Measurement 175 (2021) 109094

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism

Ling Xiang *, Penghe Wang, Xin Yang, Aijun Hu, Hao Su

School of Mechanical Engineering, North China Electric Power University, Baoding 071003, Hebei Province, China

ARTICLE INFO

Keywords:
Wind turbine
Deep learning networks
Fault detection
Supervisory control and data acquisition (SCADA)
Attention mechanism (AM)

ABSTRACT

The complex and changeable working environment of wind turbine often challenges the condition monitoring and fault detection. In this paper, a new method is proposed for fault detection of wind turbine, in which the convolutional neural network (CNN) cascades to the long and short term memory network (LSTM) based on attention mechanism (AM). Supervisory control and data acquisition (SCADA) data are used from wind turbine as input variables and build CNN architecture to extract dynamic changes of data. AM is applied to strengthen the impact of important information. AM can assign different weighs for concentrating the characteristics of LSTM to increase the learning accuracy through mapping weigh and parameter learning. The proposed model can execute early warning for anomaly state and deduce the faulted component by prediction residuals. Finally, through the cases the early failure of the wind turbine is predicted, which verifies the effectiveness of the proposed method.

1. Introduction

Due to its clean and environmentally friendly characteristics, wind energy has been paid the attention by all countries in the world. In recent years, wind energy has shown strong competitiveness in all types of energy, so the installed capacity of wind turbines has been continuously increasing [1]. However, due to the harsh, complex and changeable working environment, the failure rate of such components as main bearing, gearbox and generator of wind turbines is high, resulting in expensive maintenance and operation costs of wind turbines [2–4]. Therefore, the researches on the operation status detection method of wind turbine are conducive to the timely detection of potential faults of wind turbine, the formulation of maintenance plan and the reduction of economic losses [5–6]. Fault detection and diagnosing methods of key components in equipment are studied in a number of Refs. [7–10].

As the basic monitoring of wind turbine, SCADA system collects a mass of variables relevant to the operation characteristics of wind turbine, including of wind speed, power, temperature, current and voltage. If the hidden features between these data can be extracted, the operation status of wind turbine can be identified and early faults can be predicted. Guo et al. [11] analyzed the temperature trend to detect the fault of wind turbine generator. Ferguson et al. [12] presented a method for abnormal detection of wind turbine gearbox based on standardized temperature data. Zhang and Wang [13] used an artificial neural network (ANN) based on SCADA data to find the early faults of the main bearing of wind turbine. Pozo and Vidal [14] established the baseline principal component analysis (PCA) to identify early fault of wind turbine. Pandit and Infield [15] proposed a gaussian processing algorithm to evaluate the operation curve according to the state variable of the wind turbine, which was used as a reference to detect fault of wind turbine. The traditional algorithms have some limitations, such as slow convergence speed and low prediction accuracy when they are applied to process big data. The emergence and development of deep learning network speed up the convergence process and improve the prediction accuracy. The deep learning methods are widely used in fault detection and diagnosis [16–22], and they are verified to be effective for big data of multiple parameters to detect the faults.

About the field of state detection with deep learning methods for machines, Zhao et al. [23] reconstructed errors of the input and output values by using the deep automatic encoder network (DAE) as the detection index of the working state to monitor the operation state of the wind turbine. Wang et al. [24] proposed a blade damage identification method by using deep autoencoders (DA) model based on SCADA data. Teng et al. [25] provided a fault identification method through deep neural networks (DNN) model. In the model, the rotor speed prediction error was used as detection index to detect the failure of permanent magnet shedding of wind turbine generator. Afrasiabi et al. [26] presented a fault identification method for wind turbine based on DNN. In this method, the generative countermeasures network (GAN) was used

* Corresponding author.
E-mail address: ncepuxl@163.com (L. Xiang).

https://doi.org/10.1016/j.measurement.2021.109094
Received 18 May 2020; Received in revised form 19 January 2021; Accepted 24 January 2021
Available online 6 February 2021

Nomenclature

CNN	convolutional neural network
LSTM	long and short term memory network
AM	attention mechanism
SCADA	supervisory control and data acquisition
ANN	artificial neural network
PCA	principal component analysis
DAE	deep automatic encoder network
DA	deep autoencoders
R2	r-square
MAE	mean absolute error
GAN	generative countermeasures network
TCNN	time convolutional neural network
EWMA	exponential weighted moving average method
RELU	rectified linear unit
RMSE	root mean square error
MAPE	mean absolute percent error
DNN	deep neural networks

as the characteristic extraction block, and the time convolutional neural network (TCNN) was used as the fault classifier method. The dependability and precision of the classification model were verified in the wind farm data. A CNN method is applied to learn features from frequency data to recognize the fault of gearbox [27]. Guo et al. [28] provided a hierarchical adaptive deep convolution neural network (CNN) for bearing fault diagnosis. Yu et al. [29] reported the application of LSTM networks for bearing fault diagnosis by experiment based on severe working conditions. Kong et al. [30] propose a novel condition monitoring method of wind turbines based on spatio-temporal features fusion of SCADA data by CNN and GRU, which cascaded the deep learning method. As representatives among them, CNN and LSTM have been applied in the field of status monitoring and failure detection for mechanica equipment. However, the single method often ignores the time or space features of SCADA data, and the CNN -LSTM method can sufficiently extract the features. In this paper, it is focused on the combination of CNN and LSTM for status monitoring and failure detection.

At present, the selection of input variables is paid attention in the process of prediction by most researchers, but the influence of characteristics extracted from input variables to output is ignored. Therefore, a new method is proposed for fault detection of wind turbine by cascading deep learning networks of CNN and LSTM based on attention mechanism (AM). The SCADA data from wind turbine are used as input variables and CNN architecture is built to extract dynamic changes of data. In CNN-LSTM of AM model, A high-dimensional feature is constructed as LSTM input, and the internal dynamic changes of features are learned. AM could assign different weighs to the implied states of LSTM through mapping weigh and parameter learning, and it can strengthen the impact of important information. The exponential weighted moving average method (EWMA) is provided to identify the operation state and predict early faults of wind turbine. The proposed model can execute early warning for anomaly state of wind turbine and deduce the faulted component by prediction residuals.

The rest of the paper is arranged as follows. The basic methods about CNN, LSTM and AM briefly are described in Section 2. Section 3 presents the framework of the proposed model and the Evaluation criteria. Section 4 is designed for data analysis and detection results of two cases. The conclusions based on the performance are reported in Section 5.

2. Methodoly

In the presented method, CNN is combined with LSTM and AM is introduced in networks to enhance useful information. CNN is applied to extract characteristics of state space from wind turbine, and LSTM can better fusion time characteristics of different parts status. AM through the input of the model feature gives different weights, highlights the influence of the more critical factor, helps model make more accurate judgment [31]. Based on CNN-LSTM of AM, the predicted value of the

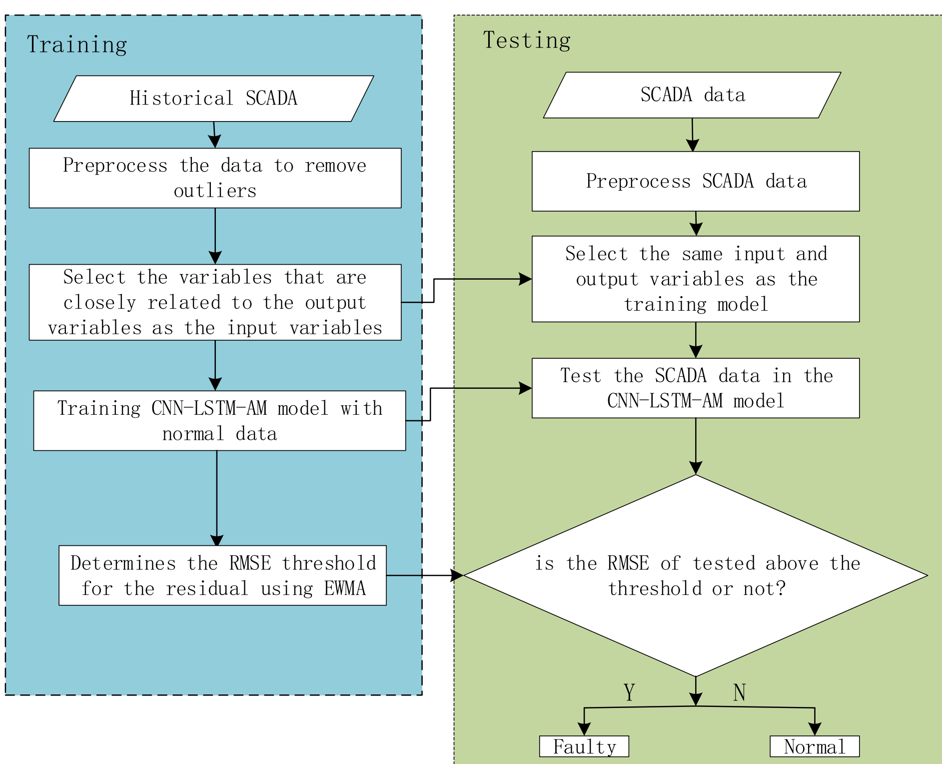

Fig. 1. Flowchart of fault detection for wind turbine.

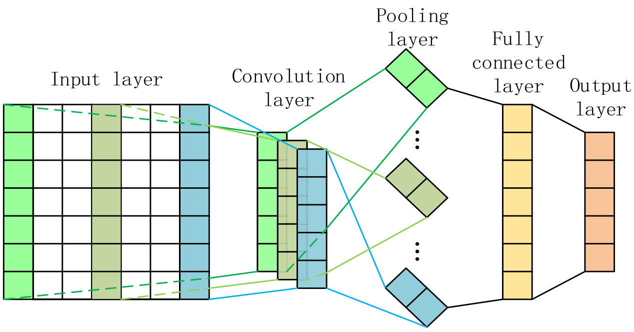

Fig. 2. Structure of CNN.

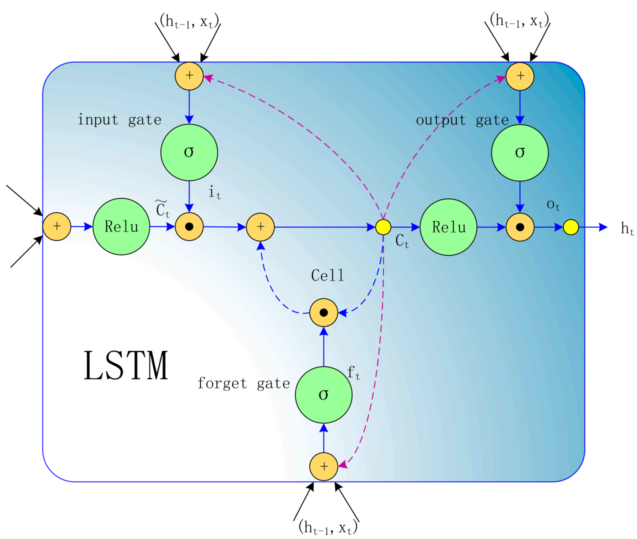

Fig. 3. Schematic diagram of LSTM.

target variable is output, and the residual of the predictive value is compared with that of real value. The flowchart of fault detection is showed as Fig. 1.

2.1. Convolutional neural network (CNN)

The essence of CNN is multiple filters so as to extract spatial features hidden in data. Through the training process, the convolution layers of the CNN are optimized for extracting highly discriminative features and the latter layers imitate a multilayer perceptron which executes the classifying work. The CNN structure is presented in Fig. 2, which

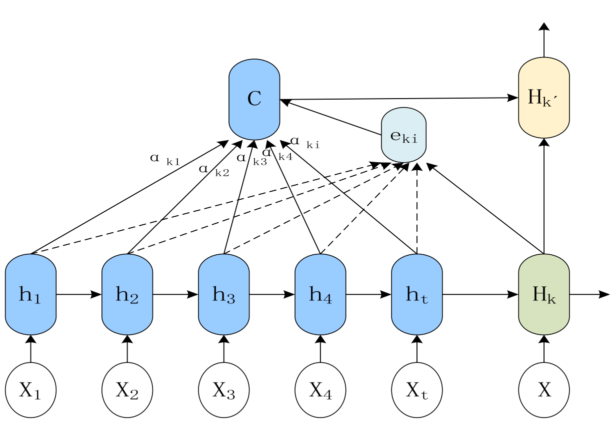

Fig. 4. Structure of AM.

includes inputting layer, convolution layer, pooling layer, full connection layer and outputting layer.

The convolution layer is designed with appropriate size to check the information in the visual field for convolution operation from the original data. This layer has the characteristics of local vision field, which can make the neurons perceive the information to generate the high-level features. In the layer the number of parameters is reduced, and the weight sharing is performed which uses the same parameters in all positions for a convolution kernel [32]. If the input data is x, the convolution layer can be represented as follows:

$$h = f(x \otimes W + b) \tag{1}$$

where $\otimes$ means convolution operation, W is the weight of the convolution kernel, and b is the offset value. $f(\cdot)$ denotes the activation function which is Rectified Linear Unit (ReLU). ReLU can add some nonlinear factors to the network, which can better deal with solve complicated issues. In addition, ReLU will make the output of some neurons equal to 0, which leads to the sparsity of the network and reduces the interdependence of parameters. Thus, the better mine relevant features and the fitter training data can be achieved, and the occurrence of overfitting problem is alleviated. The formula of ReLU is described as:

$$f(x) = max(0, x) \tag{2}$$

Pooling layer is to conduct sampling operation for the export of the convolution layer. These features of the similar area are aggregated with the maximum value. But only the important features are retained and the number of feature data is reduced in this layer. The spatial features are achieved through the source data using CNN, and the processed data are treated as inputs into the LSTM based on AM for time series forecasting.

2.2. Long and short term memory network (LSTM)

CNN is not sensitive to the time feature of time series, but LSTM can better fusion time characteristics of different parts status [33]. Then CNN is combined with LSTM to extract the time features of data from wind turbine. The output of CNN is taken as the input of LSTM. LSTM contains of inputting gate, the outputting gate and the forgot gate. Its principle diagram is shown in Fig. 3. The abandoned information is implemented by forgot gate f_t, which can be denoted by:

$$f_t = \sigma\left(\mathrm{W_f}\cdot[\mathrm{h}_{t-1}, \mathrm{x_t}] + \mathrm{b_f}\right) \tag{3}$$

where h_{t-1} is the output of the previous LSTM cell, x_t is the current input, and σ is hyperbolic tangent activation function. W_f and $b_f$are the weight matrix and the bias term, respectively.

The input gate i_t determines which information is updated, and $\widetilde{C}_t$ is immediate condition. ReLU is selected as activation function. They can be expressed by:

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

$$\widetilde{C}_t = ReLU(W_C\cdot[h_{t-1}, x_t] + b_C) \tag{5}$$

where W_i is weight matrix of the inputting gate, and b_i is offset item. $W_c$is weight matrix of the state and b_c is bias term. C_t denotes the long-term state. $o_t$is the outputting gate. The output h_t of the LSTM cell is obtained according to the updated cell state.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{6}$$

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

$$h_t = o_t * Relu(C_t) \tag{8}$$

where W_o is weight matrix of the output gate and b_o is offset item.

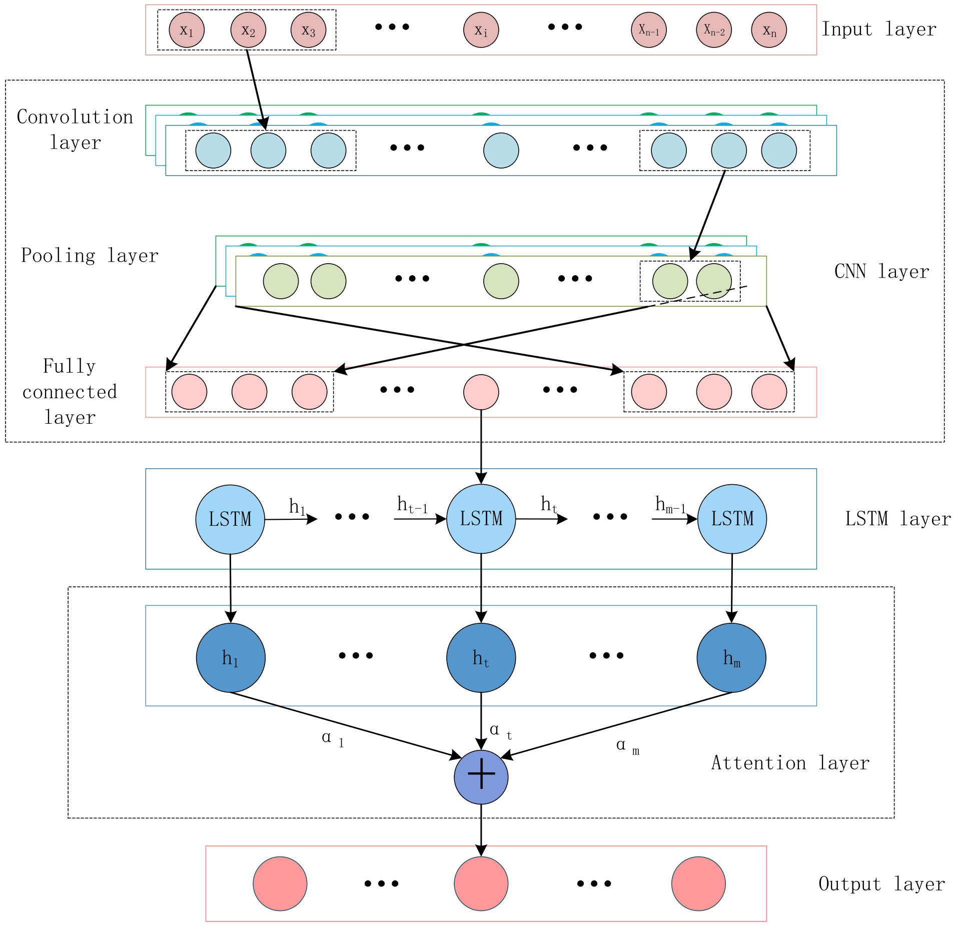

Fig. 5. Network structure of CNN-LSTM based AM.

2.3. Attention mechanism (AM)

When the human brain deals with things, more attentions often focus on the important contents, and ignore the unconcerned information. AM is introduced to strengthen the impact of important information. AM can assign different weighs to the implied states of LSTM through mapping weigh and parameter learning. AM is designed to simulate the attention allocation mechanism of human brain, which can calculate the relevance of the input and output for the distribution of the heavier weight features. AM is applied for LSTM to concentrate the characteristics that have a great impact on output variables, so as to increase the accuracy of the method. The structure of attention mechanism is presented in Fig. 4.

In Fig. 4, $X_1,X_2,\dots,X_t$ denotes the input of LSTM, and $h_1,h_2,\dots,h_t$ represents the output through hidden layer of LSTM, which is used as the input of AM to obtain the distribution of attention weight. The weight represents the importance of the state parameter. The calculation of AM is written as:

$$e_i = utanh(wh_i + b) \tag{9}$$

$$\alpha_i = \frac{exp(e_i)}{\sum_i exp(e_i)} \tag{10}$$

$$\mathrm{C} = \sum_i \alpha_i h_i \tag{11}$$

where e_i indicates the probability distribution of attention for h_i at ith moment. u and w denote the weighing coefficients. b is the bias coefficient, and Cis the weighted feature.

Table 1
Operation state variables of wind turbine.

No.	Variable	Units
1	wind speed	m/s
2	Gear box oil temperature	°C
3	Ambient temperature	°C
4	Cabin temperature	°C
5	Front axle temperature	°C
6	Rear axle temperature	°C
7	Winding temperature	°C
8	A Phase Current	A
9	A Phase Voltage	V
10	Active power	Kw
11	Reactive power	Kw
12	Gear box bearing temperature	°C

3. CNN-LSTM model based on AM

3.1. Model structure

The CNN-LSTM model based in AM is consisted of inputting layer, CNN layer, LSTM layer, AM layer and outputting layer. CNN extracts the spatial features of original data, and the extracted spatial features are treated as the inputs of LSTM network. The temporal features are extracted through LSTM, and the results are input to the AM layer. The AM layer calculates the weight based on the input data. The network structure of CNN-LSTM based on attention mechanism is shown in Fig. 5.

In inputting layer, the related variables are selected as input characteristics through correlation analysis. The CNN layer sets the

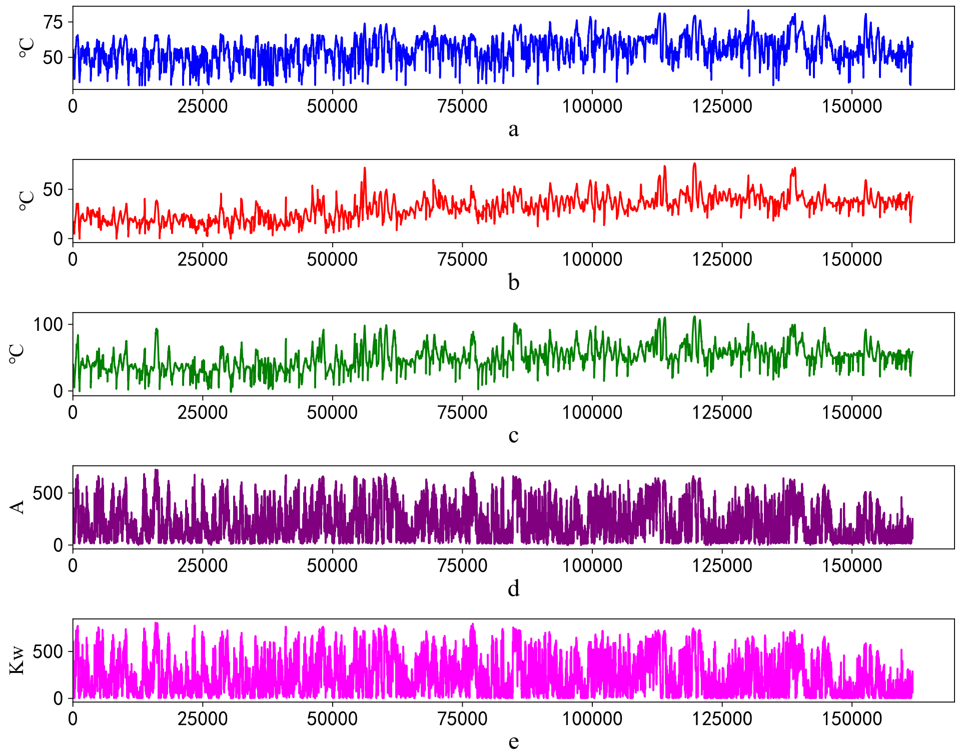

Fig. 6. Selected variables in SCADA (a) Gearbox bearing temperature (b) Front axle temperature (c) Winding temperature (d) A phase current (e) Active power.

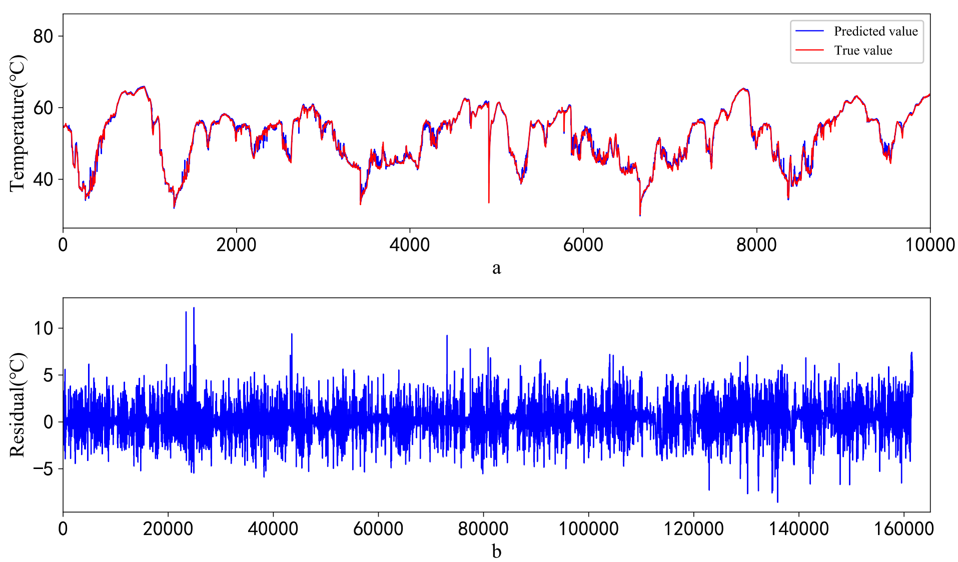

Fig. 7. Predicted values and residuals of CNN-LSTM-AM (a) Predicted and true values (b) Residual.

convolution kernel number to 64, the length of the convolution kernel to 1. In principle, the more the LSTM layer is hidden, the better the fitting degree will be. However, the training time will significantly increase with the number of layers, so layers of LSTM are set to 2. The number of LSTM neurons in first layer is 128, and there are 64 neurons in the second layer of LSTM. In AM layer, the influence on output is highlighted by learning the feature weight and assigning the input vector of AM layer. The outputting layer of the model is bearing temperature of the gearbox. The operating state of the wind turbine is determined by further analyzing the residual.

AME (Adaptive moment estimation) optimizer is used to iteratively update the network weights through training the data. The algorithm is different from the traditional stochastic gradient descent. By calculating the first and second order moment estimation of the gradient, the independent adaptive learning rate is designed for different parameters and the performance of the algorithm is also improved.

During the training of the deep learning model, SCADA data during the normal running state is selected as the training sample. After correlation analyzing, the input variables are selected, and the data characteristics of normal operation state are learned through repeatedly training for the prediction, Early stopping is adopted in the training data process to prevent overfitting. In the process of the testing, if the data input in normal running state can adapt to the characteristics of the model, the residual error of prediction is small. If the data input in abnormal running state cannot adapt to the characteristics of the model, then the residual error of prediction will increase. The working state of

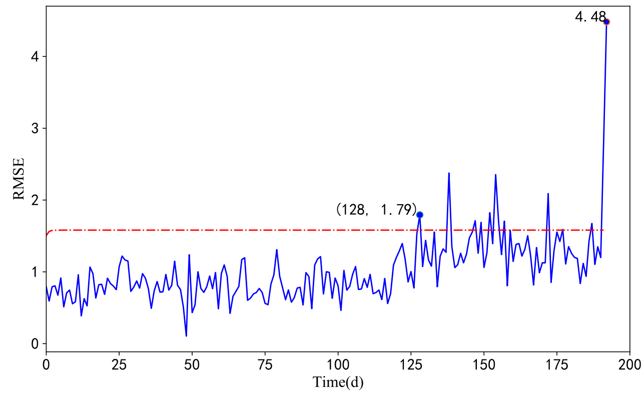

Fig. 8. Predictive result of case 1.

Table 2
Comparison of evaluation indexes of each model.

Criteria	Model			
	LSTM	BILSTM	CNN-LSTM	CNN-LSTM-AM
RMSE	1.454	1.398	1.070	1.005
MAE	0.954	0.950	0.674	0.621
R2	0.974	0.977	0.986	0.988

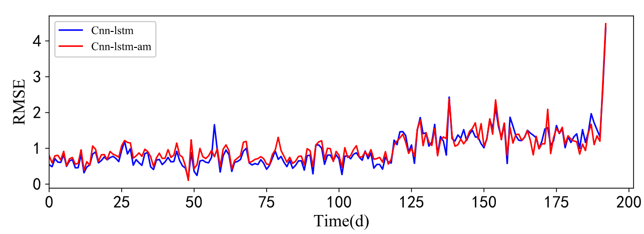

Fig. 9. RMSEs of prediction residual of models.

wind turbine is determined and the fault is detected by analyzing the results from model.

3.2. Evaluation criteria

The root mean square error (*RMSE*), mean absolute percent error (*MAPE*), mean absolute error (*MAE*) and r-square (R^2) are applied to evaluate the prediction performance of the presented model. They are written as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2} \tag{12}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y}_i - y_i\right| \tag{13}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}(\overline{y} - y_i)^{2'}} \tag{14}$$

where, n represents the predicted points; y_i and $\widehat{y}_i$ represent the real value and the predictive value of the ith point, respectively.

The working state of wind turbine is discriminated by variation trend and mutation degree of RMSE. The threshold value is set by using the exponential weighted moving average (EWMA). EWMA is a moving average weighted by exponential decreasing. If the data is close, the weight is great. And the farther the data is, the smaller the weight is. The threshold value set through EWMA can effectively detect RMSE fluctuations of residual, so the operation status of wind turbine is monitored. EWMA is expressed as:

Table 3
Operating state variables of wind turbine.

No.	Variable	Units
1	wind speed	m/s
2	Ambient temperature	°C
3	Average power	Kw
4	Maximum power	Kw
5	Generator bearing 2 temperature	°C
6	A phase temperature	°C
7	B phase temperature	°C
8	C phase temperature	°C
9	Generator speed	r/min
10	A phase current	A
11	B phase current	A
12	C phase current	A
13	Generator bearing temperature	°C

$$S_t = \lambda R_t + (1-\lambda)S_{t-1} \tag{15}$$

where λ represents the weight of historical data, R_t denotes the average value of RMSE, and the initial value of S_0 presents the mean RMSE of the predicted residual for the wind turbine in a period of time.

The threshold value for testing the operation status of wind turbine is the upper limit of EWMA, and it is obtained as:

$$U_t = \mu_R + X\sigma_R\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2t}\right]} \tag{16}$$

where μ_R and σ_R represent the mean and standard deviation of RMSE, respectively, and$X$ is a constant which is related to the position of the threshold, By training normal data, the size of X value can be determined to ensure an appropriate threshold and avoid false alarm of detection results..

4. Case and analysis

In the paper, SCADA data from two wind farms are selected as case analysis. One set of data is the SCADA data containing faults from 1st January to 14th July 2015 in a wind farm of north China. The wind turbine was shut down due to the gearbox fault on 14th July. Another set of data is taken from the SCADA data containing faults from 1st February 2013 to 2nd May 2014 of a wind farm in Zhejiang province. The wind turbine was repaired due to generator failure on 4th June 2014.

4.1. Case 1

The SCADA data of case 1 is from north China, and this wind turbine was shut down for maintenance due to gear broken of gearbox on 14th July. The previous data of 14th July are used for analysis. The data are firstly preprocessed, and the abnormal data points such as stopping and bad points are removed. Meanwhile, the operation states of wind turbine such as wind turbine startup and gearbox failure are saved. Too many variables will cause data redundancy and affect the accuracy of the prediction model. Therefore, the variables with a high correlation with output variables are selected as input variables through correlation analysis. In case 1, the selected target variable is the temperature of gear box bearing. By correlation analysis, the variables with high correlation with the temperature of gear box bearing are selected as the input variable, and the 11 variables with high correlation are selected, as shown in Table 1. Among them, the change trends of gearbox bearing temperature, front axle temperature, winding temperature, A phase current and active power, are presented in Fig. 6. It can be seen that each variable fluctuates within a certain range in each stage of wind turbine operation. Even in case of fault shutdown, there is no obvious change in

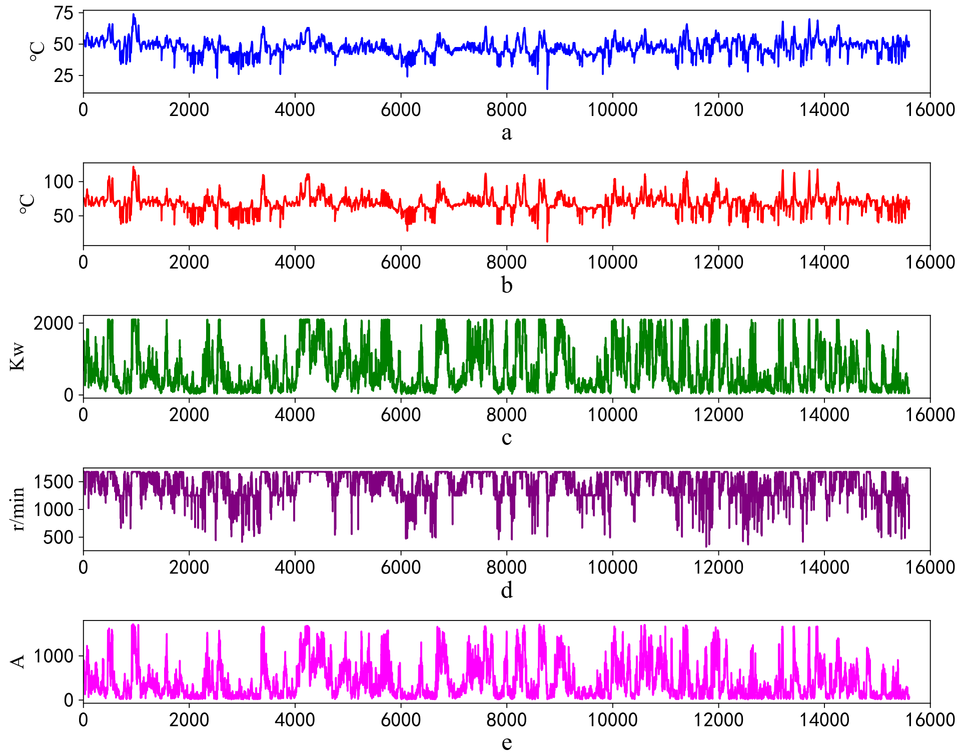

Fig. 10. Selected variables in SCADA (a) Generator bearing temperature (b) Average power (c) A phase temperature (d) Generator speed (e) A phase current.

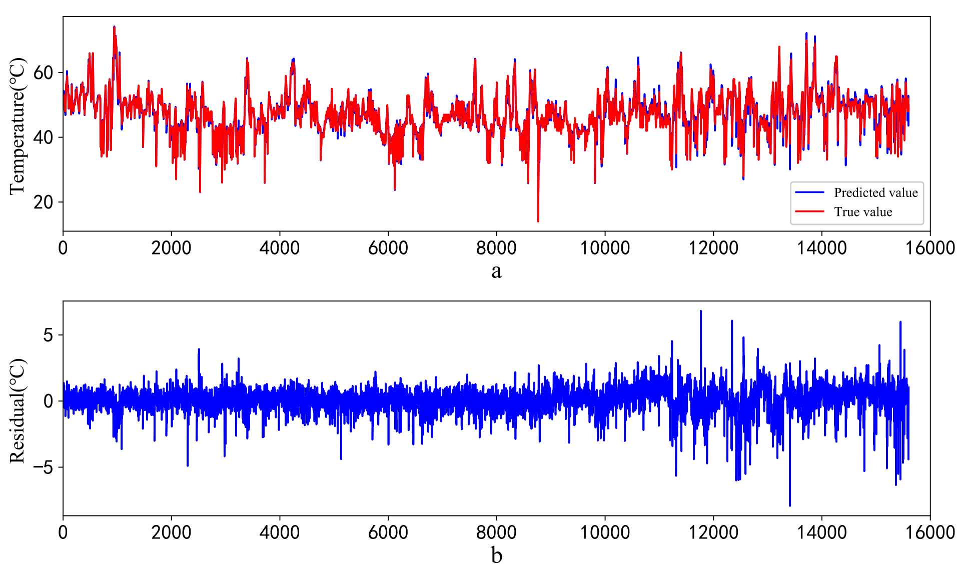

Fig. 11. Predicted values and residuals of CNN-LSTM-AM (a) Predicted and true values (b) Residual.

the trend, so the operation status of wind turbine cannot be directly determined by the change of each variable condition.

The proposed model of CNN-LSTM-AM is applied to train the pre-processed data. The data from January to April is considered as normal operation state of wind turbine. These data are treated as training samples, then all data including fault state are input to the trained model. The prediction value of the model to the bearing temperature of the output variable gearbox is obtained. The operation state of wind turbine is predicted and detected through the change trend of RMSE. A partial data comparison between real value and predictive value is described in Fig. 7 (a) through the gearbox bearing temperature of a wind turbine. The residual difference of the predictive value and the real value is presented in Fig. 7 (b). The predicted results after fault detection are shown in Fig. 8 in residual of RMSE. It can be seen that the threshold value is first violated in the 128th day. After this, the set threshold is repeatedly exceeded and has larger mutation before downtime in 14th July, with the maximum value reaching 4.48. This change can indicate that the potential failure of wind turbine has reached a certain degree, which is in line with the expected change of the detection results before the failure. The prediction time is consistent with the actual fault time, so it could be decided that the wind generator has been in fault in the 128th day. The proposed model effectively detects the fault of the wind generator.

For further verifying the superiority of this proposed model, the evaluation index of LSTM model is compared with those of BILSTM model and CNN-LSTM model, as shown in Table 2. The RMSE residuals of daily variation curves for CNN-LSTM-AM and CNN-LSTM are drawn in Fig. 8. It is clear from Table 2 that the evaluation indexes of the CNN-LSTM-AM model are superior to other models, which indicates that the proposed model can better extract the logical relationship between the inputting vector and the outputting vector. Fig. 9 shows that under the normal operation of wind turbine, the RMSE value is more stable, but under the abnormal state the RMSE has abrupt value. At the same time,

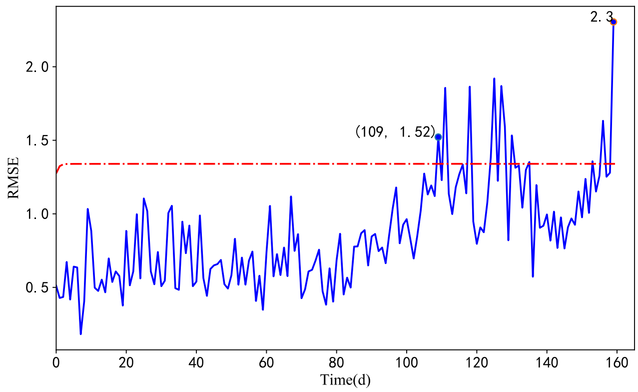

Fig. 12. Predictive result of case 2.

Table 4
Comparison of evaluation indexes of each model.

Criteria	Model			
	LSTM	BILSTM	CNN-LSTM	CNN-LSTM-AM
RMSE	3.882	3.792	0.893	0.875
MAE	2.809	2.708	0.677	0.634
R2	0.564	0.584	0.976	0.977

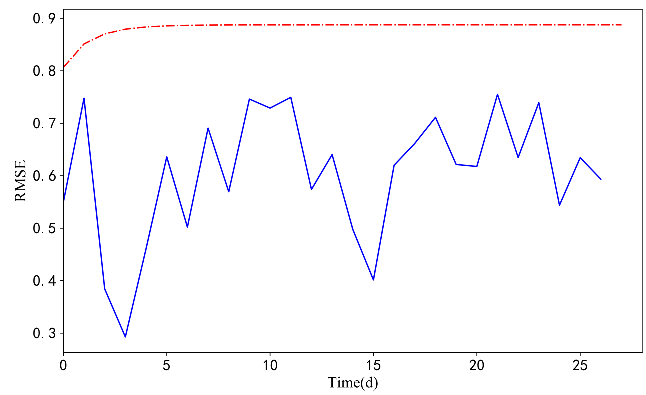

Fig. 15. Predictive result of case 3.

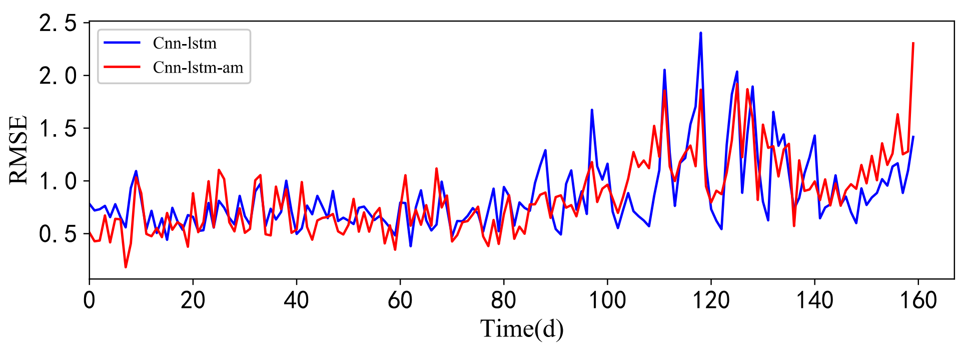

Fig. 13. RMSEs of prediction residual of models.

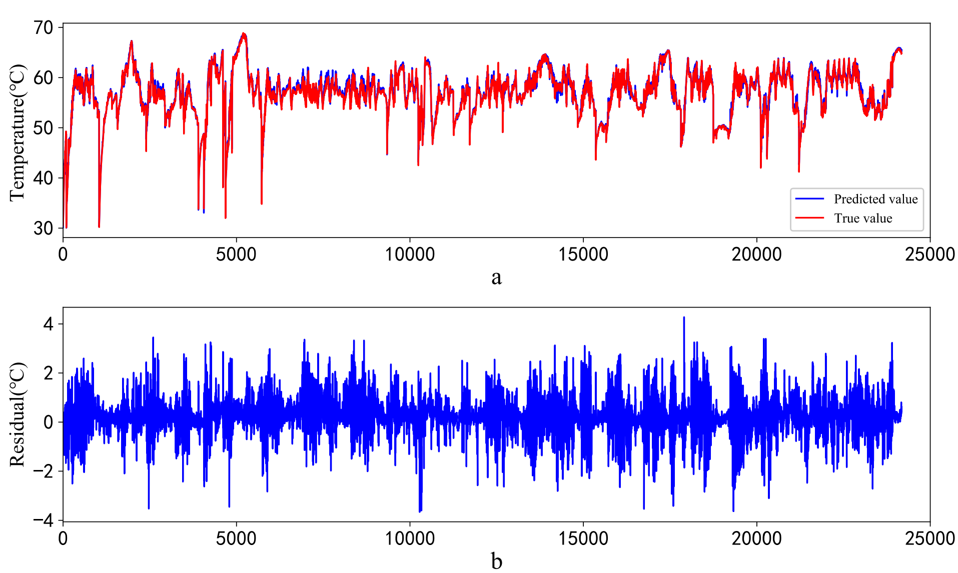

Fig. 14. Predicted values and residuals of CNN-LSTM-AM (a) Predicted and true values (b) Residual.

the evaluation index of the proposed CNN-LSTM-AM model in Table 2 is higher than other models, so CNN-LSTM-AM model is more accurate, more reliable and effective in predicting the running state of wind turbines.

4.2. Case 2

The wind turbine SCADA data collected in case 2 is on 1st November 2013 to 2nd May 2014. The wind turbine was repaired on 4th June 2014 due to the fault of alternator stator insulation and low rotor phase. The data are firstly preprocessed, the abnormal data points such as stopping and bad points are removed. Meanwhile, the operation states wind turbine such as wind turbine startup and generator failure are saved. Too many variables input into the prediction model will reduce the prediction accuracy due to data redundancy, and the generator bearing temperature will be used as the output variable. The variable with high correlation with the generator bearing temperature will be selected as the input variable. The 12 variables selected after correlation analysis are shown in Table 3. The change trends of generator bearing temperature, average power, A phase temperature, generator speed and A phase current, are presented in Fig. 10.

It can be seen from Fig. 10 that each variable fluctuates within a certain range in each stage of wind turbine operation. Even in case of fault shutdown, there is no obvious change in the trend, so the operation status of wind turbine cannot be directly determined by the change of each variable condition. Data comparison for real value and predictive value is described in Fig. 11 (a) through the generator bearing temperature. Their residual difference is presented in Fig. 11 (b). The predicted results after fault detection are shown in Fig. 12 in residual of RMSE. It can be seen that the threshold value is first violated in the 109th day. After this, the set threshold is repeatedly exceeded and has larger mutation before downtime in 4th June, with the maximum value reaching 2.3, which is in line with the expected change of the detection results before the failure. Therefore, it can be said that the detection method has detected the potential failure of the wind turbine on the 109th day. The proposed CNN-LSTM-AM method is verified through this actual case analysis.

The evaluation indexes of CNN-LSTM-AM model are compared with LSTM model, BILSTM model and CNN-LSTM model from Table 4. The RMSE residuals of daily variation curves for CNN-LSTM-AM and CNN-LSTM are drawn in Fig. 13. It is clear from Tab. 4 that the evaluation indexes of CNN-LSTM-AM model are superior to other models, especially to LSTM and BILSTM, which indicates that the proposed model can better extract the logical relationship between the inputting vector and the outputting vector and the effectiveness of the model. It can be seen from Fig. 13 that the RMSE of the proposed model is more stable when the wind turbine is in normal operation state, and there is no big mutation. The rising trend of RMSE is more obvious before the fault occurs, and the RMSE value is more prominent near the shutdown. In conclusion, the CNN-LSTM-AM model proposed in this paper is more accurate, reliable and effective in predicting the operation state of wind turbines.

4.3. Case 3

The data of case 3 is from the SCADA of wind turbine maintenance and restart. It is a completely healthy wind turbine. The data is preprocessed and the abnormal data points are removed, such as stopping points and bad points. This dataset is from September, and the data of the first 15 days is selected as the training data. Fig. 14 (a) shows the comparison of the actual temperature with the predicted value of gearbox bearing, and Fig. 14 (b) shows the residual. The predicted results of RMSE residual are presented in Fig. 15. As shown in Fig. 15, RMSE fluctuates within a certain range and its value is all below the threshold. It can be proved that the operation state of wind turbine is normal, and the CNN-LSTM-AM method proposed in this paper can effectively avoid false positive.

5. Conclusion

Considering the complex changeable working environment of wind turbine, a novel state monitoring model named CNN-LSTM-AM is proposed in this paper, which is employed for anomaly recognition and fault detecting of wind turbine. The CNN model is utilized to extract spatial features hidden in data. In training process, the convolution layers of the CNN are taken full advantage to acquire the highly differentiated features and the latter layers imitate a multilayer perceptron which executes the classifying work. LSTM can better fusion time characteristics of different parts status. AM is designed for LSTM to concentrate the characteristics that have a great impact on output variables, so as to improve the accuracy of the model. Four relevant comparative detecting models are implemented on the two cases with different faults. The analysis results indicate that the proposed method can be applied generally with a better comprehensive evaluation index and can avoid false positive, and provide strong support for decision-makers. In the future, the model will be further improved to obtain higher accurate and more stable detections. Meanwhile, transfer learning is considered to be applied to the model, which can be applied to other wind turbines through fine-tuning, so as to enhance the universality of the model.

CRediT authorship contribution statement

Ling Xiang: Conceptualization, Methodology, Software, Investigation, Writing - original draft. **: . Penghe Wang:** Writing - original draft, Data curation, Formal analysis. **Xin Yang:** Investigation, Formal analysis, Resources, Data curation. **Aijun Hu:** Resources, Writing - review & editing, Supervision. **Hao Su:** Validation, Visualization, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No.52075170).

References

[1] B.K. Sahu, Wind energy developments and policies in China: A short review, Renew. Sustain. Energy Rev. 81 (2018) 1393–1405.
[2] Y. Zhao, D. Li, A. Dong, et al., Fault Prediction and Diagnosis of Wind Turbine Generators Using SCADA Data, Energies. 10 (8) (2017) 1210–1226.
[3] L. Xiang, Z. Deng, A. Hu, Forecasting Short-Term Wind Speed Based on IEWT-LSSVM Model Optimized by Bird Swarm Algorithm, IEEE Access 7 (2019) 59333–59345.
[4] Q. He, E. Wu, Y. Pan, Multi-scale stochastic resonance spectrogram for fault diagnosis of rolling element bearings, J. Sound Vib. 420 (2018) 174–184.
[5] A. Kumar, C. Gandhi, Y. Zhou, et al., Latest developments in gear defect diagnosis and prognosis: A review, Measurement 158 (2020).
[6] A. Kumar, C. Gandhi, Y. Zhou, et al., Fault diagnosis of rolling bearing based on symmetric cross entropy of neutrosophic sets, Measurement 152 (2020).
[7] S. Lu, Q. He, J. Wang, A review of stochastic resonance in rotating machine fault detection, Mech. Syst. Sig. Process. 116 (2019) 230–260.
[8] X. Yan, Y. Liu, M. Jia, Research on an enhanced scale morphological-hat product filtering in incipient fault detection of rolling element bearings, Measurement 147 (2019).
[9] L. Cui, X. Wang, H. Wang, et al., Research on Remaining Useful Life Prediction of Rolling Element Bearings Based on Time-Varying Kalman Filter, IEEE Trans. Instrum. Meas. 69 (6) (2019) 2858–2867.
[10] A. Hu, Z. Bai, J. Lin, et al., Intelligent condition assessment of industry machinery using multiple type of signal from monitoring system, Measurement 149 (2020), 107018.
[11] P. Guo, D. Infield, X. Yang, Wind Turbine Generator Condition-Monitoring Using Temperature Trend Analysis, IEEE Trans. Sustain. Energy 3 (1) (2012) 124–133.

[12] D. Ferguson, A. McDonald, J. Carroll, et al., Standardisation of wind turbine SCADA data for gearbox fault detection, J. Eng. 2019 (18) (2019) 5147–5151.
[13] Z. Zhang, K. Wang, Wind turbine fault detection based on SCADA data analysis using ANN, Adv. Manuf. 1 (2014) 70–78.
[14] F. Pozo, Y. Vidal, Wind turbine fault detection through principal component analysis and statistical hypothesis testing, Energy. 9 (2016) 3–23.
[15] R. Pandit, D. Infield, Gaussian process operational curves for wind turbine condition monitoring, Energies. 11 (7) (2018) 1631–1650.
[16] X. Ding, Q. He, Energy-Fluctuated Multiscale Feature Learning with Deep ConvNet for Intelligent Spindle Bearing Fault Diagnosis, IEEE Trans. Instrum. Meas. 66 (8) (2017) 1926–1935.
[17] Y. Gao, X. Liu, J. Xiang, FEM simulation- based generative adversarial networks to detect bearing faults, IEEE Trans. Ind. Inf. 16 (7) (2020) 4961–4971.
[18] X. Liu, H. Huang, J. Xiang, A personalized diagnosis method to detect faults in gears using numerical simulation and extreme learning machine, Knowl.-Based Syst. 195 (2020).
[19] Q. Gao, H. Tang, J. Xiang, et al., A Walsh transform-based Teager energy operator demodulation method to detect faults in axial piston pumps, Measurement 134 (2019) 293–306.
[20] Z. Pan, Z. Meng, Z. Chen, et al., A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings, Mech. Syst. Signal Process. 144 (2020).
[21] W. Zhang, X. Li, Q. Ding, Deep residual learning-based fault diagnosis method for rotating machinery, ISA Trans. 95 (2019) 295–305.
[22] W. Zhang, X. Li, Q. Ding, et al., Machinery fault diagnosis with imbalanced data using deep generative adversarial networks, Measurement 152 (2020).
[23] H. Zhao, H. Liu, W. Hu, et al., Anomaly detection and fault analysis of wind turbine components based on deep learning network, Renew. Energy 127 (2018) 825–834.
[24] L. Wang, Z. Zhang, J. Xu, et al., Wind Turbine Blade Breakage Monitoring With Deep Autoencoders, IEEE Trans. Smart Grid 9 (4) (2018) 2824–2833.
[25] W. Teng, H. Cheng, X. Ding, et al., DNN-based approach for fault detection in a direct drive wind turbine, IET Renew. Power Gener. 12 (10) (2018) 1164–1171.
[26] S. Afrasiabi, M. Afrasiabi, B. Parang, et al., Wind Turbine Fault Diagnosis with Generative-Temporal Convolutional Neural Network, 2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe). (2019).
[27] L. Jing, M. Zhao, P. Li, et al., A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.
[28] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement 93 (2016) 490–502.
[29] L. Yu, J. Qu, F. Gao, et al., A novel hierarchical algorithm for bearing fault diagnosis based on stacked LSTM, Shock Vib. 2019 (2019) 2756284.
[30] Z. Kong, B. Tang, L. Deng, et al., Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units, Renew. Energy 146 (2020) 760–768.
[31] W. Wang, A. Wang, Q. Ai, et al., AAGAN: Enhanced Single Image Dehazing With Attention-to-Attention Generative Adversarial Network, IEEE Access 7 (2019).
[32] T. Ince, S. Kiranyaz, L. Eren, et al., Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks, IEEE Trans. Ind. Electron. 63 (11) (2016) 7067–7075.
[33] Z. Zhao, W. Chen, X. Wu, et al., LSTM network: a deep learning approach for short-term traffic forecast, IET Intel. Transport Syst. 11 (2) (2017) 68–75.