the model, and thus a new data reconstruction method is represented. CNN is used to extract the spatial correlation feature vectors of meteorological factors of different sites and the temporal correlation vectors of the meteorological features in ultra-short term, which are reconstructed in time series and used as the input data of LSTM. Then, LSTM extracts the temporal feature relationship between the historical time points for multi-step wind power forecasting. The STCM based on CNN-LSTM proposed in this study is suitable for wind farms that can collect meteorological factors at different locations. Taking the measured meteorological factors and wind power dataset of a wind farm in China as an example, four evaluation metrics of the CNN-LSTM model, CNN and LSTM individually used for multi-step wind power prediction, are obtained. The results show that the proposed STCM based on CNN-LSTM has better spatial and temporal characteristics extraction ability than the traditional structure model and can forecast the power of wind farm more accurately.

## 1 | INTRODUCTION

In recent years, wind power has developed rapidly all over the world. The fluctuation and intermission of wind power output bring unstable factors to the power system. Improving the forecasting accuracy of wind power is an effective way to reduce the instability of the power system caused by large-scale wind power integration.

The forecasting models of wind power are mainly divided into physical, statistical, machine learning and combined models [1]. The physical models convert the numerical weather forecasting (NWP) data into wind speed of the height of the wind turbine by means of microscale meteorology and computational fluid dynamics and forecast wind power indirectly by the conversion calculation [2]. The physical models can be applied to new wind farms without a large amount of historical data, and

it is easy to realise the medium and long-term forecasting of wind power. The forecasting accuracy mainly depends on the accuracy of NWP data, the information of the physical environment around the wind farm and the accuracy of the physical model [3]. However, physical models are not suitable for short-term wind power forecasting because of the high calculation cost. Based on a large number of historical data of wind farms, the statistical methods use algorithms including Kalman filter [4], autoregressive (AR) model, autoregressive moving average (ARMA) model [5] to extract the linear relation between input features (NWP, historical measured data) and wind power. Statistical models can achieve short-term wind speed forecasting, but they cannot analyse non-linear relationships between the variables [6]. Machine learning models such as backpropagation network [7], radial basis function [8], extreme learning machine [9], support vector machine (SVM) and Gaussian process [10]

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *IET Renewable Power Generation* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

establish a black-box model to fit the non-linear relationship between the input characteristics and the output wind power by learning and training a large number of historical measured data. However, the shallow machine learning models can only extract very superficial features and have weak learning ability for multi-dimensional big data.

Compared with shallow machine learning, the deep learning models have a stronger ability of computing and complex function fitting. Through non-linear optimisation of multi-layer network structure, deep learning models can automatically extract the inherent features in data from the lowest to the highest level [11]. Some scholars try to apply the deep learning models to wind power forecasting based on historical data to improve the accuracy of wind power forecasting [12]. The ultra-short-term and short-term wind speed forecasting models of three hidden layers are established by using the deep Boltzmann machine [13]. In [14], migration learning model was applied to transfer wind speed forecasting models trained by wind farms with rich historical data to wind farms with less historical data. In [15], the deep belief network model was applied to short-term wind speed forecasting to obtained high prediction accuracy in practical examples. A new forecasting model based on neural network (NN) and a novel chaotic shark smell optimisation algorithm was proposed and the effectiveness of the proposed forecasting model was tested on two real-world case studies [16]. The authors in [17] proposed a new wind power prediction approach which included an improved version of Kriging interpolation method, empirical mode decomposition (EMD), an information-theoretic feature selection method, and a closed-loop forecasting engine. In [18], a prediction approach based on the improved EMD (IEMD) in conjunction with a hybrid framework consisting of the bagging NN (BaNN), K-means clustering method, and a stochastic optimisation algorithm was proposed.

CNN and LSTM are two main deep learning models [19]. In [20], a probability forecasting model of ultra-short-term wind power based on CNN was proposed, and the accuracy of the model was verified. In [21], CNN and physical model were combined for forecasting, which further reduced the forecasting error of short-term wind power. In [22], LSTM models were applied to short-term forecasting of wind speed and wind power. In [23], the short-term wind power interval prediction based on two typical recurrent NN (RNN) models, Elman network and the nonlinear autoregressive with exogenous inputs (NARX) model and lower upper bound estimation method was investigated. Taking into account the impact of meteorological information data on wind power prediction, the authors in [24] sifted multivariate meteorological information data highly relevant to wind power with distance analysis as the input data of the LSTM model and modelled the time series from the viewpoint of time with LSTM. The authors in [25] combined CNN and LSTM to forecast wind speed and considered the influence of various meteorological factors such as temperature, wind speed, and wind direction on the wind speed in time and space. In [26], the authors investigated the combined performance of the wavelet packet decomposition and the CNN and CNN-LSTM in the wind speed multi-step

prediction to extract the hidden features of the wind speed time series.

Existing individual CNN and LSTM can establish the non-linear correlation between output and input variables through a large amount of historical data to predict wind speed or wind power. And each model has its advantages and disadvantages. The combination of CNN and LSTM can realise the complementary advantages of each model to further improve the accuracy of forecasting [25]. The power of a wind farm is related to meteorological factors such as wind speed and wind direction at various sites of the wind farm. The meteorological factors are continuous in time and space resulting in a significant cross-correlation between the factors of a target and its adjacent site [27]. However, the existing ultra-short-term forecasting methods of wind power usually ignore the influence of spatio-temporal correlation of meteorological factors at different sites of the wind farm on the wind power.

Thus, this study combines the advantage of extracting spatial features of CNN and the advantage of extracting time-series features of LSTM to extract the spatio-temporal correlation between multiple meteorological factors and wind power.

The contributions of this study are as follows. (1) Considering the influence of multiple meteorological factors of different sites throughout the target wind farm to wind power, a novel spatio-temporal correlation model (STCM) based on CNN-LSTM for ultra-short-term wind power prediction is proposed. (2) A new data reconstruction method is proposed. The input matrix is constructed by the meteorological factors at different sites of the wind farm as the vertical axis and the ultra-short-term historical time as the horizontal axis. CNN is used to extract the spatial correlation of meteorological factors on the vertical axis and the ultra-short-term temporal correlation of the features on the horizontal axis. The correlation vectors of each input matrix extracted by CNN are constructed in a long-term time series and used as the input data of LSTM to extract long-term historical temporal relationship for multi-step wind power prediction. (3) Based on the data reconstruction method, the study uses multiple independent models to share the same input matrix to achieve a multi-step prediction of wind power and reduce the time for data processing.

The organisation of this study is as follows. Section 2 introduces the structure of STCM based on CNN-LSTM. In Section 3, the spatio-temporal of multiple meteorological factors (specifically refers to wind speed and wind direction in this study) at different sites of the wind farm is analysed. The wind speeds and wind directions measured by benchmark wind turbines are reconstructed and used as the input of the model. And the STCM based on CNN-LSTM for multi-step wind power forecasting and error calculation method is presented. In Section 4, taking a wind farm in China as an example, the calculation results of four evaluation indexes show that the STCM based on CNN-LSTM established in this study can predict the wind farm power more accurately than the individually deep learning model (CNN, LSTM).

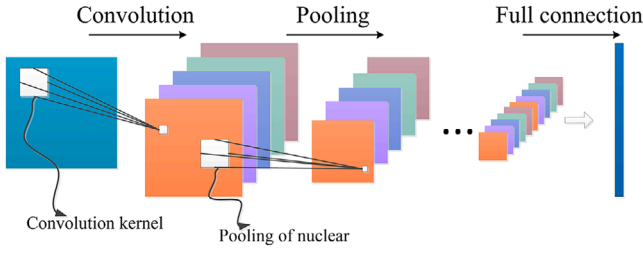


FIGURE 1 The structure of convolutional neural networks (CNN)

## 2 | STRUCTURE OF STCM BASED ON CNN-LSTM

### 2.1 | The principle of CNN algorithm

CNN is a deep feedforward NN with convolution structure using a supervised learning method [28]. The structure includes a convolutional layer, pooling layer and full connection layer as shown in Figure 1. Convolutional and pooling layers are the core modules of the CNN network feature extraction. The input features are convoluted through the convolutional layer. The pooling layer samples information from the preceding convolutional layer and minimises the spatial size [29]. Then, the full connection layer maps the two-dimensional feature to one-dimensional data for output.

1. Convolutional layer: The convolution operation of the convolutional layer can reduce the noise and enhance the key information of the original input features. Assuming that  $v$  is the input features of the original wind speeds and directions, and  $w$  is the convolution kernel of order  $J \times I$ , and the spatial correlation characteristic matrix  $y$  of the wind speeds and directions of order  $M \times N$  is output after the activation function, the element in row  $m$  and column  $n$  is defined as [30]

$$y_{mn} = f\left(\sum_{j=0}^{J-1} \sum_{i=0}^{I-1} v_{m+i, n+j} w_{ij} + b\right) \quad (1)$$

where  $m = 0, 1, \dots, M-1$ ;  $n = 0, 1, \dots, N-1$ ;  $w_{ij}$  is the element in row  $i$  and column  $j$  of  $w$ ;  $v_{m+i, n+j}$  is the element in row  $m+i$  and column  $n+j$  of  $v$ ;  $b$  is the bias;  $f$  is the activation function.

1. Pooling layer: In the pooling layer, the spatial correlation matrix  $y$  of wind speeds and directions after through the convolutional layer is further reduced by taking the average value (average pooling) or the maximum value (maximum pooling) of the area to save useful information while reducing the amount of data processing. Assuming that the dimension of the pool area is  $S_1 \times S_2$ , the dimension of the spatial correlation characteristic matrix  $C$  after output processing is  $(M/S_1) \times (N/S_2)$ , and the calculation formula of

average pooling is defined as [31]

$$C_{ab} = \frac{1}{S_1 S_2} \sum_{j=0}^{S_2-1} \sum_{i=0}^{S_1-1} y_{aS_1+i, bS_2+j} \quad (2)$$

where  $C_{ab}$  is the element in row  $i$  and column  $j$  of  $C$ ;  $a = 0, 1, \dots, M/S_1 - 1$ ;  $b = 0, 1, \dots, N/S_2 - 1$ ;  $y_{aS_1+i, bS_2+j}$  is the element in row  $aS_1+i$  and column  $bS_2+j$  of  $y$ .

1. Fully connected layer: In the full connection layer, the spatial correlation characteristic matrix  $C$  processed by the pooling layer is expanded into one-dimensional data output, which is expressed as [31]

$$p = f\left(\sum_{i=1}^n k_i c_i + b\right) \quad (3)$$

where  $c = [c_1, c_2, \dots, c_i, \dots, c_n]$  is the  $n$ -dimensional input variable;  $k = [k_1, k_2, \dots, k_i, \dots, k_n]$  is the connection weight;  $p$  is the one-dimensional output value of the spatial correlation characteristics of wind speeds and directions.

### 2.2 | The principle of LSTM algorithm

LSTM is a special RNN with stronger feature extraction ability for processing sequence data [32]. LSTM introduces memory unit on the basis of RNN, which is controlled by input, output and forgetting gates. It can better realise the storage, screening and control of information flow under the time feedback mechanism, effectively avoid information loss and solve the problem of gradient disappearance and explosion.

The structure of LSTM is shown in Figure 2. The symbol  $\sigma$  represents sigmoid activation function, and  $\tanh$  is tanh activation function. The input information flow enters from the output variable  $p_{t-1}$  at the previous time and the input variable  $v_t$  at the current time. Through the control of input, output and forgetting gates, the memory unit  $C_{t-1}$  is updated to  $C_t$ , and the output value at the current time is  $p_t$ . The output values of input, output and forgetting gates are  $i_t$ ,  $o_t$  and  $f_t$ , respectively.

The transformation equation is defined as [33]

$$\begin{aligned} i_t &= \sigma(k_{xi} v_t + k_{hi} p_{(t-1)} + k_{ci} c_{(t-1)} + b_i) \\ f_t &= \sigma(k_{xf} v_t + k_{hf} p_{(t-1)} + k_{cf} c_{(t-1)} + b_f) \\ o_t &= \sigma(k_{xo} v_t + k_{ho} p_{(t-1)} + k_{co} c_t + b_o) \\ c_t &= f_t \cdot c_{(t-1)} + i_t \cdot \tanh(k_{xc} v_t + k_{hc} p_{(t-1)} + b_c) \\ p_t &= o_t \cdot \tanh(c_t) \end{aligned} \quad (4)$$

where  $k_{xi}$ ,  $k_{hi}$ ,  $k_{ci}$  are weight matrixes of input, output at the previous moment and memory unit to input gate, respectively;  $k_{xf}$ ,  $k_{hf}$ ,  $k_{cf}$  are weight matrixes of input, output at the previous moment and memory unit to forgetting gate, respectively;

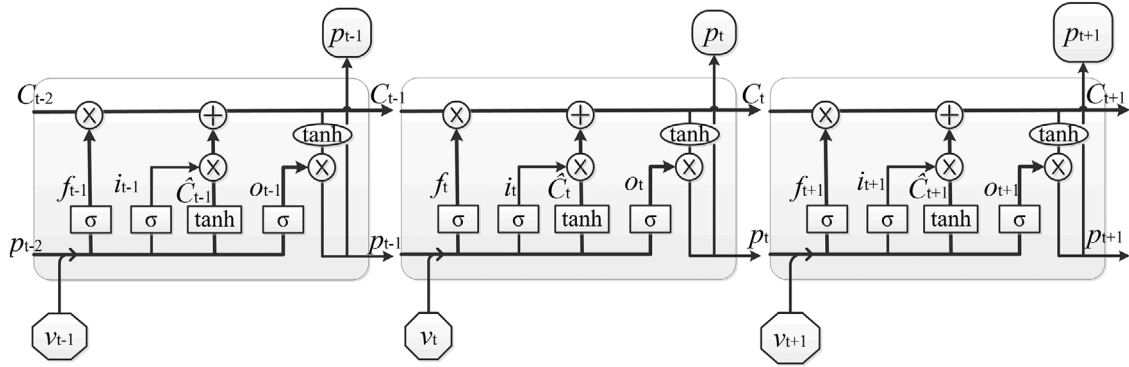


FIGURE 2 The architecture of the long short-term memory (LSTM) model

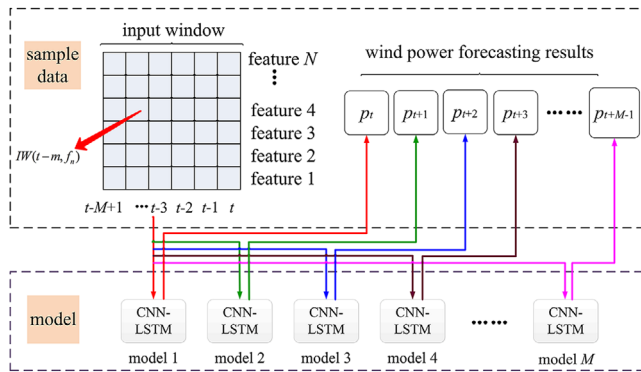


FIGURE 3 Output mode of CNN-LSTM multi-step forecasting model

$k_{x0}$ ,  $k_{y0}$ ,  $k_{z0}$  are weight matrixes of input, output at the previous moment and memory unit to output gate, respectively;  $b_f$ ,  $b_i$ ,  $b_o$ ,  $b_c$  are the bias values of input, forgetting, output gates and memory unit, respectively [33]; “ $\odot$ ” is Hadamard product.

### 2.3 | The STCM based on CNN-LSTM and data reconstruction method

Meteorological factors are closely related to wind power forecasting, including wind speed, wind direction, temperature, air pressure, and humidity; and the meteorological features of a region are similar to those of its adjacent regions [27]. Making full use of the meteorological information of different site throughout the target wind farm can improve the accuracy and reliability of wind power forecasting.

CNN-LSTM model combines the advantages of CNN and LSTM, which can extract spatial local features while time-series modelling. CNN is used to extract the spatial correlation feature vectors of meteorological factors of different sites, which are constructed in time series and used as the input data of LSTM network, and then LSTM network is used for ultra-short-term wind power forecasting. The multi-step forecasting model of wind power based on CNN-LSTM proposed in this study is shown in Figure 3.  $N$  represents the numbers of meteorological features and  $M$  represents the numbers of historical time points. The  $M \times N$  input window of the CNN-LSTM model

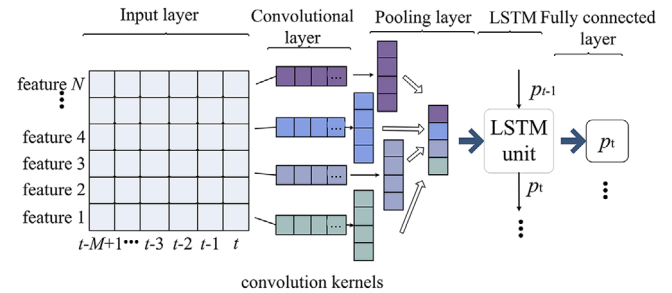


FIGURE 4 The architecture of the single CNN-LSTM model

includes multiple correlations of  $N$  meteorological factors in different sites and at  $M$  historical time points. And each point can be defined as  $IW(t-m, f_n)$ , which represents the value of the  $n$ -th meteorological factor of at historical time point  $t-m$ . The purpose of the model is to calculate the wind power of target wind farm at the next  $M$  moment from the correlation coefficient between  $N$  meteorological factor values at  $M$  historical time points. The formula is defined as

$$P(p_t, p_{t+1}, \dots, p_{t+m} \dots p_{t+M-1}) = \sum_{n=1}^N \sum_{m=0}^{M-1} (IW(t-m, f_n) \cdot C(IW(t-m, f_n), p_{t+m}) + \xi_{nm}) \quad (5)$$

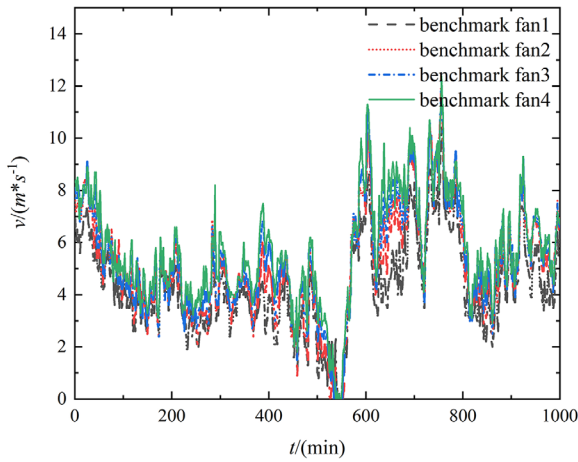
Here,  $P(p_t, p_{t+1}, \dots, p_{t+m} \dots p_{t+M-1})$  is the value of the wind powers of the target wind farm at the next  $M$  moment,  $C(IW(t-m, f_n), p_{t+m})$  is the correlation coefficient between the value of the  $n$ th meteorological factor at historical time point  $t-m$  and wind power value at the moment  $t+m$ , and  $\xi_{nm}$  is the error term.

The CNN-LSTM forecasting model with an output of  $M$  steps has  $M$  CNN-LSTM structures that share the same input, simplifying the process of data preprocessing. The  $M$  models are trained independently and do not interfere with each other. The structure of a single CNN-LSTM model is shown in Figure 4,  $N$  meteorological factors at different positions of the wind farm at  $M$  times form an input window. After convolution by multiple two-dimensional convolution kernels, CNN can extract the spa-



**TABLE 1** Programming platform and model parameters

Programming platform	Development platform	Keras's functions				Development language	Basic frequency	
	Google Keras	Conv2D, MaxPooling2D, Dense, LSTM, Activation, Dropout, plot_model				Python	3.2 GHz	
Parameters	Iterations	Convolution kernel size	Number of convolution kernels	Number of long short-term memory (LSTM) neurons	Time window	Batch size	Random inactivation rate	$L^2$
	30	2×2	4	128	254	14	0.2	0.005

**FIGURE 5** Wind speeds measured by four benchmark wind turbines in a wind farm

tial correlation of  $N$  meteorological factors on the vertical axis and the ultra-short-term temporal correlation of the features at  $M$  times on the horizontal axis. After pooling, the correlation vectors of each input matrix extracted by CNN are constructed in a long-term time series and used as the input data of LSTM to extract long-term historical temporal relationship for multi-step wind power prediction. Combining time window and long-term historical, the power forecasting results of the wind farm are achieved.

### 3 | THE STCM BASED ON CNN-LSTM FOR MULTI-STEP WIND POWER FORECASTING

#### 3.1 | Spatio-temporal correlation of wind speeds and wind directions at different positions of the wind farm

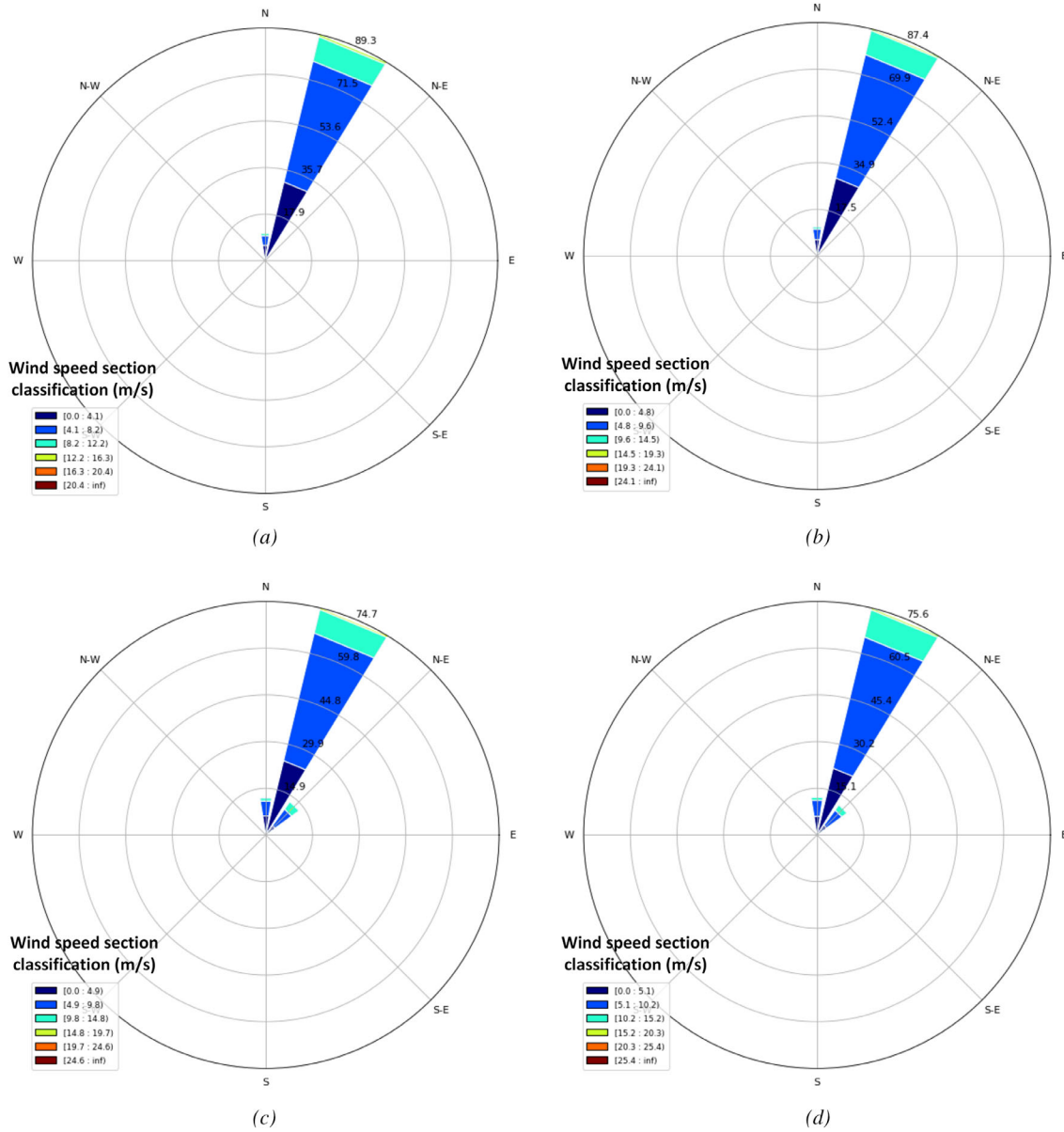
Wind farm has dozens to hundreds of wind turbines, usually influenced by distance and eddy current effect of wind turbine blades, the wind speed, and direction of each wind turbine are not exactly the same but have the correlation. The change of wind speeds with time, measured by four benchmark wind turbines in a wind farm in China is shown in Figure 5, where  $v$  is

the wind speed value,  $t$  is the time, and the data resolution is 5 min. The distribution rule of wind directions is represented by wind direction rose map, and the wind direction rose maps of the four benchmark wind turbines based on the wind speed and wind direction data of 2017 are shown in Figure 6. The wind direction in the figure is represented by eight directions, namely east (E), west (W), south (S), north (N), southeast (S-E), southwest (W-S), northeast (N-E), northwest (N-W). Different colours represent different wind speed ranges, and the number above each colour represents the frequency statistics of the corresponding wind speed section in this direction.

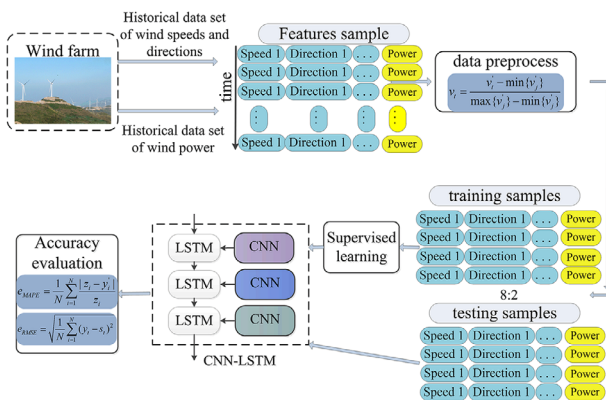
From the figures above, it can be seen that the wind speeds and wind directions distribution of four benchmark wind turbines in a wind farm have an obvious similarity. Therefore, it is very necessary to explore the wind speeds and wind directions temporal and spatial correlation of different wind turbines in the wind farm so as to achieve the power forecasting of the wind farm.

The wind speeds and wind directions measured at different benchmark wind turbines of wind farm are taken as input variables features, and the power values of wind farm are taken as output variables to construct the learning model. In order to extract the spatial correlations of each input features and consider the temporal correlation of historical input data, the relationship between wind speeds and wind directions of the benchmark wind turbines in the previous hour and the power value of wind farm in the next hour is established by CNN-LSTM algorithm (the data resolution is 5 min, and the input features of the previous hour are used to predict the wind power of the next hour). The input characteristic matrix of each data sample is represented as

$$V_t = \begin{bmatrix} v_{t-11}^1 & d_{t-11}^1 & v_{t-11}^2 & d_{t-11}^2 & \cdots & v_{t-11}^b & d_{t-11}^b \\ v_{t-10}^1 & d_{t-10}^1 & v_{t-10}^2 & d_{t-10}^2 & \cdots & v_{t-10}^b & d_{t-10}^b \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{t-i}^1 & d_{t-i}^1 & v_{t-i}^2 & d_{t-i}^2 & \cdots & v_{t-i}^b & d_{t-i}^b \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ v_{t-1}^1 & d_{t-1}^1 & v_{t-1}^2 & d_{t-1}^2 & \cdots & v_{t-1}^b & d_{t-1}^b \\ v_t^1 & d_t^1 & v_t^2 & d_t^2 & \cdots & v_t^b & d_t^b \end{bmatrix} \quad (6)$$



**FIGURE 6** Wind rose maps of four benchmark wind turbines in a wind farm. (a) Benchmark wind turbine 1, (b) benchmark wind turbine 2, (c) benchmark wind turbine 3, (d) benchmark wind turbine 4



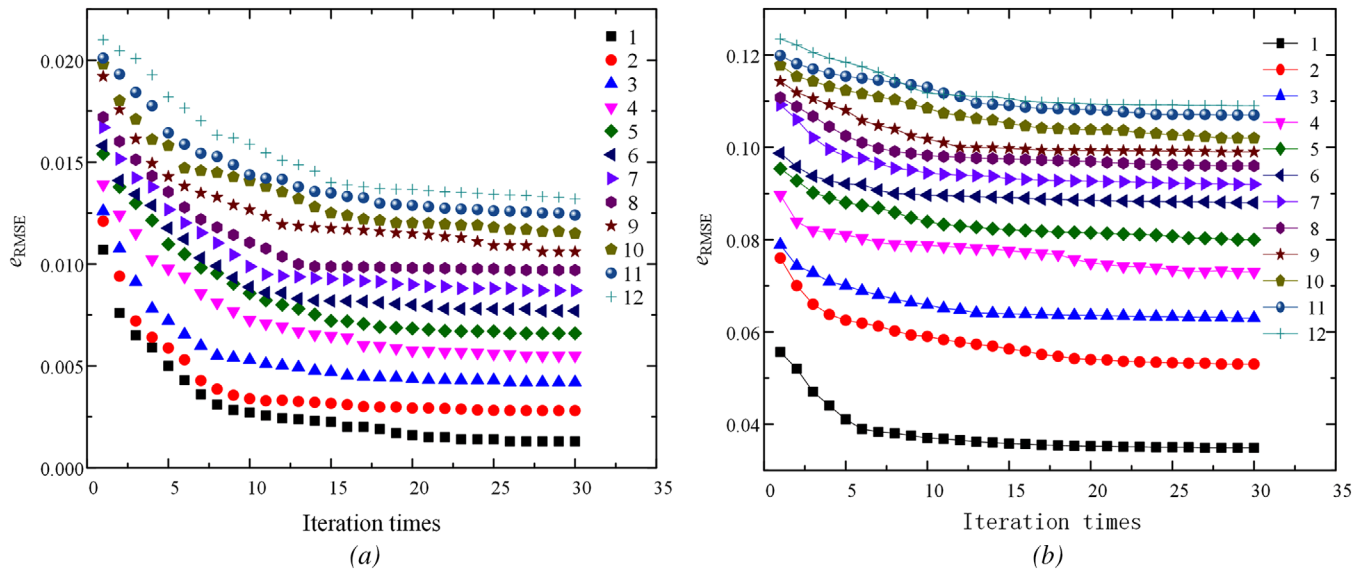
**FIGURE 7** The wind prediction model based on CNN-LSTM

where  $t$  is the current time,  $v_{t-i}^b$  is the wind speed of the  $b$ -th benchmark wind turbine at  $t-i$ , and  $d_{t-i}^b$  is the wind direction of the  $b$ -th benchmark wind turbine at  $t-i$ .

The corresponding output is the wind farm power in the next hour, which is shown below as

$$P_t = [p_t \ p_{t+1} \ \dots \ p_{t+11}]^T \quad (7)$$

The wind power forecasting is to find the mapping relationship between  $V_t$  and  $P_t$  and consider the influence of historical sequence data. It is a high-dimensional regression problem, which needs CNN-LSTM algorithm to mine the temporal and spatial correlation of features.



**FIGURE 8** The analysis of model generalisation ability. (a) the errors of the 12 networks with 1–12, (b) the errors of the 12 networks with 1–12

**TABLE 2** Mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE) and normalised RMSE (NRMSE) for 5 min–1 h (1–12 steps) ahead prediction of convolutional neural networks (CNN), LSTM and CNN-LSTM

Time	$e_{MAE}$ (MW)			$e_{MAPE}$ (%)			$e_{RMSE}$ (MW)			$e_{NRMSE}$ (%)		
	CNN	LSTM	CNN-LSTM	CNN	LSTM	CNN-LSTM	CNN	LSTM	CNN-LSTM	CNN	LSTM	CNN-LSTM
5 min	0.648	0.510	0.452	1.309	1.030	0.911	1.483	1.038	0.836	2.996	2.097	1.689
10 min	1.162	0.912	0.721	2.347	1.842	1.452	2.294	1.769	1.542	4.634	3.573	3.115
15 min	1.701	1.398	1.288	3.436	2.824	2.478	2.873	2.564	1.978	5.804	5.179	3.996
20 min	2.276	1.922	1.589	4.598	3.883	3.398	3.212	2.833	2.461	6.489	5.732	4.971
25 min	2.832	2.310	1.984	5.721	4.667	3.934	3.791	3.263	2.968	6.91	6.63	6.68
30 min	3.211	2.705	2.372	6.487	5.465	4.848	4.138	3.715	3.469	8.359	7.658	7.008
35 min	3.731	3.013	2.705	7.537	6.087	5.752	4.684	4.173	3.735	9.462	8.430	7.545
40 min	4.133	3.521	3.156	8.349	7.113	6.681	5.047	4.834	4.121	10.19	9.765	8.325
45 min	4.590	3.995	3.601	9.273	8.071	7.897	5.392	5.087	4.327	10.89	10.27	8.741
50 min	4.998	4.313	3.976	10.09	8.713	8.018	5.661	5.229	4.889	11.43	10.56	9.876
55 min	5.679	4.645	4.089	11.47	9.383	8.791	5.993	5.564	5.025	12.10	11.64	10.15
60 min	6.337	5.286	4.923	12.80	10.68	9.545	6.524	6.099	5.136	13.10	12.13	10.98
Average errors	3.442	2.879	2.573	6.948	5.813	5.316	4.258	3.864	3.398	8.53	7.80	6.92

### 3.2 | Wind farm power forecasting model and error calculation method based on CNN-LSTM

The construction process of wind power prediction model based on CNN-LSTM is shown in Figure 7.

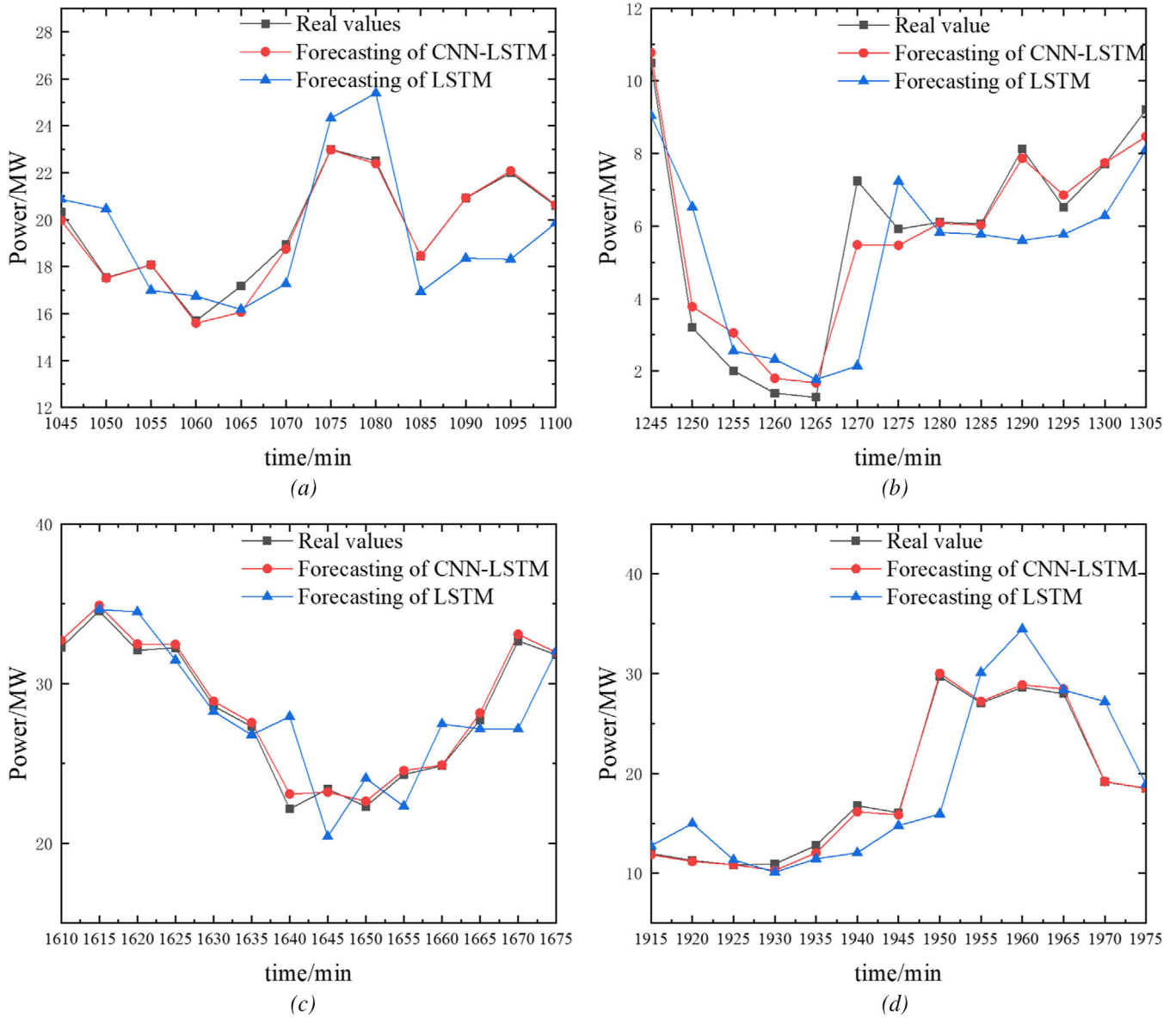
The specific steps are as follows:

#### 1. Data preprocessing

The importance of normalisation is actually so that one feature is not weighted more than another because it has a higher variance magnitude. In this scenario, the

wind direction would naturally be weighted more, and thus wind speeds and direction must be normalised. Different attribute input features have a different correlation with wind power, and the orders of magnitude of features are different. Therefore, the input features are processed by the method of normalisation, and the wind power feature data is mapped to 0~1 to form the initial features set [34]:

$$v'_t = \frac{v_t - \min\{v_j\}}{\max\{v_j\} - \min\{v_j\}}, \quad j = 1, 2, \dots, k \quad (8)$$



**FIGURE 9** The comparison for prediction results of next hour (12 steps) in four samples between CNN-LSTM, CNN and LSTM. (a) The 5000th sample, (b) the 10,000th sample, (c) the 15,000th sample, (d) the 20,000th sample

where  $v_t$  is the actual value of the input feature,  $v_t'$  is the normalised input feature, and the number of samples is  $k$ .

The final wind power forecasting data shall be achieved by inverse normalisation according to Equation (9) [34]:

$$p_t^{pre} = p_t^{pre'} \times (\max\{p_j\} - \min\{p_j\}) + \min\{p_j\}, \quad (9)$$

$$j = 1, 2, \dots, k$$

where  $p_t^{pre}$  is the predicted value after inverse normalisation;  $p_t^{pre'}$  is the predicted value of wind power,  $p_t$  is the actual value of wind power, and the number of sample points is  $k$ .

### 1. Model training

The normalised data is divided into a training set and test set according to 8:2. The training set is used for model training, and the error prediction is obtained by post-departure operation. The mean squared error (MSE) is used as the objective function, and the parameters of the whole model are updated from the back to the front by using the Adam optimisation algorithm to find a set of parameters that minimise the MSE. Adam optimisation algorithm is a random objective function-step optimisation algorithm based on low-order moment adaptive estimation [35]. In order to



**TABLE 3** The forecasting power values of CNN, LSTM and CNN-LSTM for the four samples of next hour (12 steps)

Sample	Time (min)	Wind farm power (kW)			
		Real values	CNN	LSTM	CNN-LSTM
1	5	34,720.41	36,380.652	34,368.746	33,937.188
	10	31,119.301	35,480.43	33,185.14	32,741.406
	15	30,705.721	35,168.082	33,803.734	31,940.156
	20	29,876.85	35,083.605	34,133.953	32,985.223
	25	32,880.91	31,598.648	34,355.18	32,112.92
	30	31,392.381	33,140.293	33,629.34	31,194.94
	35	29,753.711	34,199.05	33,884.062	30,657.139
	40	30,271.99	34,769.38	32,084.904	32,161.428
	45	29,890.359	31,584.9	33,973.277	31,764.613
	50	31,280.33	34,558.62	33,142.387	31,635.4
	55	31,546.961	33,692.637	32,215.523	29,640.443
	60	29,098.99	33,436.055	34,129.875	31,754.287
2	5	12,716.36	9453.897	10,412.582	14,127.441
	10	11,395.87	10,747.163	9846.695	14,485.338
	15	9569.8	10,052.157	10,286.979	13,551.66
	20	11,200.85	10,841.17	10,920.691	14,157.615
	25	8012.09	10,622.405	10,959.345	14,323.592
	30	9911.2	9954.355	10,223.284	12,792.364
	35	9654.22	8679.003	9572.729	15,278.311
	40	18,176.539	10,885.018	9644.081	15,593.974
	45	16,207.82	11,181.835	10,554.634	16,554.01
	50	17,934.699	10,975.403	10,056.064	16,016.255
	55	18,670.631	10,002.928	11,194.049	14,377.775
	60	16,668.301	10,330.784	11,382.005	12,437.075
3	5	11,414.45	10,034.183	10,903.952	12,437.857
	10	13,295.92	10,047.826	10,967.445	12,544.46
	15	12,332.99	11,117.005	11,358.424	12,956.259
	20	11,913.91	11,434.607	11,600.791	13,085.642
	25	10,530.96	10,777.605	11,857.121	13,700.035
	30	9599.46	11,184.91	11,671.862	12,625.193
	35	8368.22	11,091.872	10,623.982	13,039.483
	40	7587.65	12,063.932	10,858.019	13,511.718
	45	6144.46	11,219.689	11,978.992	13,685.547
	50	5900.34	12,035.183	11,729.29	13,867.837
	55	6129.53	12,239.117	12,965.97	13,407.783
	60	6060.73	11,132.08	13,039.952	14,411.142
4	5	6933.62	3853.386	3707.6467	7190.1113
	10	7400.749	3690.3926	3814.5557	6733.379
	15	6943.53	4185.029	4076.5503	8260.487
	20	6922.69	4670.879	4390.6025	7143.3296
	25	6042.29	4080.2886	4210.668	8417.876
	30	5127.18	4113.7373	3863.0166	8129.9614
	35	2813.58	3262.7854	3324.9043	8510.74

(Continues)

**TABLE 3** (Continued)

Sample	Time (min)	Wind farm power (kW)			
		Real values	CNN	LSTM	CNN-LSTM
	40	3953.06	4744.015	3969.952	9167.915
	45	5025.35	4008.8337	4438.9307	7972.131
	50	5289.72	5355.1353	3498.628	8273.88
	55	6619.679	5108.5923	3498.628	7546.9165
	60	6162.95	5946.305	5182.143	8813.828

prevent overfitting,  $L^2$  parameter regularisation is used in the full connection layer. The objective function of adding regular terms is shown in Equation (10) [36]:

$$e_{MSE} = \frac{1}{n} \sum_{t=1}^n (p_t^{pre'} - p_t')^2 + \frac{\alpha}{2} w^T w \quad (10)$$

where  $p_t'$  is the normalised wind power,  $\alpha$  is the regular term parameter,  $w$  is the connection weight vector of the full connection layer, and  $n$  is the number of training samples.

In addition, the optimisation method of random deactivation (dropout) is introduced. In the learning process, part of the weight or output of the hidden layer is randomly zeroed to reduce the interdependence between the hidden layer neurons so as to realise the regularisation of the NN.

### 1. Evaluating index

The test set is used to evaluate the accuracy of the model. In this study, the MAE, the mean absolute percentage error (MAPE), root mean square error (RMSE) and the normalised RMSE (NRMSE) are used as the evaluation indexes [36]:

$$e_{MAE} = \frac{1}{N} \sum_{i=1}^N |p_t^{pre} - p_t| \quad (11)$$

$$e_{MAPE} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{p_t^{pre} - p_t}{p_t} \right| \quad (12)$$

$$e_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_t^{pre} - p_t)^2} \quad (13)$$

$$e_{NRMSE} = \frac{1}{P_N} \sqrt{\frac{1}{N} \sum_{i=1}^N (p_t^{pre} - p_t)^2} \quad (14)$$

Here,  $N$  is the total sample number of the forecasting series, and  $P_N$  is the rated installed capacity.  $p_t$  and  $p_t^{pre}$  are the actual and forecasted values of the  $i$ -th wind power samples, respectively.

## 4 | EXAMPLE ANALYSIS

### 4.1 | Sample processing and parameters setting

In order to verify the validity of the model proposed in this study, data samples of a wind farm in China are selected for wind power prediction. The wind farm installed 33 wind turbines with a total installed capacity of 49.5 MW. It is located in the middle and low mountain landform with undulating terrain and different height distribution of wind turbine. The historical power values and meteorological factors of the target wind farm in 2017 were measured, a total of 105,120 time points of data with a resolution of 5 min. After deleting the missing data, a total of 104,800 time points were collected. After data reconstruction (12 time point data for each sample), there are a total of 8732 samples. The first 6732 samples are taken as the training set, and the remaining 2000 samples as the test set. The training set is used to calculate the gradient and update the weights, and the test set gives the final evaluation index.

The multi-model method is selected for multi-step prediction. Based on the input features of the historical sequences length of 12, the 5 min—1 h (1–12 step) ultra-short-term forecasting of the power time series of the wind farm is carried out. Twelve CNN-LSTM multi-step models with the same network structure share the same input, and each model is trained independently without interference. The platform used for programming and the set super parameters is shown in Table 1.

### 4.2 | Generalisation capability of the model

In order to prevent overfitting, the regularisation coefficient of  $L^2$  parameter and dropout are set to reduce network complexity and improve the generalisation ability of the model. By comparing the variation of 5 min—1 h (1–12 steps) RMSE  $e_{RMSE}$  for training and test sets with the iteration times, the generalisation ability of the model is analysed as shown in Figure 8. All the errors of the 12 networks with 1–12 predicted values in the training set decrease with the increase of iteration times, and there is still good convergence in the test set, thus the model has strong generalisation ability.

### 4.3 | Comparison of forecasting performance between CNN, LSTM and CNN-LSTM

The CNN, LSTM can significantly improve the forecasting accuracy of wind farm power compared with ANN and SVM and other conventional machine learning algorithms [37, 38]. Therefore, this study only analyses the advantages of CNN-LSTM compared with CNN and LSTM in extracting spatial correlation features of multivariate time series. The CNN and LSTM network uses the same input data to predict the wind farm power for the next hour. The comparison of the predic-

tion errors  $e_{MAE}$ ,  $e_{MAPE}$ ,  $e_{RMSE}$  and  $e_{NRMSE}$  of the CNN, LSTM and CNN-LSTM models for the test set in the next hour (12 steps) is shown in Table 2.

It can be seen from Table 2 that the  $e_{MAE}$ ,  $e_{MAPE}$ ,  $e_{NRMSE}$  and  $e_{RMSE}$  of CNN-LSTM model are lower than that of CNN and LSTM model in every step in the next hour. The average MAE, MAPE, RMSE and NRMSE of the total model decrease by 33.77%, 30.69%, 25.3% and 23.3% (compared with CNN), 12.0%, 10.6%, 14% and 12.7% (compared with LSTM), respectively. The comparison results of the wind farm power forecasting values and the real power values of the test sets for the three algorithms are shown in Figure 9. Four samples of prediction results of next hour (12 steps) are randomly selected and the specific power values are shown in Table 3. It can be seen from Figure 9 that CNN-LSTM has better forecasting effect than CNN and LSTM.

## 5 | CONCLUSIONS

Considering the temporal and spatial correlation of meteorological factors of different sites throughout the wind farm, a novel STCM based on CNN-LSTM for ultra-short-term wind power prediction is proposed. In order to deeply extract the temporal and spatial correlations for the STCM based on CNN-LSTM, a new data reconstruction method is proposed, saving data processing time while extracting spatio-temporal correlations for multi-step wind power prediction. Specifically, the wind speeds and wind directions measured by benchmark wind turbines are reconstructed and used as the input of the model. The STCM for wind farm power forecasting model and error calculation method based on CNN-LSTM is presented. To verify the accuracy and superiority of the proposed model, the measured wind speeds and wind directions by four benchmark wind turbines and the wind power data from a wind farm in China are used in the experiments. Moreover, the effectiveness and superiority of the proposed CNN-LSTM model are verified by the comparison of four evaluation metrics with CNN and LSTM individually used. The experiment results show that the average MAE, MAPE, RMSE and NRMSE of the total model decrease by 33.77%, 30.69%, 25.3% and 23.3% (compared with CNN), 12.0%, 10.6%, 14% and 12.7% (compared with LSTM), respectively. The proposed STCM based on CNN-LSTM for multi-step wind power forecasting fully considers the spatio-temporal correlation of meteorological factors throughout the wind farm and can forecast the power of wind farm more accurately.

However, there are still some shortcomings that need to be completed. The different contribution degree of meteorological factors of different sites to the power of the wind farm needs to be explored further. Assigning different weights to different contribution degree of factors in the learning process will help further improve the forecasting accuracy of wind power based on the deep learning algorithm, which is the next research focus. In addition, it is necessary to improve the optimiser algorithm of the NN.

## REFERENCES

1. Hu, Q., et al.: Short-term wind speed or power forecasting with heteroscedastic support vector regression. *IEEE Trans. Sustainable Energy* 7(1), 241–249 (2015)
2. Feng, S., et al.: Study on the physical approach to wind power prediction. *Proc. CSEE* 30(2), 1–6 (2010)
3. Xue, Y., et al.: A review on short-term and ultra-short-term wind power prediction. *Autom. Electr. Power Syst.* 39(6), 141–151 (2015)
4. Xiu, C., et al.: Short-term prediction method of wind speed series based on kalman filtering fusion. *Trans. China Electrotech. Soc.* 29(2), 253–259 (2014)
5. Sun, G., Wei, Z., Zhai, W.: Short term wind speed forecasting based on RVM and ARMA error correcting. *Trans. China Electrotech. Soc.* 27(8), 187–193 (2012)
6. Quan, H., Srinivasan, D., Khosravi, A.: Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Trans. Neural Networks Learn. Syst.* 25(2), 303–315 (2013)
7. Santamaría-Bonfil, G., Reyes-Ballesteros, A., Gershenson, C.: Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy* 85, 790–809 (2016)
8. Lee, D., Baldick, R.: Short-term wind power ensemble prediction based on Gaussian processes and neural networks. *IEEE Trans. Smart Grid* 5(1), 501–510 (2013)
9. Haque, A.U., Nehrir, M.H., Mandal, P.: A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting. *IEEE Trans. Power Syst.* 29(4), 1663–1672 (2014)
10. Noorollahi, Y., Jokar, M.A., Kalhor, A.: Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. *Energy Convers. Manage.* 115, 17–25 (2016)
11. Zhao, Z., et al.: LSTM network: A deep learning approach for short-term traffic forecast. *IET Intel. Transport Syst.* 11(2), 68–75 (2017)
12. Lu, Z., et al.: New internet based operation pattern design of wind power forecasting system. *Power Syst. Technol.* 40(1), 125–131 (2016)
13. Zhang, C., et al.: Predictive deep Boltzmann machine for multiperiod wind speed forecasting. *IEEE Trans. Sustainable Energy* 6(4), 1416–1425 (2015)
14. Hu, Q., Zhang, R., Zhou, Y.: Transfer learning for short-term wind speed prediction with deep neural networks. *Renewable Energy* 85, 83–95 (2016)
15. Wang, H., et al.: Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl. Energy* 182, 80–93 (2016)
16. Abedinia, O., Amjady, N.: Short-term wind power prediction based on Hybrid Neural Network and chaotic shark smell optimization. *Int. J. Precis. Eng. Manuf. Green Technol.* 2, 245–254 (2015)
17. Amjady, N., Abedinia, O.: Short term wind power prediction based on improved kriging interpolation, empirical mode decomposition, and closed-loop forecasting engine. *Sustainability* 9(11), 2104 (2017)
18. Abedinia, O., et al.: Shafie-khah, M., Catalão, J.P.S., Improved EMD-based complex prediction model for wind power forecasting. *IEEE Trans. Sustainable Energy* 11(4), 2790–2802 (2020)
19. Yao, G., Lei, T., Zhong, J.: A review of convolutional-neural-network-based action recognition. *Pattern Recognit. Lett.* 118, 14–22 (2019)
20. Wang, H., et al.: Deep learning based ensemble approach for probabilistic wind power forecasting. *Appl. Energy* 188, 56–70 (2017)
21. Mi, X., Liu, H., Li, Y.: Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Convers. Manage.* 180, 196–205 (2019)
22. Liu, H., Mi, X., Li, Y.: 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.* 156, 498–514 (2018)
23. Shi, Z., Liang, H., Dinavahi, V.: Direct interval forecast of uncertain wind power based on recurrent neural networks. *IEEE Trans. Sustainable Energy* 9(3), 1177–1187 (2017)
24. Zhu, Q., et al.: Short-term wind power forecasting based on LSTM. *Power Syst. Technol.* 41(12), 3797–3802 (2017)
25. Chen, Y., et al.: Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. *Energy Convers. Manage.* 185, 783–799 (2019)
26. Liu, H., Mi, X., Li, Y.: Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Convers. Manage.* 166, 120–131 (2018)
27. Ak, R., Vitelli, V., Zio, E.: An interval-valued neural network approach for uncertainty quantification in short term wind speed prediction. *IEEE Trans. Neural Networks Learn. Syst.* 26(11), 2787–2800 (2015)
28. LeCun, Y., et al.: Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1(4), 541–551 (1989)
29. LeCun, Y., Bottou, L.: Gradient-based learning applied to document recognition. *Proc. IEEE* 86(11), 2278–2324 (1998)
30. Huang, H., et al.: Pure electric vehicle nonstationary interior sound quality prediction based on deep CNNs with an adaptable learning rate tree. *Mech. Syst. Sig. Process.* 148, 107170 (2021)
31. Li, Y., et al.: Small-signal stability assessment of power system based on convolutional neural network. *Autom. Electr. Power Syst.* 43(2), 50–57 (2019)
32. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* 9(8), 1735–1780 (1997)
33. Feng, X., Qiu, L., Guo, X.: Student feedback text academic emotion recognition method based on LSTM model. *Open Educ. Res.* 2, 1–13 (2019)
34. Lin, Z., Liu, X.: Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network. *Energy* 201, 117693 (2020)
35. Diederik, P.K., Jimmy, L.B.: Adam: A method for stochastic optimization. In: *International Conference on Learning Representations*, San Diego, USA, pp. 1–15 (2015)
36. Zang, H., et al.: Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renewable Energy* 160, 26–41 (2020)
37. Wang, K., Qi, X., Liu, H.: Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy* 189, 116225 (2019)
38. Qin, Y., et al.: Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. *Appl. Energy* 236, 262–272 (2019)

**How to cite this article:** Wu Q, Guan F, Lv C, Huang Y. Ultra-short-term multi-step wind power forecasting based on CNN-LSTM. *IET Renewable Power Generation*. 2021;15:1019–1029.

<https://doi.org/10.1049/rpg2.12085>