REVIEW

WILEY

A comprehensive review on deep learning approaches in wind forecasting applications

Zhou Wu[1] | Gan Luo[1] | Zhile Yang[2] | Yuanjun Guo[2] | Kang Li[3] | Yusheng Xue[4]

[1]College of Automation, Chongqing University, Chongqing, China

[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China

[3]School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK

[4]State Grid Electric Power Research Institute, Nanjing, Jiangsu, China

Correspondence

Zhile Yang, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong 518055, China.
Email: zl.yang@siat.ac.cn

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 52077213, 62003332; State Key Laboratory of Power Transmission Equipment and System Security and New Technology

Abstract

The effective use of wind energy is an essential part of the sustainable development of human society, in particular, at the recent unprecedented pressure in shaping a low carbon energy environment. Accurate wind resource and power forecasting play a key role in improving the wind penetration. However, it has not been well adopted in the real-world applications due to the strong stochastic characteristics of wind energy. In recent years, the application boost of deep learning methods provides new effective tools in wind forecasting. This paper provides a comprehensive overview of the forecasting models based on deep learning in the field of wind energy. Featured approaches include time-series-based recurrent neural networks, restricted Boltzmann machines, convolutional neural networks as well as auto-encoder-based approaches. In addition, future development directions of deep-learning-based wind energy forecasting have also been discussed.

KEYWORDS

deep learning, deep neural networks, learning (artificial intelligence)

1 | INTRODUCTION

Given the increasing growth of energy demands, it is critical to incorporate renewable energy into the power supply. Demand for renewable energy is expected to increase on account of lower operating costs and preferential use of many power systems [1]. As a kind of renewable energy with the characteristics of green, clean, environmentally friendly and high economic benefits, wind energy is very important for the sustainable development of human society. Due to the superiority of wind energy, it has developed by leaps and bounds in the past 10 years and become one of the most cost-competitive energy sources in the world. In 2020, the global installed capacity of wind energy was 93 GW. China and the United States are the world's largest onshore wind energy markets, which together accounts for more than 60% of new installed capacity in 2020 [2]. Up to 2020, China's cumulative installed capacity of wind energy exceeded 216 million kilowatts, accounting for around 40% of the world total amount. China has become one of the leaders in the development of global wind power [3]. The cumulative installed capacity of new energy power generation in the State Grid Operating Area is 350 million kilowatts, of which the installed capacity of wind power generation is 169 million kilowatts with a yearly increase of 16%. New energy's annual power generation is 510.2 billion kWh, accounting for 9.2% of the total power generation, of which wind power generation is 315.2 billion kilowatt-hour (kWh) with a yearly increase of 11% [4]. American wind power reached an important milestone in 2019, reaching an operating capacity of 100 GW. Since 2008, the wind power generation

capacity has quadrupled and it has become the largest source of the renewable energy generation capacity in the United States, which will account for 7.2% of US electricity by 2019 [5]. The top countries in the global wind power installed capacity, such as Germany, India, Italy, Spain, the United Kingdom, France, Brazil, and Canada, are also vigorously developing wind energy [6].

Despite the advantages of wind energy, the smooth integration of large-scale wind power into the grid still faces many challenges. Due to the randomness, volatility and intermittency of wind, large-scale wind power grid connection makes it very difficult to balance power supply and demand, and also brought a universal curtailment of wind power. One possible solution to balance the challenge is to increase wind speed and power prediction. Improving wind forecast accuracy can help to optimize the overall planning and scheduling of the power grid, find the optimal combination of wind turbines, and ensure the safe and stable operation of the power system, thereby further increase the economic benefits of wind. Meanwhile, accurate wind forecasting is also one of the key prerequisites for providing wind power absorption capabilities.

Wind energy forecasting has been an intractable problem in the energy system, where numerous reviews have been proposed broadly covering data processing, power and resource forecasting. Jung et al. [7] reviewed the potential technologies that can improve the performance of wind energy forecasting models, and emphasized the promising knowledge system in the forecasting. Tascikaraoglu et al. [8] outlined the combined wind energy forecasting methods and focussed on the various model combinations. Wang et al. [9] summarized eight multi-step wind speed forecasting strategies, where 48 hybrid models were compared based on these eight strategies. Bokde et al. [10] compared the existing method with empirical mode decomposition (EMD) and its improved versions in terms of pre-processing technologies. Liu et al. [11] provided a detailed review and classification of data processing techniques in wind energy forecasting, and an in-depth study of each mentioned data processing method including purpose, function, details and performance was also provided. Liu et al. [12] reviewed eight kinds of intelligent predictors for shallow and deep learning in the wind energy prediction field and auxiliary methods that can improve the predictive ability of the predictive model that include integrated learning and optimization algorithms. Vargas et al. [13] demonstrated a new literature review method called systematic literature network analysis, which was used to summarize the development of wind energy analysis in the decision-making process in the past 30 years. The authors pointed out that the most commonly used methods this year are Monte Carlo simulation and artificial neural network methods. Wang et al. [14] reviewed applications of artificial intelligent algorithms in wind energy forecasting. Gonzalez et al. [15] summarized the commonly used performance indicators for deterministic and probabilistic short-term wind power forecasting and explained the performance of these indicators on different data sets, time resolutions and certain specific model attributes. Yang et al. [16] provided a comprehensive summary and comparison of more than one hundred wind forecasting methods from three perspectives: wind speed and power prediction, uncertainty prediction, and slope time prediction.

Though numerous reviews have been proposed in terms of wind forecasting, the emerging artificial intelligence technology, in particular, deep learning methods, has boosted in recent years and provides a number of new technologies in wind forecasting. However, the previous reviews mainly focussed on classification issues but did not discuss the development trends in detail. This paper attempts to summarize the methods of wind forecasting based on deep learning in the past 5 years, providing a comprehensive survey for researchers in developing new effective wind forecasting tools.

The remainder of the paper is organized as follows: Section 2 described some basic concepts in the wind energy forecasting field. Section 3 presents wind forecasting models based on deep learning. Section 4 discusses the possible future research direction of wind energy forecasting. Section 5 concludes this paper. Further, the prediction framework based on deep learning is shown in Figure 1, which summarized the categories of each technique.

2 | OVERVIEW OF WIND ENERGY FORECASTING

The wind is the movement of the atmosphere and a featured form of solar energy. When there is an atmospheric pressure difference, the air moves from the higher pressure area to the lower pressure area. It is caused by three concurrent events: the uneven heating of the Earth's atmosphere by the sun, irregularities found on the Earth's surface, and the rotation of the Earth. The wind flows across the wind turbine blades, and the blades with a special structure produce an air pressure difference that produces lift and drag. When the lift is stronger than the drag, the rotor shafts rotates to drive the generator to generate electricity [17, 18]. Wind power P can be calculated as follows:

$$P = \frac{1}{2}\rho A v^3 \qquad (1)$$

where P represents the wind power, ρ denotes the density of air, A is the swept area of the wind turbine, and v is the wind speed.

Wind power exhibits a highly non-linear cubic dependence on wind speed, and accurate wind speed prediction can provide higher power [19]. Besides, studies have shown that if the accuracy of wind speed forecasting is increased by 10%, wind power generation will increase by about 30% than expected [20].

2.1 | Wind time-series forecasting classifications and applications

Up to date, there is no uniform and strict standard for the forecasting term limits. They are separated strongly according
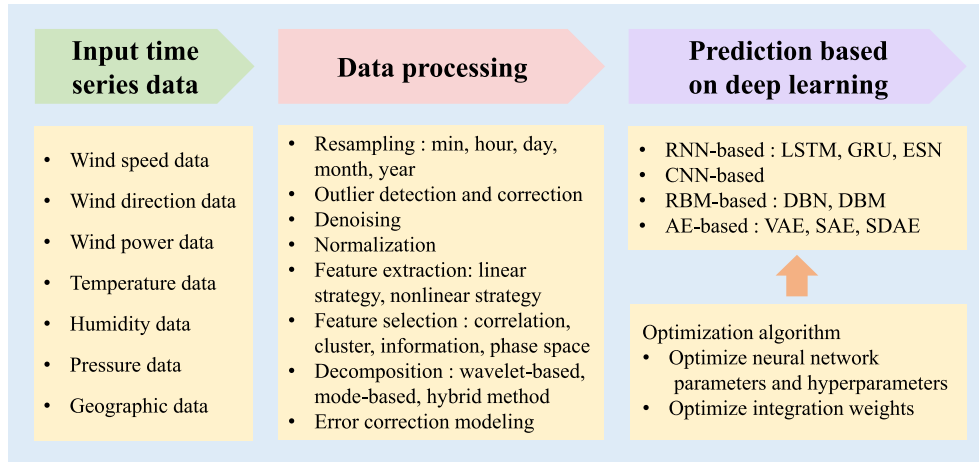
FIGURE 1 Wind energy prediction framework based on deep learning. AE, auto-encoder; CNN, convolutional neural network; DBM, deep Boltzmann machines; DBN, deep belief network; ESN, echo state network; GRU, gated recurrent unit; LSTM, long short-term memory; RBM, restricted Boltzmann machine; RNN, recurrent neural network; SAE, stacked auto-encoder; SDAE, stacked denoising auto-encoders

to the applications. Soman et al. [21] divided the forecast period into four categories: very short-term, short-term, medium-term and long-term, as shown in Figure 2.

The forecasting period is equal to the time resolution multiplied by the predicted steps, usually referring to the period of the test set rather than the training set, which is calculated as follows:

$$T_p = t_i * s_t \qquad (2)$$

where T_p is the forecasting period, t_i is the time unit of the data, s_t is the time step.

The corresponding applications are as follows:

(1) Very short-term: electricity market clearing, electricity regulations, real-time grid operations, wind turbine control, power quality research, load following and distribution
(2) Short-term: economic load dispatch planning, load increment/decrement decisions, load sharing, and operational security in the electricity market
(3) Medium-term: energy allocation, economic dispatch, reserve requirement decisions, generator online/offline decisions, coordination of wind farm and storage device, planned maintenance on network lines, transmission network planning, congestion management, day-ahead energy and reserve scheduling, wind farm maintenance and troubleshooting
(4) Long-term: wind energy resource assessment, wind farm construction planning, optimal operating cost, annual maintenance plan, operation and maintenance of conventional generation, operation management, feasibility study for wind farm, design of wind farm operation plan, energy trading strategy, and coordinate optimal unit portfolio [10, 21–26]
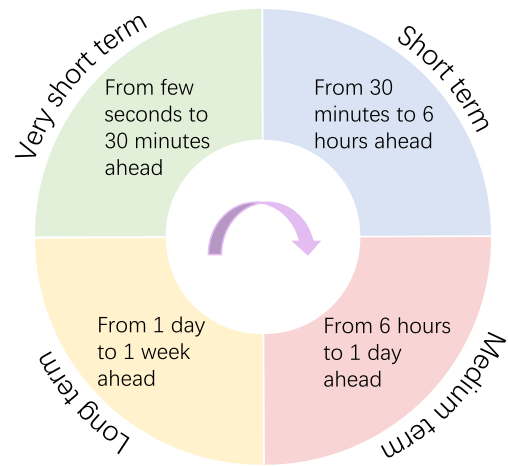

FIGURE 2 Forecasting period classifications

2.2 | Wind energy forecasting goals and results

In order that more effective energy planning and decision-making, wind energy forecasting is indispensable. From the perspective of the forecasting process, there are two types of forecasting, namely direct forecasting and indirect forecasting, respectively. Direct forecasting refers to direct forecasting through historical wind speed or wind power data. Indirect prediction first predicts the future wind speed and then converts the predicted wind speed into wind power forecast according to the power curve of the wind turbine [10]. Indirect methods are more accurate and, therefore, more popular.

According to wind forecasting results, wind forecasting models can also be divided into two categories, deterministic forecasting and probabilistic forecasting [27, 28]. Deterministic forecasting is also called point forecasting and the forecasting result is a deterministic value. The result of probabilistic

forecasting is usually an interval, and the probability distribution of interval values can be given. A single deterministic method cannot reflect the uncertainty and randomness of wind speed. Many applications in the field of wind energy need to consider the uncertainty and randomness so that probabilistic forecastings have attracted an increasing attention in recent years [29].

2.3 | Wind energy forecasting models

From the most basic types, wind forecasting methods can be divided into five categories: persistence method, physical method, conventional statistical method, machine learning method with shallow structure, machine learning method with deep structure, that is, deep learning [25].

The persistence method is fairly straightforward. It is assumed that wind speed or power at a certain future time will be the same as it is when the forecast is made [24]. The expression of this method is as follows:

$$P(t + k) = P(t) \qquad (3)$$

This model performs well in very short-term forecasting, but as the time scale increases, its accuracy gradually decreases. Hence, it is usually used as a benchmark model to compare with new models [30].

The physical method usually refers to the numerical weather prediction (NWP) model. The NWP model establishes a complex physical and mathematical model to simulate the changing process of wind by comprehensively considering meteorological and geographic factors such as temperature, humidity, air pressure, and terrain [31]. NWP models are usually used for weather forecasts in larger areas, and wind speed predictions are only part of it. There are two types of NWP models: global and regional models.

An overview of NWP global models and NWP regional models, and all the commercial and operational wind power forecasting systems and their main features are provided in Ref. [26]. The physical method can reflect the essence of atmospheric motion so that the accuracy is higher. However, this method needs to process an extremely large amount of data and carry out complex calculations. There are extremely high requirements for computing power, which leads to significant hinders for ordinary researchers [32]. Meanwhile, due to the chaotic nature of the partial differential equations in the mathematical model, it is impossible to obtain an accurate solution, and the error will be multiplied with the increase of time. In light of this, NWP models are not suitable for short forecast times but more suitable for medium-term or long-term forecasting [8, 21]. In recent research works, the forecasting periods are generally focussed on very short-term or short-term predictions [13], so the applications of NWP are less.

The conventional statistics method uses the collected wind speed time-series data to deliver predictions. After many years of development, there have been many statistical models for wind speed forecasting. Poggi et al. [33] started to utilize auto-regressive (AR) to simulate wind speed time series, and Nielsen et al. [34] used quantile regression (QR) to make predictions independently. In order to improve the forecasting performance, many auto-regressive moving average models have been developed [35–38]. In addition, numerous AR-based models have also been developed for wind speed prediction, such as vector auto-regressive [39], auto-regressive with exogenous input (ARX) [40], auto-regressive conditional heteroskedasticity [41, 42], auto-regressive integrated moving average (ARIMA) [43–46], seasonal ARIMA [47], fractional ARIMA [48], and ARFIMA [49]. In order to improve the accuracy of prediction and the robustness of the model, researchers have also developed many hybrid models based on the ARIMA model, such as WT-ARIMA [50], RWT-ARIMA [43], and VMD-ARIMA [51]. However, these models only analyse the superficial relationship between the variables in the time series, and it is difficult to deal with the complicated and non-linear relationship.

For obtaining more satisfactory prediction results, numerous non-linear statistic models have been proposed [52]. Zhang et al. [53] combined AR and Gaussian process regression (GPR) to improve prediction accuracy. In Karakucs et al. [54], polynomial auto-regressive is proposed, which is a non-linear model with linear parameters. Due to the non-linear term of the Hammerstein model, the Hammerstein auto-regressive model is superior to the ARIMAs [55]. Some enhanced models such as smooth transition auto-regressive, self-exciting threshold auto-regressive [56] and Markov switching auto-regressive [57] have also been proposed. Furthermore, the researchers also used some unusual models, for example, non-linear auto-regressive with exogenous input [58], generalized auto-regressive conditional heteroskedasticity (GARCH) [59], multiple-kernel relevance vector regression [60], threshold seasonal auto-regressive conditional heteroscedasticity [61], Bayesian-based adaptive robust multi-kernel regression [62]. However, with the increasing complexity of time-series data, it is not easy to meet the requirements of prediction accuracy because traditional statistic models have little ability to extract the features of data.

The shallow machine learning methods include neural networks with a couple of layers. Marugan et al. [63] summarize most of the shallow neural network models. Compared with the persistence method and the traditional statistical method, the shallow machine learning method has higher prediction accuracy and better effect in practice. Nevertheless, these models can only learn the shallow features in the wind time-series data and need extensive feature engineering [64].

Deep learning is a machine learning method for deep network architecture. The characteristics of input data are learnt through a computational model composed of multiple non-linear processing layers. Compared with shallow machine learning models and traditional statistical models, deep learning

TABLE 1 Summary of models with LSTM predictor

Ref.	Input data	Time interval	Model	Prediction performance
Yang et al. [68]	Wind speed	1-h	Peephole LSTM	MAPE 0.0421
Yu et al. [69]	Wind power	10-min	LSTM-EFG	MSE 5.531
Zhang et al. [70]	Wind speed	15-min	SWLSTM-GPR	RMSE 0.174
Araya et al. [72]	Wind speed	1-h	LSTM-Ms	MAE 1.081
Pei et al. [73]	Wind speed	15-min	EWT-NCULSTM	MAE 0.10 MAPE 5.45 RMSE 0.15
Zhang et al. [74]	Wind speed	1-h	NLSTM	RMSE 0.45 MAPE 8.16
Qu et al. [76]	Wind power	5-min	PCA-LSTM	NMAE 0.5672 NRMSE 1.0732
Xu et al. [77]	Wind power	10-min	Adaptive LSTM	MAPE 2.7
Huang et al. [78]	Wind speed	1-h	EEMD-LSTM-GPR-vc	MAPE 14.28 RMSE 0.83 MAE 0.71
Liu et al. [79]	Wind speed	1-h	EWT-LSTM-Elman	MAPE 3.24 MAE 0.28 RMSE 0.37
Li et al. [80]	Wind speed	10-min	mm-LSTM	MAE 1.8682 RMSE2.8484
Liu et al. [81]	Wind power	15-min	DWT-LSTM	MAE 5.49 MAPE 1.75 RMSE 8.64
Liu et al. [82]	Wind speed	10-min	SDAE-LSTM	MAPE 7.1323 MAE 0.3066 RMSE 0.3880
Wu et al. [83]	Wind speed	10-min	DBSCAN-SDAE-LSTM	NMAE 0.0798
Li et al. [85]	Wind speed	10-min	EWT-LSTM-RELM -IEWT	MAPE 0.3202 MAE 0.0359 RMSE 0.0465
Jaseena et al. [86]	Wind speed	10-min	SAE-LSTM	MAE 0.3982 RMSE 0.5969
Moreno et al. [87]	Wind speed	10-min	VMD-SSA-LSTM	MSE 0.021 RMSE 0.145 MAPE 1.75
Ma et al. [88]	Wind speed	15-min	CEEMDAN-error-VMD-LSTM	RMSE 0.111 MAPE 3.21 MAE 0.0831
Liang et al. [89]	Wind speed	5-min	MSLSTM	MSE 0.00297 RMSE 0.05448 MAE 0.04275
Liu et al. [90]	Wind speed	1-h	VMD-SSA-LSTM-ELM	MAPE 3.04 MAE 0.42 RMSE 0.53
Wang et al. [91]	Wind speed	1-h	VMD-OFE-LSTMN-ECS	MAE 0.37 RMSE 0.73 MAPE 27.15
Chen et al. [92]	Wind speed	15-min	SSA-CEEMDAN-ConvLSTM	MSE 0.006 MAE 0.052 MAPE 1.135
Lu et al. [93]	Wind power	15-min	E-D LSTM	RMSE 2.6
Yin et al. [94]	Wind power	1-h	EMD-VMD-CNN-LSTM	MAE 0.222 RMSE 0.296
Wang et al. [95]	Wind power	15-min	LW-CLSTM	MSE 12.7 RMSE 3.5 MAE 1.36
Chen et al. [96]	Wind speed	5-min	MFSTC-CNN-LSTM	MAE 1.0384 RMSE 1.4003
Zhu et al. [97]	Wind speed	10-min	PSTN	RMSE-A 0.531 MAPE-A 3.674
Xiang et al. [98]	Wind speed	15-min	WT-bi-LSTM	MAE 0.074 MAPE 0.551 RMSE 0.157

Abbreviations: CEEMDAN, complete ensemble empirical mode decomposition with adaptive noise; CNN, convolutional neural network; DBSCAN, density-based spatial clustering of applications with noise; EEMD, ensemble empirical mode decomposition; EMD, empirical mode decomposition; GPR, Gaussian process regression; ELM, extreme learning machine; IEWT, inverse empirical wavelet transform; LSTM, long short-term memory; SAE, stacked auto-encoder; SDAE, stacked denoising auto-encoders; SSA, singular spectrum analysis; VMD, variational mode decomposition.

methods can extract more abstract and hidden features in data, so as to obtain better accuracy in prediction tasks. The effectiveness and accuracy of prediction models based on deep learning have been widely recognized.

3 | DEEP-LEARNING-BASED WIND FORECASTING

There are usually three steps in wind speed prediction: wind energy data processing, predictor prediction and model performance evaluation. The deep neural network (DNN) is generally used as a feature extractor and a predictor. At present, many DNNs have been applied to wind forecasting. The basic prediction structures based on deep learning mainly include recurrent neural network (RNN), convolutional neural network (CNN), restricted Boltzmann machine (RBM) and so on. Additionally, there are some other deep networks such as generative adversarial network, extreme learning machine (ELM), stacked auto-encoder (SAE), stacked denoising auto-encoders (SDAE) etc. Table 1 provides the summary of models with long short-term memory (LSTM) predictor, and Table 2 shows the summary of other forecasting models.

T A B L E 2 Summary of other forecasting models

Ref.	Input data	Time interval	Model	Prediction performance
Peng et al. [104]	Wind speed	1-h	WSTDGRU	RMSE 0.3757
Liu et al. [105]	Wind speed	15-min	SSA-CNNGRU-SVR	MAPE 0.97 MAE 1.72 MAE 2.41
Liu et al. [106]	Wind speed	1-h	STNN-VB	RMSE 0.226
Niu et al. [107]	Wind power	1-h	AGRU	NRMSE 6.26 MAPE 4.25 MAPE 14.94
Xian et al. [108]	Wind speed	10-min	SD-BiGRU	MAE 0.1456 MAPE 4.78 RMSE 0.1871
Chitsazan et al. [109]	Wind speed	10-min	NESN-MP	MAE 0.3 RMSE 0.43 MSE 0.19
Gouveia et al. [110]	Wind power	1-h	WTESN	NMAE 0.03
Hu et al. [111]	Wind power	1-month	DeepESN	MAE 1.71 RMSE 2.2567 MAPE 5.23
Khan et al. [112]	Wind power	1-h	DBN-wp	MAE 0.0754 RMSE 0.1010 SDE 0.1007
Wang et al. [113]	Wind power	10-min	kmeans-DBN	MAPE 1.1739
Wang et al. [114]	Wind speed	5-min	WT-DBN	MAE 0.4947 RMSE 0.6671 MAPE 4.76
Yu et al. [115]	Wind speed	10-min	DBNLP	MAE 2.3355 RMSE 3.0609
Zhu et al. [116]	Wind speed	5-min	PDCNN	A-MAPE 4.414 A-RMSE 0.400
Mujeeb et al. [117]	Wind power	1-h	EDCNN	MAPE 2.43 NRMSE 0.096 MAE 2.24
Yildiz et al. [118]	Wind power	1-h	RBCNN	RMSE 0.0499 MAE 0.0376 SMAPE 0.2532
Shivam et al. [119]	Wind speed	10-min	ResAUnet	MAE 1.035 RMSE 1.342 MAPE 0.272
Yan et al. [120]	Wind power	10-min	SDAE-mm	NRMSE 0.14
Chen et al. [121]	Wind speed	1-h	SDAE-ELM	RMSE 0.21 MAE 0.28

Abbreviations: DBN, deep belief network; ELM, extreme learning machine; SDAE, stacked denoising auto-encoders.

T A B L E 3 Summary of present reviews

Ref.	Key contribution
Bokde et al. [10]	Hybrid EMD or EEMD models, the evolution of the EMD-based method, novel techniques of treating inherent mode functions, error measures EMD-based research works
Liu et al. [11]	Review and categorization of data processing techniques, a general evaluation with different perspectives, research directions and challenges
Jung et al. [7]	State-of-the-art forecasting approaches, general forecasting accuracy, techniques to improve the forecasting performance
Tascikaraoglu et al. [8]	Combined forecasting approaches, up-to-date annotated bibliography
Wang et al. [9]	Comparison of eight multi-step forecasting strategies, two improvements on prediction
Liu et al. [12]	Classification of some intelligent predictors, summarizes their merits and limitations, two auxiliary methods: Ensemble learning and metaheuristic optimization algorithms
Vargas et al. [13]	The evolution of wind energy analysis over the last 30 years, an innovative literature review approach
Wang et al. [14]	Applications of artificial intelligent algorithms in wind farms
Gonzalez et al. [15]	The recently proposed forecasting model, performance evaluation methods
Yang et al. [16]	Three novel technologies, four classifications of wind data, 37 evaluation criteria, 100 methods, 22 sub-categories of forecasting approaches in three perspectives

Abbreviations: EEMD, ensemble empirical mode decomposition; EMD, empirical mode decomposition.

3.1 | RNN-based models

RNN originated from a feed-forward neural network. Unlike conventional feed-forward neural networks, it adopts a cyclic connection structure that reuses the calculation result of the previous iteration of the loop, gaining a memory function [65]. RNN has a great learning advantage for the non-linear characteristics of sequence data.

3.1.1 | Models with long short-term memory predictor

LSTM network is designed to solve the vanishing gradient problem that occurs when RNN learns sequences with long-term dependence [66]. Compared to the simple structure of RNN, LSTM is far more complicated. Due to its versatility, its principle will not be introduced in detail. It consists of input gate i_t, forget gate f_t, update gate g_t and output gate o_t. Figure 3 illustrates a single-LSTM cell. The calculation formulas of LSTM are as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{4}$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{5}$$

$$g_t = \tanh(W_g[h_{t-1}, x_t] + b_g) \tag{6}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

$$c_t = c_{t-1} \odot f_t + g_t \odot i_t \tag{8}$$

$$h_t = \tanh(c_t) \odot o_t \tag{9}$$

where $W_{i,f,g,o}$ is the weight matrices, $b_{i,f,g,o}$ is the bias vectors, c_t is the memory cell, and σ is the sigmoid activation function.

Wu et al. [67] adopted a CNN to extract features and then used LSTM for short-term prediction. However, there are shortcomings such as long training time and insufficient prediction accuracy. For optimizing the performance of LSTM, researchers have also made many improvements on the basis of it. Extending the LSTM cell through peephole connections solves the problem that when the LSTM closes the output gate, the gate cannot obtain any information from the output of the storage unit, bringing better prediction effects [68]. Yu et al. [69] proposed LSTM-EFG, which enhances the effect of forgetting the door and improves the activation function. The shared weight long short-term memory network model is introduced to reduce the training time and the variables that need to be optimized [70].

For further controlling the over-fitting problem of LSTM, Eze et al. [71] designed an oLSTM model based on the mixed regularization of LSTM and dropout. The proposed model is an energy-based regression method that captures the cooperative adaptation of input variables. This method can effectively control the vanishing gradient problem of mapping input and output wind data. An LSTM-Ms model was designed to use feed-forward neural networks to construct rougher time-scale sequences than the original model and then used LSTM to process these sequences [72]. Through LSTM-Ms, it is easier to learn the long-term dependence of wind speed sequences. Pei et al. [73] proposed an EWT-NCULSTM. Compared with the traditional LSTM, the proposed model combines the input gate and the forget gate as an update gate and improves the update method of the storage unit with reference to the gated recurrent unit (GRU). The empirical wavelet transform (EWT) strategy is employed to decompose wind speed data to achieve the purpose of noise reduction. After that, the new cell update long short-term memory network model is adopted to predict each sub-sequence and lastly sum up to get the final result. Many methods only consider the correlation of meteorological factors but do not consider their causality. Zhang et al. [74] employed a new method, namely long short-term memory network based on neighbourhood gates (NLSTM), which dynamically adjusts the network structure according to the specific equivalent tree causality to handle the complex causality in wind speed prediction, thereby improving the accuracy of prediction.

Excessive stacking of LSTM units may lead to a decrease in training accuracy and efficiency. Lopez et al. [75] found a better starting point for training by evaluating a number of instances and using these output signals to perform a ridge regression to obtain the output layer weights. Generally, the high-frequency wind speed sub-series has short-term dependence, whereas the low-frequency sub-series has short-term and long-term dependence. Liu et al. [12] proposed models with different characteristics to predict sub-sequence with different frequencies are more likely to achieve the satisfying result. To further improve the accuracy of predictions, researchers have developed many hybrid models. The basic idea is to use various signal processing and analysis methods to refine the input data, and then use one or more predictors to make predictions.

Qu et al. [76] employed a principal components analysis (PCA) to extract valid information from NWP and input it into LSTM for prediction. It is proposed that Adaptive LSTM uses the Pearson analysis to extract strong correlation factors and input them into LSTM for prediction [77]. Huang et al. [78] designed an EEMD-GPR-LSTM method, where ensemble empirical mode decomposition (EEMD) is adopted to decompose the original data of the wind speed. Afterwards, the LSTM and GPR methods are used to predict the inherent mode functions, respectively. Finally, determine the weight of the two prediction results by the variance-covariance method and provide combined prediction results.

Liu et al. [79] designed a new hybrid model that mixes two RNNs. The proposed EWT-LSTM-Elman model uses EWT to get multiple sub-signals and uses LSTM to predict low-frequency sub-signals and ElmanNN to predict high-frequency sub-signals. The experimental results are satisfactory. Li et al. [80] adopted MM to process the wind speed sequence into a stationary long-term baseline and a non-
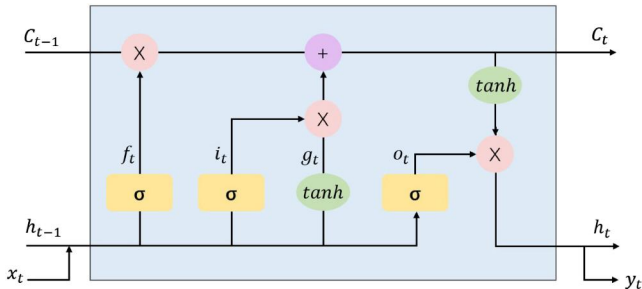

FIGURE 3 The structure of long short-term memory

stationary short-term residue and then use LSTM to make predictions. Liu et al. [81] introduced a DWT-LSTM model for short-term wind power forecasting. The DWT is utilized to handle the non-stationary time series into multiple highly stationary components, then use LSTM to independently predict each component and finally obtain the final prediction result by linearly summing the prediction values of each component. Liu et al. [82] proposed a deep architecture SDAE-LSTM with feature selection. In this model, a feature selection framework based on mutual information was first developed to determine the most suitable input for the prediction model. Then, the authors used SDAE to capture the inherent features contained in the original data and used LSTM to output the results.

Wu et al. [83] proposed a DBSCAN-SDAE-LSTM model, which first selected representative training samples from NWP data by density-based spatial clustering of applications with noise (DBSCAN), used SDAE together with batch normalization for deep feature extraction and finally utilized LSTM for prediction. Liu et al. [84] utilized wavelet packet decomposition (WPD) to process the original data into two levels of high and low frequencies, 1D-CNN is adopted to predict high-frequency sub-sequences, and low-frequency sub-sequences is predicted by CNNLSTM, forming a WPD-LSTMCNN-CNN hybrid architecture. Li et al. [85] developed a combined EWT-LSTM-RELM-IEWT model. Unlike other models, the hybrid model used regularized extreme learning machine to model the error sequence of each sub-signals and adopted an inverse empirical wavelet transform (IEWT) to construct the final prediction sequence and filter outliers. Jaseena et al. [86] proposed an SAE-LSTM model, which made use of SAE to recognize the deep features of the input series and then employed StackedLSTM to make predictions. Moreno et al. [87] proposed a four-step forecasting framework: (1) AM-FM demodulation; (2) VMD-SSA (singular spectrum analysis) decomposition; (3) Ensemble forecasting and reconstruction; (4) Model accuracy verification.

The literature considered preliminary prediction errors and proposed CEEMDAM-error-VMD-LSTM that used a multi-step decomposition prediction strategy. First of all, the original data is processed into sub-sequences and residual sequences by complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) algorithm and then each sequence is predicted using LSTM. The error sequence is obtained by subtracting the original sequence prediction result and the original observation value. Variational mode decomposition (VMD) is employed to process the error signal into a series of sub-sequence and then use LSTM to predict each sub-sequence. Finally, the predicted error sequence is employed to correct the prediction result of the original sequence to obtain a better prediction result [88].

Single time-series data may have an impact on the prediction accuracy. Liang et al. [89] developed MSLSTM that can use multiple historical meteorological variables including wind speed, wind direction, temperature, humidity, pressure, dew point and solar radiation to make predictions. For the sake of improving the robustness of multi-step prediction, Liu et al. [90] advised a hybrid model of VMD-SSA-LSTM-ELM. VMD is applied to process raw data into several sub-signals. SSA is employed to further extract the trend information of all sub-signals; LSTM is adopted to complete the prediction of the low-frequency sub-signals; and finally, ELM is utilized to complete the prediction of the high-frequency sub-signals.

Although many methods decompose the data, it does not eliminate the role of irrelevant information in the input wind data. Therefore, for eliminating the interference of unnecessary components in the input signal and improve the prediction accuracy, Wang et al. [91] developed an EMD-OFE-LSTMN-ECS method. First of all, VMD is applied to process the non-stationary wind speed signal, and Kullback–Leibler divergence and energy measure (EM) are both adopted to capture key features. Then, sample entropy reorganization key features are applied to input LSTMN for prediction. At the same time, an error correction strategy based on GARCH is also employed to correct the prediction error without ignoring its inherent correlation and heteroscedasticity. Chen et al. [92] adopted SSA, CEEMDAN and invert-EMD to reduce noise and decompose the original data and then used the master-slave forecast model composed of ConvLSTM and BPNN to make predictions. The test results showed that the prediction accuracy is competitive. Lu et al. [93] proposed an E-D LSTM model to lower specification risk. This model uses the LSTM-based (encoder-decoder) E-D model to build an automatic encoder to map wind power time series to a fixed-length form. Then, enter multiple LSTMs together with weather forecast information to make predictions. Su et al. [99] took wind frequency components and wind turbine status into consideration and proposes a WPD-EEMD-LSTM model for very short-term wind power prediction. Yin et al. [94] developed EMD-VMD-CNN-LSTM architecture that effectively utilized the relationship between wind speed, wind energy and wind direction. The method adopted EMD-VMD to process the original data to generate sub-sequences with coupling relationship, utilized CNN-LSTM as a cascade prediction model and finally superimposed all sub-sequence prediction values to output the results.

Some studies have not taken multidimensional meteorological characteristics into account. Li et al. [100] used multi-layer perceptron (MLP) to extract meteorological features highly related to actual wind speed in multi-dimensional meteorological historical data divided into dry and rainy seasons, then applied CNN to extract features of historical data and finally input the extracted features into LSTM for prediction. In order to further utilize the information of multi-variate data, Wang et al. [95] developed the LW-CLSTM model. Firstly, wind power data, historical measurement data and turbine status data are fused, cleaned, reduced in dimension and standardized to extract the time-period characteristics of the output power. Then, the proposed method used the time sliding window algorithm to construct a data set and input it into a network composed of CNN and LSTM for prediction.

The general wind energy prediction model does not consider the spatiotemporal correlation between data dimensions. Dou et al. [101] proposed a multidimensional spatiotemporal data input modelling method based on gridded NWP

and then used a network with CNN and LSTM structure for prediction. Although some models consider spatiotemporal correlation, they ignored the influence of meteorological factors with spatiotemporal properties on wind speed. Chen et al. [96] developed a multifactor spatio-temporal correlation-CNN-LSTM combination model. CNN is adopted to learn the spatial feature relationship between meteorological elements at each site and LSTM is employed to learn the temporal feature relationship between historical time points. Zhu et al. [97] introduced a deep architecture termed predictive spatio-temporal network. Firstly, the spatial features of the wind speed matrices are extracted by CNN. Then, the LSTM captures the temporal dependence between the spatial features. Finally, input the two together into LSTM for prediction.

Some researchers have also employed LSTM networks with different structures. Xiang et al. [98] made full use of information from both past and future directions and established an auto-regressive model based on a bidirectional long-term short-term memory neural network model (WT-bi-LSTM) with wavelet decomposition to predict wind speed on multiple time scales.

3.1.2 | Models with GRU predictor

GRU is an updated version of RNN-based methods and shares a similar structure with LSTM methods. It abandons the storage unit mechanism, replaces the forget gate and input gate with an update gate z_t, and replaces the output gates with a reset gate r_t [102]. Many research works have shown that LSTM and GRU have similar experimental effects, whereas GRU is computationally cheaper and efficient [103]. The single GRU cell is illustrated as Figure 4, and the formulations of its nodes are given as follows:

$$z_t = \sigma(W_Z \cdot [s_{t-1}, x_t]) \tag{10}$$

$$r_t = \sigma(W_r \cdot [s_{t-1}, x_t]) \tag{11}$$

$$\hat{s}_t = \tanh(W \cdot [r_t \odot s_{t-1}, x_t]) \tag{12}$$

$$s_t = (1 - z_t) \odot s_{t-1} + z_t \odot \hat{s}_t \tag{13}$$

where W, W_Z, W_r are the weight matrices and σ is the sigmoid activation function.
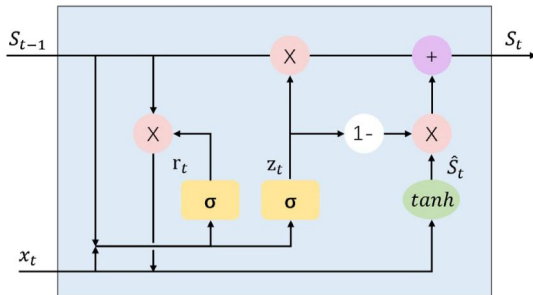

FIGURE 4 The structure of a gated recurrent unit

GRU methods have been combined with a number of data processing and prediction approaches. In Ref. [122], a bivariate EMD-GRU model is proposed and the copula function is utilized to analyse the non-linear relationship between wind energy and meteorological factors and extracted the key factors with the highest correlation with wind energy. The bivariate data composed of the two are input into bivariate EMD, decomposed to obtain sub-sequence data and finally made predictions using GRU.

To broaden the applications of the quasi-EMD method in actual prediction, wavelet soft threshold denoising is applied to wind speed time-series noise reduction, followed by GRU for prediction [104]. Such a combination not only improve the accuracy of the forecast but also reduce the volatility of the results. Although GRU can capture the dependence of the time range and is suitable for time-series data, it does not consider the spatial correlation. ConvGRU was developed to combine the advantages of both to solve the spatiotemporal prediction problem [123]. The proposed SSA-CNNGRU-SVR combined a CNNGRU for trend component prediction, a convolutional layer for capturing deep features, and a GRU layer for obtaining long-term dependencies [105]. Liu et al. [106] introduced a spatio-temporal neural network model that integrated ConvGRU and 3D CNN and used a new coding prediction structure to generate spatio-temporal results.

Advanced attention mechanisms and MIMO strategies are used for feature selection. Niu et al. [107] proposed unrolled architecture of sequence-to-sequence GRU with the attention mechanism (AGRU). The attention mechanism evaluated the importance of each input variable against the target wind energy value and then generated a weighted representation based on their correlation with the target variable. The feature selection method based on a novel attention mechanism identified the most important factors that affect the wind power generation process under different environmental conditions.

Some researchers have taken a different approach and proposed a new method based on NWP. In the first place, it extracted the standard deviation of the numerical weather forecast wind speed error as weights and reorder these weights according to the numerical weather forecast wind speed time series to obtain the weighted time series. Then, an error correction model based on a BiGRU is proposed. The numerical weather forecast wind speed, weighted time-series trends and details are used as inputs to correct the numerical weather forecast wind speed error. By using the corrected numerical weather forecast wind speed, the wind energy curve model is employed to predict short-term wind energy [124]. Similar to LSTM, GRU also has combined methods based on decomposition and prediction ideas. In [108], SSA is used to process the original data into main series and residual series, and the VMD algorithm is adopted to process the residual signal, associated with PSR for reconstructing the decomposed sequence in the high-dimensional phase space and input BiGRU for prediction.

3.1.3 | Models with echo state network predictor

Echo state network (ESN) is another important cluster in the deep learning areas. Unlike other RNNs, ESN use the reservoir as a hidden layer, consisting of input layer x, reservoir u and output layer y [125]. The reservoir maps input data from a relatively low-dimensional input space to a high-dimensional state space. It contains a great many sparsely connected neurons with initial random weights. The initial random weights remain unchanged during the training process. The training process of ESN is to learn the connection weights from the reservoir to the output layer. The update equation of the network is expressed as follows:

$$u(t+1) = f_1\big(W^{in}x(t+1) + W^h u(t)\big) \qquad (14)$$

$$y(t+1) = f_2\big(W^{out} x(t+1)\big) \qquad (15)$$

where f_1, f_2 are the activation functions, W^{in}, W^h, W^{out} denote the input weight matrix, reservoir weight matrix, and output weight matrix, respectively.

Chitsazan et al. [109] developed NESN-P (polynomial) and NESN-MP (multivariable polynomial) to improve the learning ability and the computational efficiency of ESN. The authors designed a reservoir containing a linear internal state and a reading whose output is a non-linear function of the internal state. MP is a cubic multivariate polynomial to reduce the number of internal states relative to the classic ESN.

In the field of time-series forecasting, decomposition and forecasting are more common ideas. WT is applied to eliminate the irregular fluctuation of the sequence and then utilized PCA to reduce the redundant information of the input series. Moreover, SC is utilized to select the appropriate sample set and input ESN for prediction [126]. A WTESN model is proposed, which used the WT-based multi-resolution analysis method to decompose time series into different time scales [110].

Most studies use a single model to make predictions, whereas suffering poor stability and low prediction accuracy. The literature adopted ESN to integrate the intermediate results of four mixed models to output the final prediction results [127]. On the one hand, the higher computational costs may arise from using LSTM, which are overfitting due to the hyperparameters and structures. On the other hand, the simplicity of traditional ESN will lead to poor generalization ability. Lopez et al. [128] combined the two to propose an LSTM + ESN model, using LSTM as the neuron in the hidden layer of ESN. Deeper neural networks may also improve prediction performance. Hu et al. [111] developed a DeepESN with multiple reservoirs for wind power forecasting.

3.2 | RBM-based models

Restricted boltzmann machine (RBM) is a generative stochastic neural network that learns the probability distribution of its inputs. It consists of a visible layer and a hidden layer, and the units in one layer are connected to all units in the other layer. It is worth noting that there is no connection between nodes within the RBM visible layer or hidden layer. The training process is to learn the connection weight between the display layer and the hidden layer. A deep belief network (DBN) and deep boltzmann machines (DBM) can be formed by stacking RBM. The top two layers of DBN are undirected graphs, and the remaining layers form a top-down directed connection. DBM and DBN have the same structure but all connections are undirected [129].

It is a hard problem for general networks to extract advanced features of original sequences. A PDBM is proposed, which adds a predictive layer composed of several inference values on top of the DBM. PDBM forecasts the wind speed by high-level features extracted from low-level features of input series [130]. Tao et al. [131] used DBN for wind power forecasting and have achieved relatively good results. Khan et al. [112] combined ARIMA and DBN to make predictions, used ARIMA to predict the conventional components after VMD decomposition, adopted DBN to forecast the irregular components and finally combined the techniques to generate the final prediction result. Wang et al. [113] introduced the K-means clustering method to select NWP sample data that has an influence on the prediction accuracy and enters the DBN for prediction. A WT-DBN-QR model is proposed, using DBN to extract deep invariant structures and hidden non-linear features in the sequence decomposed by WT [114].

As the training sample size increases, the computational complexity becomes higher. Yu et al. [115] proposed the DBNLP technique, which maps a one-dimensional time series to a high-dimensional space, then selected training samples with the same pattern as the predicted samples based on the Euclid distance and finally extracted DBN for prediction.

3.3 | CNN-based models

CNN is a kind of feed-forward neural network that consists of convolution layer, sampling layer and fully connected layer [132]. CNN is often used for tasks such as hand gesture classification, object detection and time-series prediction [133]. CNN takes the convolution layer as the core. The operation of a convolution kernel effectively learns the complex spatial features and invariant structure in the data. Its calculation expression is as follows:

$$y^k = f\big(\omega^k \otimes x + b^k\big) \qquad (16)$$

where f represents the activation function and w, b denote the weight and deviation of the kth layer.

The CNN method act as the core forecaster in the wind prediction. Wang et al. [134] employed WT to decompose the input data into multiple frequencies and then utilized DeepCNN to predict each frequency. Zhu et al. [116] proposed a PDCNN, where CNN was applied to extract spatial

features and MLP was applied to extract the spatial correlation of spatial features. For improving the prediction performance of CNN, researchers make improvements on the basis of the traditional CNN. Mujeeb et al. [117] proposed efficient deep convolution neural networks with one modified output layer, which is called the enhanced regression output layer. Yildiz et al. [118] designed an improved residual-based deep CNN. The proposed model has more competitive performance than many current state-of-the-art networks.

Simply increasing the depth of CNN does not necessarily make the model have better learning ability. In the literature [119], the residual dilated convolutional network based on U-net with nonlinear attention (ResAUnet) model composed of dilated causal convolutional network as the basic unit is proposed. The U-net architecture was used to copy the low-level features to the corresponding high-level features to recover or enhance the temporal information. It also applied a residual attention block to combine the feature mapping of lower-level residual blocks and higher-level one, and then, the data are fed into the residual block with the same dilation value as the corresponding lower-level residual block.

3.4 | AE-based models

Auto-encoder (AE) is a featured variant of neural networks for unsupervised learning, composed of input layer, hidden layer and output layer. It uses encoders and decoders to map input to output to reconstruct data [135].

In [136], a two-stage prediction model was constructed by AE. In the pre-training stage, the model network is composed of three AEs, and in the fine-tuning stage, another layer is added at the end of the pre-training network. Yan et al. [120] established a multi-to-multi mapping network combined with a stacked denoised auto-encoder (SDAE) for multi-scale wind power forecasting. First, the input NWP data based on SDAE is corrected, and a number of SDAEs with diverse model parameters and input features are integrated into ensemble SDAE for predicting. Chen et al. [121] proposed an SDAE-ELM model for multi-period forecasting. Variance analysis is applied to reduce the impact of time-series fluctuations and then SDAE is employed to process low-level non-linear features and denoise. The ELM-based integrated learner is used to optimize the SDAE fine-tuning process.

To enhance the generalization performance of AE, the sparsity is introduced as a regularization item in the paper retraining process. The deep sparse AEs are used as base-regressors [137] to improve the predictability of wind speed uncertainty and eliminate data noise. Khodayar et al. [138] designed a DNN architecture with stacked auto-encoder SAE and SDAE for forecasting. At the same time, rough neurons are also used to extend AE and DAE to form a robust DNN with rough regression layers. Jahangir et al. [139] designed a multi-modal method, using SDAE to reduce the noise of the input data and a rough neural network with a sinusoidal activation function for prediction.

4 | DISCUSSION AND FUTURE DIRECTION

With the continuous development of wind energy forecasting, many research works show that the prediction accuracy of the hybrid models combining several techniques outperforms the non-hybrid model [7]. The recent research works mostly use the hybrid model. Table 3 shows the performance of the models mentioned in the article. Summarizing the researchers' work, we found that the wind energy hybrid prediction framework consists of four main steps: data pre-processing, predictor prediction, error post-processing, and model performance evaluation. The technologies that may be used in the entire forecasting process can be divided into 10 categories: denoising, outlier detection and correction, resampling, normalization, decomposition, feature engineering, residual error modelling, filter-based correction, predictor and optimization algorithm [7, 11, 140–142]. There has been a detailed review of these technologies by scholars, and we will not repeat them here. Moreover, we provide some research recommendations for challenges and open issues in the wind energy forecasting field as follows.

The prospect of multi-modal learning that integrates multi-source heterogeneous data is broad. Proper fusion of multi-modal data can effectively use the abstract information existing in the data to achieve integrated perception and prediction [143]. The geographic information system (GIS) can provide detailed information about the geographical space of wind farms, and fusion of the data provided by GIS can link the location information of wind turbines for centralized prediction. Interferometric synthetic aperture radar is a radar technology used in surveying, mapping and remote sensing. Images obtained by synthetic aperture radar are used for coherent processing to generate a digital elevation model. We can extract spatial information from the model for wind energy prediction.

Enough emphasis should be put on the frontiers of deep learning. Transformers have great advantages in processing sequence data, modelling long-term dependencies between input sequence elements and supporting parallel processing of sequences [144]. We can use models developed by researchers based on transformers, such as DeepTransformer [145] and Informer [146], for wind energy prediction tasks, which may have incredible results. Neural networks of different topology structures are feasible. The graph neural network can be used to process the dependence relationship between wind energy multivariate data and then make predictions [147]. The data processed by the multi-resolution method can take advantage of the information hidden by the data [148]. The neural network based on the attention mechanism has better performance than the traditional RNN, and it brings very good results on many sequence processing tasks [149, 150]. It is a good choice for learning the internal temporal correlation of the data. Graph convolutional networks can be applied to learn the spatial correlation between neighbouring sites [151]. Deterministic forecasting cannot

reflect the uncertainty in the real world, and probabilistic forecasting is the development direction of wind energy forecasting in the future. Conditional GAN can be utilized to learn the conditional probability distribution of wind energy data sets [152].

Although the wind energy prediction model developed by the researchers can already be used in practice, its interpretability still lacks systematic research. Interpretability means that when solving a problem, we can get enough information that we need and understand [153]. In the field of wind energy prediction, interpretability for deep learning models can give the basis for decision-making for each prediction. As a result, explanatory models are more secure and their predictions are more reliable. Interpretability research is necessary for risk control and management in the wind energy field.

5 | CONCLUSION

Wind energy is a kind of the renewable energy source with the largest installed capacity in the world and the most promising development in the future. The accuracy of wind energy forecasting has a great impact on the stability and security of the grid. Improving the accuracy of wind energy forecasting can bring higher economic and environmental benefits. As an important method in the wind energy forecasting field, deep learning has been developed rapidly in recent years, and many scholars have also reviewed this field. However, the existing review does not pay attention to the development logic of deep learning in the field of wind energy prediction.

This article introduces various wind energy prediction models based on deep learning. Deep learning predictors mainly include CNN, RNN, DBN etc. These hybrid prediction models based on DNNs have their own advantages and disadvantages under different prediction tasks. For example, RNN-based models are better at extracting dependencies within time series, and based on the CNN model is better at extracting the correlation of multiple time series. Some possible future development trends are also provided for researchers' reference. For example, methods in other time-series forecasting fields can be migrated to the wind energy forecasting field, and advanced neural network architectures developed in recent years can also be used. In the literature reviewed, it does not involve the use of optimization algorithms to optimize neural network parameters or hyperparameter models but mainly focusses on the architecture and methods of predictive models.

DATA AVAILABILITY STATEMENT
Yes.

ORCID
Zhile Yang https://orcid.org/0000-0001-8580-534X

REFERENCES
1. IEA: Global Energy Review 2020. International Energy Agency (2020)
2. GWEC: Global Wind Report 2019. Global Wind Energy Council (2020)
3. Zhang, S., et al.: China in global wind power development: role, status and impact. Renew. Sustain. Energy Rev. 127, 109881 (2020)
4. Grid, S.: State Grid Co. Ltd. Service New Energy Development Report 2020 (2020)
5. AWEA: Wind Powers America Annual Report 2019. American Wind Energy Agency (2019)
6. WWEA: Installed capacity by the end of 2019 (MW). World Wind Energy Agency (2019)
7. Jung, J., Broadwater, R.P.: Current status and future advances for wind speed and power forecasting. Renew. Sustain. Energy Rev. 31, 762–777 (2014)
8. Tascikaraoglu, A., Uzunoglu, M.: A review of combined approaches for prediction of short-term wind speed and power. Renew. Sustain. Energy Rev. 34, 243–254 (2014)
9. Wang, J., et al.: Analysis and application of forecasting models in wind power integration: a review of multi-step-ahead wind speed forecasting models. Renew. Sustain. Energy Rev. 60, 960–981 (2016)
10. Bokde, N., et al.: A review on hybrid empirical mode decomposition models for wind speed and wind power prediction. Energies. 12(2), 254 (2019)
11. Liu, H., Chen, C.: Data processing strategies in wind energy forecasting models and applications: a comprehensive review. Appl. Energy. 249, 392–408 (2019)
12. Liu, H., et al.: Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. Energy Convers. Manag. 195, 328–345 (2019)
13. Vargas, S.A., et al.: Wind power generation: a review and a research agenda. J. Clean. Prod. 218, 850–870 (2019)
14. Wang, Y., et al.: A review of applications of artificial intelligent algorithms in wind farms. Artif. Intell. Rev. 53(5), 3447–3500 (2020)
15. González-Sopeña, J., Pakrashi, V., Ghosh, B.: An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renew. Sustain. Energy Rev. 183, 110515 (2020)
16. Yang, B., et al.: State-of-the-art one-stop handbook on wind forecasting technologies: an overview of classifications, methodologies, and analysis. J. Clean. Prod. 283, 124628 (2020)
17. Martinez, C., et al.: Predicting wind turbine blade erosion using machine learning. SMU Data Sci. Rev. 2(2), 17 (2019)
18. Letcher, T.M.: Wind energy engineering: a handbook for onshore and offshore wind turbines. Academic Press (2017)
19. Kalmikov, A.: Wind power fundamentals. In: Jain, P. (ed.) Wind Energy Engineering, pp. 17–24. Elsevier (2017)
20. Ackermann, T., Söder, L.: Wind energy technology and current status: a review. Renew. Sustain. Energy Rev. 4(4), 315–374 (2000)
21. Soman, S.S., et al.: A review of wind power and wind speed forecasting methods with different time horizons. In: North American Power Symposium 2010, pp. 1–8. IEEE (2010)
22. Colak, I., Sagiroglu, S., Yesilbudak, M.: Data mining and wind power prediction: a literature review. Renew. Energy. 46, 241–247 (2012)
23. Wang, Y., et al.: A review of applications of artificial intelligent algorithms in wind farms. Artif. Intell. Rev. 53(5), 3447–3500 (2019)
24. Zhao, X., Wang, S., Li, T.: Review of evaluation criteria and main methods of wind power forecasting. Energy Proc. 12, 761–769 (2011)
25. Chang, W.Y., et al.: A literature review of wind forecasting methods. J. Power Energy Eng. 2(04), 161–168 (2014)
26. Monteiro, C., et al.: Wind power forecasting: state-of-the-art 2009. Technical report, Argonne National Lab.(ANL), Argonne, IL (United States) (2009)

27. Jiang, Y., et al.: A new hybrid framework for probabilistic wind speed prediction using deep feature selection and multi-error modification. Energy Convers. Manag. 199, 111981 (2019)

28. Giebel, G., et al.: The state-of-the-art in short-term prediction of wind power: a literature overview. ANEMOS. plus (2011)

29. 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)

30. Sfetsos, A.: A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renew. Energy. 21(1), 23–35 (2000)

31. Liu, H., Tian, H.q., Li, Y.f.: Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. Appl. Energy. 98, 415–424 (2012)

32. Zhou, Q., Wang, C., Zhang, G.: Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. Appl. Energy. 250, 1559–1580 (2019)

33. Poggi, P., et al.: Forecasting and simulating wind speed in Corsica by using an autoregressive model. Energy Convers. Manag. 44(20), 3177–3196 (2003)

34. Nielsen, H.A., Madsen, H., Nielsen, T.S.: Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. Wind Energy. 9(1-2), 95–108 (2006)

35. Rajagopalan, S., Santoso, S.: Wind power forecasting and error analysis using the autoregressive moving average modeling. In: 2009 IEEE Power & Energy Society General Meeting, pp. 1–6. IEEE (2009)

36. Erdem, E., Shi, J.: ARMA based approaches for forecasting the tuple of wind speed and direction. Appl. Energy. 88(4), 1405–1414 (2011)

37. Cao, Y., et al.: Wind power ultra-short-term forecasting method combined with pattern-matching and ARMA-model. In: 2013 IEEE Grenoble Conference, pp. 1–4. IEEE (2013)

38. Torres, J.L., et al.: Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Sol. Energy. 79(1), 65–77 (2005)

39. Liu, Y., Roberts, M.C., Sioshansi, R.: A vector autoregression weather model for electricity supply and demand modeling. J. Mod. Power Syst. Clean Energy. 6(4), 763–776 (2018)

40. Tastu, J., et al.: Spatio-temporal analysis and modeling of short-term wind power forecast errors. Wind Energy. 14(1), 43–60 (2011)

41. Lv, P., Yue, L.: Short-term wind speed forecasting based on non-stationary time series analysis and ARCH model. In: 2011 International Conference on Multimedia Technology, pp. 2549–2553. IEEE (2011)

42. Wang, M.D., Qiu, Q.R., Cui, B.W.: Short-term wind speed forecasting combined time series method and arch model. In: 2012 International Conference on Machine Learning and Cybernetics, vol. 3, pp. 924–927. IEEE (2012)

43. Singh, S., Mohapatra, A., et al.: Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. Renew. Energy. 136, 758–768 (2019)

44. Shukur, O.B., Lee, M.H.: Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. Renew. Energy. 76, 637–647 (2015)

45. Yunus, K., Thiringer, T., Chen, P.: ARIMA-based frequency-decomposed modeling of wind speed time series. IEEE Trans. Power Syst. 31(4), 2546–2556 (2015)

46. Liu, H., Tian, H.q., Li, Y.f.: An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. J. Wind Eng. Ind. Aerod. 141, 27–38 (2015)

47. Wang, J., et al.: Techniques of applying wavelet de-noising into a combined model for short-term load forecasting. Int. J. Electr. Power Energy Syst. 62, 816–824 (2014)

48. Kavasseri, R.G., Seetharaman, K.: Day-ahead wind speed forecasting using f-ARIMA models. Renew. Energy. 34(5), 1388–1393 (2009)

49. Yuan, X., et al.: Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. Energy. 129, 122–137 (2017)

50. Fang, C., Gao, Y., Ruan, Y.: Improving forecasting accuracy of daily energy consumption of office building using time series analysis based on wavelet transform decomposition. In: IOP Conference Series: Earth and Environmental Science, vol. 294, p. 012031. IOP Publishing (2019)

51. Deng, Y., Wang, B., Lu, Z.: A hybrid model based on data preprocessing strategy and error correction system for wind speed forecasting. Energy Convers. Manag. 212, 112779 (2020)

52. Lydia, M., et al.: Linear and non-linear autoregressive models for short-term wind speed forecasting. Energy Convers. Manag. 112, 115–124 (2016)

53. Zhang, C., et al.: A Gaussian process regression based hybrid approach for short-term wind speed prediction. Energy Convers. Manag. 126, 1084–1092 (2016)

54. Karakuş, O., Kuruoğlu, E.E., Altınkaya, M.A.: One-day ahead wind speed/power prediction based on polynomial autoregressive model. IET Renew. Power Gener. 11(11), 1430–1439 (2017)

55. Maatallah, O.A., et al.: Recursive wind speed forecasting based on Hammerstein Auto-Regressive model. Appl. Energy. 145, 191–197 (2015)

56. Pinson, P., et al.: Regime-switching modelling of the fluctuations of offshore wind generation. J. Wind Eng. Ind. Aerod. 96(12), 2327–2347 (2008)

57. Ailliot, P., Monbet, V.: Markov-switching autoregressive models for wind time series. Environ. Model. Software. 30, 92–101 (2012)

58. Cadenas, E., et al.: Wind speed prediction using a univariate ARIMA model and a multivariate NARX model. Energies. 9(2), 109 (2016)

59. Jiang, Y., et al.: A novel wind speed prediction method: hybrid of correlation-aided DWT, LSSVM and GARCH. J. Wind Eng. Ind. Aerod. 174, 28–38 (2018)

60. Fei, S.w.: A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. Int. J. Electr. Power Energy Syst. 78, 910–915 (2016)

61. Ambach, D., Schmid, W.: A new high-dimensional time series approach for wind speed, wind direction and air pressure forecasting. Energy. 135, 833–850 (2017)

62. Wang, Y., et al.: Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model. Appl. Energy. 208, 1097–1112 (2017)

63. Marugán, A.P., et al.: A survey of artificial neural network in wind energy systems. Appl. Energy. 228, 1822–1836 (2018)

64. Hu, Q., Zhang, R., Zhou, Y.: Transfer learning for short-term wind speed prediction with deep neural networks. Renew. Energy. 85, 83–95 (2016)

65. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)

66. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (1997)

67. Wu, W., et al.: Probabilistic short-term wind power forecasting based on deep neural networks. In: 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–8. IEEE (2016)

68. Yang, T., et al.: A novel method of wind speed prediction by peephole LSTM. In: 2018 International Conference on Power System Technology (POWERCON), pp. 364–369. IEEE (2018)

69. Yu, R., et al.: LSTM-EFG for wind power forecasting based on sequential correlation features. Future Generat. Comput. Syst. 93, 33–42 (2019)

70. Zhang, Z., et al.: Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression. Appl. Energy. 247, 270–284 (2019)

71. Eze, E.C., Chatwin, C.R.: Enhanced recurrent neural network for short-term wind farm power output prediction. IJRDO J. Appl. Sci. 5(2), 28–35 (2019)

72. Araya, I.A., Valle, C., Allende, H.: LSTM-based multi-scale model for wind speed forecasting. In: Iberoamerican Congress on Pattern Recognition, pp. 38–45. Springer (2018)

73. Pei, S., et al.: Wind speed prediction method based on empirical wavelet transform and new cell update long short-term memory network. Energy Convers. Manag. 196, 779–792 (2019)

74. Zhang, Z., et al.: Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction. Energy Convers. Manag. 192, 37–51 (2019)

75. López, E., et al.: Efficient training over long short-term memory networks for wind speed forecasting. In: Iberoamerican Congress on Pattern Recognition, pp. 409–416. Springer (2016)

76. Xiaoyun, Q., et al.: Short-term prediction of wind power based on deep long short-term memory. In: 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1148–1152. IEEE (2016)

77. Xu, G., Xia, L.: Short-term prediction of wind power based on adaptive LSTM. In: 2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2), pp. 1–5. IEEE (2018)

78. Huang, Y., Liu, S., Yang, L.: Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM. Sustainability. 10(10), 3693 (2018)

79. Liu, H., Mi, X.w., Li, Y.f.: 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 (2018)

80. Li, M., et al.: Ultra-short term wind speed prediction using mathematical morphology decomposition and long short-term memory. CSEE J. Power Energy Syst. 6(4), 890–900 (2020)

81. Liu, Y., et al.: Wind power short-term prediction based on LSTM and discrete wavelet transform. Appl. Sci. 9(6), 1108 (2019)

82. Liu, X., et al.: Wind speed forecasting using deep neural network with feature selection. Neurocomputing. 397, 393–403 (2020)

83. Wu, Y.X., Wu, Q.B., Zhu, J.Q.: Data-driven wind speed forecasting using deep feature extraction and LSTM. IET Renew. Power Gener. 13(12), 2062–2069 (2019)

84. 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. Manag. 166, 120–131 (2018)

85. Li, Y., Wu, H., Liu, H.: Multi-step wind speed forecasting using EWT decomposition, LSTM principal computing, RELM subordinate computing and IEWT reconstruction. Energy Convers. Manag. 167, 203–219 (2018)

86. Jaseena, K., Kovoor, B.C.: A hybrid wind speed forecasting model using stacked autoencoder and LSTM. J. Renew. Sustain. Energy. 12(2), 023302 (2020)

87. Moreno, S.R., et al.: Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. Energy Convers. Manag. 213, 112869 (2020)

88. Ma, Z., et al.: Application of hybrid model based on double decomposition, error correction and deep learning in short-term wind speed prediction. Energy Convers. Manag. 205, 112345 (2020)

89. Liang, S., Nguyen, L., Jin, F.: A multi-variable stacked long-short term memory network for wind speed forecasting. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 4561–4564. IEEE (2018)

90. Liu, H., Mi, X., Li, Y.: Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers. Manag. 159, 54–64 (2018)

91. Wang, J., Li, Y.: Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl. Energy. 230, 429–443 (2018)

92. Chen, G., et al.: Short-term wind speed forecasting with principle-subordinate predictor based on conv-LSTM and improved BPNN. IEEE Access. 8, 67955–67973 (2020)

93. Lu, K., et al.: Short-term wind power prediction model based on encoder-decoder LSTM. In: IOP Conference Series: Earth and Environmental Science, vol. 186, p. 012020 (2018)

94. Yin, H., et al.: A cascaded deep learning wind power prediction approach based on a two-layer of mode decomposition. Energy. 189, 116316 (2019)

95. Wang, Y., et al.: A short-term output power prediction model of wind power based on deep learning of grouped time series. Eur. J. Elec. Eng. 22(1), 29–38 (2020)

96. 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. Manag. 185, 783–799 (2019)

97. Zhu, Q., et al.: Learning temporal and spatial correlations jointly: a unified framework for wind speed prediction. IEEE Trans. Sustain. Energy. 11(1), 509–523 (2019)

98. Xiang, J., et al.: Multi-time scale wind speed prediction based on WT-bi-LSTM. In: MATEC Web of Conferences, vol. 309, p. 05011. EDP Sciences (2020)

99. Su, Y., et al.: A LSTM based wind power forecasting method considering wind frequency components and the wind turbine states. In: 2019 22nd International Conference on Electrical Machines and Systems (ICEMS), pp. 1–6. IEEE (2019)

100. Li, G., et al.: The application of deep neural network algorithm considering multi-dimensional meteorological feature extraction in short-term wind speed prediction. In: 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 2432–2437. IEEE (2019)

101. Dou, J., Liu, C., Wang, B.: Short-term wind power forecasting based on spatiotemporal correlations. In: 2018 International Conference on Mechanical, Electrical, Electronic Engineering & Science (MEEES 2018). Atlantis Press (2018)

102. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)

103. Wang, W., et al.: R-NET: machine reading comprehension with self-matching networks. Natural Language Computer Group, Microsoft Reserach. Asia, Beijing', Technical report, China, Technical Report 5, (2017)

104. Peng, Z., et al.: A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. Energy Convers. Manag. 207, 112524 (2020)

105. Liu, H., et al.: Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. Renew. Energy. 143, 842–854 (2019)

106. Liu, Y., et al.: Probabilistic spatiotemporal wind speed forecasting based on a variational Bayesian deep learning model. Appl. Energy. 260, 114259 (2020)

107. Niu, Z., et al.: Wind power forecasting using attention-based gated recurrent unit network. Energy. 196, 117081 (2020)

108. Xiang, L., et al.: Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Convers. Manag. 220, 113098 (2020)

109. Chitsazan, M.A., Fadali, M.S., Trzynadlowski, A.M.: Wind speed and wind direction forecasting using echo state network with nonlinear functions. Renew. Energy. 131, 879–889 (2019)

110. Gouveia, H.T.V., De Aquino, R.R.B., Ferreira, A.A.: Enhancing short-term wind power forecasting through multiresolution analysis and echo state networks. Energies. 11(4), 824 (2018)

111. Hu, H., Wang, L., Lv, S.X.: Forecasting energy consumption and wind power generation using deep echo state network. Renew. Energy. 154, 598–613 (2020)

112. Khan, A., et al.: Deep belief networks based feature generation and regression for predicting wind power. arXiv preprint arXiv:1807.11682 (2018)

113. Wang, K., et al.: Deep belief network based k-means cluster approach for short-term wind power forecasting. Energy. 165, 840–852 (2018)

114. Wang, H., et al.: Deep belief network based deterministic and probabilistic wind speed forecasting approach. Appl. Energy. 182, 80–93 (2016)

115. Yu, Y., et al.: Forecasting a short-term wind speed using a deep belief network combined with a local predictor. IEEJ Trans. Electr. Electron. Eng. 14(2), 238–244 (2019)

116. Zhu, Q., et al.: Wind speed prediction with spatio–temporal correlation: a deep learning approach. Energies. 11(4), 705 (2018)

117. Mujeeb, S., et al.: Wind power forecasting based on efficient deep convolution neural networks. In: International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, pp. 47–56. Springer (2019)

118. Yildiz, C., et al.: An improved residual-based convolutional neural network for very short-term wind power forecasting. Energy Convers. Manag. 228, 113731 (2021)

119. Shivam, K., Tzou, J.C., Wu, S.C.: Multi-step short-term wind speed prediction using a residual dilated causal convolutional network with nonlinear attention. Energies. 13(7), 1772 (2020)

120. Yan, J., et al.: Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping. IEEE Trans. Power Syst. 33(3), 3276–3284 (2018)

121. Chen, L., Li, Z., Zhang, Y.: Multiperiod-ahead wind speed forecasting using deep neural architecture and ensemble learning. Math. Probl. Eng. 1–14 (2019)

122. Liu, H., Lin, W., Li, Y.: Ultra-short-term wind power prediction based on copula function and bivariate EMD decomposition algorithm. Arch. Electr. Eng. 69(2), 271–286 (2020)

123. Liu, Y., et al.: Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. Appl. Energy. 253, 113596 (2019)

124. Ding, M., et al.: A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. Neurocomputing. 365, 54–61 (2019)

125. Jaeger, H., Haas, H.: Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science. 304(5667), 78–80 (2004)

126. Liu, D., Wang, J., Wang, H.: Short-term wind speed forecasting based on spectral clustering and optimised echo state networks. Renew. Energy. 78, 599–608 (2015)

127. Chen, Y., et al.: A novel combined model based on echo state network for multi-step ahead wind speed forecasting: a case study of NREL. Energy Convers. Manag. 179, 13–29 (2019)

128. López, E., et al.: Wind power forecasting based on echo state networks and long short-term memory. Energies. 11(3), 526 (2018)

129. Salakhutdinov, R., Hinton, G.: An efficient learning procedure for deep Boltzmann machines. Neural Comput. 24(8), 1967–2006 (2012)

130. Zhang, C.Y., et al.: Predictive deep Boltzmann machine for multiperiod wind speed forecasting. IEEE Trans. Sustain. Energy. 6(4), 1416–1425 (2015)

131. Tao, Y., Chen, H., Qiu, C.: Wind power prediction and pattern feature based on deep learning method. In: 2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), pp. 1–4. (2014)

132. LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. Handbook Brain Theory Neural Netw. 3361(10), 1995 (1995)

133. Gadekallu, T.R., Mamoun, A., et al.: Hand gesture classification using a novel CNN-crow search algorithm. Complex Intell. Syst. 19(1), 1–14 (2021)

134. Wang, H.z., et al.: Deep learning based ensemble approach for probabilistic wind power forecasting. Appl. Energy. 188, 56–70 (2017)

135. Bengio, Y., et al.: Greedy layer-wise training of deep networks. In: Advances in Neural Information Processing Systems, pp. 153–160. (2007)

136. Jiao, R., et al.: A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. IEEE Access. 6, 17851–17858 (2018)

137. Qureshi, A.S., et al.: Wind power prediction using deep neural network based meta regression and transfer learning. Appl. Soft Comput. 58, 742–755 (2017)

138. Khodayar, M., Kaynak, O., Khodayar, M.E.: Rough deep neural architecture for short-term wind speed forecasting. IEEE Trans. Ind. Inf. 13(6), 2770–2779 (2017)

139. Jahangir, H., et al.: Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ANN. Sustain. Energy Technol Assessments. 38, 100601 (2020)

140. Deng, X., et al.: Wind power forecasting methods based on deep learning: a survey. Comput. Model. Eng. Sci. 122(1), 273–301 (2020)

141. Wang, Y., et al.: A prediction model for ultra-short-term output power of wind farms based on deep learning. Int. J. Comput. Commun. Control. 15(4) (2020)

142. Lei, M., et al.: A review on the forecasting of wind speed and generated power. Renew. Sustain. Energy Rev. 13(4), 915–920 (2009)

143. Ramachandram, D., Taylor, G.W.: Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. 34(6), 96–108 (2017)

144. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008. (2017)

145. Wu, N., et al.: Deep transformer models for time series forecasting: the influenza prevalence case. arXiv preprint arXiv:2001.08317 (2020)

146. Zhou, H., et al.: Informer: beyond efficient transformer for long sequence time-series forecasting. arXiv preprint arXiv:2012.07436 (2020)

147. Wu, Z., et al.: Connecting the dots: multivariate time series forecasting with graph neural networks. arXiv preprint arXiv:2005.11650 (2020)

148. Liu, H., Yang, R., Duan, Z.: Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction. Energy Convers. Manag. 217, 112995 (2020)

149. Dai, Z., et al.: Transformer-xl: attentive language models beyond a fixed-length context. arXiv preprint arXiv:1901.02860 (2019)

150. Lan, Z., et al.: Albert: a lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019)

151. Zhu, J., et al.: A3T-GCN: attention temporal graph convolutional network for traffic forecasting. arXiv preprint arXiv:2006.11583 (2020)

152. Koochali, A., Dengel, A., Ahmed, S.: If you like it, GAN it. Probabilistic multivariate times series forecast With GAN. arXiv preprint arXiv:2005.01181 (2020)

153. Zhang, Q.s., Zhu, S.C.: Visual interpretability for deep learning: a survey. Front. Inf. Technol. Electron. Eng. 19(1), 27–39 (2018)

How to cite this article: Wu, Z., et al.: A comprehensive review on deep learning approaches in wind forecasting applications. CAAI Trans. Intell. Technol. 7(2), 129–143 (2022). https://doi.org/10.1049/cit2.12076