



OPEN ACCESS

EDITED BY

Bin Zhou,
Hunan University, China

REVIEWED BY

Wei Yao,
Huazhong University of Science and
Technology, China
Chuan-Ke Zhang,
China University of Geosciences
Wuhan, China
Bo Yang,
Kunming University of Science and
Technology, China

*CORRESPONDENCE

Haoxuan Xu,
haoxuanxu@zuel.edu.cn

SPECIALTY SECTION

This article was submitted to Process
and Energy Systems Engineering,
a section of the journal
Frontiers in Energy Research

RECEIVED 14 September 2022

ACCEPTED 30 September 2022

PUBLISHED 09 January 2023

CITATION

Ai X, Li S and Xu H (2023), Wind speed
prediction model using ensemble
empirical mode decomposition, least
squares support vector machine and
long short-term memory.
Front. Energy Res. 10:1043867.
doi: 10.3389/fenrg.2022.1043867

COPYRIGHT

© 2023 Ai, Li and Xu. This is an open-
access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Wind speed prediction model using ensemble empirical mode decomposition, least squares support vector machine and long short-term memory

Xueyi Ai¹, Shijia Li¹ and Haoxuan Xu^{2*}

¹Evergrande School of Management, Wuhan University of Science and Technology, Wuhan, China,

²School of Business Administration, Zhongnan University of Economics and Law, Wuhan, China

Due to the randomness and intermittency of wind, accurate and reliable wind speed prediction is of great importance to the safe and stable operation of power grid. In this paper, a novel hybrid wind speed forecasting model based on EEMD (Ensemble Empirical Mode Decomposition), LSSVM (Least Squares Support Vector Machine), and LSTM (Long Short-Term Memory) is proposed, aiming at enhancing the forecasting accuracy of wind speed. The original data series is firstly processed by EEMD and SE into a series of components with different frequencies. Subsequently, a combined mechanism composed of LSSVM and LSTM is presented to train and predict the high-frequency and low-frequency sequences, respectively. Finally, the predicted values of all the data sequences are superimposed to obtain the ultimate wind speed forecasting results. In order to respectively illustrate the superiority of data feature processing and combined prediction mechanism in the proposed model, two experiments are performed on the two wind speed datasets. In accordance with the four performance metrics of the forecasting results, the EEMD-LSTM-LSSVM model obtains a higher accuracy in wind speed prediction task.

KEYWORDS

wind speed forecasting, long short-term memory, least squares support vector machine, ensemble empirical mode decomposition, sample entropy

1 Introduction

In recent years, with the improvement of people's environmental awareness, renewable energy resources such as wind energy have attracted more and more attention worldwide. Unlike traditional fossil fuels that results in environmental pollution and global warming, wind energy, as a clean and pollution-free form of practical renewable energy, has become a research hotspot around the world (Huang et al., 2022). According to a wind report published by the Global Wind Energy Council, more than 90 GW of new wind power capacity was installed in 2020, raising the total installed capacity to 743 GW, up 14.3% from the previous year. In particular, China's total

installed capacity accounted for 38.5% of global wind power capacity with the largest proportion (Wang et al., 2021). Despite such surprising growth, the randomness, uncertainty and intermittency of wind greatly affect the stable operation of large-scale power grid-integrated systems (Sun et al., 2022). An effective way to address this issue is to accurately predict wind speed.

Recently, a lot of prediction methods have been proposed, which are roughly divided into two categories, namely physical methods and statistical methods (Liu et al., 2018a). The physical methods, such as numerical weather prediction (NWP) and weather researcher forecasting (WRF), approximately simulate future wind speed using meteorological factors (atmospheric pressure, temperature, density, etc.) for comprehensive analysis (Xiao et al., 2018). One advantage of physical methods is to predict wind speed directly from real-time data, which, however, requires a large amount of physical information with high precision, thus leading to significant economic cost.

As the data science increasingly develops, statistical methods including conventional statistical models and artificial intelligence (AI)-based statistical models, have been widely regarded for time series forecasting (Hao and Tian, 2019). This type of method can effectively mine historical information and explain signal features (Aslam and Albassam, 2022). The conventional statistical models, like autoregressive (AR) (Poggi et al., 2003) and AR-integrated moving average (ARIMA) (Yunus et al., 2015), can predict wind speed on the basis of historical data and characterize linear fluctuation trend, but fail to adequately capture nonlinear features within wind speed data. To avoid nonlinearity, the AI-based statistical models have emerged continuously, including back-propagation neural network (BPNN) (Sun and Wang, 2018), support vector machine (SVM) (Liu et al., 2016) and least squares support vector machine (LSSVM) (Yuan et al., 2015). Among them, the structures of neural network are difficult to determine and the calculation process of traditional SVM is relatively complicated. In contrast, as an improved effective technology, LSSVM, which requires fewer parameters and owns higher convergence speed, has been widely utilized for predicting wind speed. For example, Du et al. (2018) proposed a novel robust hybrid system using LSSVM for electrical power system forecasting, and verified the satisfactory accuracy of the proposed hybrid model. The previous examples demonstrated that traditional AI-based methods are capable of extracting nonlinear and uncertain features of wind speed data but more suitable for handling series forecasting issues with low complexity.

To the best of our knowledge, the volatility and nonlinearity of wind speed series are attributed to the interference of various natural factors, thus shallow AI-based prediction models can hardly learn the sophisticated features from massive historical data, especially for the high-frequency sequence characterized by randomness and short-term dependency. Because of the hierarchical and distributed feature representations, deep

learning (DL) network possesses strong capability to predict high-frequency sequence and robustness to parameters (Hu and Chen, 2018). Deep learning methods including long short-term memory (LSTM) (Shahid et al., 2021), convolutional neural network (CNN) (Yu et al., 2020), deep belief network (DBN) (Wang et al., 2016a), and gated recurrent unit (GRU) (Li et al., 2022) have drawn much attention recently. Liu et al. (2018b) proposed a deep learning framework based on LSTM neural network for one-step forecasting of wind speed. In terms of multi-step forecasting, Moreno et al. (2020) forecast each component resulting from a two-stage signal decomposition strategy by employing LSTM network. As described earlier, DL-based LSTM neural network has been extensively used to forecast highly complex wind speed series due to its remarkable ability of learning and remembering both short and long-term features.

In recent years, it is widely recognized that hybrid or ensemble models based on signal processing have achieved admirable results in wind speed forecasting. Signal processing, namely data decomposition and reconstruction, is regarded as a key step to enhance the forecasting performance and efficiency of time series. Among these decomposition strategies, ensemble empirical mode decomposition (EEMD), as a modified version of empirical mode decomposition (EMD), can effectively avoid mode aliasing by adding normally distributed white noise into raw signal. Wang et al. (2016b) have proved that the prediction accuracy for wind speed can be improved by using EEMD.

Summarily, the aforementioned filter methods can decompose an original data signal into multiple subseries with various frequency-scales, making them more stationary and regular for further prediction. Nevertheless, the computation time of signal processing-based models will be inevitably prolonged with the increase in the number of components. In order to balance forecasting accuracy and economical cost, entropy has been developed to calculate the time complexity of data series as well as aggregate the decomposed components.

Due to the non-stationary nature of wind, wind data series contains complex characteristic information. Although filter methods can effectively reduce the volatility of wind series to some extent, components with different frequencies have different characteristics that can be matched with different learning mechanism to adequately exert the advantages of different models. Chen et al. (2022) used two different predictors, namely LSTM and improved BPNN, to respectively predict sequences with high-complexity and low-complexity. Hu and Chen (2018) similarly proposed a LSTM-based nonlinear combined mechanism to avoid the influence of single model on the prediction results. Additionally, Fu et al. (2020) proposed a composite prediction method containing KELM and ConvLSTM in the light for the frequency scales of each component to further improve the generalization capability and robustness of a single model. When it comes to data sequences mixed with different frequency scales, combining the strengths of different prediction models will be a

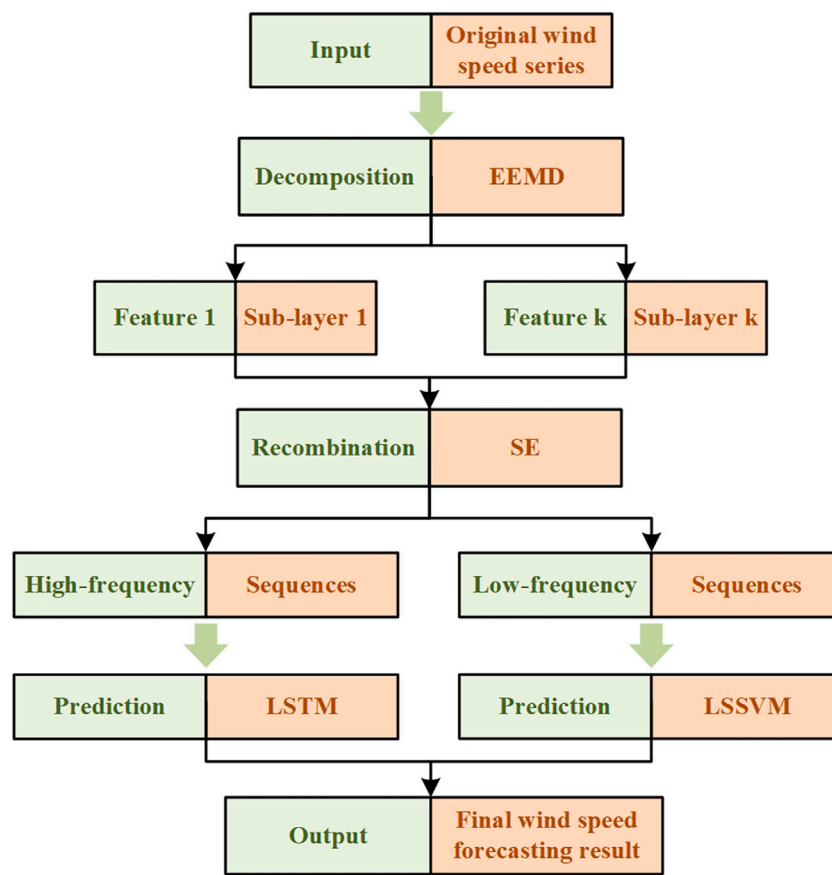


FIGURE 1
Flowchart of the EEMD-LSTM-LSSVM model.

promising way to achieve accurate prediction results. Thus, this paper integrates signal analysis method and different predictors together for wind speed forecasting.

The main contributions in this paper are described as follows:

- A hybrid wind speed prediction model (EEMD-LSTM-LSSVM) is proposed to overcome the disadvantage of single prediction model by using EEMD, LSSVM, and LSTM network.
- The original data signal is decomposed into a series of sub-layers applying EEMD to extract the fluctuation features, and then reconstructed into several frequency scales of sequences using sample entropy (SE) to reduce workload.
- To avoid the influence of single method on forecasting ability, a combined mechanism containing LSTM and LSSVM is developed. The high-frequency sequences are predicted by fitting into LSTM network, while remaining sequences (low-frequency ones) are predicted by LSSVM.
- The superiority of the developed combined mechanism is verified on two wind speed datasets by comparing single

benchmark models and hybrid benchmark models based on BPNN, SVR, LSSVM, LSTM, and persistence model.

The remainder of this paper is arranged as follows: [Section 2](#) describes hybrid EEMD-LSTM-LSSVM model. In [Section 3](#), two experiments are conducted on two datasets and comparative analysis is performed. [Section 4](#) summarizes the conclusions and future research directions.

2 The EEMD-LSTM-LSSVM model

2.1 Framework of EEMD-LSTM-LSSVM model

The flow chart of the EEMD-LSTM-LSSVM model is presented in [Figure 1](#), which is generally described as follow:

- The EEMD is firstly employed to decompose the original signal into different sub-layers with different complex

feature information, also known as intrinsic mode functions (IMFs).

- (2) To reduce the number of IMFs to be predicted, SE is used to calculate the sample entropy value of every sub-layer, and reconstruct it into several new high-frequency and low-frequency components.
- (3) Aiming at all components with different frequencies, LSTM network is adopted to forecast the high-frequency sequences, while LSSVM model to forecast low-frequency sequences.
- (4) The final forecasting results of original data series are obtained by superimposing the predicted values of all the sequences. In order to measure the prediction accuracy of all the experimental models, four performance evaluation metrics are adopted in this paper.

2.2 Ensemble empirical mode decomposition

As an effective signal-processing tool, EMD, proposed by Huang et al. (1998), has been widely implemented for time-frequency analysis. Its role is to adaptively decompose the original nonstationary sequence into multiple subsequences with different frequencies, also known as intrinsic mode functions (IMFs). However, the disadvantages of separation and intermittency results in the poor performance of separating similar frequency components and thus modal aliasing. To tackle this drawback and enhance the signal-processing ability of EMD, a modified version EEMD was introduced by Huang and Wu (2009) to effectively ameliorate mode aliasing by adding normally distributed white noise into raw non-stationary wind series. Great progress has been made on EEMD, an improvement and optimization of EMD in operation speed and computational efficient. In this paper, the difficulty in accurate forecasting of wind speed is attributed to its nonlinearity and non-stationarity. Accordingly, EEMD is employed to decompose raw wind speed series to make prediction less difficult and serve for subsequent aggregation.

2.3 LSTM-LSSVM-based combined mechanism

The data sequences of different frequencies obtained from EEMD and SE contain different characteristic information, which can be substituted into different learning mechanisms to avoid the influence of a single model on the forecasting performance. Consequently, hybrid models based on the combined mechanism are generally superior to single prediction models. Here, LSTM and LSSVM models are used to respectively predict high-frequency and low-frequency sequences. In order to validate the feasibility of the combined mechanism, two kinds of above sequences are selected from the

TABLE 1 Comparison of two prediction models using different sequences.

	Model	MAE	MSE	RMSE	MAPE
High-frequency	LSTM	0.1976	0.0686	0.2619	0.5662
	LSSVM	0.2403	0.0996	0.3156	0.6216
Low-frequency	LSTM	0.0090	0.00017	0.0132	0.0009
	LSSVM	0.0056	0.00004	0.0066	0.0005

following datasets for comparative experiment. The comparison results are shown in Table 1.

It can be intuitively observed from Table 1 that LSTM outperforms LSSVM in the processing of sequences with higher frequency, but inferior to LSSVM in the processing of sequence with lower frequency. Thus, this paper puts forward the combined mechanism of LSTM-LSSVM, whose strength is to avoid the preference of single predictor to data feature processing and minimize the prediction error (Hu and Chen, 2018).

3 Experiments and analysis

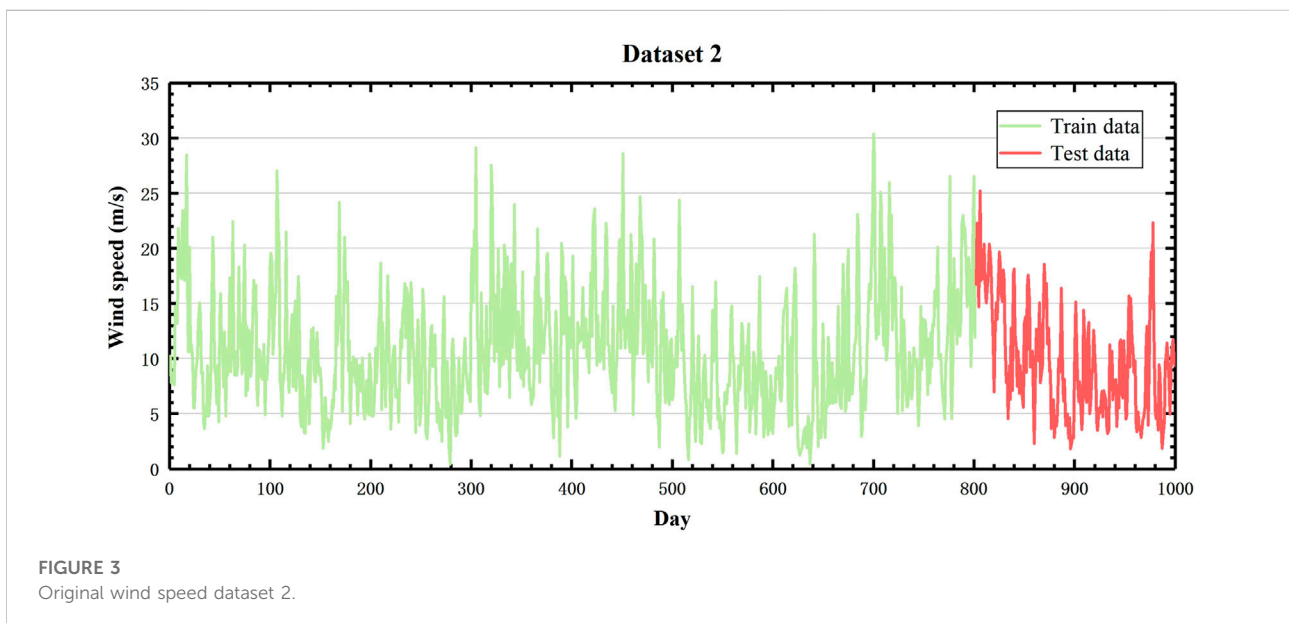
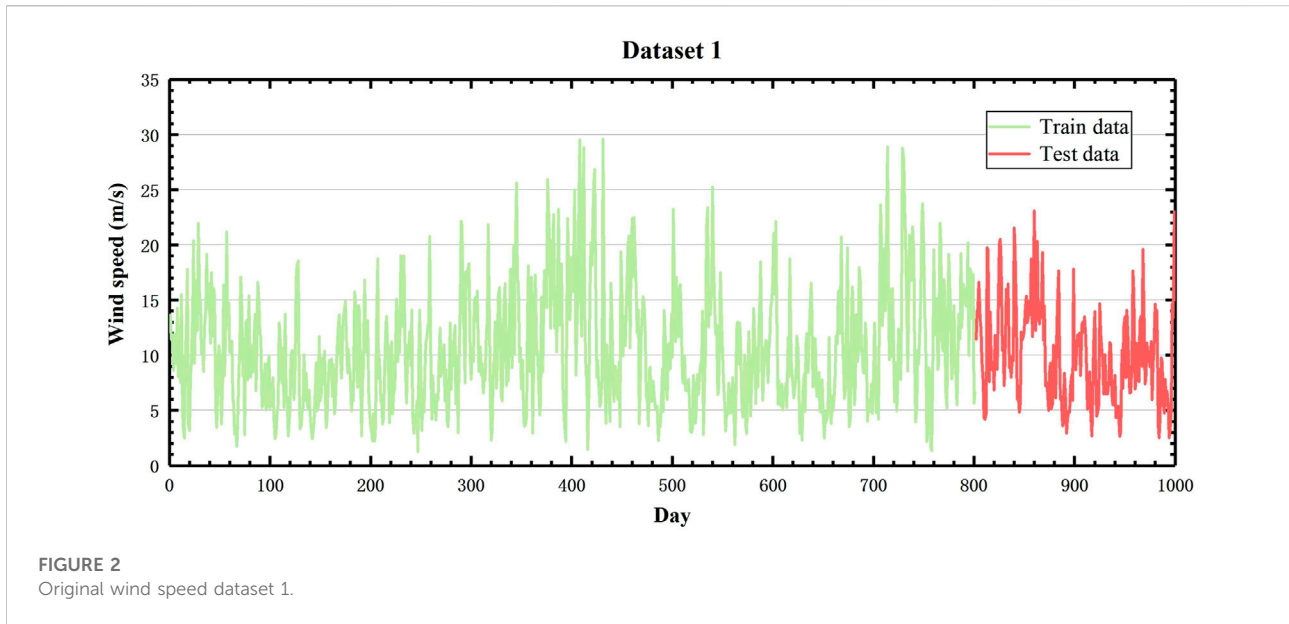
In this section, the two used wind speed datasets are firstly described in Section 3.1, and then the forecasting accuracy of the EEMD-LSTM-LSSVM model is to be demonstrated on two comparative experiments using two datasets. All the employed models are performed in MATLAB R2018a on a server equipped with 2.31 GHz CPU and 16 GB RAM.

3.1 Data description

The original data series of wind speed used in this paper are available on websites (<https://www.kaggle.com/datasets>). To validate the effectiveness of the proposed model, two groups of wind speed datasets containing 1,000 data points were selected as the experiment data. Taking five consecutive wind speed data as the input of models, each group of 1,000 data was constructed into a data matrix of 995 rows and five columns, in which the first 80% was used as training sets and the remaining 20% as test sets, as shown in Figures 2, 3. The statistics of two wind data series are given in Table 2.

3.2 Experimental setting

In order to further verify the superiority of the proposed model, two groups of comparative models were divided into single benchmark models and hybrid benchmark models for different purposes. Five single benchmark models, namely persistence model, BPNN model, support vector regression



(SVR), LSSVM, and LSTM model, were used for the first experiment, aiming to illustrate the superior capability of data preprocessing to improve the forecasting accuracy of wind speed by comparing the model after preprocessing with other five single models. The role of the second group of models was to investigate the impact of such combined mechanism on the forecasting performance of wind speed by comparing the multi-predictor model with single predictor model. Four benchmark prediction models, EEMD-BPNN, EEMD-SVR, EEMD-LSSVM, and EEMD-LSTM, were used for the second experiment.

According to Peng et al. (2020) and Moreno et al. (2020), the key parameters of the above employed models are set by the trials, as listed in Table 3. Specifically, the two parameters for EEMD algorithm, namely Nstd and NE, are set as 0.2 and 100 respectively, according to Sun and Wang (2018). For BPNN model, the number of hidden neurons is 10, the maximum training number is 2,000, and the target training error is 0.005, according to Peng et al. (2020) and Cui et al. (2021). In terms of LSSVM, one key parameter r is set to 200 and another parameter δ to 12 by the trials (Yuan et al., 2015). The

TABLE 2 The statistical information of two data sets.

Data set	Samples	Statistical information (m/s)			
		Max	Min	Mean	Standard deviation
Data set 1	Entire data (995)	29.58	1.25	10.51	5.23
	Train data (796)	29.58	1.25	10.61	5.40
	Test data (199)	23.09	2.54	10.16	4.52
Data set 2	Entire data (995)	30.37	0.42	10.55	5.35
	Train data (796)	30.37	0.42	10.76	5.41
	Test data (199)	26.54	1.79	9.88	5.18

TABLE 3 Key parameter setting for each model.

Model	Parameter	Value
EEMD	Nstd	0.2
	NE	100
BPNN	Number of hidden neurons	10
	Maximum of epochs	2000
	Goal error of training	0.005
LSSVM	Regularization parameter r	200
	Kernel bandwidth δ	12
LSTM	Optimizer	Adam
	Number of hidden units	200
	Initial learn rate	0.009
	Learn rate drop period	125
	Learn rate drop factor	0.2
	Maximum of epochs	500

TABLE 4 EEMD subsequences and recombination.

	IMFs	SE value	Result	New subs
Data set 1	IMF1	1.8208	IMF1	Sub1
	IMF2	1.0934	IMF2	Sub2
	IMF3	0.6482	IMF3&IMF4&IMF5	Sub3
	IMF4	0.5490		
	IMF5	0.4199		
	IMF6	0.1718	IMF6	Sub4
	IMF7	0.0379	IMF7&IMF8&IMF9	Sub5
	IMF8	0.0345		
	IMF9	0.0120		
Data set 2	IMF1	1.7741	IMF1	Sub1
	IMF2	1.2256	IMF2	Sub2
	IMF3	0.6511	IMF3&IMF4&IMF5	Sub3
	IMF4	0.5450		
	IMF5	0.4074		
	IMF6	0.1825	IMF6	Sub4
	IMF7	0.0446	IMF7&IMF8&IMF9	Sub5
	IMF8	0.0339		
	IMF9	0.0049		

parameters for SVR model are also set by the trials. One hidden layer with 200 neurons is adopted for LSTM model, with 500 epochs at most, according to Moreno et al. (2020). In addition, four typical performance metrics, namely mean absolute error (MAE), root mean square error (RMSE), mean absolute percent error (MAPE) and coefficient of determination (R^2), are used to measure the prediction accuracy of the employed models in this paper. The equations for these metrics are given as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

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

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

where N is the number of test samples, \bar{y} is the mean value of test samples, and y_i and \hat{y}_i are the i th actual value and predicted value respectively. Note that the greater the value of R^2 is, the better the forecasting performance is.

3.3 Data decomposition and reconstruction

As an adaptive signal analysis method, EEMD was firstly used to decompose the original signal into nine sub-layers including eight IMFs and one residual, and then SE was used to calculate the sample entropy values of each IMF component. The greater the SE value is, the higher frequency the time series

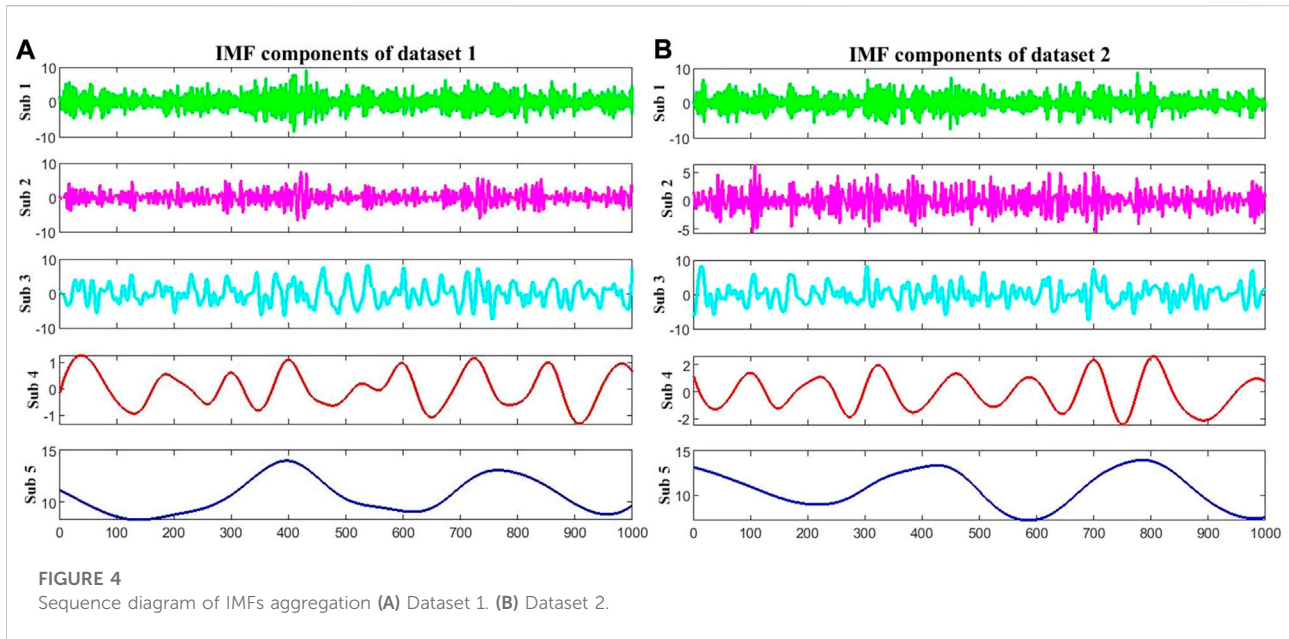


FIGURE 4 Sequence diagram of IMFs aggregation (A) Dataset 1. (B) Dataset 2.

TABLE 5 Performance metrics of six models on experiment I.

	Dataset 1				Dataset 2			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
Persistence	3.1536	4.2552	0.3468	0.0991	3.0819	3.8675	0.3825	0.4016
BPNN	2.9501	3.7575	0.3669	0.2975	3.0079	3.6693	0.4447	0.4613
SVR	2.9203	3.7602	0.3403	0.2965	2.9071	3.6102	0.4064	0.4786
LSSVM	2.8843	3.7399	0.3459	0.3040	2.9690	3.6391	0.4324	0.4702
LSTM	2.8844	3.6851	0.3561	0.3243	2.8885	3.5409	0.4183	0.4984
Proposed	0.9775	1.2712	0.1127	0.9196	0.8720	1.0921	0.1132	0.9523

The bold values mean the best results among all the comparative models.

has. In order to reduce workload and avoid excessive decomposition, nine sub-layers were reconstructed into two high-frequency and three low-frequency sequences according to similar SE values, as presented in Table 4 and Figure 4.

3.4 Result analysis

In this section, the performance of EEMD-LSTM-LSSVM model is to be proved by comparing with single benchmark models and hybrid benchmark models respectively using two wind speed datasets.

3.4.1 Comparative experiment I: EEMD-LSTM-LSSVM and single benchmark models

In order to demonstrate the impact of EEMD-based preprocessing method in the proposed hybrid model, single

persistence model, BPNN, SVR, LSSVM, and LSTM models are employed for comparison. The prediction results of the six models applied on two datasets are presented in Table 5, where the optimal result is highlighted in bold. Figure 5 shows the wind speeds predicted by the six models. Figure 6 further displays the line comparison between the actual data and the predicted data obtained by EEMD-LSTM-LSSVM for two datasets.

It can be clearly observed from Table 5 and Figures 5, 6 that the EEMD-LSTM-LSSVM model proposed in this paper is obviously superior to five single benchmark models, with the MAE, RMSE, MAPE, and R² values of 0.9775, 1.2712, 0.1127, and 0.9196 for dataset 1 and with the MAE, RMSE, MAPE, and R² values of 0.8720, 1.0921, 0.1132, and 0.9523 for dataset 2, respectively, which suggests that the hybrid prediction model based on data preprocessing is more reliable to extract different characteristic information. Among the five compared models, LSTM model performs the best especially for dataset 2, which

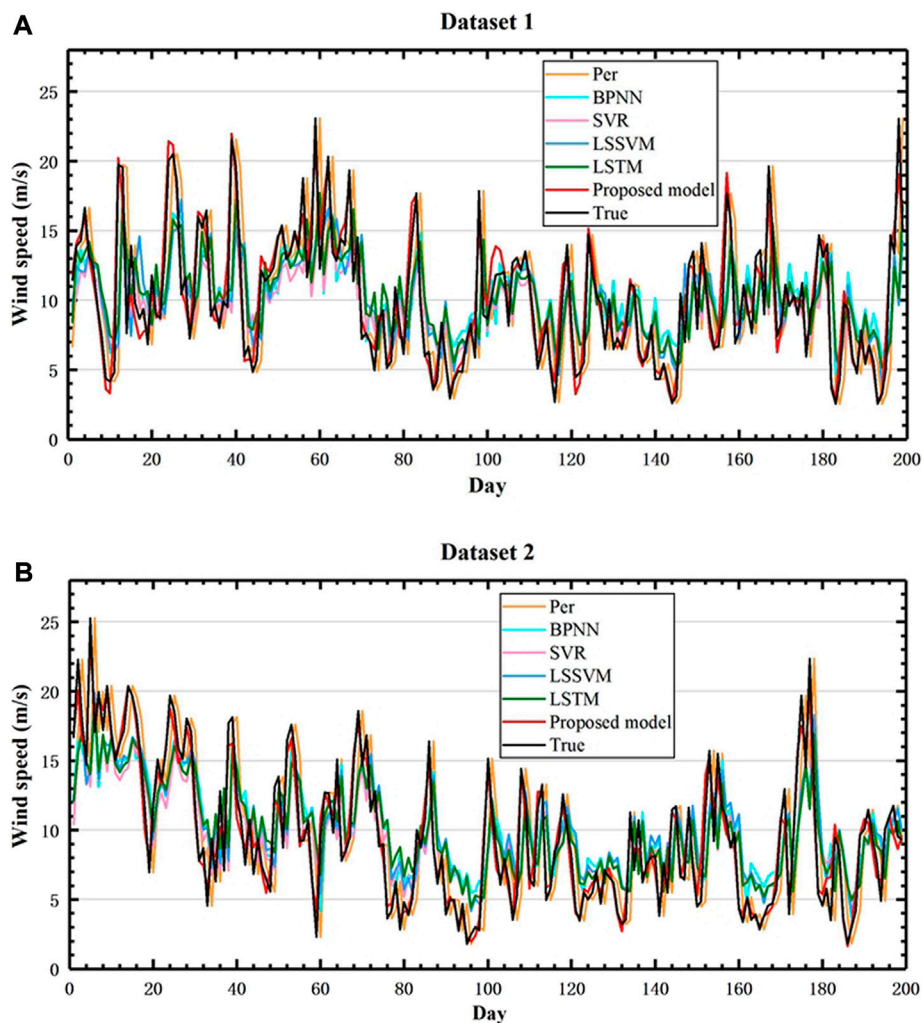


FIGURE 5
The line chart of forecasting results on experiment I (A) Dataset 1. (B) Dataset 2.

indicates that DL-based prediction methods outperform traditional models.

3.4.2 Comparative experiment II: EEMD-LSTM-LSSVM and hybrid benchmark models

The second experiment is devoted to verify the superiority of combined mechanism of two predictors to EEMD-BPNN, EEMD-SVR, EEMD-LSSVM, and EEMD-LSTM. Table 6 shows the performance metrics of five hybrid models, and Figure 7 further displays the 3D bar graphs of forecasting results using two datasets respectively. In addition, Figure 8 shows the wind speeds forecast by the five models.

As can be clearly seen from Table 6 and Figures 7, 8, whether dataset 1 or dataset 2, the EEMD-LSTM-LSSVM model favorably ranks among the four hybrid compared

models, which demonstrates the superiority of combined mechanism for two predictors to improve the accuracy of forecasting results. Additionally, the employed hybrid models on experiment II have obvious advantages in four performance metrics over the single benchmark models on experiment I. Thus, it can be reasonably concluded that data preprocessing strategy has obvious influence on the forecasting performance.

3.4.3 Summary of experiment I-II

Based on the results of experiment I-II, the conclusions can be drawn as follows:

- (1) The combined EEMD-LSTM-LSSVM model based on EEMD, LSSVM, and LSTM network not only is capable to more effectively forecast wind speed than other

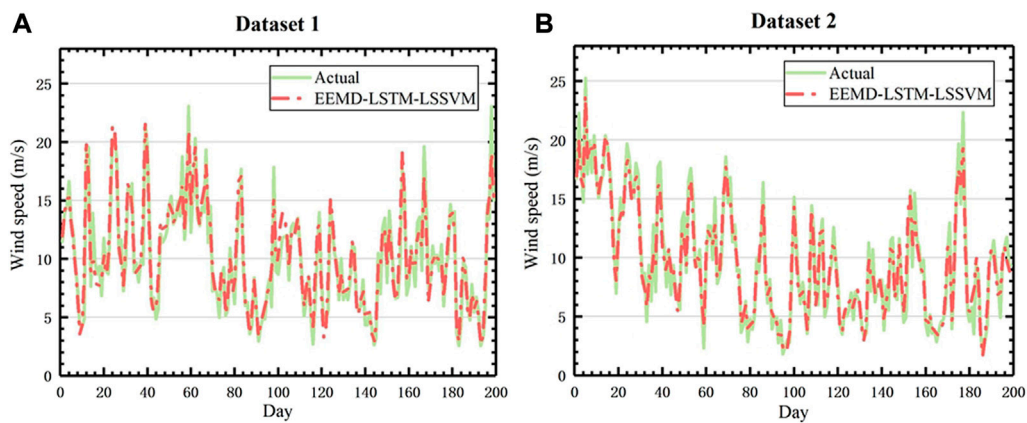


FIGURE 6 Predicted results obtained by EEMD-LSTM-LSSVM (A) Dataset 1. (B) Dataset 2.

TABLE 6 Performance metrics of five models on experiment II.

	Dataset 1				Dataset 2			
	MAE	RMSE	MAPE	R ²	MAE	RMSE	MAPE	R ²
EEMD-BPNN	1.4805	1.8815	0.1746	0.8239	1.2553	1.5806	0.1819	0.9000
EEMD-SVR	1.6078	1.9726	0.1967	0.8064	0.9053	1.1240	0.1243	0.9495
EEMD-LSSVM	1.0273	1.3856	0.1194	0.9045	0.8837	1.0980	0.1158	0.9518
EEMD-LSTM	1.0084	1.2887	0.1148	0.9174	0.8824	1.1058	0.1148	0.9511
Proposed	0.9775	1.2712	0.1127	0.9196	0.8720	1.0921	0.1132	0.9523

The bold values mean the best results among all the comparative models.

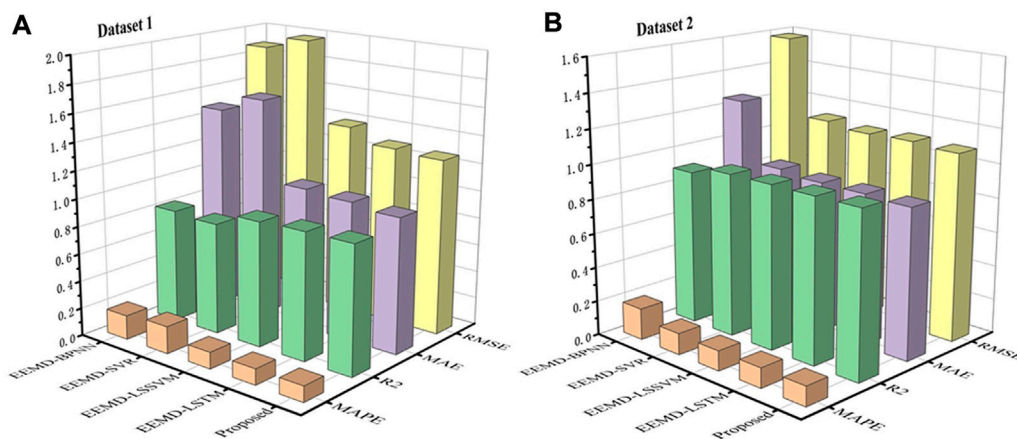


FIGURE 7 Result comparison of five models on experiment II (A) Dataset 1. (B) Dataset 2.

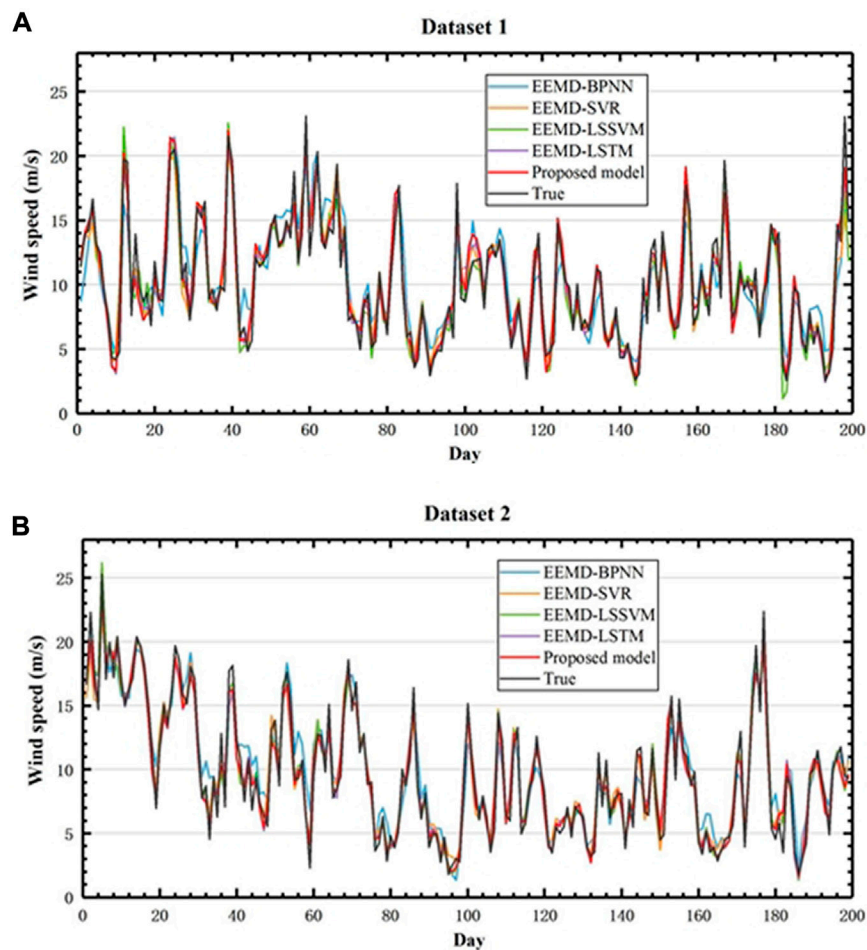


FIGURE 8
The line chart of forecasting results on experiment II (A) Dataset 1. (B) Dataset 2.

comparative models, but also largely improves the accuracy of forecasting results.

- (2) The forecasting accuracy of processing-based models is obviously higher than that of other single models, implying that data preprocessing algorithms included in the combined model can greatly contribute to better forecasting performance.
- (3) By integrating the high-frequency ability of LSTM and low-frequency ability of LSSVM model, the combined LSTM-LSSVM predictor can remarkably improve the forecasting accuracy, which further verifies the effectiveness of the developed combined mechanism in improving the prediction accuracy and ability.
- (4) For both two datasets, the values of four performance metrics obtained by single benchmark models are all worse than those of four metrics obtained by hybrid models, suggesting that the hybrid model has great advantages over the single one.

4 Conclusion

Accurate wind speed prediction plays a vital role in the security and stability of wind power grid-connected system. This paper proposed a novel combined prediction model based on EEMD, LSSVM, and LSTM neural network. To make data sequence more stable and certain, EEMD was first employed to decompose original signal data into a series of sub-layers. At the same time, in order to reduce workload and accumulated error, sample entropy was used to reconstruct the IMF components according to the SE values of sub-layers to obtain new subsequences with different frequencies. Since the subsequences of different frequencies contain different characteristic information, two prediction models, namely LSSVM and LSTM neural networks, were selected to respectively train and forecast high-frequency and low-frequency components. The ultimate predicted values were obtained by superimposing all the predicted values of two models. To investigate the impacts of signal processing method and combined mechanism on the forecasting

accuracy respectively, two group different models, namely single benchmark models and hybrid benchmark models, were compared using two wind speed datasets.

By analyzing experiment results, conclusions can be drawn as follows: 1) Whether dataset 1 or dataset 2 with volatility and uncertainty, experiments I-II indicated that the proposed hybrid EEMD-LSTM-LSSVM model can obtain a satisfactory forecasting performance in terms of four performance metrics. 2) Compared with single models, the superior ability of decomposition-based models has been evidenced, that is, the hybrid model based on signal processing method has significant influence on improving the forecasting performance. 3) The proposed model in this paper respectively trains and predicts high-frequency and low-frequency sequences mainly based on two different prediction models, and the combined mechanism of different predictors adopts different learning methods and training rules for different feature sequences to improve the prediction accuracy.

Despite the above mentioned interesting conclusion, the practical inspirations of this paper are worth mentioning. To the best of our knowledge, wind power generation has served for smart grid, smart microgrids, smart buildings and smart homes, and has a significant impact on electricity supply. However, the difficulty in wind power generation is attributed to unsatisfied wind speed forecasting results restricted by the nonlinearity and non-stationarity of wind speed data. In other words, wind speed prediction has become an important component of wind power generation and simultaneously affects the reliability and stability of wind farm and wind power grid connection system. As a result, the hybrid model proposed in this paper not only improves the accuracy and stability of wind speed prediction, but also has significant reference value for the scheduling, management and optimization of wind farms. For future research, there are two directions remaining to be explored. On the one hand, meteorological factors related to wind power generation, such as temperature, humidity and wind direction, have influence on wind speed, making it interesting to incorporate these meteorological factors instead of single wind speed time series. On the other hand, it is a promising subject to design advanced intelligence optimization algorithms to search for the key

parameters of deep learning network so as to further improve forecasting accuracy.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/datasets>.

Author contributions

XA: writing—review and editing, supervision, and funding acquisition. SL: methodology, software, and writing—original draft. HX: project administration, validation, and funding acquisition.

Funding

This work was jointly supported by National Natural Science Foundation of China (grant numbers 71901167 and 71701212) and the Project of Science and Technology Research Program of Hubei Provincial department of Education (grant number Q20191102).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Aslam, M., and Albassam, M. (2022). Forecasting of wind speed using an interval-based least square method. *Front. Energy Res.* 469, 896217. doi:10.3389/fenrg.2022.896217
- Chen, G., Tang, B., Zeng, X., Zhou, P., Kang, P., and Long, H. (2022). Short-term wind speed forecasting based on long short-term memory and improved BP neural network. *Int. J. Electr. Power & Energy Syst.* 134, 107365. doi:10.1016/j.ijepes.2021.107365
- Cui, L. G., Tao, Y. Q., Deng, J., Liu, X. L., Xu, D. Y., and Tang, G. F. (2021). Bbo-bpnn and ampso-bpnn for multiple-criteria inventory classification. *Expert Syst. Appl.* 175, 114842. doi:10.1016/j.eswa.2021.114842
- Du, P., Wang, J., Yang, W., and Niu, T. (2018). Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. *Renew. Energy* 122, 533–550. doi:10.1016/j.renene.2018.01.113
- Fu, W. L., Wang, K., Tan, J. W., and Zhang, K. (2020). A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting. *Energy Convers. Manag.* 205, 112461. doi:10.1016/j.enconman.2019.112461
- Hao, Y., and Tian, C. (2019). A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Appl. Energy* 238, 368–383. doi:10.1016/j.apenergy.2019.01.063
- Hu, Y. L., and Chen, L. (2018). A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. *Energy Convers. Manag.* 173, 123–142. doi:10.1016/j.enconman.2018.07.070

- Huang, H., Hong, Y., and Wang, H. (2022). Probabilistic prediction intervals of wind speed based on explainable neural network. *Front. Energy Res.* 970. doi:10.3389/fenrg.2022.934935
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. A* 454, 903–995. doi:10.1098/rspa.1998.0193
- Huang, N. E., and Wu, Z. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* 01, 1–41. doi:10.1142/S1793536909000047
- Li, Y., Tang, F., Gao, X., Zhang, T., Qi, J., Xie, J., et al. (2022). Numerical weather prediction correction strategy for short-term wind power forecasting based on bidirectional gated recurrent unit and XGBoost. *Front. Energy Res.* 1014, 836144. doi:10.3389/fenrg.2021.836144
- Liu, H., Duan, Z., Li, Y., and Lu, H. (2018a). A novel ensemble model of different mother wavelets for wind speed multi-step forecasting. *Appl. Energy* 228, 1783–1800. doi:10.1016/j.apenergy.2018.07.050
- Liu, H., Mi, X. W., and Li, Y. F. (2018b). 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. Manag.* 156, 498–514. doi:10.1016/j.enconman.2017.11.053
- Liu, Y., Sun, Y., Infield, D., Zhao, Y., Han, S., and Yan, J. (2016). A hybrid forecasting method for wind power ramp based on orthogonal test and support vector machine (OT-SVM). *IEEE Trans. Sustain. Energy* 8 (2), 451–457. doi:10.1109/TSTE.2016.2604852
- Moreno, S. R., da Silva, R. G., Mariani, V. C., and dos Santos Coelho, L. (2020). Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. *Energy Convers. Manag.* 213, 112869. doi:10.1016/j.enconman.2020.112869
- Peng, L., Zhu, Q., Lv, S. X., and Wang, L. (2020). Effective long short-term memory with fruit fly optimization algorithm for time series forecasting. *Soft Comput.* 24 (19), 15059–15079. doi:10.1007/s00500-020-04855-2
- Poggi, P., Muselli, M., Notton, G., Cristofari, C., and Louche, A. (2003). Forecasting and simulating wind speed in Corsica by using an autoregressive model. *Energy Convers. Manag.* 44 (20), 3177–3196. doi:10.1016/S0196-8904(03)00108-0
- Shahid, F., Zameer, A., and Muneeb, M. (2021). A novel genetic LSTM model for wind power forecast. *Energy* 223, 120069. doi:10.1016/j.energy.2021.120069
- Sun, W., and Wang, Y. (2018). Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Convers. Manag.* 157, 1–12. doi:10.1016/j.enconman.2017.11.067
- Sun, Z., Shen, Y., Chen, Z., Teng, Y., and Qian, X. (2022). Interval prediction method for wind speed based on ARQEA optimized by beta distribution and SWLSTM. *Front. Energy Res.* 10, 927260. doi:10.3389/fenrg.2022.927260
- Wang, H. Z., Wang, G. B., Li, G. Q., Peng, J. C., and Liu, Y. T. (2016a). Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl. Energy* 182, 80–93. doi:10.1016/j.apenergy.2016.08.108
- Wang, S., Zhang, N., Wu, L., and Wang, Y. (2016b). Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* 94, 629–636. doi:10.1016/j.renene.2016.03.103
- Wang, Y., Zou, R., Liu, F., Zhang, L., and Liu, Q. (2021). A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* 304, 117766. doi:10.1016/j.apenergy.2021.117766
- Xiao, L., Dong, Y., and Dong, Y. (2018). An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. *Energy Convers. Manag.* 160, 273–288. doi:10.1016/j.enconman.2018.01.038
- Yu, Y., Han, X., Yang, M., and Yang, J. (2020). Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. *IEEE Trans. Ind. Appl.* 56 (6), 6117–6127. doi:10.1109/TIA.2020.2992945
- Yuan, X., Chen, C., Yuan, Y., Huang, Y., and Tan, Q. (2015). Short-term wind power prediction based on LSSVM-GSA model. *Energy Convers. Manag.* 101, 393–401. doi:10.1016/j.enconman.2015.05.065
- Yunus, K., Thiringer, T., and Chen, P. (2015). ARIMA-based frequency-decomposed modeling of wind speed time series. *IEEE Trans. Power Syst.* 31 (4), 2546–2556. doi:10.1109/TPWRS.2015.2468586