# Research on Ultra-Short-Term Prediction Model of Wind Power Based on Attention Mechanism and CNN-BiGRU Combined

Yuyu Meng[1,2], Chen Chang[1,2], Jiuyuan Huo[1,2,3]*, Yaonan Zhang[2], Hamzah Murad Mohammed Al-Neshmi[1], Jihao Xu[1] and Tian Xie[1]

[1]School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China, [2]National Cryosphere Desert Data Center (NCDC), Lanzhou, China, [3]Lanzhou Ruizhiyuan Information Technology Co. LTD, Lanzhou, China

With the rapid development of new energy technologies and aiming at the proposal of the "DOUBLE CARBON" goal, the proportion of wind energy and other new sustainable energy power solutions in the power industry continues to increase and occupy a more critical position. However, the instability of wind power output brings serious challenges to safe and stable power grid operations. Therefore, accurate ultra-short-term wind power prediction is of great significance in stabilizing power system operations. This paper presents an ACNN-BiGRU wind power ultra-short-term prediction model based on the Attention mechanism, the fusion of convolutional neural network (CNN), and bidirectional gated recurrent unit (BiGRU). The model takes a single wind turbine as the prediction unit and uses the real-time meteorological data in the wind farm, the historical power data of the wind turbine, and the real-time operation data for parallel training. Then, it extracts the key features of the input data through CNN and uses the BiGRU network to conduct bidirectional modeling learning on the dynamic changes of the features proposed by CNN. In addition, the Attention mechanism is introduced to give different weights to BiGRU implicit states through mapping, weighting, and learning parameter matrix to complete the ultra-short-term wind power prediction. Finally, the actual observation data of a wind farm in Northwest China is used to verify the feasibility and effectiveness of the proposed model. The model provides new ideas and methods for ultra-short-term high-precision prediction for wind power.

Keywords: wind power prediction, ultra-short-term prediction, hybrid model, the attention mechanism, convolutional neural network, bidirectional gated recurrent unit

## 1 INTRODUCTION

With the proposal of "Carbon peaking" and "Carbon neutrality" dual carbon goals and the rapid development of wind power technology (Zou et al., 2021), the proportion of wind energy in the power supply in all countries in the world continues to increase and occupies a more and more important position. However, wind power has the characteristics of fluctuation, randomness, and intermittence (Fu et al., 2021), and large-scale wind power grid connection has brought some challenges to the safe and stable power system operation (Chen et al., 2021). In order to solve these problems, it is necessary to collect numerical weather prediction (NWP), real-time weather station

data, real-time output power data, inverter unit status, and other data (Sun et al., 2021), establish relevant prediction models and complete the short-term power or ultra-short-term power prediction of wind power plants, so as to ensure the safe and stable operation of power system and realize efficient power generation through power dispatching. Wind power prediction is mainly divided into four categories according to the prediction time scale: ultra-short-term, short-term, medium-term, and long-term (Zheng et al., 2018). According to the prediction function specification of the State Grid, the ultra-short-term wind power shall be able to predict the wind power output power in the next 0–4 h, and the time resolution shall not be less than 15 min (Liu et al., 2015; Abdollah et al., 2016; Dou, 2018). Compared with other prediction categories, ultra-short-term prediction is greatly affected by uncertain factors in a short time, the given prediction time is relatively short, and the required accuracy is high. It is necessary to predict high-precision wind power value in a very short time, which brings significant challenges to hardware and software. The ultra-short-term prediction of wind power with high accuracy can effectively avoid the impact of short-term fluctuation of wind power on the power supply stability of the power system. The power system dispatching department can timely adjust the dispatching plan, reduce the power grid's rotating reserve capacity, improve the power system's absorption capacity to wind power, and realize the safe grid connection of wind power.

At present, wind power prediction methods are mainly divided into physical methods, statistical methods, and learning methods (Wang et al., 2021). The physical method is to find the corresponding internal function relationship through various physical quantities and historical data to determine the predicted value of wind power at a particular time in the future (Zhao et al., 2017). However, due to the transition from physical state to numerical state and the fact that the prediction model parameters are not directly related to the previous data, there will be relatively high errors in the ultra-short-term prediction based on physical methods. The statistical method is to make statistics on a large number of wind power's historical data and weather monitoring data collected by the wind farm, and obtain the corresponding relationship between the above data and the predicted value of wind power through common statistical methods, so as to calculate the predicted value of wind power at a particular time in the future (Natapol and Thananchai, 2019; Yirtici et al., 2019). However, historical data often dramatically affects this kind of wind power prediction. Once the historical data is insufficient or the data is mixed with inaccurate data, the prediction results will not be accurate enough.

The core of the learning method is to build a reliable nonlinear mapping relationship between the input value and output value through a deep learning algorithm, which has a robust nonlinear mapping ability (Li et al., 2020). Moreover, the learning method has a good self-correction ability so that the incorrect data found in the historical data can not cause interference to the model. Therefore, wind power prediction research based on the learning method has become the mainstream and focus of research in recent years (He et al., 2019). For example, Liu et al. (2018) built a wind power prediction framework through a convolution neural

network. Li et al. (2018) established a wind speed prediction model using long short-term memory (LSTM) network as the main predictor. In reference (Niu et al., 2020), a new wind power prediction model is established by using a gated recurrent unit (GRU) network. Liang et al. (2021) proposed CNN-LSTM combined wind power prediction model in combination with CNN and LSTM. Fan et al. (2021) introduce wind speed data at different heights and combine CNN and a two-way gating unit to predict wind speed. However, it is difficult for a single model or simple model combination to adapt to various complex situations. Selecting and combining the advantages of each model and organically integrating multiple models according to the actual operation of the wind farm can effectively improve the prediction accuracy and speed.

Therefore, this paper proposes an ACNN-BiGRU wind power ultra-short-term prediction model based on the Attention mechanism and the fusion of convolutional neural network and bidirectional gated recurrent unit. The model uses CNN to compress the hidden state in the BiGRU model, extract the temporal and spatial correlation between wind power data, shorten the calculation time, and solve gradient disappearance and explosion problems. The BiGRU network is used to model and learn the dynamic changes of the features proposed by CNN. The Attention mechanism is introduced to give different weights to the implied states of BiGRU through the mapping weighting and learning parameter matrix so as to reduce the loss of historical information and strengthen the influence of important information, and then complete ultra-short-term power prediction. In this paper, the model uses the real-time meteorological data, historical power data, and real-time wind turbine operation data of all wind turbines in the wind farm for parallel training, model parameters are optimized, and the predicted power of a single wind turbine is accumulated, which can effectively predict the total power of the wind farm accurately. Finally, the actual observation data of a wind farm in Northwest China are used for experimental verification and compared with various prediction methods for prediction accuracy and speed. The results show that this method can effectively improve prediction accuracy and training speed and has the value of popularization and application.

# 2 RELATED WORKS

## 2.1 Research on Wind Power Prediction

In the research on wind power prediction, Li et al. (2020) proposed the physical modeling method of wind power prediction based on pre-calculation. Ye and Zhao. (2014) used the combined prediction method of wind power combined with statistical and physical models to predict wind power, which verified that the combined model effectively improved the accuracy and made up for the shortcomings of the single model. However, the physical model needs to consider too many uncertain factors, resulting in large errors and considerable computational resources. Compared with the physical model, the statistical method has a single category of demand data, which depends on the correlation between the

historical data of wind power output and the historical data of meteorological indicators. Hodge et al. (2011) used the autoregressive integral moving average (ARIMA) model to predict the future wind power output with historical data. Shi. (2020) considered the limited processing capacity of the least squares support vector machine (LSSVM) model for unsteady components of wind power generation and re-proposed the wind power prediction method based on the combined model of LSSVM and ARIMA, which further improved the accuracy of wind power prediction. Although the demand data category of the statistical model is single, it can not better adapt to the mutation information and needs to collect a large amount of historical data, which has certain limitations.

In recent years, with the rapid development of Artificial Intelligence technology, more and more scholars have applied it to the field of wind power prediction and put forward a variety of wind power prediction models. Such as Wang et al. (2018) proposed a short-term wind power prediction method based on a deep belief network, with numerical meteorological data as input. Lin and Liu. (2020) took the data in supervisory control and data acquisition (SCADA) database with a sampling rate of 1-s as the input and constructed a five-layer feedforward neural network (FNN) to realize wind power prediction. In addition, due to the excellent feature extraction ability of deep convolution neural network and its successful application in the field of image classification, it has been applied in the field of wind power prediction. Hong and Rioflorido. (2019) established a combined prediction model using CNN and radial basis function neural network with double Gaussian function as the activation function to predict the short-term wind power. Wang et al. (2017) proposed a wind power prediction method based on wavelet transform (WT) and CNN. The wavelet transform was adopted to decompose the original wind power data into different frequencies, and then the CNN model was used to predict each frequency.

In the deep learning methods, a recurrent neural network (RNN) is suitable for dealing with the time series problems, but it has a long-term dependency problem. With the development of RNN, its variants, such as LSTM and GRU, appear to overcome the problem and further improve the accuracy of time series prediction. Therefore, Yin et al. (2019) proposed a dual-mode decomposition method composed of empirical mode decomposition (EMD) and variable mode decomposition (VMD) to decompose the original wind force and wind speed time series. Then, the cascade model combining CNN and LSTM is used to extract the meteorological and temporal features of the decomposed subsequence. Compared with LSTM, GRU is more simplified and efficient, and its superiority has been verified. Liu et al. (2019) proposed a multi-step wind speed prediction model combining CNN-GRU and support vector regression (SVR) and decomposed the data by singular spectrum analysis (SSA). Liu C. et al. (2021) proposed a regional wind power prediction method based on adaptive zoning and long-term and short-term matching. This method adds the predicted power of each sub-region to evaluate the power of the whole region in each period.

The above literature research results show that deep learning has good application value in the field of wind power prediction.

Nevertheless, it is mainly applied to the short-term prediction of wind power, and some application results have been achieved. For the ultra-short-term prediction of wind power, if the short-term prediction method is directly applied to the ultra-short-term prediction, the prediction effect is often unsatisfactory (Peng et al., 2016; Wang et al., 2021).

## 2.2 Research on Ultra-Short-Term Prediction of Wind Power

More and more scholars have widely studied wind power ultra-short-term prediction research, but there are still some deficiencies to be further improved in this field. For example, Zhu et al. (2017) select input variables by the Pearson correlation coefficient method and propose an ultra-short-term wind power prediction method based on the LSTM network to reduce the complexity of the prediction model. However, its research only depends on historic data and does not fully consider the factors affecting the ultra-short-term power value of wind farms, so the prediction accuracy still needs to be improved. In order to improve the prediction accuracy of ultra-short-term wind power, Zhang Q. et al. (2021) proposed a multi-variable long short-term memory (MLSTM) network algorithm based on deep learning by comprehensively using wind power history data and wind speed history data for wind power prediction. However, its research relies on historical wind power and wind speed data. The prediction effect is often poor for wind farms lacking historical data or poor data quality. On this basis, Wang et al. (2019) used the good timing memory characteristics of the LSTM network, combined wavelet decomposition technology with the LSTM depth network, and proposed an ultra-short-term probability prediction model of wind power based on small wavelength long short-term memory network. The combination of wavelet decomposition and the deep learning method does improve the prediction accuracy to a certain extent. Nevertheless, using a single model to predict the decomposed waveform lacks certain generalization ability, and it is not easy to achieve the same effect for wind farm power prediction in different geographical locations.

In addition, Zhong et al. (2021) used numerical weather forecast data and wind power history data as the input characteristics of extreme learning machine (ELM) and LSTM network, respectively, and generated prediction data and proposed an ultra-short-term wind power combination prediction method based on historical similarity weighting. This method uses NWP data to solve the problem of over-dependence on historical data, but the prediction speed and feature screening and extraction need to be improved. Xue et al. (2019) propose an ultra-short-term wind power prediction model combining CNN and GRU. Compared with the LSTM model, this model is superior to the latter in prediction speed. This study only uses NWP data and does not consider the role of historical and real-time data, so the prediction accuracy is expected to improve. Yang et al. (2021) propose an ultra-short-term wind power prediction based on multi-location NWP and GRU. This research not only takes NWP data and historical power data as the characteristics affecting the predicted power but

also screens the relevant characteristics to reduce the redundancy of input information. However, the prediction results rely too much on NWP meteorological data, which has the problems of specific errors, a large prediction range, and a long prediction time interval that directly affect the final power prediction accuracy.

Through the study of relevant literature, it is found that, at present, most of the ultra-short-term prediction studies of wind power use NWP data and historical power data, in which the prediction range of NWP can only be accurate to local regions or locations. For ultra-short-term power prediction, the prediction interval is too long. The prediction often takes one or more electric fields as the prediction unit. Thus, the prediction accuracy is poor for the electric fields with long operation times and frequent maintenance of some wind turbines. In addition, each wind turbine has a separate anemometer, wind vane, and thermometer, which can measure the meteorological data of a single wind turbine in real-time. After the comparative study, it is found that compared with NWP data, the real-time meteorological data is closer to the real meteorological data of a single wind turbine in the next ultra-short time. Moreover, the blade rotation of the wind turbine has inertia and has a certain delay in response to the change in wind speed. The power value of the wind turbine in the next ultra-short-time is also greatly affected by the current wind turbine speed.

Therefore, this paper uses the real-time meteorological data, historical power data, and real-time wind turbine operation data of each wind turbine in the wind farm as the input data, carries out parallel training for each wind turbine, optimizes the model parameters, and finally accumulates the predicted power of each single wind turbine. Through the accurate prediction of the power value of a single fan, the model can significantly improve the prediction accuracy of the whole wind farm. Moreover, some wind turbines in the wind farm often cannot work due to maintenance for practical applications. The overall prediction accuracy of the prediction model will not be affected, so it has a strong generalization ability.

# 3 ULTRA SHORT TERM PREDICTION MODEL OF WIND POWER

## 3.1 Convolutional Neural Network

A convolutional neural network (CNN) is a feedforward neural network with convolution calculation and depth structure (Wang, 2020). Its internal neural network layer is mainly composed of the convolution layer, pool layer, and full connection layer. Due to the unique convolution structure of the network, it has been widely and deeply applied in the field of image analysis and processing in recent years. In addition, it can also be used for feature extraction of the Spatio-temporal correlation feature matrix set so that it can compress the input data to the greatest extent without losing the original data features, reduce the input of redundant information and obtain the corresponding key Spatio-temporal correlation features. Its principle is to obtain effective information by using the convolution layer and pooling layer, automatically extracting

feature vectors in data, effectively reducing the complexity of feature extraction and data reconstruction and improving the quality of data features (Zhao et al., 2019; Yildiz et al., 2021).

The data input in this paper is time-series data. In the process of data feature extraction, firstly, a one-dimensional sequence is input to the input layer. After the input data is acted by the first convolution layer and pooling layer, the data features of the input data are extracted. Then the extracted data features are then input to the second convolution layer and pooling layer for processing to obtain the final data features. Compared with the original input data, the feature map has a certain reduction in the vector dimension and carries more apparent characteristics of wind power data. Therefore, it can be better used by the subsequent bidirectional gated recurrent unit (BiGRU) network. In this study, rectified linear unit (ReLU) function is selected as the activation function of the convolution layer. As an unsaturated nonlinear function, ReLU can accelerate the training process's convergence speed and significantly improve CNN's performance. The definition of ReLU is shown in **Eq. 1**. Where $x$ is the input variable, if the input $x$ is less than 0, make the output equal to 0. If the input $x$ is greater than 0, make the output equal to the input.

$$g(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{1}$$

## 3.2 Bidirectional Gated Recurrent Unit

A Gated recurrent unit (GRU) neural network is a kind of recurrent neural network (RNN) and one of many variants of long short-term memory (LSTM) (Chung et al., 2014). LSTM can capture long-term dependence and is suitable for analyzing time series data, but the complex internal structure leads to a long training time. GRU optimizes and improves LSTM to reduce training parameters and ensure prediction accuracy. Compared with the structure of LSTM, GRU combines the forgetting gate and the input gate into an update gate, which has fewer structural parameters and a faster convergence speed than the three gates of LSTM. Therefore, GRU has only two gate structures: the update gate and the reset gate. The update gate is responsible for certain memory and selective forgetting of the data at the previous moment, while the reset gate processes the current information and transmits it to the neural network unit. The GRU based element calculates $h_t$ by **Eqs 2–5**. Where σ is the sigmoid activation function, which is used as a gating signal and can control the value within [0, 1]. The closer the gating signal is to 1, the more data is remembered; otherwise, the more data is forgotten. $W_r$, $W_z$ and $W$ are trainable parameter matrices. Reset the gating to obtain the reset data $r_t \cdot h_{t-1}$, then splice the value with $x_t$, through a *tanh* activation function, the output value is controlled at $[-1, 1]$ to obtain the hidden state $\tilde{h}_t$. Update the gating performs forgetting and selective memory operations simultaneously, among which $(1 - z_t) \cdot h_{t-1}$ selectively forgets the state of the previous node and $z_t \cdot \tilde{h}_t$ selectively remembers the hidden state.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{2}$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{3}$$

$$\tilde{h}_t = tanh(W_r \cdot [r_t * h_{t-1}, x_t]) \tag{4}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{5}$$

Due to the reduction of some matrix operations, the training time of GRU is significantly reduced because it maintains the basic structure of LSTM. GRU can still overcome the disappearance of a gradient in traditional RNN training and maintain good training performance and has reasonable practicability for ultra-short-term wind power prediction with timing characteristics. It should be noted that the recurrent bidirectional network as a whole has always been better than the recurrent feedforward network in the sequence annotation task. GRU unit can only obtain the historical time information but can not obtain the information characteristics of future time. The bidirectional gating recurrent unit is composed of forward GRU and reverse GRU. Finally, the network's output is obtained by the superposition of forward output and reverse output (Lin, 2019; Yang et al., 2021). Therefore, this paper adopts BiGRU to learn the timing relationship between the previous time, the next time, and the current state. The unit based on BiGRU calculates $h_t$ through **Eqs 6–8**. Where $\overrightarrow{h_t}$ is the output of forward GRU and $\overleftarrow{h_t}$ is the output of reverse GRU. Finally, $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are combined to obtain the output $h_t$ of BiGRU layer at time $t$.

$$\overrightarrow{h_t} = \overrightarrow{GRU}\left(x_t, \overrightarrow{h_{t-1}}\right) \tag{6}$$

$$\overleftarrow{h_t} = \overleftarrow{GRU}\left(x_t, \overleftarrow{h_{t-1}}\right) \tag{7}$$

$$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right] \tag{8}$$

## 3.3 Attention Mechanism

The attention mechanism is a resource allocation mechanism that simulates the human brain's attention. It assigns different weights to the input features so that the essential features will not disappear with the increase of step size and then highlight the role of important information so that the model is easier to deal with the dependence on long-time series. Its core goal is to enhance the key information and weaken the role of redundant information on the current target task (Feng et al., 2020; Meng et al., 2021). The Attention mechanism pays enough Attention to the distribution of key information through probability distribution and realizes the dynamic distribution of weight coefficients for different inputs, thus improving the prediction accuracy of the model.

BiGRU can solve the problem of long-term memory to a certain extent, but when learning too long sequence data, there will be problems such as low efficiency, long time, loss of local feature information, etc. In order to make up for the defects of BiGRU learning process, this paper introduces the Attention mechanism acting on the time dimension of the input sequence. Firstly, we can selectively focus on the information at different positions in the sequence data to reduce the length of the input data. Secondly, the Attention mechanism can assign weights to the features that affect the prediction results, help the

model learn the potential features more effectively, improve the detection accuracy and robustness of the model, and make it easier to obtain the long-distance interdependent features in the wind power time series data.

## 3.4 CNN-BiGRU Prediction Model Based on Attention Mechanism

This article proposes a CNN-BiGRU wind power ultra short-term prediction model based on the Attention mechanism to take a single wind turbine as the prediction unit. The model structure and workflow are shown in **Figure 1**. It is mainly divided into the input layer, CNN layer, CNN-BiGRU layer, Attention layer, and output layer. Historical power data, real-time meteorological data, and wind turbine operation data are used as inputs to extract features through the CNN layer. The BiGRU and Attention layers learn the proposed features' internal variation law of power to realize the prediction function. Finally, the prediction results are obtained through the output layer.

Each layer in the model is described as follows:

1) Input layer. The historical power data, real-time meteorological data, and operation data of a single wind turbine in the wind farm are taken as the input of the prediction model and expressed by $X = [x_1 \cdots x_t \cdots x_n]^T$. The wind turbine data with length $n$ is input into the prediction model after preprocessing and then enters the CNN layer for processing.

2) CNN layer. Spatio-temporal feature extraction is carried out on the input wind turbine data, which is composed of two convolution layers, two pool layers, and full connection layers. Where $Y$ and $E$ represent two one-dimensional convolution layers, and $Max$ represents the maximum pooling layer. According to the input wind turbine data characteristics, the convolution layer one and convolution layer two are designed as one-dimensional convolution, and the ReLU activation function is selected for activation. In order to retain more data fluctuation characteristics, the pooling method of the pooling layer is selected as maximum pooling. The data processed by the convolution and pooling layers are mapped to the hidden layer feature space, the full connection layer structure is built and transformed into one-dimensional structure output, and the corresponding feature vector is extracted and input to the BiGRU layer.

3) BiGRU layer. The feature vector extracted from the CNN layer is bi-directional learned. Build a single-layer and two-way GRU structure, fully learn the proposed features, capture the change law of its internal information features, and input them to the Attention mechanism layer.

4) Attention layer. The corresponding weight is dynamically assigned to the output vector after learning and processing at the BiGRU network layer. According to the weight distribution principle, the corresponding probabilities of different eigenvectors are calculated, and the better weight parameter matrix is continuously updated and iterated.
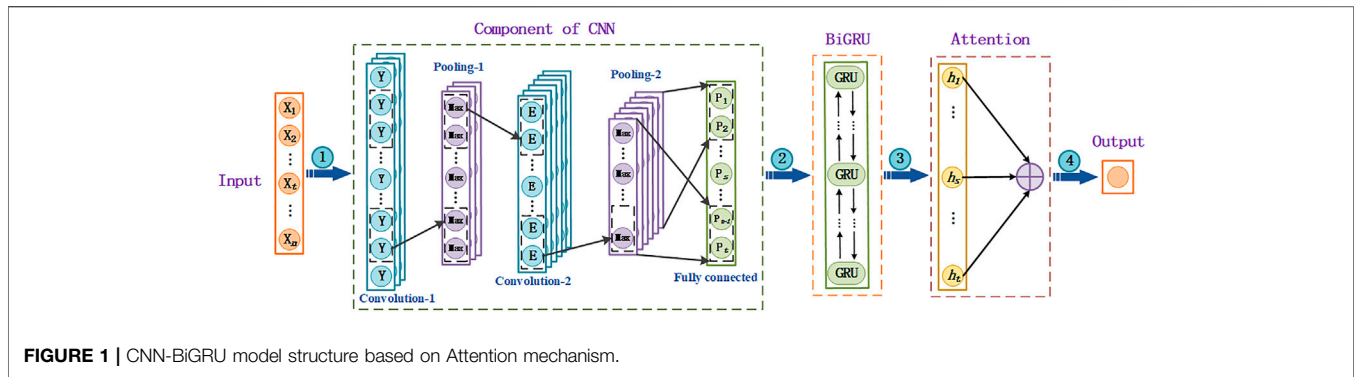
**FIGURE 1 |** CNN-BiGRU model structure based on Attention mechanism.

5) Output layer. The input of the output layer is the output of the Attention mechanism layer, and the final predicted value of wind power is output through the full connection layer.

In the training process of the prediction model, this paper uses the Adam algorithm as the optimization algorithm of the model gradient to continuously optimize the model parameters. Adam is a first-order optimization algorithm that can replace the traditional stochastic gradient descent (SGD) process. The algorithm can continuously iteratively update the parameters of the neural network based on the training data to optimize the output value of the loss function, has high computational efficiency and low memory occupancy, and the diagonal scaling of the algorithm gradient is invariant (Ling et al., 2013; Kingma and Ba, 2014). The model's loss function uses the mean square error function, and its calculation formula is shown in **Eq. 9**, where $n$ is the number of samples trained by a single wind turbine; $y_i$ is the actual power value; $\overline{y_i}$ is the output power value of the model.

$$L_{oss} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y_i}\right)^2 \tag{9}$$

# 4 EXPERIMENTAL COMPARISON AND ANALYSIS

## 4.1 Data Description and Data Preprocessing

This paper takes 100 wind farms with a rated power of 200 MW in Northwest China as the research object. In the experiments, a single wind turbine is taken as a prediction unit, the historical power data, real-time operation data, and real-time meteorological data of each wind turbine are taken as the research data, and the sampling interval is 5 min. A total of 10, 538, 900 wind turbine data from 1 June 2019, to 30 May 2020, were used as a dataset, in which the training set and test set accounted for the first 70 and 20% of the dataset, respectively, and the last 10% of the dataset was used as a validation set to evaluate the generalization ability of the model.

In this experiment, the prediction model is trained and established based on the real-time speed, pitch angle, wind

**TABLE 1 |** Original characteristic variable range of wind power.

| Number | Characteristic Variable | Range |
|---|---|---|
| 1 | Wind speed (m/s) | 0.00~49.70 |
| 2 | Rotate speed (r/min) | 0.00~16.00 |
| 3 | Wind direction (°) | −250.70~502.74 |
| 4 | Temp air (°C) | −21.88~36.80 |
| 5 | Pitch angle (°) | −79.57~116.44 |
| 6 | Wind power (KW) | −56.20~2,115.60 |

speed, wind direction, ambient temperature, and the historical power data of the wind turbine. The speed and pitch angle of the wind turbine can reflect the real-time operation state of the wind turbine, and the wind speed is a direct variable in wind energy production, but it is affected by the wind direction and ambient temperature. In addition, the historical power information of the wind turbine can effectively reflect its change trend. The original data range of characteristic variables related to wind power is shown in **Table 1**.

In the actual operation of the wind farm, due to the uncertainty of measurement equipment and data transmission equipment, there are noise signals and many abnormal values in the original data of almost every wind farm. Therefore, this paper preprocesses the original data of the wind farm:

1) Replace the power data greater than the rated installed capacity with the installed capacity value.
2) Replace the power data less than zero with zero value.
3) The data with continuous missing less than or equal to three sampling points shall be supplemented with adjacent power data.
4) If the data of more than three sampling points is missing continuously, the continuous method shall be used to supplement.
5) Using the method of sampling interval transformation, the unified time interval is set to 15 min.

Wind turbines #20, #50, and #80 were selected for analysis according to different geographical locations and altitudes in the wind farm. Their location distribution is shown in **Figure 2**, and the fluctuation information of the first 1,000 sampling points is shown in **Figure 3**.
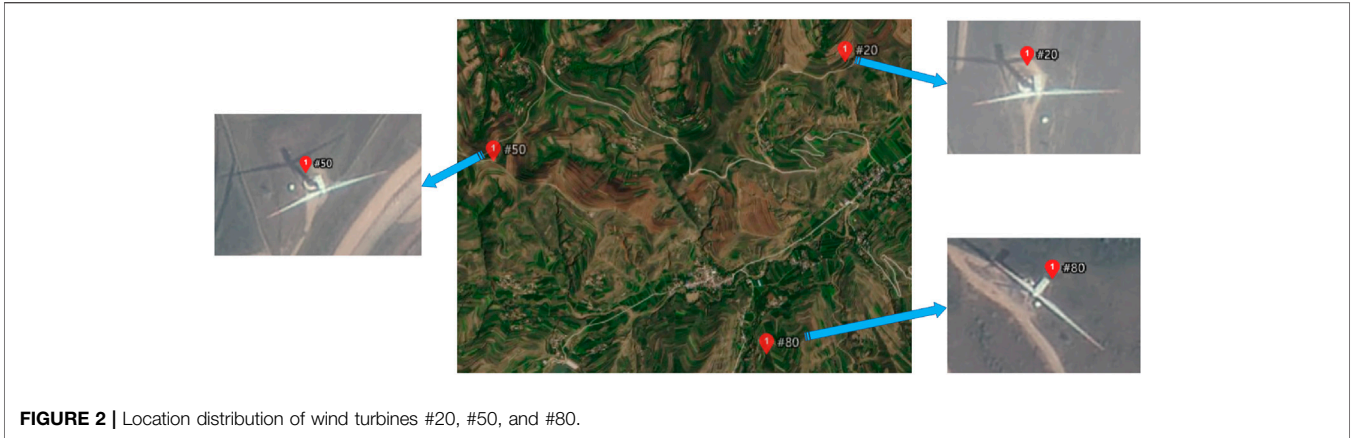
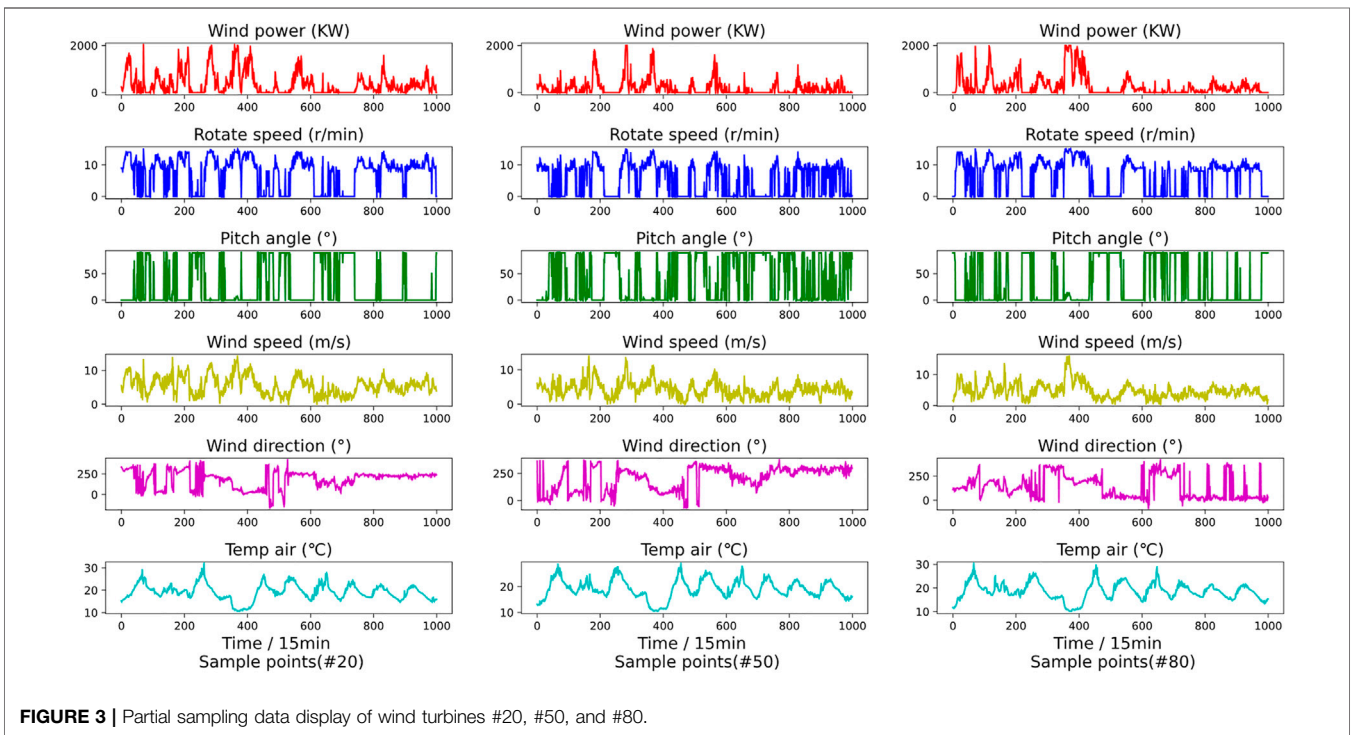**FIGURE 2 |** Location distribution of wind turbines #20, #50, and #80.



**FIGURE 3 |** Partial sampling data display of wind turbines #20, #50, and #80.

Because the dimensions of different features are different, in order to fully consider the correlation degree between each feature and power, it is necessary to normalize the data used to the maximum and minimum and normalize the data to the interval [0, 1]. That is **Eq. 10**. Where $x_c$ is the input data of the standardized model, $x_s$ is the original input data; $x_{max}$ and $x_{min}$ are the maximum and minimum values of the original data, respectively.

$$x_c = \frac{x_s - x_{min}}{x_{max} - x_{min}} \tag{10}$$

## 4.2 Prediction Accuracy Evaluation Criteria
In this study, the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are

used as the error evaluation indexes of wind power prediction. The expression is shown in **Eqs 11–13**. Where $n$ is the total number of prediction results; $y_i$ and $\overline{y_i}$ are the actual power value and predicted power value of the $i$ sampling point, respectively. MAPE can measure the quality of the model prediction results. RMSE and MAE can evaluate the prediction accuracy and are sensitive to the maximum or minimum error in the results. In wind power prediction, the smaller the value of MAE and RMSE, the more accurate the power prediction result, and the smaller the MAPE, the better the prediction effect of the model (Liu X. et al., 2021; Zhang J. et al., 2021).

$$R_{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y_i}\right)^2} \tag{11}$$

**FIGURE 4 |** Wind ultra-short-term-power prediction framework.
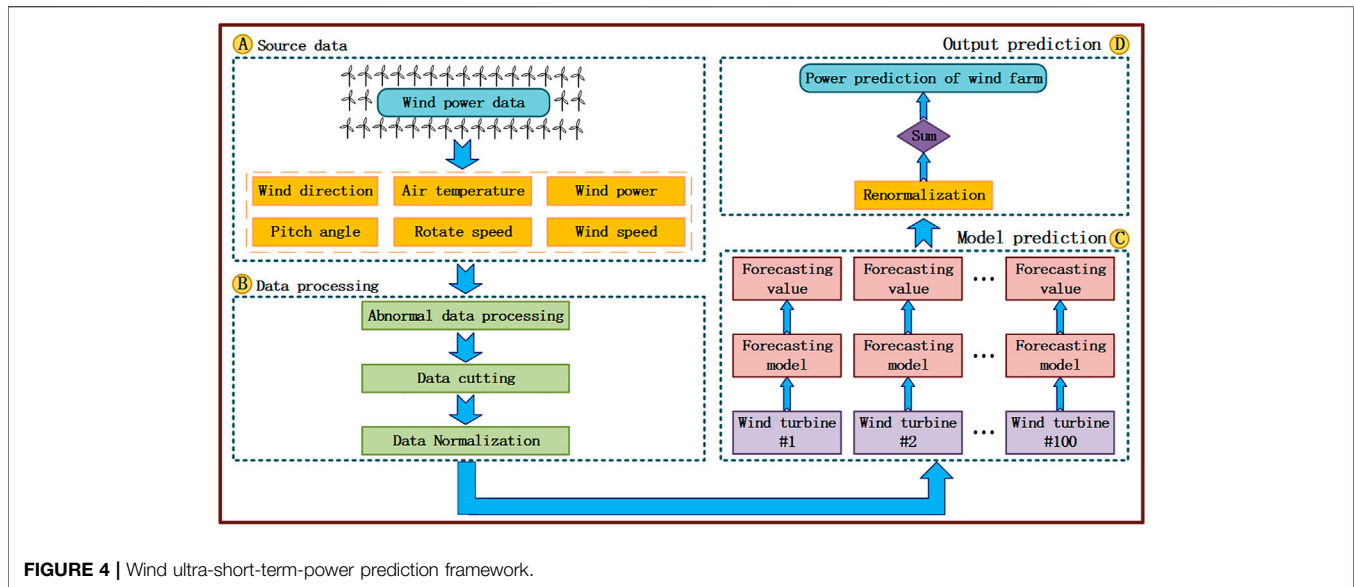
$$M_{AE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \overline{y_i} \right| \tag{12}$$

$$M_{APE} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_i - \overline{y_i} \right|}{y_i} \times 100\% \tag{13}$$

## 4.3 Wind Power Prediction Framework

The structure of the power generation of the wind farm predicted in this paper is shown in **Figure 4**, which includes an active data collection module, data processing module, model prediction module, and prediction output module. The specific steps of its model are as follows:

**Step 1.** Source data module. The main function of the source data collection module is to complete the collection and integration of wind farm data, including wind turbine speed, pitch angle, wind speed, wind direction, and ambient temperature data.

**Step 2.** Data processing module. The data processing module processes the data collected by the source data collection module. First, replace the abnormal values and fill in the missing values in the source data, then cut and split the data according to the fan number, and finally normalize the data before model prediction according to the fan.

**Step 3.** Model prediction module. The model prediction module inputs the normalized data into each model according to the fan for training and prediction and transmits the predicted value of each model to the prediction output module for processing.

**Step 4.** Output prediction module. The prediction output module performs inverse normalization on the predicted values of each model and finally accumulates and sums the predicted power of each wind turbine so as to obtain the ultra-short-term predicted total power of the wind farm.

## 4.4 Model Training and Result Evaluation Analysis

In this paper, the wind power prediction of the wind farm is verified. The prediction effect is the best through many experiments when the model input time series dimension is 8. According to the demand of the actual application of the field, the prediction period is set as 30 min. That is, for a specific time $t$, the real wind power data of 15 ~ 120 min in front of each wind turbine, the wind speed, wind direction, and temperature at time t, as well as the speed and pitch angle of the wind turbine, are used as the input data to predict the wind power after 30 min.

In addition, in terms of single wind turbine ultra-short-term power prediction, the proposed model is also analyzed and compared with the prediction results of the support vector regression (SVR) model, the backpropagation (BP) neural network model, and the GRU model. In the ultra-short-term prediction of wind farm total power, the prediction accuracy of the GRU model, BiGRU model, and BiGRU model based on the Attention mechanism are analyzed, and the prediction accuracy and speed of ACNN-LSTM model based on the Attention mechanism are analyzed, and compared. The results show that the proposed model has obvious advantages in prediction performance and efficiency for single wind turbine ultra-short-term power prediction and wind farm total power ultra-short-term prediction and has popularization and application value.

In this experiment, we choose to build an ACNN-BiGRU prediction model based on the Attention mechanism in the Windows environment by using Python language and the TensorFlow framework. A CNN network with two-layer convolution and a two-layer maximum pooling layer is built at the head of the prediction model. The two-layer convolution layer has 6 and 16 convolution cores, respectively, with a step of one to extract the features of the original data. The model prediction part, including bidirectional GRU layer neurons, is constructed. According to the test feedback of the network, the

**FIGURE 5 |** Power prediction results of different models' wind turbines #20, #50, and #80.
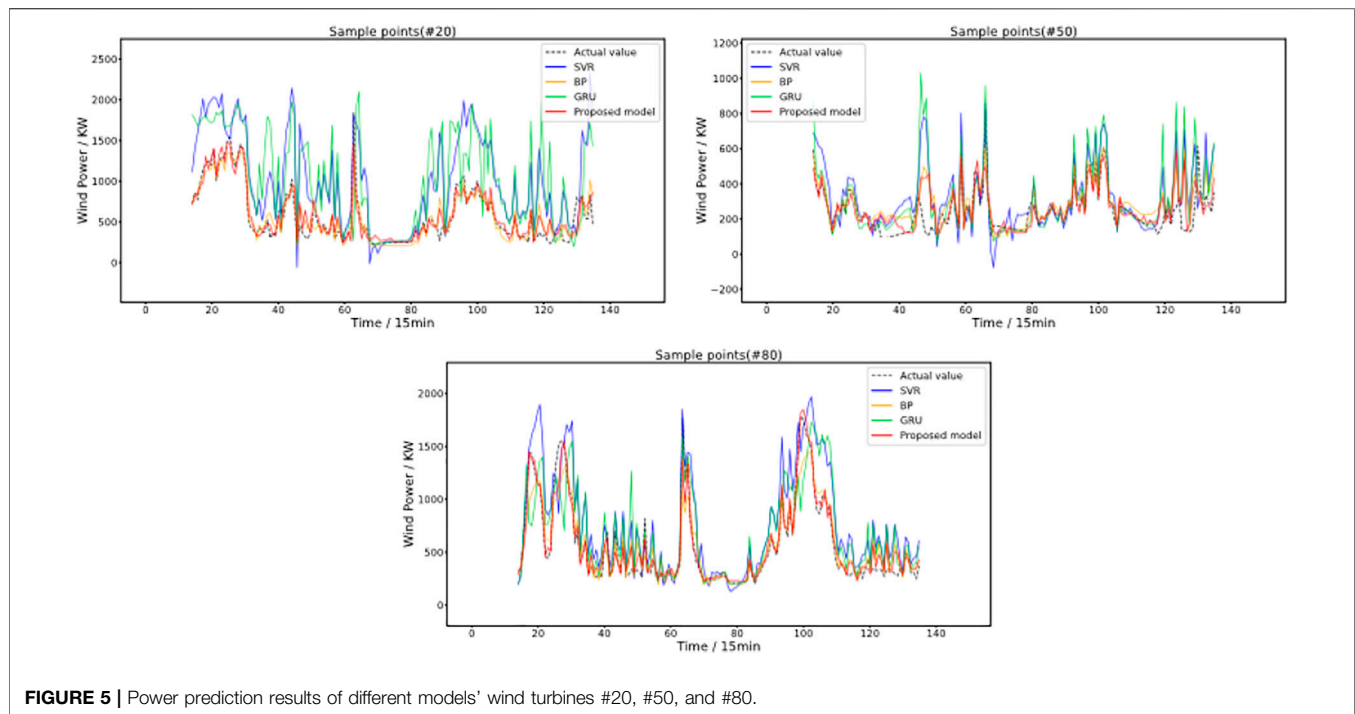
number of neurons in this layer is set to 20. After passing through the Attention layer, the wind power values predicted by the t time model are output through the fully connected network.

### 4.4.1 Ultra Short Term Power Prediction of Single Wind Turbine

The power prediction model of each wind turbine in this paper is the same as that of the CNN Attention model, and the number of iterations of each wind turbine prediction model is 200. In order to fully illustrate the generalization ability of the prediction model in this paper, the prediction results of wind turbines #20, #50, and #80 located in different geographical locations and altitudes in the wind farm are evaluated and analyzed. It is compared with the SVR, BP neural network, and GRU models. The prediction results of each wind turbine are shown in **Figure 5**, and the prediction error indexes of each model are shown in **Figure 6**. The following conclusions can be drawn from **Figures 5**, **6**:

1)  Compared with the traditional machine learning SVR model, traditional deep learning BP model, and single GRU model, the prediction curve of the wind power prediction model proposed in this paper can better fit the actual wind power curve of each wind turbine and has the minimum RMSE and MAE values. It shows that the model has a good prediction effect for wind turbines in different geographical environments.
2)  In the error evaluation index of prediction results, it can be seen that for the power prediction of different wind turbines, the fluctuation range of RMSE, MAE, and MAPE values of the model proposed in this study is the smallest compared with other models, indicating that the model has a relatively stable prediction effect on the prediction of different wind

turbines in the wind farm and has good generalization performance.

### 4.4.2 Ultra Short Term Prediction of the Total Power of Wind Farm

In order to further verify the prediction ability and practicability of the model proposed in this paper, the prediction results of the whole wind farm are compared and analyzed. Some prediction results and error indicators of the GRU model, BiGRU model, and BiGRU model based on the Attention mechanism are shown in **Figure 7**. The following conclusions can be drawn from **Figure 7**:

1)  The fitting effect of the single GRU model is the worst, and the values of RMSE, MAE, and MAPE are the largest. The BiGRU model showed a significant improvement compared with the GRU model, and the evaluation index values RMSE, MAE, and MAPE decreased by 10.00, 9.20, and 17.50, respectively. It shows that the BiGRU model plays a significant role in improving the accuracy of wind power prediction than the single GRU model.
2)  Compared with the BiGRU model, the prediction results of the BiGRU + Attention model are improved by introducing the Attention mechanism, and the evaluation index values RMSE, MAE, and MAPE are reduced by 2.40, 0.80, and 1.70, respectively. It can be seen that the addition of the Attention mechanism enhances the key information and weakens the redundant information, helps the model learn potential features more effectively, and improves the prediction accuracy and robustness of the model.
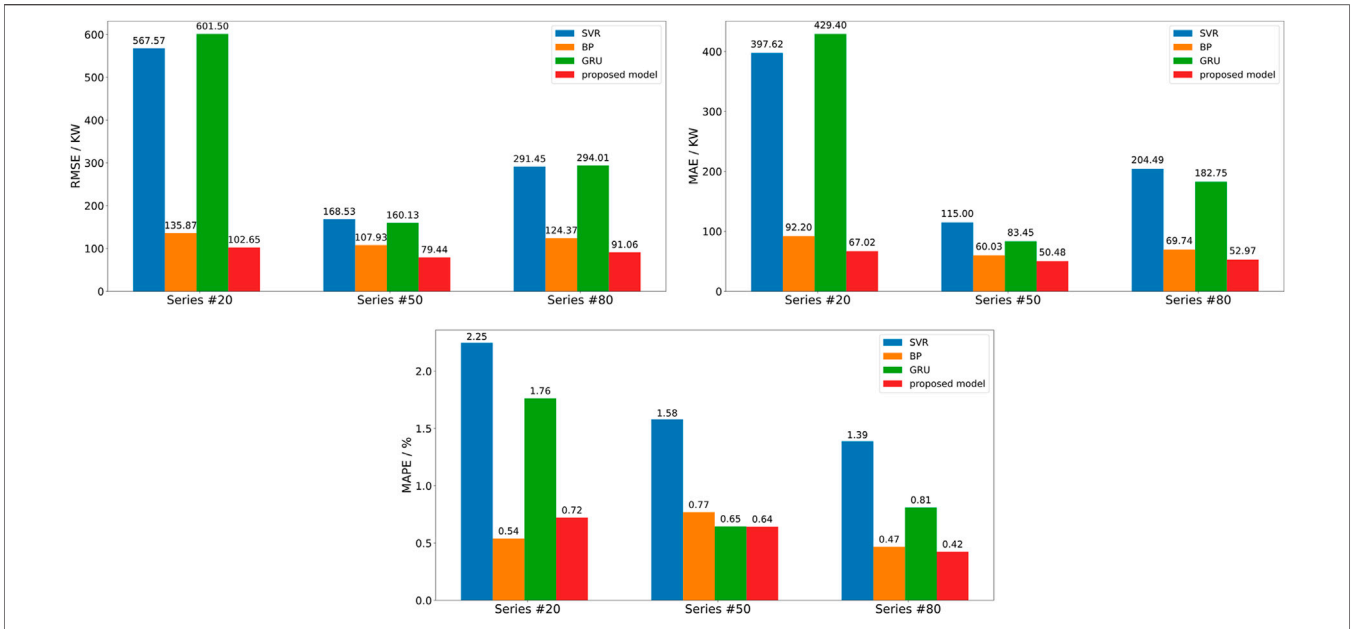3)  Based on the BiGRU model based on the Attention mechanism, CNN is introduced to form the model CNN-

**FIGURE 6 |** Power prediction and evaluation indexes of different models of wind turbines #20, #50, and #80.
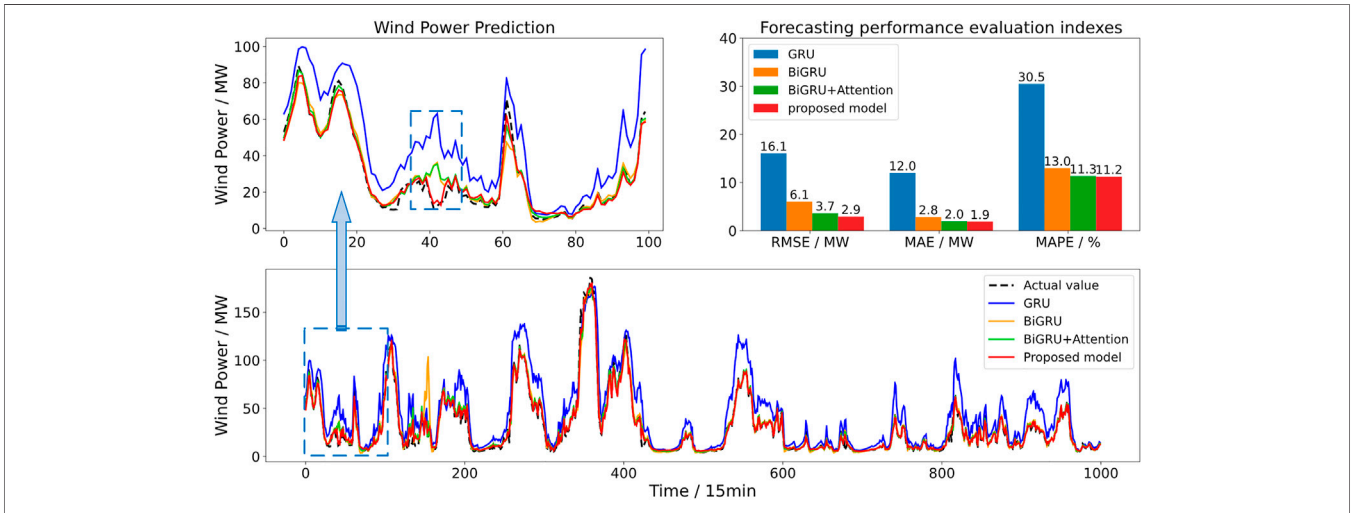


**FIGURE 7 |** Partial prediction results and prediction error indexes of different models of the whole wind farm.

BiGRU proposed in this paper. Compared with the BiGRU model based on the Attention mechanism, the evaluation index values of RMSE, MAE, and MAPE are reduced by 0.80, 0.10, and 0.10, respectively, which further improves the prediction accuracy of wind power. Combined with the comparison of prediction results, it can be seen that the introduction of CNN plays a role in feature extraction of the input data set and reduces the input of redundant information. In particular, the power trough between prediction point 40 and prediction point 50 is pronounced. Other models do not capture the fluctuation characteristics of power decline in a short time, while the model in this paper

can effectively capture the relevant characteristics of downward power fluctuation, showing the ability of wind power prediction in more extreme cases.

4) The power prediction value of the whole wind farm is obtained by accumulating the prediction values of each wind turbine, and the final prediction accuracy of the wind farm is high. Combined with the comparative analysis and research of wind turbines #20, #50, and #80, it shows that the CNN-BiGRU model proposed in this paper has more accurate results for the prediction of each wind turbine in the wind farm, and this model has strong prediction ability and practicability.
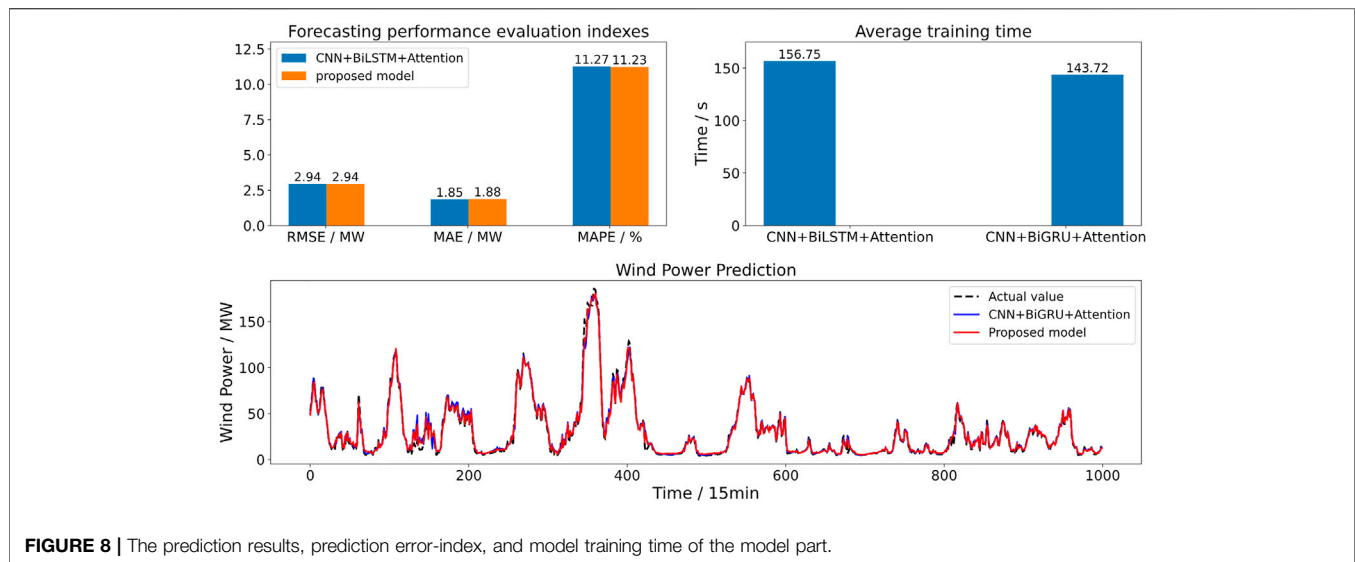
**FIGURE 8 |** The prediction results, prediction error-index, and model training time of the model part.

Since both LSTM and GRU are improved versions of RNN, they have similar predictive performance. Therefore, the prediction performance and efficiency of the proposed model are compared with that of the CNN-LSTM model based on the Attention mechanism. In order to ensure the accuracy and rationality of the experiment, the two models adopt the same neural network structure and the number of training iterations. The model's prediction result, error evaluation index, and training time are shown in **Figure 8**. It can be seen that the prediction effect and error of the two models are almost the same, but the efficiency of the former in training time is slightly higher than that of the latter. This indicates that the model presented in this paper has higher time efficiency while ensuring better prediction performance and is more valuable for promotion and application in ultra-short-term wind power prediction.

## 5 CONCLUSION

Aiming at the problem of low accuracy and efficiency in the current ultra-short-term wind power prediction, a new ACNN-BiGRU multi turbine wind power prediction model based on the Attention mechanism is proposed. On the one hand, a convolutional neural network with double convolution layers is constructed and effectively mines the feature of input uncertain time series through its multi-layer feature catcher and maximum pooling structure. On the other hand, the Attention mechanism is introduced to enhance the influence weight of key information, which effectively helps the model learn potential features and ultimately improves the model's prediction accuracy. In the experimental comparison stage, the experiments of ultra-short-term power prediction of a single wind turbine and total power prediction of the whole wind farm are designed. Compared with the advanced mainstream model, the experimental results show that the model effectively improves the accuracy and speed of wind power ultra short-term power prediction and has obvious advantages in prediction performance

and efficiency. Furthermore, for the practical application of wind farms, some fans often cannot work due to maintenance, yet their prediction accuracy will not be affected. It effectively solves the problems existing in the practical application of current wind farm power prediction and has strong generalization ability and high engineering application value.

The model used in this paper needs to carry out parallel training and prediction for multiple wind turbines, which requires high hardware operation ability and high application cost for scale. The next step is to analyze further and study the wind turbines of the wind farm to reduce the computing pressure of hardware and save costs. In addition, the model proposed in this paper is not only applicable to the prediction of wind power generation but also applicable to applied research in other fields. Such as the study proposed by Diaz et al. (2021), the diagnosis and classification of Parkinson's disease based on sequential and one-dimensional convolution BiGRUs; and Lucas et al. (2022), who proposed the application of BiGRU-CNN neural network in the detection of electric theft, etc., which also achieved good results. Therefore, this model can be extended to other fields for research and application.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

YM put forward the main research points and designed the framework; CC completed the manuscript writing and revision; JH and YZ provided significant suggestions on the methodology and structure of the manuscript; HM and JH revised grammar and expression; TX collected

relevant background information; JX analyzed the data of this study.

National Cryosphere Desert Data Center (NCDC), and the Foundation of a Hundred Youth Talents Training Program of Lanzhou Jiaotong University.

## REFERENCES

Abdollah, K. F., Abbas, K., and Saeid, N. (2016). A New Fuzzy-Based Combined Prediction Interval for Wind Power Forecasting. *IEEE Trans. Power Syst.* 31, 18–26. doi:10.1109/TPWRS.2015.2393880

Chen, X. J., Zhang, X. Q., Dong, M., Huang, L. S., Guo, Y., and He, S. Y. (2021). Deep Learning-Based Prediction of Wind Power for Multi-Turbines in a Wind Farm. *Front. Energy Res.* 9, 1–6. doi:10.3389/fenrg.2021.723775

Chung, J., Gulcehre, C., Cho, K. H., and Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. Eprint Arxiv. doi:10.48550/arXiv.1412.3555

Diaz, M., Moetesum, M., Siddiqi, I., and Vessio, G. (2021). Sequence-based Dynamic Handwriting Analysis for Parkinson's Disease Detection with One-Dimensional Convolutions and BiGRUs. *Expert Syst. Appl.* 168, 114405. doi:10.1016/j.eswa.2020.114405

Dou, J. L. (2018). Wind Power Prediction Technology Based on Deep Learning Algorithm. Beijing: China Electric Power Research Institute. [Master's thesis].

Fan, H., Zhang, X. M., Mei, S. W., and Yang, Z. L. (2021). Ulatra-short-term Wind Speed Prediction Model for Wind Farms Based on Spatiotemporal Neural Network. *Automation Electr. Power Syst.* 45, 28–38. doi:10.7500/AEPS20190831001

Feng, B., Zhang, Y. W., Tang, X., Guo, C. X., Wang, J. J., Yang, Q., et al. (2020). Power Equipment Defect Record Text Mining Based on BiLSTM-Attention Neural Network. *Proc. CSEE* 40, 1–10. doi:10.13334/j.0258-8013.pcsee.200530

Fu, Y., Ren, Z. X., Wei, S. R., Wang, Y., Huang, L. L., and Jia, F. (2021). "Ultra-Short Term Power Prediction of Offshore Wind Power Based on Improved LSTM-TCN Model. *Proc. CSEE*, 1–10. doi:10.13334/j.0258-8013.pcsee.210724

He, Y., Yan, Y., and Xu, Q. (2019). Wind and Solar Power Probability Density Prediction via Fuzzy Information Granulation and Support Vector Quantile Regression. *Int. J. Electr. Power & Energy Syst.* 113, 515–527. doi:10.1016/j.ijepes.2019.05.075

Hodge, B.-M., Zeiler, A., Brooks, D., Blau, G., Pekny, J., and Reklatis, G. (2011). Improved Wind Power Forecasting with Arima Models. *Comput. Aided Chem. Eng.* 29, 1789–1793. doi:10.1016/B978-0-444-54298-4.50136-7

Hong, Y.-Y., and Rioflorido, C. L. P. P. (2019). A Hybrid Deep Learning-Based Neural Network for 24-h Ahead Wind Power Forecasting. *Appl. Energy* 250, 530–539. doi:10.1016/j.apenergy.2019.05.044

Kingma, D., and Ba, J. (2014). Adam: a Method for Stochastic Optimization. *Comput. Sci.*, 1–15. doi:10.48550/arXiv.1412.6980

Korprasertsak, N., and Leephakpreeda, T. (2019). Robust Short-Term Prediction of Wind Power Generation under Uncertainty via Statistical Interpretation of Multiple Forecasting Models. *Energy* 180, 387–397. doi:10.1016/j.energy.2019.05.101

Li, L. L., Zhao, X., and Tseng, M. L. (2020). Shot-term Wind Power Forecasting Based on Support Vector Machine with Improved Dragonfly Algorithm. *J. Clean. Prod.* 242, 38–47. doi:10.1016/j.jclepro.2019.118447

Li, S. H. (2020). Research on Short-Term Wind Speed and Wind Power Prediction Based on Historical Data of Wind Farm. Gansu: Lanzhou University of Technology. [Master's thesis]. doi:10.27206/d.cnki.ggsgu.2020.000667

Li, Y., Shi, H., Han, F., Duan, Z., and Liu, H. (2019). Smart Wind Speed Forecasting Approach Using Various Boosting Algorithms, Big Multi-step Forecasting Strategy. *Renew. Energy* 135, 540–553. doi:10.1016/j.renene.2018.12.035

Liang, C., Liu, Y. Q., and Zhou, J. K. (2021). Wind Speed Prediction at Multi-Locations Based on Combination of Recurrent and Convolutional Neural Networks. *Power Syst. Technol.* 45, 534–542. doi:10.13335/j.1000-3673.pst.2020.0767

Lin, J. Y. (2019). Research on Unbalanced Text Classification Algorithm Based on Improved BiGRU. Guangdong: Guangdong University of Technology. [Master's thesis]. doi:10.27029/d.cnki.ggdgu.2019.000619

Lin, Z., and Liu, X. (2020). Wind Power Forecasting of an Offshore Wind Turbine Based on High-Frequency Scada Data and Deep Learning Neural Network. *Energy* 201, 117693. doi:10.1016/j.energy.2020.117693

Ling, W. N., Hang, N. S., and Li, R. Q. (2013). Short-term Wind Power Forecasting Based on Cloud SVM Model. *Electr. Power Automa-tion Equip.* 33, 34–38. doi:10.3969/j.issn.1006-6047.2013.07.006

Liu, C., Pei, Z. U., Wang, B., Dong, C., and Shi, Y. (2015). *Function Specification of Wind Power Forecasting*. Beijing: China Electric Power Press.

Liu, C., Zhang, X., Mei, S., and Liu, F. (2021a). Local-pattern-aware Forecast of Regional Wind Power: Adaptive Partition and Long-Short-Term Matching. *Energy Convers. Manag.* 231, 113799. doi:10.1016/j.enconman.2020.113799

Liu, H., Mi, X., Li, Y., Duan, Z., and Xu, Y. (2019). 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. doi:10.1016/j.renene.2019.05.039

Liu, H., Mi, X., and Li, Y. (2018). 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. doi:10.1016/j.enconman.2018.04.021

Liu, X., Zhang, L., Zhang, Z., Zhao, T., and Zou, L. (2021b). Ultra-short-term Wind Power Prediction Model Based on VMD Decomposition and LSTM. *IOP Conf. Ser. Earth Environ. Sci.* 838, 012002. doi:10.1088/1755-1315/838/1/012002

Lucas, D., Altamira, S., Gloria, P., Edgar, M., Jesús, M., and Nicolás, M. (2022). BiGRU-CNN Neural Network Applied to Electric Energy Theft Detection. *Electronics* 11, 693. doi:10.3390/ELECTRONICS11050693

Meng, A. B., Chen, S., Wang, C. E., Ding, W. F., Cai, Y. F., Fu, J. J., et al. (2021). Ultra-short Term Wind Power Prediction Based on Chaotic CSO Optimized Time Series Attention GRU Model. *Power Syst. Technol.* 45, 1–9. doi:10.13335/j.1000-3673.pst.2021.0787

Niu, Z., Yu, Z., Tang, W., Wu, Q., and Reformat, M. (2020). Wind Power Forecasting Using Attention-Based Gated Recurrent Unit Network. *Energy* 196, 117081. doi:10.1016/j.energy.2020.117081

Peng, X. S., Xiong, L., Wen, J. Y., Cheng, S., and Wang, B. (2016). A Summary of the State of the Art for Short-Term and Ultra-short-term Wind Power Prediction of Regions. *Proc. CSEE* 36, 6315–6326. doi:10.13334/j.0258-8013.pcsee.161167

Shi, K. F. (2020). Ultra-short Term Power Prediction of Wind Power Based on Combined Model. Hebei: Yan Shan University. [Master's thesis]. doi:10.27440/d.cnki.gysdu.2020.001122

Sun, R. F., Zhang, T., He, Q., and Xu, H. X. (2021). Summary of Key Technologies and Applications of Wind Power Prediction. *High. Volt. Eng.* 47, 1129–1143. doi:10.13336/j.1003-6520.hve.20201780

Wang, H.-z., Li, G.-q., Wang, G.-b., Peng, J.-c., Jiang, H., and Liu, Y.-t. (2017). Deep Learning Based Ensemble Approach for Probabilistic Wind Power Forecasting. *Appl. Energy* 188, 56–70. doi:10.1016/j.apenergy.2016.11.111

Wang, K., Qi, X., Liu, H., and Song, J. (2018). Deep Belief Network Based K-Means Cluster Approach for Short-Term Wind Power Forecasting. *Energy* 165, 840–852. doi:10.1016/j.energy.2018.09.118

Wang, P., Sun, Y. H., Zhai, S. W., Hou, D. C., and Wang, S. (2019). Ultra-short Term Probability Prediction of Wind Power Based on Small Wave Long Short-Term Memory Network. *J. Nanjing Univ. Inf. Sci. Technol. Nat. Sci. Ed.* 11, 460–466. doi:10.13878/j.cnki.jnuist.2019.04.015

Wang, Y. (2021). Combined Model of Short-Term Wind Speed Prediction of Wind Farm Based on Deep Learning. [Master's thesis]. Anhui: University of Science and Technology of China. doi:10.27517/d.cnki.gzkju.2021.000735

Wang, Y. C. (2020). Wind Power Prediction Model Based on Deep Neural Network. Shanghai: Shanghai DianJi University. [Master's thesis]. doi:10.27818/d.cnki.gshdj.2020.000014

Wang, Y. H., Shi, Y. X., Zhou, X., Zeng, Q., Fang, B., and Bi, Y. (2021). Ultra-short Term Power Prediction of BiLSTM Multi Wind Turbine Based on Time Mode Attention Mechanism. *High. Volt. Eng.*, 1–9. doi:10.13336/j.1003-6520.hve.20211561

Xue, Y., Wang, L., Wang, S., Zhang, Y. F., and Zhang, N. (2019). An Ultra-short Term Wind Power Prediction Model Combining CNN and GRU Network. *Renew. Energy Resour.* 37, 144–150. doi:10.13941/j.cnki.21-1469/tk.2019.03.023

Yang, H. Y., Zhang, Z. Z., and Zhang, L. (2021). Network Security Situation Assessment Based on Parallel Feature Extraction and Improved BiGRU. *J. Tsinghua Univ. Technol.*, 1–7. doi:10.16511/j.cnki.qhdxxb.2022.22.006

Yang, M., and Bai, Y. Y. (2021). Ultra-short Term Prediction of Wind Power Based on Multi Position NWP and Gated Recurrent Unit. *Automation Electr. Power Syst.* 45, 177–183. doi:10.7500/AEPS20200521007

Ye, L., and Zhao, Y. N. (2014). Review of Wind Power Prediction Based on Spatial Correlation. *Automation Electr. Power Syst.* 38, 126–135. doi:10.7500/AEPS20130911004

Yildiz, C., Acikgoz, H., Korkmaz, D., and Budak, U. (2021). An Improved Residual-Based Convolutional Neural Network for Very Short-Term Wind Power Forecasting. *Energy Convers. Manag.* 228, 113731. doi:10.1016/j.enconman.2020.113731

Yin, H., Ou, Z., Huang, S., and Meng, A. (2019). A Cascaded Deep Learning Wind Power Prediction Approach Based on a Two-Layer of Mode Decomposition. *Energy* 189, 116316. doi:10.1016/j.energy.2019.116316

Yirtici, O., Ozgen, S., and Tuncer, I. H. (2019). Predictions of Ice Formations on Wind Turbine Blades and Power Production Losses Due to Icing. *Wind Energy* 22, 945–958. doi:10.1002/we.2333

Zhang, J., Liu, D., Li, Z., Han, X., Liu, H., Dong, C., et al. (2021b). Power Prediction of a Wind Farm Cluster Based on Spatiotemporal Correlations. *Appl. Energy* 302, 117568. doi:10.1016/j.apenergy.2021.117568

Zhang, Q., Tang, Z. H., Wang, G., Yang, Y., and Tong, Y. (2021a). Ultra-short Term Wind Power Prediction Model Based on Long Short-Term Memory Network. *Acta Energiae Solaris Sin.* 42, 275–281. doi:10.19912/j.0254-0096.tynxb.2019-1193

Zhao, B., Wang, Z. P., Ji, W. J., Gao, X., and Li, X. B. (2019). CNN-GRU Short-Term Power Load Forecasting Method Based on Attention Mechanism. *Power Syst. Technol.* 43, 4370–4376. doi:10.13335/j.1000-3673.pst.2019.1524

Zhao, X., Wang, S. X., and Liu, R. J. (2017). Research on Wind Power Combination Prediction Based on Grey Correlation and Cointegration Theory. *Acta Energiae Solaris Sin.* 38, 1299–1306.

Zheng, D., Semero, Y. K., Zhang, J., and Wei, D. (2018). Short-term Wind Power Prediction in Microgrids Using a Hybrid Approach Integrating Genetic Algorithm, Particle Swarm Optimization, and Adaptive Neuro-Fuzzy Inference Systems. *IEEJ Trans. Elec Electron Eng.* 13, 1561–1567. doi:10.1002/tee.22720

Zhong, W. Z., Li, C. G., Cui, Y., Li, F., and Wang, D. D. (2021). Ultra-short Term Wind Power Combination Forecasting Considering Historical Similarity Weighting. *Acta Energiae Solaris Sin.* 1, 9. doi:10.19912/j.0254-0096.tynxb.2021-0308

Zhu, Q. M., Li, H. Y., Wang, Z. Q., Chen, J. F., and Wang, B. (2017). Ultra-short-term Prediction of Wind Farm Power Generation Based on Long Short-Term Memory Networks. *Power Syst. Technol.* 41, 3797–3802. doi:10.13335/j.1000-3673.pst.2017.1657

Zou, C. N., Xiong, B., Xue, H. Q., Zheng, D. W., Songtao, W., Ge, Z. X., et al. (2021). Position and Role of New Energy in Carbon Neutralization. *Petroleum Explor. Dev.* 48, 411–420. doi:10.1016/s1876-3804(21)60039-3