

Research Article

Attention Mechanism-Based CNN-LSTM Model for Wind Turbine Fault Prediction Using SSN Ontology Annotation

Yuan Xie ¹, Jisheng Zhao,² Baohua Qiang ¹, Luzhong Mi,² Chenghua Tang,¹ and Longge Li¹

¹School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China

²Beijing Huadian Tianren Electric Power Control Technology Co., Ltd., Beijing 100039, China

Correspondence should be addressed to Baohua Qiang; qiangbh@guet.edu.cn

Received 30 December 2020; Revised 22 February 2021; Accepted 10 March 2021; Published 27 March 2021

Academic Editor: Xingsi Xue

Copyright © 2021 Yuan Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The traditional model for wind turbine fault prediction is not sensitive to the time sequence data and cannot mine the deep connection between the time series data, resulting in poor generalization ability of the model. To solve this problem, this paper proposes an attention mechanism-based CNN-LSTM model. The semantic sensor data annotated by SSN ontology is used as input data. Firstly, CNN extracts features to get high-level feature representation from input data. Then, the latent time sequence connection of features in different time periods is learned by LSTM. Finally, the output of LSTM is input into the attention mechanism module to obtain more fault-related target information, which improves the efficiency, accuracy, and generalization ability of the model. In addition, in the data preprocessing stage, the random forest algorithm analyzes the feature correlation degree of the data to get the features of high correlation degree with the wind turbine fault, which further improves the efficiency, accuracy, and generalization ability of the model. The model is validated on the icing fault dataset of No. 21 wind turbine and the yaw dataset of No. 4 wind turbine. The experimental results show that the proposed model has better efficiency, accuracy, and generalization ability than RNN, LSTM, and XGBoost.

1. Introduction

In recent years, with the development of human beings, the exploitation and utilization of petroleum and fossil fuels have promoted the development of various fields. However, due to the over exploitation of these resources, the nonrenewable energy is reduced gradually, which makes the energy crisis and climate change become an urgent problem [1]. Countries of the world begin to turn their attention to renewable energy. Wind energy as a renewable energy has the characteristics of clean and easy to use, which has aroused great concern in the world. At present, although wind energy develops rapidly, how to predict accurately the occurrence of wind turbine fault and reduce effectively the maintenance cost of wind farm are still urgent problems to be solved.

Sensor technology has been applied widely in various fields. A large number of sensors are also installed on the wind turbine equipment to collect the operation data of the wind turbine. Based on the collected mass operation data,

wind turbine state data analysis method is used to predict the fault of wind turbine. This method analyzes the state data of the wind turbine from multiple layers and excavates the potential valuable fault information of data by intelligent algorithm, realizing the fault prediction of the wind turbine. In recent years, there have been some related research results. There are mainly fault prediction models based on traditional machine learning methods. Kusiak and Verma [2] used data mining technology to analyze the bearing overtemperature fault and established a bearing fault prediction model to predict the fault. Although the overtemperature fault can be predicted accurately, the false alarm rate is relatively high, which increases the burden of maintenance personnel. Kusiak and Li [3] layer upon layer processed SCADA data of wind turbine. Based on the processed data, the wind fault prediction model was constructed and predicted successfully for wind turbine fault, but the accuracy of the model needs to be improved. Zhong et al. [4] proposed a rapid data-driven fault diagnostic method, which integrates data preprocessing

and machine learning techniques. In this method, fault features are extracted by using the modified Hilbert-Huang transforms and correlation techniques. Pairwise-coupled sparse Bayesian extreme learning machine is applied to build a real-time multifault diagnostic system. The experimental results show that the method can diagnose quickly the fault, but the generalization ability of the model is poor. Chen and Zhang [5] proposed a fault prediction method of wind turbine blade cracking based on RF-LightGBM algorithm. In this method, firstly, random forest algorithm ranks the importance of features to select important features. Then, the classification model is trained by the data after feature selection and optimized by K -fold cross validation. Finally, the method predicted successfully the wind turbine blade fault. Wang et al. [6] put forward the XGBoost algorithm-based fault prediction model of wind turbine main bearing. Wu et al. [7] presented a fault diagnosis method of wind turbine based on ReliefF and XGBoost algorithm. Hsu et al. [8] proposed a novel fault diagnosis technique for wind turbine gearbox. While statistical process control was applied to fault diagnosis, random forest and decision tree algorithms were employed to construct the predictive models for wind turbine anomalies. Fan and Tang [9] proposed a wind turbine pitch anomaly recognition system based on AdaBoost-SAMME. The proposed models in literature [5–9] have high accuracy, but the generalization ability of these models needs to be improved. Wang et al. [10] proposed a fault diagnosis method of wind turbine gearbox based on Riemannian manifold, which is fast in model training, but the accuracy of the model is not high. Zhang et al. [11] proposed a wind turbine fault diagnosis method based on the Gaussian process meta-model, which has excellent performance. But the model has high dependence on dataset, which leads to poor generalization ability of the model.

With the development of deep learning, there are many models for wind turbine fault prediction based on deep learning. Chen et al. [12] proposed a fault diagnosis method of wind turbine gearbox based on wavelet neural network. This method can predict faults accurately. But it cannot learn deep features in data. Lu et al. [13] proposed a wind turbine planetary gearbox fault diagnosis method based on self-powered wireless sensor and deep learning. This method has high accuracy, but it cannot learn the temporal relationship between data. Kavaz and Barutcu [14] proposed a wind turbine sensor fault detection method based on artificial neural network. This method combines with SCADA system for data acquisition, which makes the model have the advantages of low cost, but the model lacks the ability to process time series data. Chen et al. [15] proposed a model of the relationship between root cause and symptom of wind turbine fault based on Bayesian network to predict wind turbine fault. However, the complexity of the model increases exponentially with the increase of parent nodes of Venn diagram, which makes the training time too long. Shi et al. [16] proposed an intelligent fault diagnosis method of wind turbine bearing based on convolution neural network. This method can extract features effectively, but the accuracy of the method needs to be improved. Zhang et al. [17] proposed a fault diagnosis method of wind turbine gearbox based on 1DCNN-PSO-SVM. This method has a good performance

in feature extraction, but it lacks the ability to learn time series features. In order to solve the problems of low efficiency and poor accuracy of manual detection method for wind turbine blade defect, Zhang and Wen [18] proposed an improved Mask R-CNN detection method for wind turbine blade defect. Firstly, ResNet-50 combined with FPN is used to generate feature map. Then, it is input into RPN to screen out the ROI. Finally, the size of the feature map is determined by ROIAlign, which is input into the full connection layer network for blade defect detection. The experimental results show that this method has fast detection efficiency and high accuracy. However, this method only considers the characteristics of a single data graph and does not mine the temporal relationship between data. Chang et al. [19] proposed a concurrent convolution neural network for fault diagnosis. The method has high accuracy and strong generalization ability. However, the network structure parameters and calculational speed of this method need to be further optimized. Jiang et al. [20] proposed a fault diagnosis model of wind turbine gearbox based on multiscale convolution neural network. The proposed MSCNN incorporates multiscale learning into the traditional CNN architecture. It improves greatly the feature learning ability and diagnosis performance of the model. But the model cannot learn the long-term correlation between the data. Yin et al. [21] proposed a temperature fault early warning method for wind turbine main bearing based on Bi-RNN, but there are problems of gradient disappearance and gradient explosion for processing time series data. Yin et al. [22] proposed a fault diagnosis method of wind turbine gearbox based on the optimized LSTM neural network with cosine loss. However, this method cannot learn features directly from the original vibration sign. Zheng et al. [23] put forward the fault prediction method for wind turbine gearbox based on K -means clustering and LSTM. Although it can process effectively time series data, it inputs directly the original features of the wind turbine into the model for training, resulting in model training too long. On this basis, a fault prediction method for wind turbine based on deep trust network was proposed, which has a good effect on feature extraction of wind turbine data, but it has the problem of big error [24]. Lei et al. [25] presented an end-to-end LSTM model for fault prediction. The model uses the data fusion strategy of IDENTITY function to extract multisensor features. LSTM captures long-term dependencies through recurrent behaviour and gates mechanism. The experimental results show that this method is able to do fault classification effectively from raw time series signals collected by single or multiple sensors. However, the strategy used in the data fusion stage of this method needs to be improved.

In accordance with shortcomings of traditional machine learning in processing time series data and advantages of CNN [26] in feature extraction and LSTM [27] in processing time series data, an attention mechanism-based CNN-LSTM model for wind turbine fault prediction is constructed. The contributions of this method are summarized as follows:

- (i) This method proposes a joint training mode of CNN and LSTM, which can process effectively the time series data and ensure the training speed. In addition,

considering that the irrelevant features of data will lead to the degradation of the model performance, therefore, attention mechanism [28] is used to redistribute feature weights to ensure the performance of the model and improve the generalization ability of the model

- (ii) In order to further eliminate the influence of irrelevant features and improve the construction speed of the model, random forest algorithm [29] is used to select features in the data preprocessing stage, which further improves the generalization ability of the model

The rest of this paper is organized as follows. In Section 2, the data acquisition process of wind turbine is introduced. Section 3 introduces the model of this paper. Section 4 presents the experimental procedure and experimental results. Section 5 is the conclusion including summary of the method and future work.

2. Data Acquisition

In view of the isomerization of the data transmitted by the sensor in the form and description information, in this paper, SSN ontology annotation method is used to format and unify sensor data of wind turbine. The specific process of ontology annotation method is shown in Figure 1.

Firstly, the method needs to analyze the sensor data of wind turbine to find its concept and properties. Secondly, it is necessary to analyze the structure of SSN ontology and compare sensor data with SSN ontology structure to extract useful information. Based on these, the annotation information of sensor data can be obtained by the semiautomatic annotation mapping method. Then, the annotation information of data is generated into a mapping file in XML format by R2RML mode. The mapping file is used mainly to store the annotation information of wind turbine sensor data. Then, the mapping file of wind turbine sensor data is transformed into ontology instance to generate RDF semantic sensor data by the semantic conversion algorithm of sensor data. Finally, the RDF semantic sensor data is stored in HBase by Spark Streaming.

Based on the ontology semantic annotation method of SSN, SCADA system is used to collect data. The icing fault dataset and yaw fault dataset of a wind farm in Yunnan Province are collected. According to the SSN ontology structure diagram as shown in Figure 2, the corresponding wind turbine features are generated. Among them, the icing fault data involves 26 features, as shown in Table 1; the yaw fault data involves 28 features, as shown in Table 2.

3. Introduction to the Model

3.1. Model Structure. The structure of the attention mechanism-based CNN-LSTM wind turbine fault prediction model is shown in Figure 3. The model can be divided into three parts. In the first part, the processed data are input into CNN and the high-level feature representation is obtained by feature extraction of CNN. In the second part, the output of

CNN is input into LSTM; the deeper temporal relationship between features is obtained by LSTM. In the third part, attention mechanism is introduced to acquire more related information of fault. This model is also called as the CLA model.

For a given training dataset $D = \{(x_i, y_i) \mid i = 1, \dots, p\}$, x_i represents the input sample data, $x_i \in R^q$, where q is the number of features, and $y_i \in \{0, 1\}$ is the label of the i -th sample data. The sample data is input into CNN, and the high-level feature representation is obtained by extracting the original features:

$$J_i = f(\omega \times x_{i:i+g-1} + b), \quad (1)$$

where ω is the convolution kernel, g is the size of convolution kernel, $i : i + g - 1$ is the i -th feature to the $(i + g - 1)$ -th feature, and b is the offset term. After the calculation of convolution layer, the characteristic matrix J is acquired:

$$J = [c_1, c_2, \dots, c_{n-g+1}]. \quad (2)$$

Then, the maximum pooling technique [30] is used to process the local characteristic matrix J of the fault to retain the key information of the features and reduce the parameters; finally, the local optimal solution is as follows:

$$M = \max(c_1, c_2, \dots, c_{n-g+1}) = \max\{J\}. \quad (3)$$

Then, the M vector is connected to form the H vector by the full connection layer:

$$H = \{M_1, M_2, \dots, M_n\}. \quad (4)$$

The output H of convolution network is taken as the input of LSTM. After receiving the output of CNN, the forgetting gate, memory gate, and output gate are calculated by the hidden state h_{t-1} of the last time and the current input x_t to form a memory unit, which runs through all processes. It can retain important information and remove unimportant information; the specific process is as follows:

- (1) Firstly, the data passes through the forgetting gate. The sigmoid unit in the forgetting gate generates a vector between 0 and 1 by calculating h_{t-1} and x_t . According to this vector, LSTM can determine what information needs to be retained and what information needs to be discarded in memory unit C

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

where f_t is the forgetting gate, σ is the activation function, and h_{t-1} and x_t are the inputs.

- (2) The next step is to determine which part of the new information is added to the memory unit, which involves two operations. Firstly, input gate is obtained by calculating h_{t-1} and x_t ; input gate can determine which part of the information needs to be updated. Then, let h_{t-1} and x_t pass through a

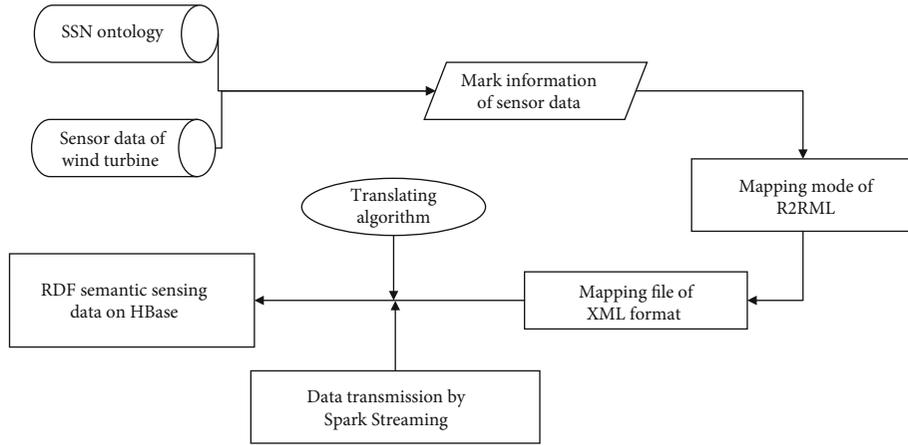


FIGURE 1: The process of SSN ontology annotation.

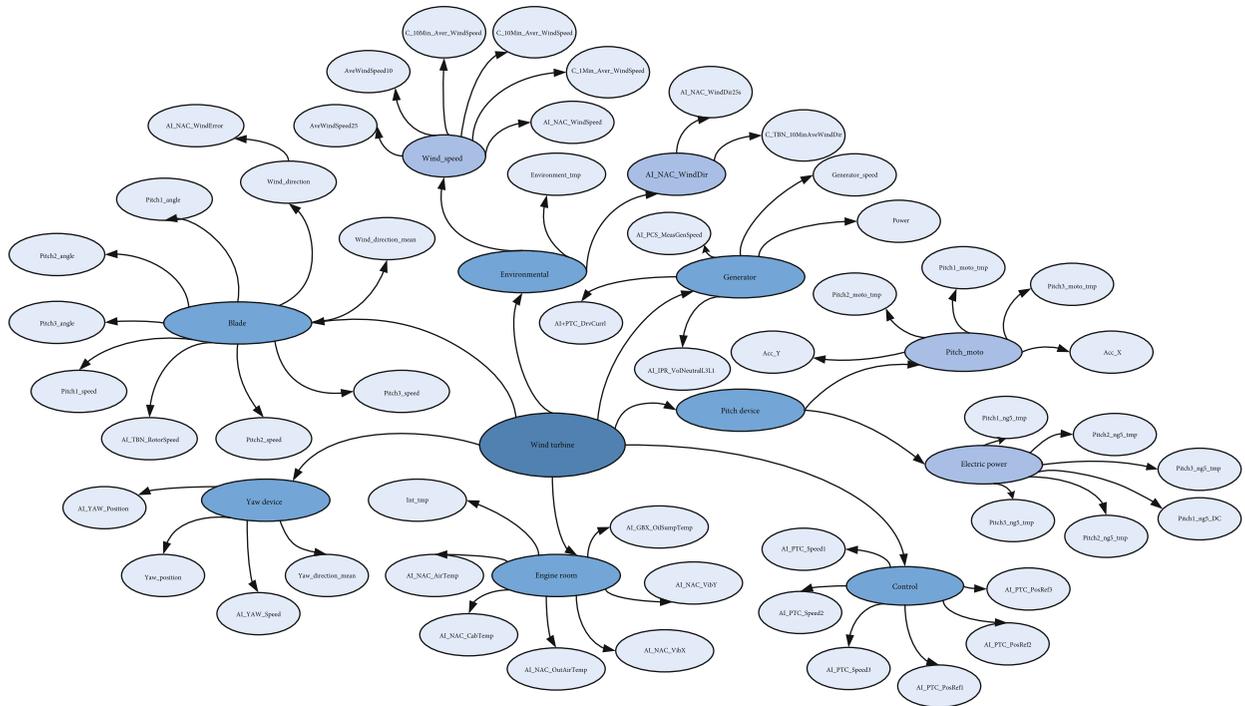


FIGURE 2: SSN ontology structure diagram.

TABLE 1: Original characteristics of wind turbine icing fault.

Features	Features	Features	Features
wind_speed	pitch1_angle	pitch2_moto_tmp	pitch2_ng5_tmp
generator_speed	pitch2_angle	pitch3_moto_tmp	pitch3_ng5_tmp
power	pitch3_angle	acc_X	pitch1_ng5_DC
wind_direction	pitch1_speed	acc_Y	pitch1_ng5_DC
wind_direction_mean	pitch2_speed	environment_tmp	pitch1_ng5_DC
yaw_position	pitch3_speed	int_tmp	
yaw_speed	pitch1_moto_tmp	pitch1_ng5_tmp	

TABLE 2: Original characteristics of wind turbine yaw fault.

Features	Features	Features	Features
aveWindSpeed25	AI_NAC_WindDir25s	AI_TBN_RotorSpeed	AI_PTC_PosRef1
aveWindSpeed10	C_TBN_10MinAveWindDir	AI_NAC_Position	AI_PTC_PosRef2
AI_NAC_WindSpeed	AI_GBX_OilSumpTemp	AI_YAW_Speed	AI_PTC_PosRef3
C_10Min_Aver_WindSpeed	AI_NAC_WindError	AI_PTC_Speed1	AI_NAC_VibX
C_15Min_Aver_WindSpeed	AI_NAC_AirTemp	AI_PTC_Speed2	AI_NAC_VibY
C_1Min_Aver_WindSpeed	AI_NAC_CabTemp	AI_PTC_Speed3	AI_PTC_DrvCurr1
AI_NAC_WindDir	AI_NAC_OutAirTemp	AI_PCS_MeasGenSpeed	AI_IPR_VoltNeutralL3L1

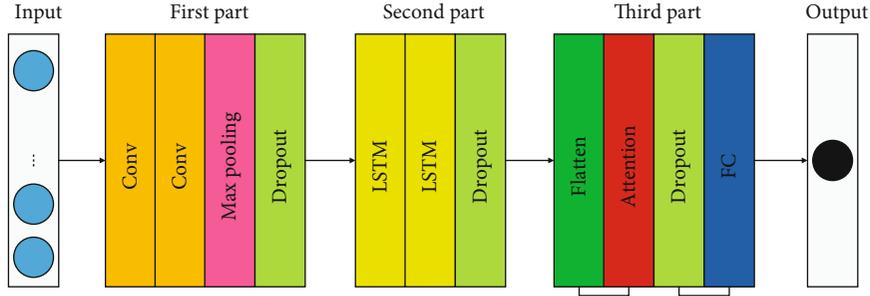


FIGURE 3: Structure of the CLA model.

tanh layer to get new candidate memory units \tilde{C}_t ; this information of candidate memory may be updated into the memory unit

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \end{aligned} \quad (6)$$

where i_t is the input gate, \tanh is the activation function, and \tilde{C}_t is the candidate memory unit.

- (3) LSTM will update memory unit; memory unit C_{t-1} will be updated to memory unit C_t . The process of updating is as follows: firstly, forget some information of old memory unit through the forgetting gate. Secondly, increase part of the information of candidate memory unit \tilde{C}_t through input gate. Finally, get a new memory unit C_t

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \quad (7)$$

where C_t is the new memory unit.

- (4) Finally, the LSTM determines which state characteristics need to be output through the output gate; let the input h_{t-1} and x_t pass through the sigmoid layer of the output gate to get a judgment condition, and then let the memory unit pass through the tanh layer to get a vector between -1 and 1, which is multiplied by the judgment condition of the output gate to get the final output of the memory unit

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (8)$$

$$h_t = o_t * \tanh(C_t), \quad (9)$$

where O_t is the output gate and h_t is the final output.

In the appeal of Equations (1) to (9), W_f, W_i, W_c, W_o is the weight matrix and b_f, b_i, b_c, b_o is the offset vector.

- (5) The data is processed by LSTM to get the output $H_t = [h_1, \dots, h_t]$ of each time step; then, input H_t into the attention module. Through the attention mechanism, different weights can be assigned to the fault characteristics of wind turbine. The formulas are as follows:

$$w_t = \tanh(H_t), \quad (10)$$

where h_t is the input and w_t is the target weight.

Then, the softmax function is used to probabilistically affect the attention weight:

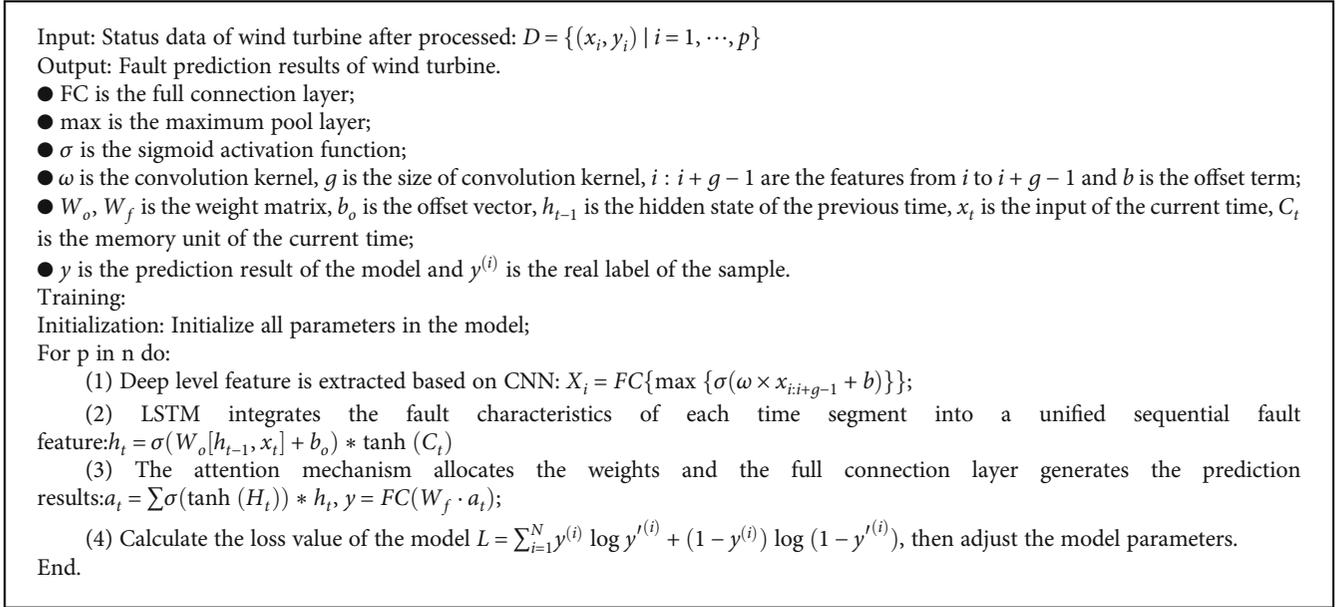
$$P_t = \frac{\exp(w_t)}{\sum_{t=1}^m \exp(w_t)}, \quad (11)$$

where P_t is the weight probability vector.

The generated attention weight is assigned to the corresponding hidden layer state code h_t :

$$a_t = \sum_{t=1}^m P_t \cdot h_t, \quad (12)$$

where a_t is the weighted average of h_t and P_t is the weight.



ALGORITHM 1: CLA model algorithm.

Finally, the results are output through the full connection layer:

$$y_t = \sigma(W_{\text{final}} \cdot a_t), \quad (13)$$

where y_t is the predication results, W_{final} is the weight matrix, and σ is the sigmoid activation function.

3.2. Model Training and Testing. Algorithm 1 shows pseudocode of CLA model algorithm. Firstly, all the trainable parameters in the model are initialized. Secondly, the training dataset $D = \{(x_i, y_i) \mid i = 1, \dots, p\}$ is put into CNN through the input layer of the model. CNN uses convolution layer and pool layer to get the deep feature matrix $X = [X_1, \dots, X_i]$ of data. Then, the output of CNN is used as the input of LSTM. LSTM can further learn the feature of wind turbine data at multiple continuous times to obtain the time series dependence between features. The output of each time step for LSTM recorded as $H_t = [h_1, \dots, h_t]$ is input to the attention mechanism module to redistribute the feature weight. Finally, according to the weighted features a_t , the full connection layer generated the prediction results y . According to the error between the predicted results and the real values, the model updates the trainable parameters. The model judges whether the training times reach the upper limit of the maximum number of iterations. If not, the number of iterations is increased by 1 and the training process is repeated. Otherwise, the trained model can be obtained.

The training and testing flow of the CLA model is shown in Figure 4. Firstly, the processed data are divided into training dataset and test dataset. Secondly, the training dataset is input into the constructed model and the features of data are extracted by convolution neural network. Then, the results are input into LSTM for time series feature transformation; the fault characteristics of each time segment are integrated into a unified sequence fault features; besides,

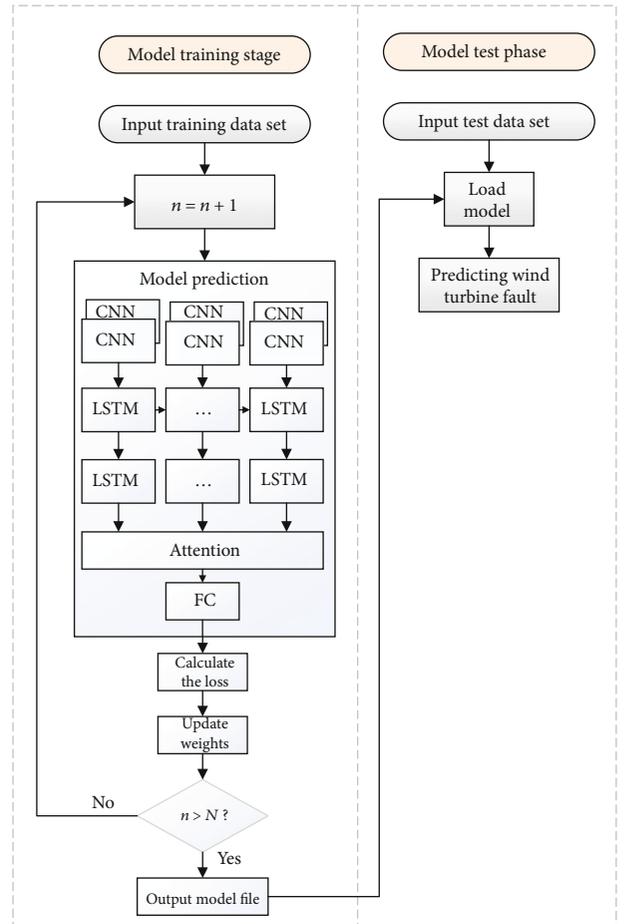


FIGURE 4: Process of the CLA model training and testing.

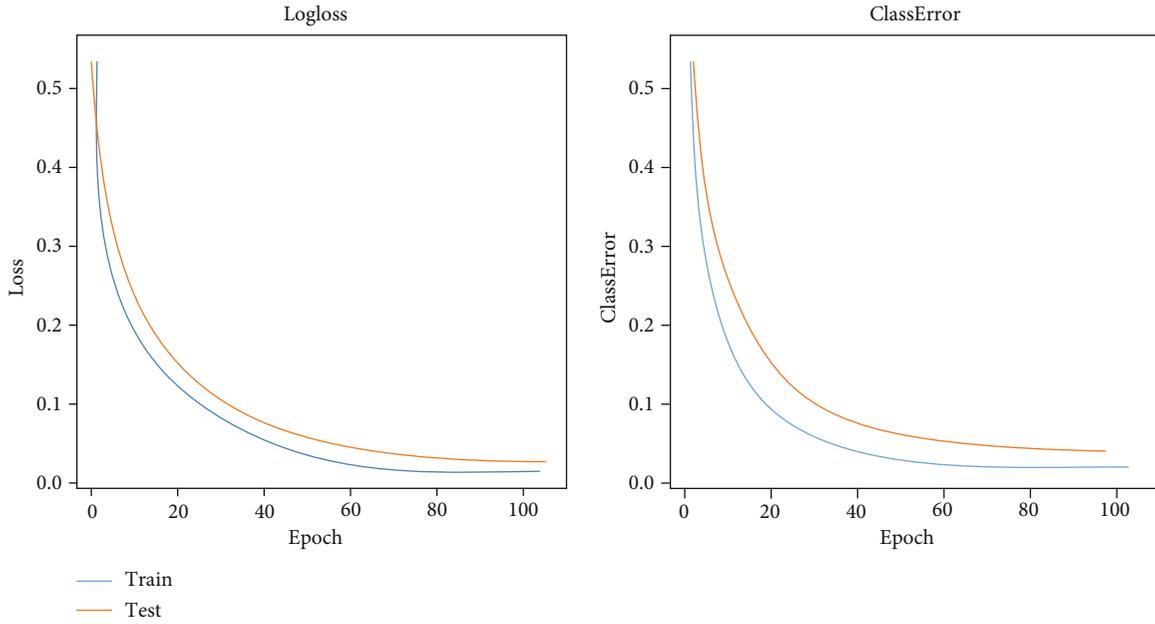


FIGURE 5: Logloss and ClassError of the CLA model on ice fault dataset of No. 15 wind turbine.

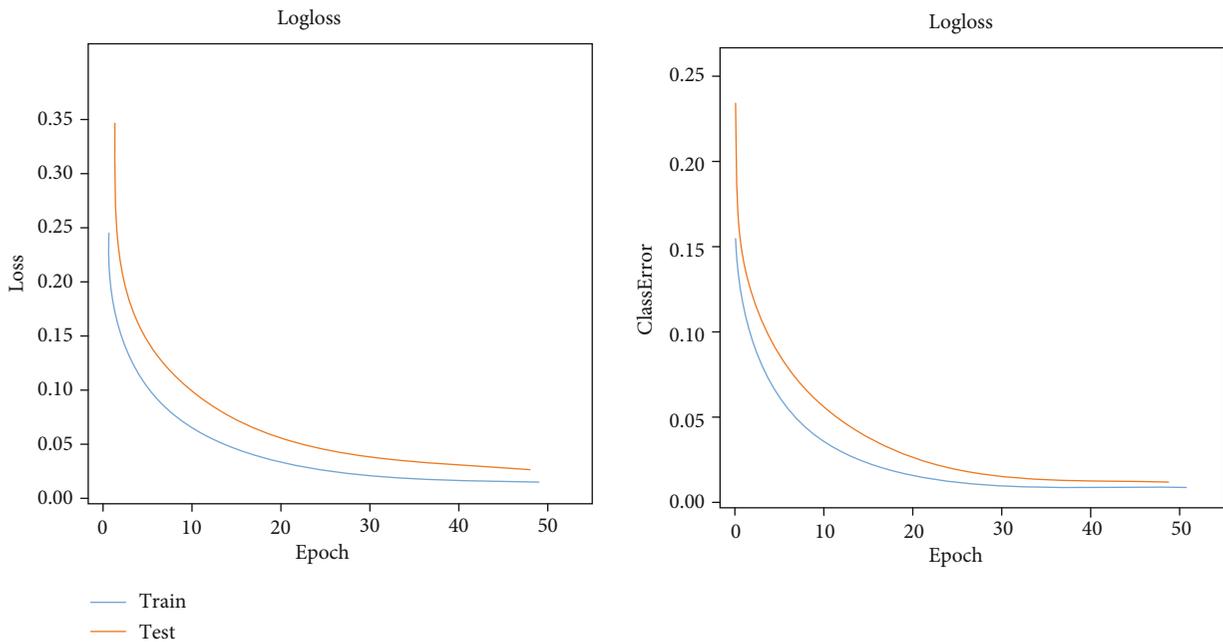


FIGURE 6: Logloss and ClassError of the CLA model on yaw fault dataset of No. 3 wind turbine.

attention mechanism is introduced in the model. Finally, the prediction results are output through the full connection layer and the model parameters are constantly updated according to the loss value. In this experiment, the number of iterations is used as the end condition of training; when the model reaches the preset number of iterations, the fault prediction model is obtained. In the test phase, firstly, the trained model is loaded. Then, the test dataset is input into the model for fault prediction. Finally, the experimental results are obtained.

The CLA model integrates CNN, LSTM, and attention mechanism. Among them, CNN can reduce model parameters by sharing convolution kernel parameters; thus, it can accelerate the calculation speed of the model. Through memory unit, LSTM can deal with the long time sequence dependence of data and solve the problem of gradient disappearance caused by too long time step. Attention mechanism can make the model pay more attention to the features that have a high correlation with the fault. Besides, attention mechanism also can reduce the influence of nonimportant

TABLE 3: Experimental results of the model on the test datasets of No. 15 and No. 3 wind turbine.

Algorithm	Wind turbine icing fault test dataset No. 15 wind turbine				Wind turbine yaw fault test dataset No. 3 wind turbine			
	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>
CLA	0.9646	0.9730	0.9634	0.9712	0.9821	0.9805	0.9754	0.9819
LSTM	0.9388	0.9227	0.9580	0.9400	0.9793	0.9803	0.9781	0.9792
RNN	0.7986	0.8083	0.7833	0.7956	0.9672	0.9609	0.9738	0.9673
XGBoost	0.9805	0.9875	0.9733	0.9804	0.9855	0.9906	0.9852	0.9873

TABLE 4: Experimental results of the model on the dataset of No. 21 and No. 4 wind turbine.

Algorithm	Wind turbine icing fault dataset No. 21 wind turbine				Wind turbine yaw fault dataset No. 4 wind turbine			
	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>
CLA	0.7492	0.8171	0.7421	0.7591	0.7709	0.8143	0.7429	0.7737
LSTM	0.7242	0.7279	0.7163	0.7220	0.7455	0.8000	0.6546	0.7201
RNN	0.6793	0.6793	0.6792	0.6793	0.7248	0.7637	0.6511	0.7029
XGBoost	0.7038	0.7120	0.6847	0.6981	0.7038	0.7120	0.6847	0.6981

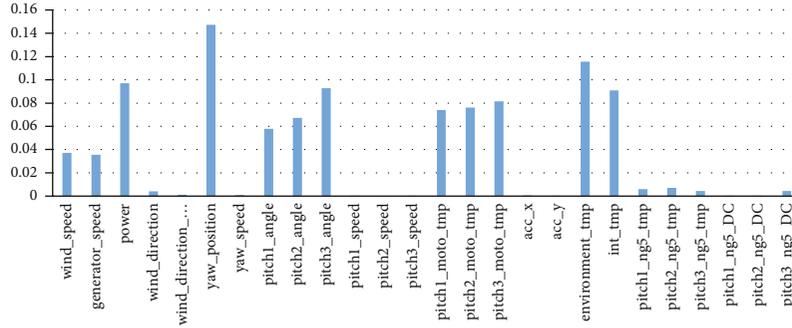


FIGURE 7: Importance of icing fault characteristic.

features, so as to improve the generalization ability and accuracy of the model.

4. Experiment and Analysis

The construction of the CLA model is based on Keras deep learning framework. In the experiment of wind turbine icing fault prediction, a two-layer convolution neural network and a two-layer LSTM neural network are used, where the number of neurons in each layer of convolution neural network and LSTM neural network is set to 64. In the experiment of wind turbine yaw fault prediction, a two-layer convolution neural network and a two-layer LSTM neural network are also used, where the number of neurons in each layer of convolution neural network is set to 32 and the number of neurons in each layer of LSTM neural network is set to 16. In order to prevent overfitting, dropout is introduced into the model and the value of dropout is set to 0.5.

4.1. Selection of Evaluation Indexes. In the CLA model, sigmoid is used as the activation function and binary_crossentropy is

TABLE 5: Characteristics of wind turbine icing fault after screening.

Features	Score
yaw_position	0.147249
environment_tmp	0.115538
power	0.097059
pitch3_angle	0.092650
int_tmp	0.090929
pitch3_moto_tmp	0.081600
pitch2_moto_tmp	0.076074
pitch1_moto_tmp	0.073881
pitch2_angle	0.067191
pitch1_angle	0.057764
wind_speed	0.037283
generator_speed	0.035642

used as the loss function. We select the commonly used evaluation indicators in the field of fault prediction, including accuracy (*A*), precision (*P*), recall (*R*), and *F1* value (*F1*).

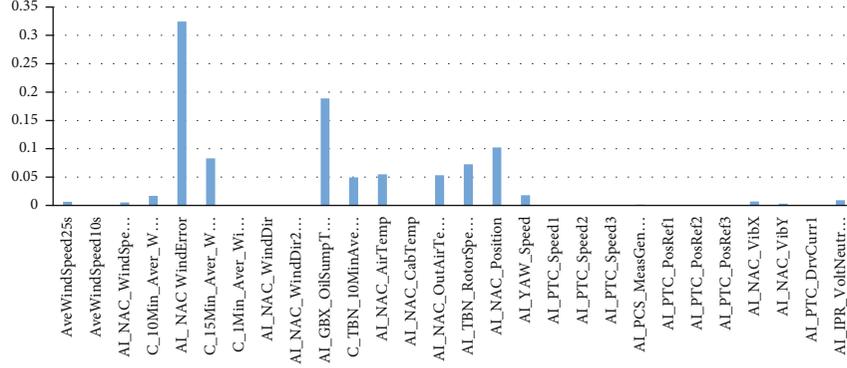


FIGURE 8: Importance of yaw fault characteristics.

Accuracy refers to the ratio between the number of samples classified correctly by the model and the total number of samples for a given test dataset:

$$A = \frac{TP + TN}{TP + FP + TN + FN}. \quad (14)$$

Precision refers to how many of the samples predicted as positive by the model are real positive samples:

$$P = \frac{TP}{TP + FP}. \quad (15)$$

Recall rate indicates how many positive samples in the dataset are predicted correctly:

$$R = \frac{TP}{TP + FN}. \quad (16)$$

$F1$ value represents the combined value of precision and recall rate. Because P and R are a pair of contradictory variables, it is difficult to simultaneously improve them in the experimental process. So, it is necessary to comprehensively evaluate them. Therefore, $F1$ value is proposed to reconcile them; $F1$ value is the comprehensive average of P and R . The closer the precision rate and the recall rate are, the greater the $F1$ is:

$$F1 = 2 \frac{P \cdot R}{P + R}. \quad (17)$$

In Equations (14)–(17), TP is the number of samples that predict the wind turbine samples in normal state as normal state, FP is the number of samples that predict the wind turbine samples in fault state as normal state, FN is the number of samples that predict the wind turbine in normal state as failure state, and TN is the number of samples that predict the wind turbine in fault state as failure state.

4.2. Experimental Scheme. The hardware environment of the experiment is Intel Xeon e5-2698v4 (20 Core) processor and 128 G of memory, the operating system is 64 bit Ubuntu 16.04, Python 3.6 is used as software programming language, and PyCharm 2017.1.2 is used as a software development tool. The icing fault dataset of No. 21 and No. 15 wind tur-

TABLE 6: Characteristics of wind turbine yaw fault after screening.

Features	Score
AI_NAC_WindError	0.324536
AI_GBX_OilSumpTemp	0.188854
AI_NAC_Position	0.102554
C_15Min_Aver_WindSpeed	0.082901
AI_TBN_RotorSpeed	0.072518
AI_NAC_AirTemp	0.055106
AI_NAC_OutAirTemp	0.053102
C_TBN_10MinAveWindDir	0.049441
AI_YAW_Speed	0.018063
C_10Min_Aver_WindSpeed	0.016823

bine and the yaw fault dataset of No. 4 and No. 3 wind turbine are used in the experiment.

In order to verify the effectiveness of the proposed wind turbine fault prediction model in dealing with time series data and generalization problems, two groups of experiments are carried out, using RNN [21], LSTM [25], and XGBoost [6] algorithm as a comparison.

The first group of experiments, the data of No. 15 wind turbine and No. 3 wind turbine, is, respectively, divided into training dataset and corresponding test dataset. Two groups of training datasets are used directly to train the CLA model and other reference models; then, the corresponding test datasets are used to verify the model.

The second group of experiments, the two CLA models and other reference models trained in the first group, was tested on the data of No. 21 wind turbine and No. 4 wind turbine; the experimental results were compared and analyzed.

4.3. Experimental Results and Analysis. As shown in Figures 5 and 6, Logloss and ClassError of the CLA model on the dataset of Experiment 1 are shown, respectively. It can be found from the graph that Logloss and ClassError decrease continuously during the process of model iteration, which indicates that the model is converging with the increase of iteration times.

On the two datasets of Experiment 1, the results of the CLA model and other comparison algorithm models are shown in Table 3. On the test dataset of No. 15 wind turbine

TABLE 7: Experimental results after feature screening.

Algorithm	Wind turbine icing fault dataset				Wind turbine yaw fault dataset			
	No. 21 wind turbine				No. 4 wind turbine			
	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>	<i>A</i>	<i>P</i>	<i>R</i>	<i>F1</i>
CLA	0.7776	0.7899	0.7565	0.7728	0.8622	0.8228	0.9234	0.8702
LSTM	0.7664	0.7310	0.7431	0.7531	0.8401	0.8025	0.9023	0.8495
RNN	0.6855	0.6571	0.7756	0.7115	0.7751	0.7912	0.7475	0.7687
XGBoost	0.7345	0.7286	0.7571	0.7426	0.7345	0.7286	0.7571	0.7426

icing fault, the accuracy rate of the CLA model is 96.46%. Compared with the LSTM and RNN algorithm models, the accuracy rate of the CLA model is improved, respectively, by 2.58% and 16.60% and that of the other three indicators *P*, *R*, and *F1* of the CLA model is also the best. Besides, the four indicators of the CLA model are all above 96%. But compared with the XGBoost algorithm model, the accuracy of the CLA model is reduced by 1.59%. On the test dataset of No. 3 wind turbine yaw fault, the accuracy of the CLA model is 98.21%. Compared with the LSTM and RNN algorithm models, the accuracy rate of the CLA model is improved, respectively, by 0.28% and 1.49% and that of the other three indicators *P*, *R*, and *F1* of the CLA model is also the best. Besides, the four indicators are all above 96%. But the accuracy of the CLA model is reduced by 0.34% compared with the XGBoost algorithm model.

In Experiment 2, the trained model in Experiment 1 is applied, respectively, to the dataset of No. 21 wind turbine and No. 4 wind turbine; the experimental results are shown in Table 4. On the ice fault dataset of No. 21 wind turbine, the accuracy rate of the CLA model is 74.92%, which is 2.50%, 6.99%, and 4.54%, respectively, higher than the LSTM, RNN, and XGBoost algorithm models, and that of the other three indicators *P*, *R*, and *F1* of the CLA model is also the best. On the yaw fault dataset of No. 4 wind turbine, the accuracy rate of the CLA model is 77.09%, which is 2.54%, 4.61%, and 6.71%, respectively, higher than the LSTM, RNN, and XGBoost algorithm models, and that of the other three indicators *P*, *R*, and *F1* of the CLA model is also the best.

According to the experimental results of Experiment 1 and Experiment 2, the CLA model has the best performance in *A*, *P*, *R*, and *F1* compared with the LSTM and RNN algorithm models; the results show that the temporal relationship of fault features can be learned to improve the accuracy and generalization ability of the model by fusion of CNN and LSTM and introducing attention mechanism. It proves the effectiveness of the CLA model. However, on the test dataset of No. 15 and No. 3 wind turbine, the accuracy rate of the CLA model is reduced, respectively, by 1.59% and 0.34% compared with the XGBoost algorithm model. Because the CLA model needs more data for deep learning compared with XGBoost, due to the lack of data, the model is over fitted, which makes the performance of the CLA model worse than the XGBoost algorithm model on the test dataset of No. 15 and No. 3 wind turbine, but the CLA model is more able to mine the deep relationship of time series data, so the generalization ability of the CLA model is better than XGBoost.

5. Optimization of CLA Model

In order to further improve the generalization ability of the CLA model, the original features of the icing fault dataset of No. 15 and No. 21 wind turbine are input into the random forest algorithm; the original features are screened through many experiments. Finally, the histogram of feature importance is obtained as shown in Figure 7. Based on this histogram, 12 features of the highest correlation degree with the icing fault of the wind turbines are selected, as shown in Table 5.

In the same way, random forest algorithm is used in the yaw fault dataset of No. 3 and No. 4 wind turbine; the feature importance histogram is obtained as shown in Figure 8. Based on this histogram, 10 features of the highest correlation degree with the yaw fault of the wind turbines are selected, as shown in Table 6.

Two feature datasets of No. 15 and No. 3 wind turbine screened by random forest algorithm were input into the CLA model for training; the trained model is used to test the feature dataset of No. 21 and No. 4 wind turbine screened by random forest algorithm for verifying the screening effect. The experimental results are shown in Table 7. For the icing fault dataset of No. 21 wind turbine after screening, the CLA model is still the best and the accuracy rate is improved by 2.84%. For the yaw fault dataset of No. 4 wind turbine after screening, the CLA model is still the best and the accuracy rate is improved by 9.13%. It shows that feature screening by random forest algorithm can improve effectively the generalization ability of the model.

6. Conclusions

In this paper, an attention mechanism-based CNN-LSTM model for wind turbine fault prediction is proposed. The model is trained by the icing fault dataset and yaw fault dataset of wind turbine annotated by SSN ontology. The trained model can predict accurately the occurrence of wind turbine fault. The experimental results show that the model is more effective and better than some of the current mainstream models. In this method, the model can learn effectively the deep-seated temporal characteristics among samples and focus on the features of high correlation with wind turbine fault through the joint training of CNN-LSTM and the introduction of attention mechanism. These strategies improve successfully the accuracy and generalization ability of the model. In addition, the generalization ability of the model can be further improved by using the random forest algorithm in the data preprocessing stage. In the later stage,

we will continue to study how to improve the model so that the prediction accuracy and generalization ability of the model can be further improved. Besides, we will also research how to introduce more excellent machine learning algorithms into the model for adapting to the complex working environment of wind turbine; it means that the model will be applied truly to practical engineering.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61762025 and 62062028), the Guangxi Key Research and Development Program (AB18126053, AB18126063, and AD18281002), the Natural Science Foundation of Guangxi of China (2017GXNSF AA198226, 2019GXNSFDA185007, 2019GXNSFDA185006, and 2018GXNSFAA294058), the Guangxi Key Science and Technology Planning Project (Nos. AA18118031 and AA18242028), the National Energy Technology Environmental Protection Group Co., Ltd (IKY.2019.0002), the Innovation Project of GUET Graduate Education (2019YCXS051 and 2020YCXS052), and the Guangxi Colleges and Universities Key Laboratory of Intelligent Processing of Computer Image and Graphics (No. GIIP201603).

References

- [1] M. Bhattacharya, S. R. Paramati, I. Ozturk, and S. Bhattacharya, "The effect of renewable energy consumption on economic growth: evidence from top 38 countries," *Applied Energy*, vol. 162, pp. 733–741, 2016.
- [2] A. Kusiak and A. Verma, "Analyzing bearing faults in wind turbines: a data-mining approach," *Renewable Energy*, vol. 48, pp. 110–116, 2012.
- [3] A. Kusiak and W. Y. Li, "The prediction and diagnosis of wind turbine faults," *Renewable Energy*, vol. 36, no. 1, pp. 16–23, 2011.
- [4] J. H. Zhong, J. Zhang, J. J. Y. Liang, and H. Q. Wang, "Multi-fault rapid diagnosis for wind turbine gearbox using sparse Bayesian extreme learning machine," *IEEE Access*, vol. 7, pp. 773–781, 2019.
- [5] W. G. Chen and H. L. Zhang, "Application of RF-LightGBM algorithm in early warning of wind turbine blade cracking," *Electronic Measurement Technology*, vol. 43, no. 1, pp. 162–168, 2020.
- [6] G. L. Wang, H. S. Zhao, and Z. Q. Mi, "Application of XGBoost algorithm in prediction of wind motor main bearing fault," *Electric Power Automation Equipment*, vol. 39, no. 1, pp. 73–77, 2019.
- [7] Z. D. Wu, X. L. Wang, and B. C. Jiang, "Fault diagnosis for wind turbines based on ReliefF and eXtreme gradient boosting," *Applied Sciences*, vol. 10, no. 9, 2020.
- [8] J. Y. Hsu, Y. F. Wang, K. C. Lin, M. Y. Chen, and J. H. Y. Hsu, "Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning," *IEEE Access*, vol. 8, pp. 23427–23439, 2020.
- [9] S. Fan and Q. X. Tang, "Wind turbine pitch anomaly recognition system based on AdaBoost-SAMME," *Power System Protection and Control*, vol. 48, no. 21, pp. 31–40, 2020.
- [10] S. B. Wang, X. G. Sun, and C. W. Li, "Wind turbine gearbox fault diagnosis method based on Riemannian manifold," *Mathematical Problems in Engineering*, Article ID 153656, 2014.
- [11] D. M. Zhang, J. Yuan, J. Zhu, Q. Ji, X. Zhang, and H. Liu, "Fault diagnosis strategy for wind turbine generator based on the Gaussian process metamodel," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4295093, 2020.
- [12] H. T. Chen, S. X. Jing, X. H. Wang, and Z. Y. Wang, "Fault diagnosis of wind turbine gearbox based on wavelet neural network," *Journal of Low Frequency Noise Vibration and Active Control*, vol. 37, no. 4, pp. 977–986, 2018.
- [13] L. Lu, Y. G. He, T. Wang, T. C. Shi, and Y. Ruan, "Wind turbine planetary gearbox fault diagnosis based on self-powered wireless sensor and deep learning approach," *IEEE Access*, vol. 7, pp. 119430–119442, 2019.
- [14] A. G. Kavaz and B. Barutcu, "Fault detection of wind turbine sensors using artificial neural networks," *Journal of Sensors*, vol. 2018, Article ID 5628429, 11 pages, 2018.
- [15] B. D. Chen, P. J. Tavner, Y. H. Feng, W. W. Song, and Y. N. Qiu, *Bayesian Network for Wind Turbine Fault Diagnosis*, 2012, <http://dro.dur.ac.uk/11029/1/11029>.
- [16] G. Y. Shi, J. Xu, and Q. Yang, "Intelligent fault diagnosis on wind turbine bearing based on convolutional neural network," *Journal of North China Electric Power University*, vol. 47, no. 4, pp. 71–79, 2020.
- [17] X. L. Zhang, P. Han, L. Xu, F. Zhang, Y. Wang, and L. Gao, "Research on bearing fault diagnosis of wind turbine gearbox based on 1DCNN-PSO-SVM," *IEEE Access*, vol. 8, pp. 192248–192258, 2020.
- [18] C. Zhang and C. B. Wen, "Fault detection of wind turbine blade based on improved Mask R-CNN," *Renewable Energy Resources*, vol. 38, no. 9, pp. 1181–1186, 2020.
- [19] Y. H. Chang, J. L. Chen, C. Qu, and T. Y. Pan, "Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels," *Renewable Energy*, vol. 153, pp. 205–213, 2020.
- [20] G. Q. Jiang, H. B. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196–3207, 2019.
- [21] S. Yin, G. L. Hou, X. D. Yu, N. Li, Q. Wang, and L. J. Gong, "Research on temperature early warning method of wind turbine main bearing based on Bi-RNN," *Journal of Zhengzhou University*, vol. 40, no. 5, pp. 45–51, 2019.
- [22] A. J. Yin, Y. H. Yan, Z. Y. Zhang, C. Li, and R. V. Sanchez, "Fault diagnosis of wind turbine gearbox based on the optimized LSTM neural network with cosine loss," *Sensors*, vol. 20, no. 8, p. 2339, 2020.
- [23] H. F. Zheng, Y. R. Dai, and Y. Q. Zhou, "Fault prediction of fan gearbox based on K-means clustering and LSTM," *Materials Science and Engineering*, vol. 631, no. 2, pp. 32–43, 2019.
- [24] H. F. Zheng and Y. R. Dai, "Fault prediction of fan gearbox based on deep belief network," *Journal of Physics*, vol. 1449, no. 1, 2020.

- [25] J. H. Lei, C. Liu, and D. X. Jiang, "Fault diagnosis of wind turbine based on long short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.
- [26] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, Computer Science, 2014.
- [27] Q. S. Liu, F. Zhou, R. L. Hang, and X. T. Yuan, "Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification," *Remote Sensing*, vol. 9, no. 12, pp. 1330–1335, 2017.
- [28] J. R. Zhang, F. A. Liu, W. Z. Xu, and H. Yu, "Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism," *Future Internet*, vol. 11, no. 11, pp. 237–249, 2019.
- [29] H. Y. Lv and Q. Feng, "A review of random forests algorithm," *Journal of Hebei Academy of Sciences*, vol. 36, no. 3, pp. 37–41, 2019.
- [30] X. Yao, "Evolving artificial neural networks," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1423–1447, 2002.