IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 26, 2021, accepted September 4, 2021, date of publication September 7, 2021, date of current version September 15, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110909

Study of Wind Turbine Fault Diagnosis and Early Warning Based on SCADA Data

YILONG SHI[1], YIRONG LIU[2], AND XIANG GAO[1]
[1]School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China
[2]Binzhou Power Supply Company, State Grid Shandong Electric Power Company, Jinan 250001, China

Corresponding author: Xiang Gao (gaoxiang@sdu.edu.cn)

ABSTRACT Wind turbine fault diagnosis and early warning are important to reduce wind farm operation and maintenance costs and improve power generation efficiency. In this paper, we take the Supervisory Control and Data Acquisition (SCADA) data as the research object and research wind turbine health data purification, fault diagnosis model building, and unit operation status monitoring from a completely data-driven perspective. Firstly, for the problem that Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm cannot identify high-density anomalous data. An anomaly data processing scheme combining a density clustering algorithm and normal power interval estimation is proposed. The accuracy of extracting health data from wind turbines is improved. Secondly, to address the problem that the eXtreme Gradient Boosting (XGBoost) algorithm has more hyperparameters, we propose an optimization scheme based on the Bayesian Optimization Algorithm (BOA) and tree model for feature weight measurement, which improves the efficiency and accuracy of intuitive mapping from SCADA system monitoring data to fault features. Finally, a wind turbine condition monitoring scheme based on the information fusion of multi-characteristic monitoring parameters is designed. The wind turbine condition monitoring scheme proposed in this paper can warn generator system failure 3.67 hours, gearbox system failure 5.17 hours in advance, and hydraulic system failure 2.33 hours in advance.

INDEX TERMS Wind turbine, fault diagnosis, fault warning, XGBoost, SCADA.

I. INTRODUCTION

In recent years, with the continuous deterioration of the global ecological environment and the gradual depletion of fossil fuels, wind power generation has gradually become a new power generation mode replacing the traditional power generation mode in the world [1], [2]. Because most wind turbines are installed in remote areas rich in wind energy, such as mountains, wilderness, islands, or even the sea, they are subject to extreme temperature differences and strong wind gusts throughout the year, resulting in a much higher failure rate than other electromechanical equipment [3], [4]. The traditional wind farm unit maintenance strategy is highly dependent on regular maintenance and post-maintenance and can only deal with the monitoring and early warning of part of the wind farm unit [5], [6]. At the same time, due to the long deployment cycle of spare parts, the fault maintenance cost of large-scale wind farms remains high, which has a significant impact on the economic benefits of wind farms [7], [8]. Due to the complex structure, variable operation conditions, and strong coupling between components of wind turbines, failures occur frequently and even chain phenomena, leading to significant accidents such as combustion and collapse of wind turbines [9], [10]. Suppose the fault of the wind turbine can be accurately diagnosed and estimated and its development trend and the potential fault symptoms can be found early [11], [12]. In that case, the optimal maintenance strategy can be developed to reduce the failure rate to ensure the safe and efficient operation of the wind turbine. At the same time, through the fault trend warning, to avoid major property damage, to protect personal and equipment safety. However, the precise and general predictive wind power operation and maintenance technology are not yet mature, and the contradiction between the growing wind power installed scale and the relatively lagging wind field operation and maintenance technology is becoming more and more acute [13], [14].

At present, many results have been achieved in the study of fault diagnosis and prediction of wind turbines [15],[16].

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

Adouni uses multi-layer ANN technology to build a fault detection and identification method for wind turbines with low voltage traverse, which speeds up fault detection and enhances the robustness and anti-interference ability of fault identification [17]. However, the large amount of training data and long modeling time of the ANN algorithm, and the drawback that the convergence and reliability of the ANN algorithm are difficult to guarantee to restrict its further development. Qin uses the principal component analysis algorithm to reduce the dimension of multi-dimensional target characteristic parameters to one-dimensional as the classification label of SVM to build a concise fault diagnosis model and realizes the diagnosis of gearbox bearing faults and generator bearing faults [18]. Although SVM has outstanding advantages in solving decision problems with small, high-dimensional nonlinear data samples, it is not good at handling multi-classification problems with large data samples, and the SVM algorithm does not have uncertainty management capability. Based on SCADA data, Wang built a prediction model of gearbox oil pressure with the deep neural network, which could give an early warning of gearbox faults according to the change of gearbox oil pressure [19]. Roshan Kumar provides a review of the signal processing method for detecting wind turbine damage, verifying the advantages and disadvantages of the signal processing method for specific types of wind turbine damage [20]. The fault diagnosis method based on signal processing has achieved good results, but in the process of implementation, the signal processing method will become very complicated when considering complex operating conditions. At present, for the fault warning methods of critical parts of wind turbines, most of the research is carried out based on historical experience by artificially selecting a certain characteristic monitoring quantity. However, the wind turbine is a strong coupling system with multiple sub-systems working together. The single parameter information content is limited, so it is difficult to reflect its abnormal state fully. In addition, when setting the fault warning threshold of operating characteristic parameters, most methods artificially set a fixed threshold according to expert experience, which leads to an intense subjectivity of model establishment and reduces the credibility and generalization ability of the model.

In order to realize the fault diagnosis and fault warning of the wind turbine, this paper takes SCADA system data of the actual wind field as the research object. The research on health data purification, fault diagnosis model construction, and operating state monitoring of wind turbines is carried out from a complete data drive. Firstly, taking the ''wind speed power'' curve of wind turbines as the breakthrough point, the distribution characteristics of abnormal data in the actual operation data of wind turbines are studied. Then, an anomaly data processing scheme combining density clustering algorithm and normal power interval estimation is proposed, which makes up for the failure of the DBSCAN algorithm to identify high-density anomaly data. At the same time, to determine the correlation between fault information and unit monitoring parameters without relying on prior knowledge, the feature weight measurement method based on the tree model was studied, and the manifestation forms of different faults in SCADA monitoring parameters were measured from the perspective of data analysis. Finally, to solve the problem that a single monitoring parameter has a low information content and is difficult to fully reflect the abnormal state of the system, the method of integrating characteristic parameters from different sources and different scales into operation state indicators is studied according to the typical weight. The consistent description of the operation state of the wind turbine is obtained from many monitoring parameters. To solve the problems of strong subjectivity, weak generalization ability, and easy false alarm caused by artificial fixed threshold setting, a dynamic threshold setting scheme based on adaptive principle was designed, which thoroughly considered the operation situation of the unit in the previous time, and could effectively realize the early warning of various faults.

The rest of the article is shown below. Section 2 introduces the algorithm analysis of abnormal data processing, fault diagnosis, and feature analysis, fault warning method, etc. The experimental results of each algorithm are described in Section 3. Finally, section 4 is the conclusion and the next step.

II. ALGORITHM ANALYSIS

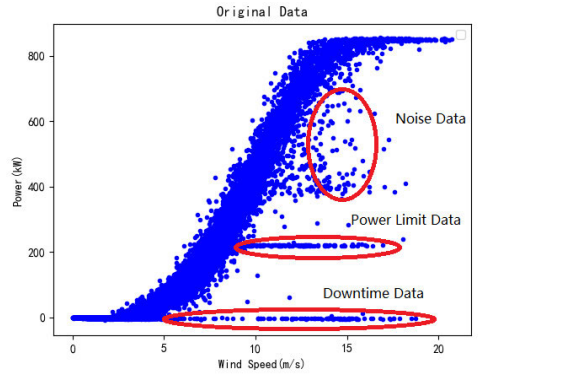
A. EXCEPTION DATA HANDLING SCHEME

As the wind turbine generator is in normal operation for a long time, there is a massive gap between the storage capacity of fault data and health data in the actual SCADA system of the wind field. In order to reduce the repetitive fault data mining work and avoid too much reliance on relevant prior knowledge, this paper uses the health data of wind turbines to establish a normal model. In this way, the high-quality health data samples screened from the original SCADA data set are the basis for subsequent studies. In the actual operation of a wind turbine, due to the uncertainty of wind speed and direction as well as the constraints of variable speed constant frequency electric control, the operating state of the wind turbine usually switches randomly and frequently between different operating conditions, which will produce abnormal data such as shutdown data, power limit data, fault data, and outlier data. As shown in Figure 1(a), the common anomalous data are as follows:

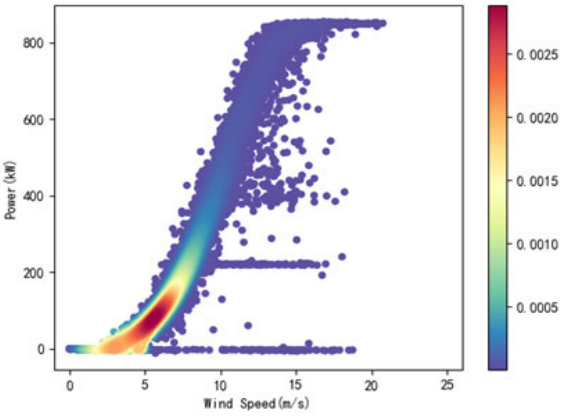
1) Downtime data: The wind turbine's measured wind turbine is greater than the cut wind speed, and the output power is 0 for a continuous period of time, which is mainly caused by artificial wind abandonment or communication failure.
2) Limited power data: The output power of the wind turbine is distributed below the ideal power curve and does not change with the change of wind speed (or changes little). Such abnormal points are mainly caused by the artificial control of the wind turbine to limit the output.

3) Noisy data: Data points are randomly distributed outside the overall data points, and such abnormal matters are generally caused by wind turbine faults or noise.

By comparing Figure 1(b), it can be seen that the distribution densities of the three kinds of abnormal data are all low. Among them, the noise data has no fixed aggregation range and presents a random and discrete distribution. The downtime data and the power limit data fluctuate around a certain power value, and the longitudinal height of the power distribution is about $20kW$; That is, the power fluctuation range is about $\pm 10kW$. Therefore, the DBSCAN clustering algorithm based on density can be considered to identify low-density discrete abnormal data.



(a) Scatter diagram of data distribution



(b) Scatter diagram of distribution density

FIGURE 1. "Wind speed–power" distribution diagram of the unit with abnormal data.

The process of abnormal data processing scheme based on normal power interval estimation proposed in this chapter is as follows:

1) DBSCAN algorithm was used to eliminate low-density discrete noise data.
2) Calculate the density midpoint of each partition according to the power partition of the data.
3) The least-square algorithm was used to fit the density midpoint [21], which was used as the ideal power curve of wind turbine operation.
4) According to the 3-Sigma criterion, the normal power range is set with the ideal power curve as the center [22].
5) Remove abnormal data outside the normal power range.

When partitioning data according to power, set the partition with 20kW as the step size, as shown in Figure 2:

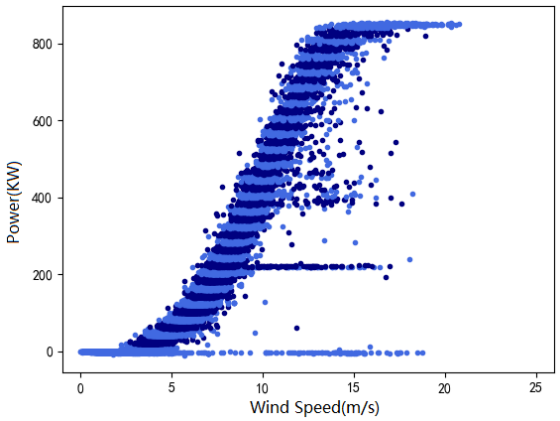


FIGURE 2. Schematic diagram of power partition.

After the partition is completed, the mean wind speed in each section is calculated and denoted as $\bar{x}_i$; the standard deviation of wind speed is indicated by σ_i, the power means within this partition is marked as $\bar{y}_i$, the density midpoint within each section is marked as $(\bar{x}_i, \bar{y}_i)$. Calculate the mean, standard deviation of each section and mark it as σ:

$$\sigma = \frac{1}{n}\sum_{i=1}^{n}\sigma_i \tag{1}$$

In Equation 1, n is the total number of partitions, i represents each partition, σ_i is the standard deviation of wind speed in each partition, and σ is the mean value of the standard deviation of the partitions.

Since the output power of the wind turbine generator is proportional to the cubic power of the wind speed, the least square method is used to carry out cubic polynomial fitting for each center point, and the ideal output power formula of the wind turbine generator is obtained.

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \tag{2}$$

In Equation 2, y is the wind speed, x is the output power, and a_0, a_1, a_2,and a_3 are the fitting coefficients.

According to the above analysis, wind turbines' "wind speed and power" data are approximately normally distributed around the ideal power curve, and the farther away from the perfect power curve, the less the data distribution. According to the 3-Sigma criterion, in a normal distribution, the probability of data distribution within three standard deviations of the mean value is about 99.73%. Therefore, the normal data distribution interval is set as follows:

$$[y(x - 3\sigma), y(x + 3\sigma)] \tag{3}$$

In equation 3, y is the wind speed, x is the output power, and σ is the mean of the standard deviation of the partition.

The interval shown in the equation estimates the normal power interval, and the probability of data exceeding the gap is less than 0.27%. Therefore, data that is not in this range can be identified as abnormal values. Finally, the normal power interval boundary is taken as the threshold, and the outliers outside the gap are deleted to obtain relatively pure healthy data samples.

B. FAULT DIAGNOSIS AND FEATURE ANALYSIS

At present, most of the wind turbine fault diagnosis methods based on machine learning generally have the problems of time-consuming training and flawed interpretation of results [23]. In order to solve this problem, this paper builds a fault diagnosis model based on the BOA-XGBOOST algorithm. It analyzes the correlation between fault information and the monitoring parameters of the SCADA system based on the tree model. This will pave the way for the follow-up fault warning research.

1) FAULT DIAGNOSIS DATA SET CONSTRUCTION

The actual output power of wind turbines is not uniformly distributed within the rated power range, and most robust data are distributed in the low power range. If only a segment of operation data is randomly intercepted from the SCADA data of wind turbine as a normal data set. In this way, the data may only be concentrated within a specific power range, resulting in the normal data set being patchy and unable to reflect wind turbines' actual operation fully. Figure 3 shows the power data distribution density histogram of a wind turbine generator in a wind farm in 2019, divided into ten intervals according to active power. The horizontal axis represents the output power, and the vertical axis represents the distribution probability of data within the power range.

In order to make the health data set genuinely reflect the actual operating conditions of all wind turbines in the wind farm, the historical SCADA data of 10 stable operating units were randomly selected from the wind farm in this chapter. First of all, abnormal data were removed to obtain health data samples. Then, according to the interval division and data distribution density, as shown in Figure 3, a moderate proportion of health data samples is selected to build a health data set with a total of 1000 data pieces.

In this chapter, according to the working principle and structural characteristics of wind turbines, based on the statistics of the fault categories with high frequency, the frequently occurring faults in wind turbines are divided into generator system faults, gearbox system faults, and hydraulic system faults for fault diagnosis and feature analysis. The fault categories of the three types of faults are shown in Table 1:

The faults of wind turbines are primarily the result of the deterioration of parts. State deterioration is a process from quantitative change to qualitative change. The data in a period of time before the occurrence of the fault contain the characteristic information of the fault. Finally, the fault

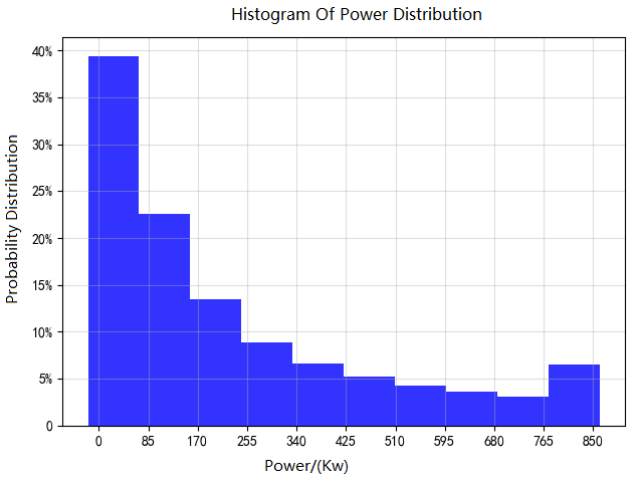


FIGURE 3. Histogram of active power distribution density.

diagnosis model modeling data set constructed after corresponding labels were marked for different data sets is shown in Table 2:

2) MULTI-CLASSIFICATION MODEL BASED ON BOA-XGBOOST ALGORITHM

The nature of the fault diagnosis model is a multi-classification model, which classifies unknown data samples by learning different data characteristics of known categories. XGBoost algorithm can efficiently deal with multiple classification problems, but it is difficult to tune due to a large number of super parameters. In order to solve this problem, this chapter adopts BOA to find the combination of super parameters with the highest classification accuracy. The process is shown in Figure 4:

The main steps of the BOA-based approach to adjusting XGBoost's hyperparameters are shown in Figure 4:

1) Set the super parameter space. The XGBoost algorithm contains three types of hyperparameters: generic parameters, model parameters, and learning task parameters. Table 3 shows the range of super

TABLE 1. Alarm contents of each fault category.

Fault Category	SCADA Fault Content
Generator system failure	Generator drive bearing temperature is high
	High generator temperature
	Wind turbine generator overload
	Excessive generator speed
	Rotor side temperature is too high
	The rotor speed does not match the generator speed
Gearbox system failure	Gearbox oil temperature is high
	Gearbox cooler overloaded
	Gearbox high speed bearing high temperature
	High spindle temperature
Hydraulic system failure	Hydraulic motor overload
	Low working pressure
	Build pressure timeout
	Hydraulic motor temperature error
	High temperature of hydraulic oil

parameters and the meaning of parameters set in this chapter that greatly impact model performance.

TABLE 2. Fault diagnosis model modeling dataset.

The Unit State	Data Volume	Data Labels
Normal	1000	0
Generator system failure	1000	1
Gearbox system failure	1000	2
Hydraulic system failure	1000	3

1) Determine the prior probability distribution. BOA is a process of constantly updating the initial distribution. The primary distribution of hyperparameters and classification accuracy need to be determined before optimization. The classification model was trained by randomly selecting multiple combinations of super parameters, and the prior distribution relationship between the varieties of various parameters and the accuracy was obtained.
2) BOA optimization process. In the process of super parameter optimization, the Gaussian process is used as the probability function to represent the unknown optimal parameters, starting from the original priori obtained in the previous step. Then, the acquisition function is used to select the unevaluated hyperparameter combinations from the parameter space around the currently found optimal hyperparameter combinations, and the information is increased through iteration, and the prior is constantly revised. Finally, at the end of the iteration, the hyperparameter of the model with the highest accuracy was selected as the optimal hyperparameter combination. Finally, the classification model is trained by various super optimal parameters to obtain the final model.

When training the classification model, the modeling data set of fault diagnosis model is divided into the training set and test set according to the ratio of (8:2), which are used for training and evaluation model respectively. During the training of the model, the XGBoost algorithm converts multiple classification problems into multiple dichotomy problems,

TABLE 3. XGBoost parameter setting range and meaning.

Parameter Name	Setting Range	Parameter Step Size	Parameter meaning
n_estimatores	(0,300)	10	Maximum number of iterations
max_depth	(3,10)	1	The tree deep
min_child_weight	(1,5)	1	Defines the minimum weight sum required for child nodes
gamma	(0,0.5)	0.1	Penalty coefficient
subsample	(0,0.8)	0.1	proportion of samples to be extracted when training the tree
learning_rate	(0.05,0.3)	0.05	learning_rate

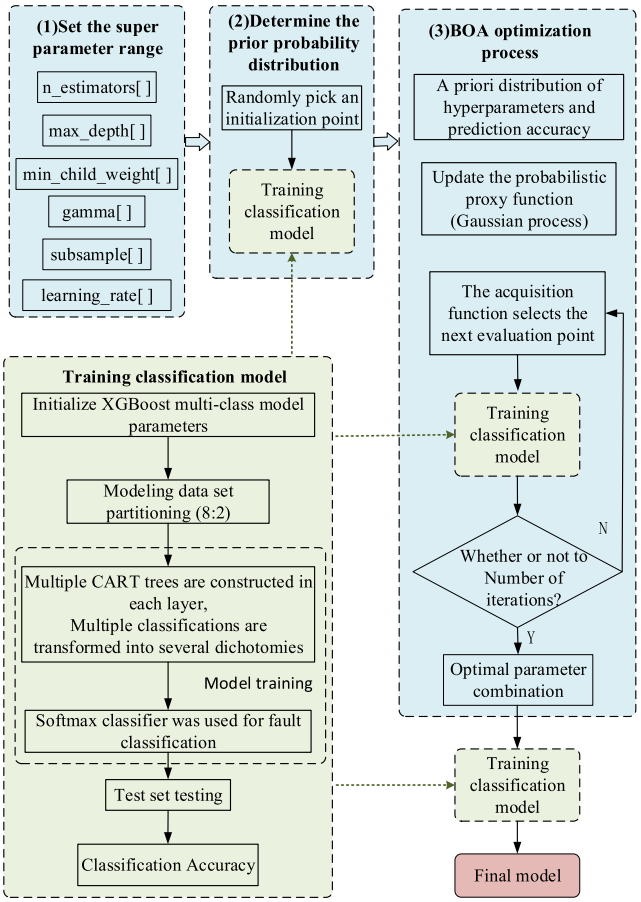


FIGURE 4. BOA-XGBoost model construction process.

calculates the predicted score values of all leaf nodes, and converts them into probability values by Softmax layer after weighted sum, and then classifies them according to the probability values. When using the test set to evaluate the model, the accuracy rate is adopted as the evaluation index:

$$Accuracy = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FP_i)} * 100\% \quad (4)$$

In Equation 4, n is the total number of categories, and in this chapter $n = 4$, TP_i denotes the number of correct classifications of a category, and FP_i denotes the number of incorrect classifications of a category. *Accuracy* denotes the ratio of the number of correctly classified instances to the number of instances in the total test set.

3) FEATURE WEIGHT MEASUREMENT METHOD BASED ON TREE MODEL

The operation condition of the wind turbine is complex, and its fault manifestation is closely related to the operation state. There is a potential correlation between the monitoring parameters of the SCADA system and the fault of the wind turbine. It is of great significance for potential fault analysis and overall health state evaluation of wind turbines to master

the fault characteristics of each component of wind turbines, which is helpful to realize the intuitive mapping from monitoring quantity to the operating state of the turbine.

XGBoost algorithm concentrates all the samples on a leaf node when creating a binary tree and gradually generates a tree through the constant splitting of leaf nodes. In the process of leaf node splitting, feature parallelism is used to select the feature to be split. Using multiple threads, first try to treat each feature as a split feature, find the optimal segmentation point of each feature. Then select the feature with the most significant gain after splitting according to different features as the splitting feature. So the number of times a feature is used as a split is a measure of the importance of a feature, and the more times a feature is used as a split, the more important that feature is. The weight $weight_f$ of defining feature f is shown in Eq.5

$$Weight_f = \sum_{i=1}^{Tree} n_i \tag{5}$$

In Equation 5, $Weight_f$ is the weight of feature f in building the XGBoost model. $Tree$ is the number of binary trees composing the model, t_k is the number of times that feature f is used as a splitting feature in the i tree. The weight of a feature is the sum of The Times that the feature is used as a splitting feature in all decision trees.

Figure 5 shows part of the branches of the 50th decision tree during fault diagnosis model training. F13 in the figure indicates the ambient wind direction. F17 denotes the oil temperature of the gearbox. F18 represents gearbox high speed bearing temperature. F29 means the converter controller temperature. F17 split twice in the figure and once with the other features, so on this branch, feature F17 (gearbox oil temperature) has a weight of 2, and the other features have a weight of 1.

C. WIND TURBINE CONDITION MONITORING SCHEME BASED ON INFORMATION FUSION OF MULTIPLE CHARACTERISTIC PARAMETERS

1) WIND TURBINE OPERATING STATE MONITORING MODEL

The process of failure of wind turbine equipment is a process in which the deterioration degree of equipment develops gradually from qualitative change to quantitative change. In the process of equipment deterioration, early signs will be generated. When the wind unit is operating in good condition, although it may be disturbed by environmental factors, there is a stable relationship between the parameters in the system, and specific monitoring data can be accurately reconstructed by using other monitoring data through an appropriate regression model. When the wind unit or the subsystem of each component of the wind unit fails, the relationship between the parameters of the system is broken. At this time, the results of a specific monitoring data reconstructed by the regression model will have a high error. With the development of wind turbine faults, the uncertainty of the relationship between the parameters in the system is further increased, leading to the gradual increase of the error between the reconstructed value and the actual value, showing a trend of climbing up or jitter rising. Therefore, the severity of deviation from the normal state of the relationship between the parameters of a wind turbine can be regarded as the critical point of the state monitoring of the wind turbine. By tracking the variation trend of the reconstruction errors of several characteristic parameters, the operating state of the wind turbine can be monitored in real-time.

The wind turbine operating state monitoring model proposed in this chapter is shown in Figure 6, which mainly consists of SCADA historical data and real-time data. The SCADA historical health data is used to determine the threshold value of wind turbine health status indicators.

2) RUNNING STATE INDEX BASED ON MULTIPLE MONITORING PARAMETERS

Among the many monitoring parameters of the SCADA system, the temperature data, electrical data, or environmental data of each part of the wind turbine may be used as fault characteristic parameters. The data dimension of each monitoring project is different, and the data scale is also very different. If the reconstruction errors of each characteristic parameter are directly fused, the dimensional chaos and the problem of large-scale parameter features drowning small-scale parameter features will occur. To solve the above issues, before calculating the operating state index, this chapter sets a window to calculate the relative residual of a monitoring parameter in a certain period of time to eliminate the influence of dimension and data range. Assuming that the actual value, predicted value, and window size of a specific characteristic parameter are y_t, $\hat{y}_t$ and l, then the relative residuals $r(k)$ of the expected parameter within l time spans at the time of k are as follows:

$$r(k) = \frac{1}{l}\sqrt{\frac{\sum_{t=t_k-l}^{t_k}\left(y_t - \hat{y}_t\right)^2}{\bar{y}_k^2}} \tag{6}$$

In Equation 6, $r(k)$ denotes the relative residual, l denotes the size of the window, the actual value is y_t, $\hat{y}_t$ denotes the predicted value, t_k denotes the k moment, and $\overline{y_k}$ denotes the

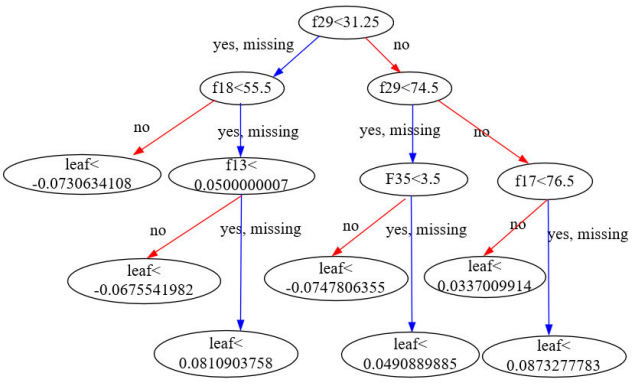


FIGURE 5. Part of the branches of the decision tree.

mean value of the actual values of the monitored parameters in the window.

$$\overline{y_k} = \frac{1}{l}\sum_{t=t_k-l}^{t_k} y_t \tag{7}$$

In Equation 7, y_k denotes the mean of the actual values of the monitored parameters within the window, l denotes the size of the window with the actual value of y_t, and t_k denotes at moment k.

The automatic shutdown protection time of the wind turbine is 5 minutes. In order to avoid the extreme abnormal point of relative residuals caused by data mutation during mechanical downtime, the window size is set as 10 minutes in this paper. At the same time, the original residual samples every 1 minute will be aggregated into a relative residual sample set every 10 minutes.

At present, the fault warning for a certain part or subsystem of the wind turbine is mostly realized by monitoring a single monitoring quantity. However, because the wind turbine is a complex nonlinear system of ''electromagnetic coupling, mechanical transmission, and energy conversion,'' the variation trend of a single monitoring quantity cannot reflect the whole operating condition of the system, which will cause false alarm or missing alarm. In the light of the problems above, this chapter takes the relative residuals of multiple fault feature monitoring parameters as evaluation indexes on the basis of the above fault feature analysis to realize the state monitoring of the wind turbine subsystem. Assume that the characteristic fault parameters of a subsystem of the wind turbine are monitoring quantity A, monitoring quantity B, and monitoring quantity C, respectively. The feature weights of the three feature monitoring quantities are w_A, w_B, and w_C, respectively. Relative residuals are r_A, 2, and 3, respectively. The state index of this subsystem is defined as follows:

$$R = \frac{w_A r_A + w_B r_B + w_C r_C}{w_A + w_B + w_C} \tag{8}$$

In Equation 8, R represents the state indicator of the system, w_A, w_B, and w_C represent the characteristic weights of the three characteristic monitoring quantities, and r_A, r_B, and r_C are the relative residuals.

The mathematical meaning of the above equation is to assign weights to the relative residuals of multiple characteristic parameters according to the feature weights and fuse the relative residuals of multiple monitoring quantities into a state indicator. This indicator is dimensionless. The larger the indicator value is, the more serious the unit deviates from the ideal working condition. If the hand exceeds a specific threshold value, it indicates that the monitoring part of the wind turbine has shown signs of failure, which needs to be maintained.

In order to further observe the development trend of state indicators, based on the calculation of state indicators, this chapter uses EWMA to calculate the trend control chart of state indicators and predict the changing trend of state indicators. EWMA control point value expression is shown in Equation 9.

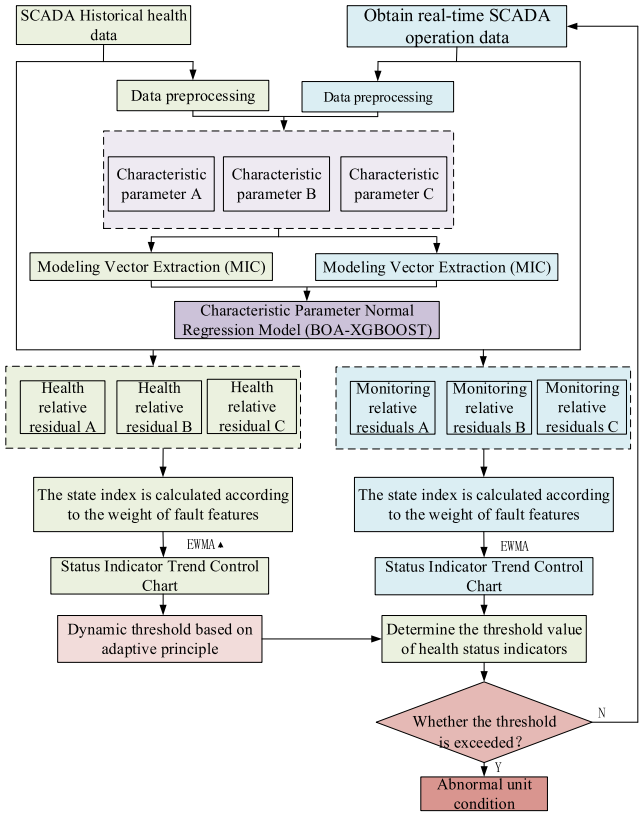


FIGURE 6. Frame of unit operation condition monitoring model.

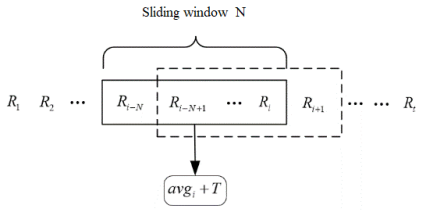


FIGURE 7. Schematic diagram of dynamic threshold setting.

Equation 9.

$$v_t = \beta R_t + (1-\beta)v_{t-1} \tag{9}$$

In Equation 9, v_t represents the trend value of the state index at time t, and R_t represents the state index at time t. Coefficient β represents the weighting factor of the EWMA control chart to historical data, $\beta \in (0, 1]$, $(1-\beta)$ represents the rate of historical weighted decline, $\beta = 0.9$. Due to the inevitable errors in model prediction, processing the residuals through EWMA can reduce the fluctuation range of the residuals and effectively eliminate the false alarm points, making the warning algorithm more stable and accurate.

3) DYNAMIC FAULT THRESHOLD SETTING BASED ON ADAPTIVE PRINCIPLE

Wind turbine operating conditions are complex. Under the influence of some uncontrollable objective factors, SCADA parameters of the wind turbine may deviate from normal

values within the alarm range, which is manifested as extreme points on residual values. If a fixed threshold is used to set the alarm threshold of the fault alarm system, the situation of false alarm may occur when the extreme point of residual error is higher than the fixed threshold. Aiming at the problem of false alarm caused by fixed alarm threshold, this paper designed a dynamic threshold setting scheme based on the adaptive principle, as shown in Figure 7. The dynamic threshold is set in segments through the sliding window.

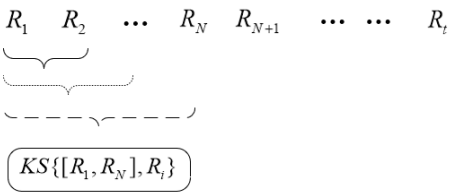


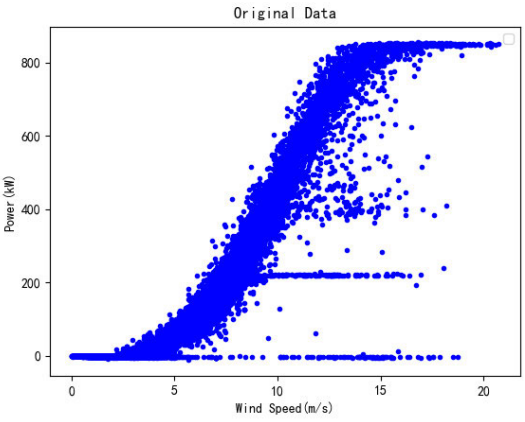
FIGURE 8. Diagram of subset selection.

Specific steps for setting fault threshold are as follows:

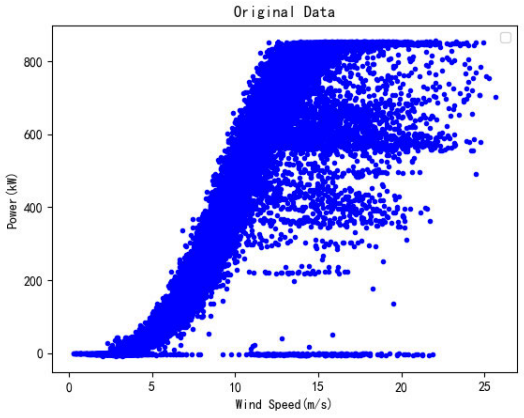
1) Set the window size and select the smallest data subset that can reflect the characteristics of the original data set as the sliding window. According to the K-S test principle, if the result value of the K-S test of two data sets is greater than 0.05, the two data sets can be considered to have the same distribution law. As shown in Figure 8, when the size of the sliding window is determined, a specific range of data is first selected from the beginning of the data sample as the sub-data set, and the K-S test is conducted with the original data set to check the similarity between the two. Then extend the range of the data subset to the right in turn until the value of K between the subset and the parent set is greater than 0.05, and the length of the recorded subset is the window size N.
2) Set the fault threshold. When setting the dynamic fault threshold, the changing trend of the state index in the previous period of time should be fully considered, and the data in the sliding window determined in Step (1) should be selected. The threshold value in the window data should be calculated according to Equation 10:

$$R = \frac{w_A r_A + w_B r_B + w_C r_C}{w_A + w_B + w_C} \tag{10}$$

In Equation 10, R_k is the state indicator at a specific moment, N is the sliding window size, and R_{th} is the upper limit of the confidence interval. R_{th} analyzes the distribution characteristics of wind turbine operating state indicators through the kernel density estimation method and sets them based on the principle of small probability events. According to the theory of interval estimation in statistics, let's say that the probability is α. If the cumulative probability distribution of state indicators within a specific range is $P\{0 \le R \le R_{th}\} = 1 - \alpha$, then interval $[0, R_{th}]$ is said to be the confidence interval of $1 - \alpha$ confidence of state indicators R. The smaller the value of α is, the smaller the probability is when the state index value of the wind turbine is $R > R_{th}$. If $1 - \alpha$ is taken as the confidence, state index R is almost all distributed in the normal interval of $[0, R_{th}]$, so R_{th} can be used as the threshold value of the abnormal state of the wind turbine.
3) Step 3: Move the data window frame by frame and set the new threshold according to Step (2).
4) Step 4: Repeat Step 3 to get the thresholds at all times, which are connected to form an adaptive threshold curve fitting the changing trend of R_t.



(a) Data sample 1



(b) Data sample 2

FIGURE 9. Distribution diagram of experimental data "wind speed–power."

III. EXAMPLE ANALYSIS

A. ABNORMAL DATA PROCESSING EXPERIMENT

In order to test the effect of the proposed data processing scheme, two groups of data are set in this section for abnormal data processing experiments, as shown in Figure 9. As shown in Figure 9 (a), there are apparent data subjects in the sample data set, in which there are fewer noise data and more shutdown and limited electric power data. As shown in Figure 9 (b), the data of the data sample has a comprehensive and disorderly distribution range. A large number of

low-density abnormal data are distributed at the lower right corner of the data body, and there are also a small amount of power limit data and more shutdown data.

Firstly, the DBSCAN clustering algorithm parameters were adjusted to identify the low-density abnormal data of experimental data samples, as shown in Figure 10(a) and (b), respectively. Then, the results after removing the anomalous data were shown in Figures 10 (c) and (d).As can be seen from Figures 10 (c) and (d), the DBSCAN clustering algorithm can well remove a large number of low-density discrete abnormal data from the original data sample. However, a small amount of downtime and power-limited data are still not removed.

Figure 11(a) and (b) are the results of setting a normal power range for data samples. The yellow star data points in the figure are the midpoint of the data density of each partition. The green curve is the ideal power curve obtained after the midpoint of the fitting density, and the red curve on the left and right sides of the data strip represents the upper and lower limits of the normal power range, respectively.

Figures 11 (c) and (d) show the results after removing abnormal data according to the normal power range. As shown in the figure, almost all types of anomalous data have been eliminated, and a full high-density normal data main band has been retained.

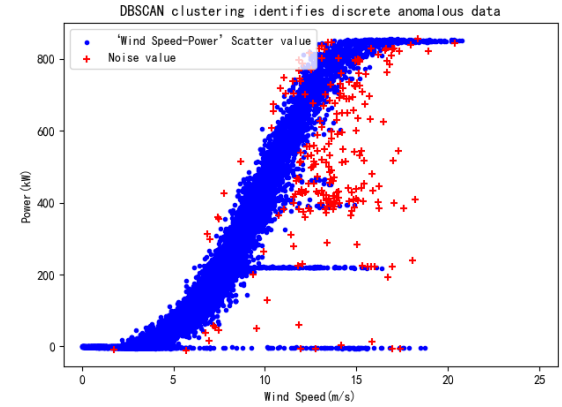
B. FAULT DIAGNOSIS AND FAULT FEATURE ANALYSIS

In order to verify the validity and reliability of the fault diagnosis model, a comparative experiment is designed in this summary to compare the performance of different hyper-parameter optimization algorithms and multi-classification algorithms with the fault diagnosis model built in this chapter. The commonly used parameter optimization methods include grid search and random search. The parameter tuning task of the XGBoost classification model is carried out according to the super-parameter range set in Table 3. The classification accuracy and optimization time of the model test set are taken as evaluation indexes, performance comparison results of the three-parameter optimization methods are shown in Table 4:

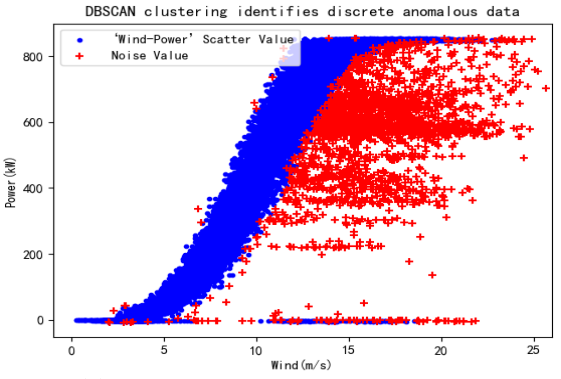
TABLE 4. Comparison of three optimization algorithms.

Optimization Algorithm	Optimization time/s	Accuracy/%
BOA	1214	99.217%
The grid search	8406	99.217%
Random search	3687	99.209%

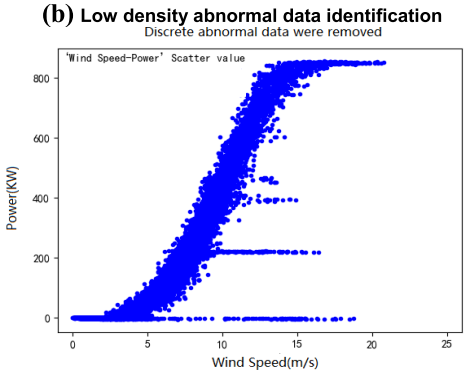
As can be seen from the above table, the accuracy of the XGBoost classification model can be adjusted to more than 99.2% by the three-parameter optimization algorithms. However, the optimization time varies greatly, and the BOA algorithm uses the least time, which is about 1/3 of the random search algorithm and 1/7 of the grid search algorithm. Therefore, the BOA algorithm has higher efficiency of super-parameter optimization. Using BOA to determine the XGBoost classifier's hyperparameters are shown in Table 5:



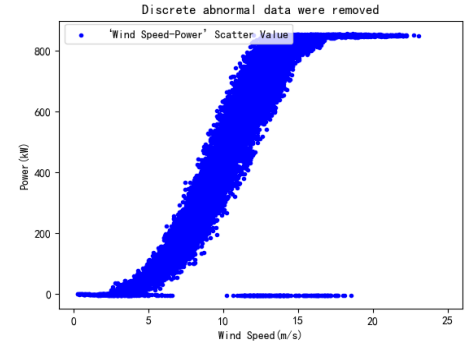
(a) Low density anomaly data identification



(b) Low density abnormal data identification



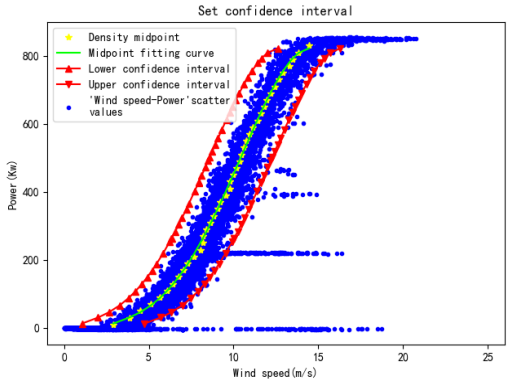
(c) Abnormal low-density data were removed



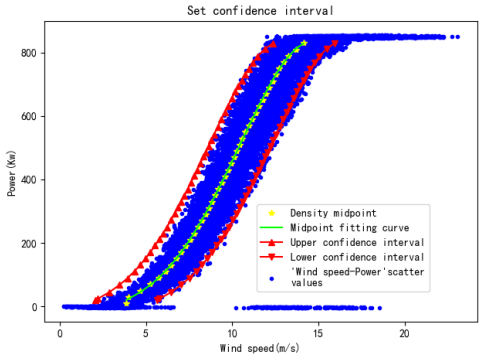
(d) Abnormal low-density data were removed

FIGURE 10. Recognition and elimination of low density abnormal data.

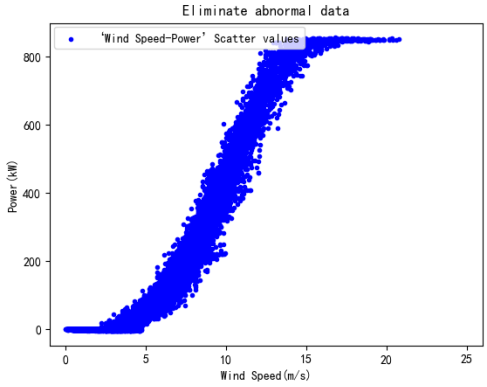
The commonly used classification algorithms include decision tree, SVM, GBDT, Adaptive Boosting (Adaboost) algorithm, and deep learning network represented by DBN.BOA



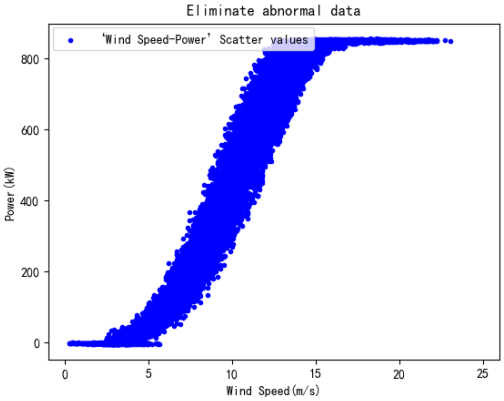
(a) Data sample 1 normal power interval setting



(b) Data sample 2 normal power interval setting



(c) Data sample 1 abnormal power data was removed



(d) Data sample 2 abnormal power data were removed

FIGURE 11. Recognition and elimination of abnormal power data.

TABLE 5. Final parameter list.

Parameter Name	The parameter value
n_estimatores	240
max_depth	6
min_child_weight	1
gamma	0.1
subsample	0.8
learning_rate	0.15

was used to build the optimal multi-classification model of each classification algorithm, in which the DBN structure was determined to be a 3-layer neural network with 1024-100-100 structure through experiments. In the test stage, in order to eliminate the influence of algorithm randomness, ten groups of different test data were randomly selected during the experiment, and the mean value of accuracy and training duration of each model was calculated as the evaluation indexes of the model reliability and effectiveness. The experimental results are shown in Table 6:

From the analysis of the above table, it can be seen that the classification model based on a decision tree takes the shortest time to train but has the lowest classification accuracy. The classification accuracy of both GBDT and XGBoost is above 99%, but the training time of XGBoost is short. Compared with XGBOOST, the effectiveness and reliability of AdaBoost and SVM algorithms to build multi-classification models are flawed. DBN-based multi-classification model's average accuracy can reach 95%, but its training time is the longest, and its effectiveness is the worst. Therefore, the multi-classification model based on the XGBoost algorithm has better efficacy and reliability compared with other classification algorithms.

In order to test the performance of the fault diagnosis model in practical application, this section uses the SCADA data of the wind farm in 2020 to simulate the real-time data stream. It tests the accuracy of the fault diagnosis model in judging the running state of wind turbines. The simulated real-time data stream data constructed, including normal data and various faults, are shown in Table 7:

After the data set is dispersed, randomly selected data are input into the fault diagnosis model to judge the state of the unit. The test results are shown in Table 8:

TABLE 6. Comparison of multiple classification algorithms.

Classification algorithm	Training time per second	Accuracy
XGBoost	0.623	99.217%
decision-making tree	0.392	82.535%
SVM	1.829	90.472%
GBDT	4.743	99.036%
AdaBoost	2.624	85.241%
DBN	26.535	95.686%

TABLE 7. Data composition of simulated real-time data stream.

SCADA system record type	Data volume	The unit state	Identifier
Normal data set	500	normal	N
High generator temperature	20	Generator system failure	F1
Excessive generator speed	20		
Speed mismatch between rotor	20		
Generator front bearing temperature is high	20		
Gearbox oil temperature is high	30	Gearbox system failure	F2
The gear cooler is overloaded	20		
High spindle temperature	30		
Low working pressure	30	Hydraulic system failure	F3
Hydraulic motor temperature error	30		
High hydraulic oil temperature	30		

TABLE 8. Test results of simulated real-time data flow diagnostic model.

The data type	Actual class	Number of classifications (entries) Predict N1	Predict F1	Predict F2	Predict F3
Normal	N1	500	0	0	0
High generator temperature	F1	0	20	0	0
Excessive generator speed	F1	0	20	0	0
Speed mismatch between rotor and generator	F1	0	20	0	0
High generator front bearing temperature	F1	0	20	0	0
High gearbox oil temperature	F1	0	19	1	0
Gear cooler overload	F2	0	0	20	0
High spindle temperature	F2	0	0	20	0
Low working pressure	F2	0	0	20	0
Hydraulic motor temperature error	F2	0	0	20	0
High hydraulic oil temperature	F3	0	0	0	20

It can be seen from the above table that the fault diagnosis model built in this chapter can basically accurately identify all fault types. The excess-high front bearing temperature of the generator is rarely misjudged as gearbox system failure. This is because in the transmission system structure of wind turbines, the front bearing of the generator is directly connected to the high-speed shaft of the gearbox, and the local high temperature of generator and gearbox often occur together, leading to some data characteristics that are not obvious.

C. FAULT WARNING EXPERIMENT

1) NORMAL REGRESSION MODEL CONSTRUCTION

According to the fault feature analysis and Eq.8, the fault feature parameters and feature weights of each subsystem are determined, as shown in Table 9:

The MIC algorithm was used to calculate the maximum mutual information coefficient between each fault characteristic parameter in Table 9 and other monitoring parameters of the SCADA system. The modeling vector of each fault characteristic parameter was selected according to the correlation.

TABLE 9. Fault characteristic parameters and feature weight ratio.

Unit subsystem	Fault characteristic parameters	Feature weight ratio
The generator system	Generator stator winding temperature V1	36%
	Generator drive side bearing temperature	32%
	Generator slip ring temperature	32%
The gearbox system	Gearbox oil temperature	49%
	Gearbox cooler power	29%
	Gearbox high speed bearing temperature	22%
The hydraulic system	Oil temperature of hydraulic system	48%
	Hydraulic system motor temperature	31%
	Working pressure of hydraulic system	21%

First, the SCADA historical data used to establish the normal model was removed from abnormal data. The health data set was obtained to build the normal model of each fault characteristic parameter. Then, the Bayesian optimization algorithm was used to find the optimal parameter combination of each regression model. Finally, the *MAE* and r^2 scores of the validation set were used as the evaluation criteria. The results are shown in Table 10. It can be seen that all kinds of regression models of characteristic fault parameters constructed in the end have similar scores, can reconstruct typical fault parameters accurately, and can well fit the variation trend of various parameters.

2) CALCULATION OF HEALTH STATUS INDICATOR AND SETTING OF FAILURE THRESHOLD

SCADA data of a wind turbine in normal operation that did not participate in regression model training and testing were selected from this wind farm. The data of uninterrupted operation for 14 days (20160 sampling points) without any failure or human intervention were intercepted as the health history data set.

Firstly, the normal regression model was used to reconstruct the characteristic parameters of each fault. Then the relative error of the reconstruction results is calculated. Finally, the status index trend control chart of the generator system, gearbox system, and hydraulic system under normal operation of the wind turbine is calculated, as shown in Figure 12:

It can be seen from the figure that most state indicators of the generator subsystem are distributed below 0.01 during operation, and there are several fluctuations in the middle, but the fluctuation range is maintained within 0.025. While the transmission system and the hydraulic system operated stably during this period, and most of the state indexes were distributed within 0.008. In addition, the state indexes of the three subsystems did not show an obvious change trend during the 14-day operation.

3) FAULT THRESHOLD SETTING

The health status index data distribution of each subsystem was counted, and the frequency distribution histogram, kernel density function curve, and cumulative probability curve are

TABLE 10. Final validation results of each model.

Regression model	MAE	r^2 score
Generator stator winding temperature V1	0.964	0.896
Generator drive side bearing temperature	0.844	0.889
Generator slip ring temperature	0.861	0.878
Gearbox oil temperature	0.749	0.894
Gearbox cooler power	0.763	0.893
Gearbox high speed bearing temperature	1.106	0.895
Oil temperature of hydraulic system	0.814	0.896
Hydraulic system motor temperature	1.318	0.894
Working pressure of hydraulic system	0.929	0.896

calculated, as shown in Figure 13. The abscissa in the figure is the state index of the system. The left ordinate is the index distribution density, which corresponds to the figure's kernel density curve and index distribution histogram. The cumulative probability curve is obtained by integrating the kernel density curve in segments and connecting probability points, which corresponds to the dotted blue line in the figure, and its coordinate is the vertical coordinate on the right.

Set the confidence of 99.5%, and find the corresponding state index when the cumulative probability density of each subsystem is 99.5% from the above figure as the upper limit of the normal state index. Finally, the upper limit of the confidence interval of the generator system is set as 0.0183, the upper limit of the confidence interval of the gearbox system is set as 0.121, and the upper limit of the confidence interval of the hydraulic system is set as 0.0125.

4) GENERATOR SYSTEM FAILURE

According to the SCADA system fault record sheet, Unit 84 in the wind farm was shut down at 17:9 on May 15, 2019, after the SCADA system sent out the fault alarm of "high temperature of generator spindle." The operation data of the wind turbine seven days before the failure time was selected to do the generator system fault warning experiment. The normal regression model was used to reconstruct the characteristic fault parameters and calculate the relative error. The state detection diagram of each subsystem was calculated, as shown in Figure 14:

As shown in Figure 14 (a), the status index of the generator system presents a protrusive amplitude near the 580th sampling point but does not exceed the fault threshold. This indicates that during this period of time, the operating state of the generator system fluctuated wildly, but it did not reach the extent of failure. However, after the 950th sampling point, the status index rose rapidly and gradually exceeded the alarm threshold. Figure 14 (b) is a partially enlarged view of the generator system state detection diagram, in which the red dotted line is the fault occurrence point (the 1008th sampling point), and the yellow dotted line is the fault warning point (the 986th sampling point). This indicates that the monitoring system can send out the generator system fault alarm about 3.67 hours in advance (22 sampling points). In comparison, the state indicators of the gearbox system and the hydraulic system, as shown in Figure 14 (c) and (d), are relatively

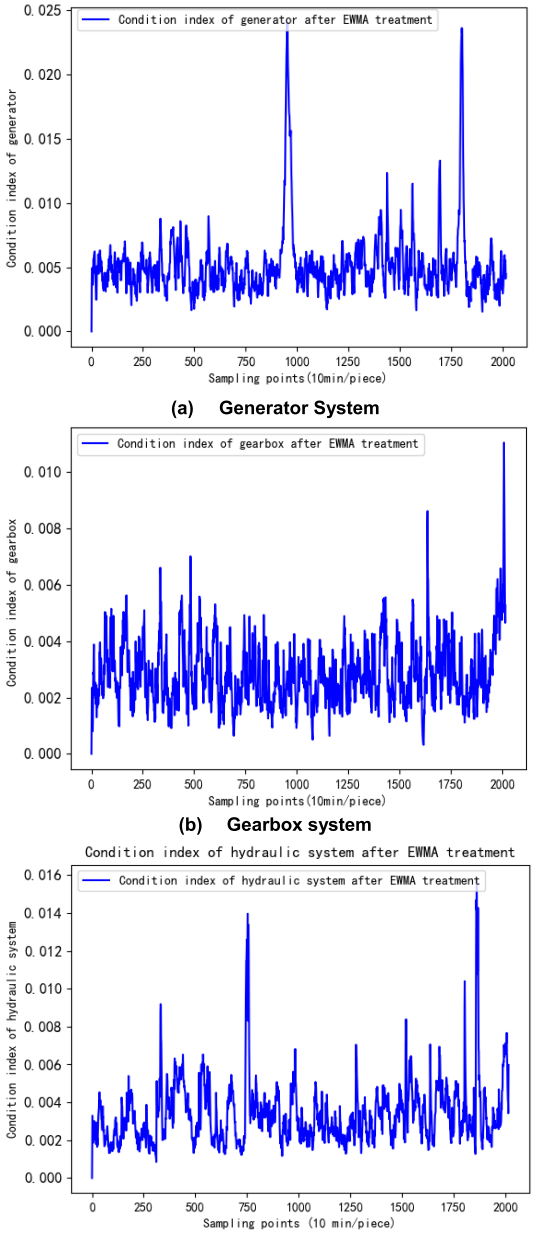


FIGURE 12. Trend chart of health status indicators of each subsystem.

smooth on the whole. After the 950th sampling point, the state indicators show a slight increase but do not exceed the fault threshold.

5) GEAR BOX SYSTEM FAILURE

The SCADA system fault record shows that Unit 87 in the wind farm was shut down at 9:39 on April 22, 2019, after the SCADA system issued a fault alarm of "gearbox cooler overload." The operation data of the wind turbine seven days before the failure time was selected to do the gearbox system fault warning experiment. The normal regression model was used to reconstruct each characteristic fault parameter and calculate the relative error. Finally, the state detection diagram of each subsystem was obtained, as shown in Figure 15:

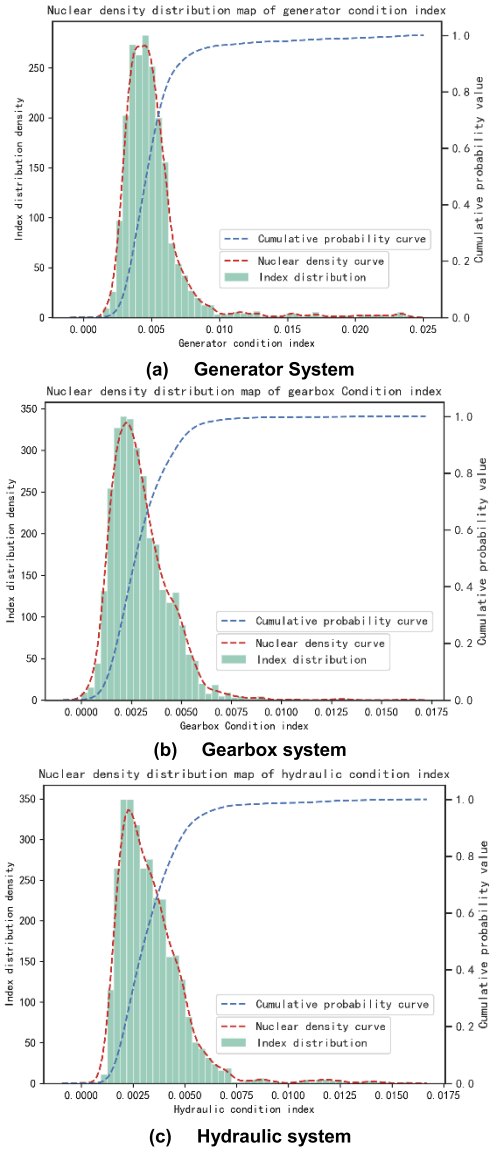


FIGURE 13. Density distribution diagram of state indicators of each subsystem.

As can be seen from Figure 15(a), the operation index of the gearbox system is relatively stable on the whole, and the state index rises rapidly from the 940th sampling point and gradually exceeds the fault threshold. Figure 15(b) is a partially enlarged view of the state detection 10 hours before the failure of the generator system. The red dotted line in the figure shows the fault occurrence point (the 1008th sampling point). The yellow dotted line shows the fault warning point (the 977th sampling point). It can be seen from the figure that the condition monitoring system can send out the generator system fault alarm about 5.17 hours in advance (31 sampling points). Figure 15 (c) is the generator system condition monitoring diagram. During the whole operation period, no-fault threshold was exceeded. Figure 15 (d) is the condition monitoring diagram of the hydraulic system. As can

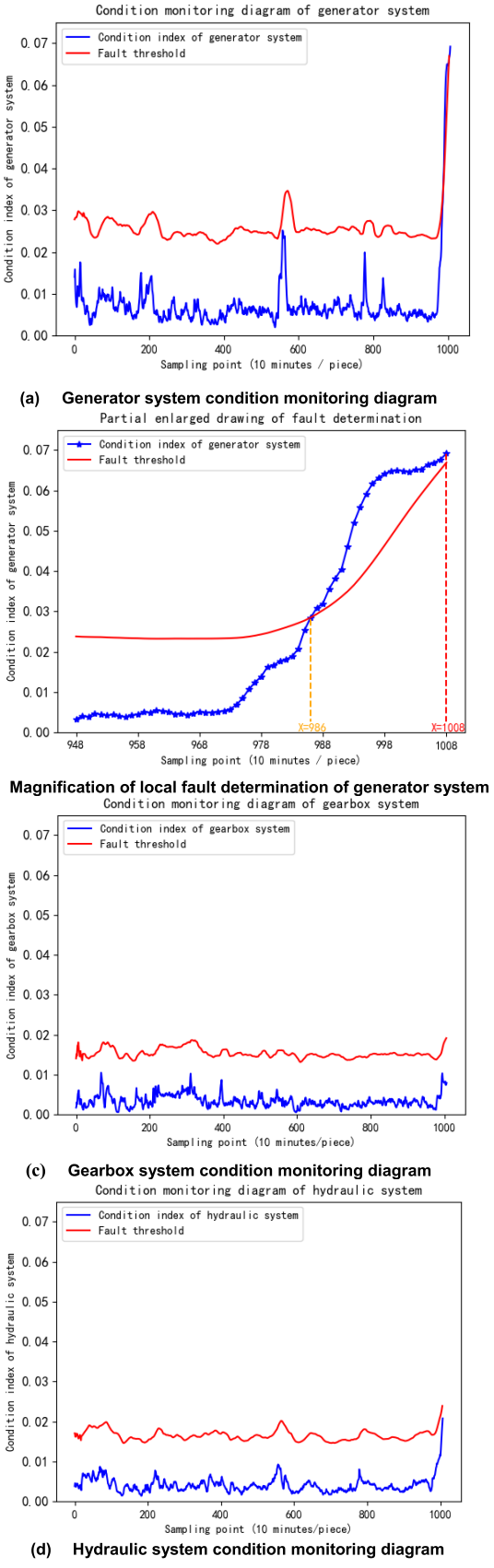
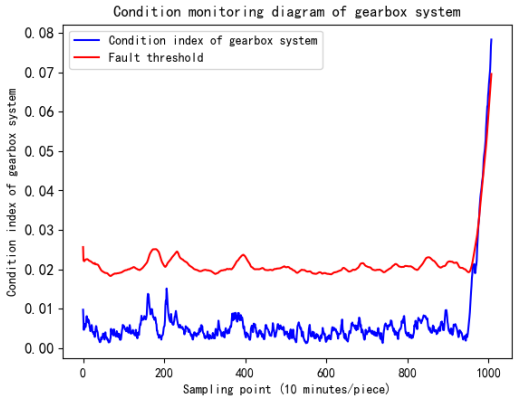
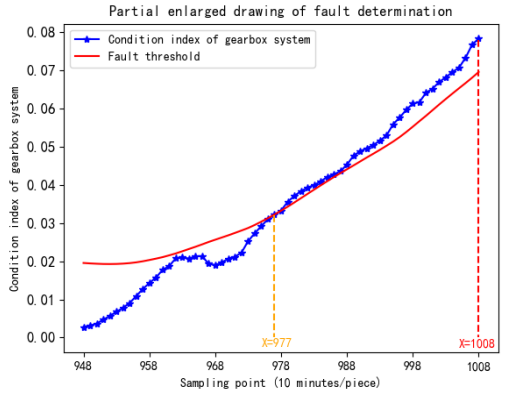


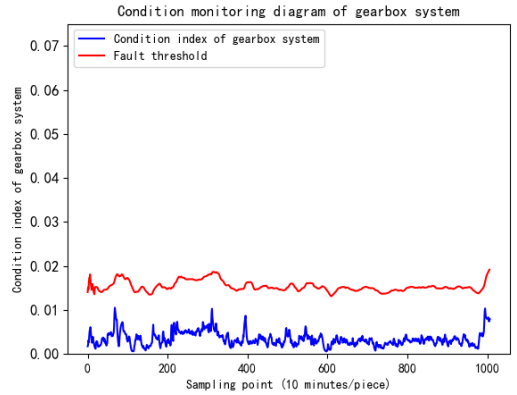
FIGURE 14. Monitoring diagram of subsystems.



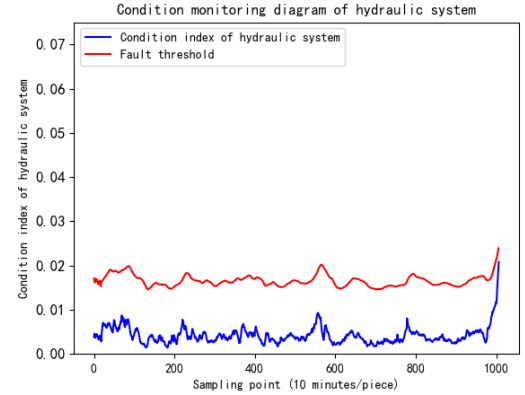
(a) Gearbox system condition monitoring diagram



(b) Local enlargement of gearbox system fault determination

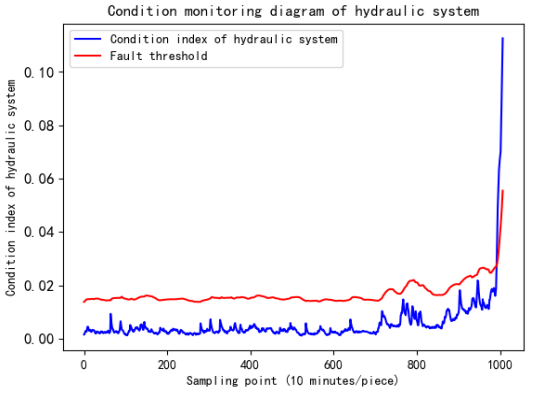


(c) Generator system condition monitoring diagram

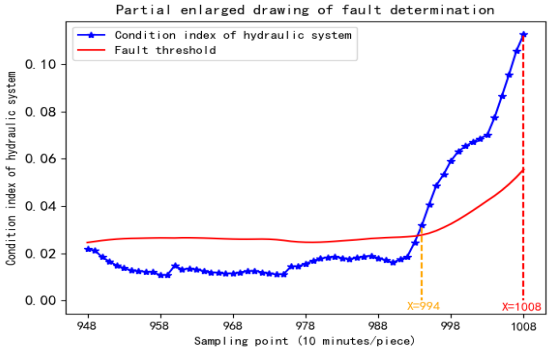


(d) Hydraulic system condition monitoring diagram

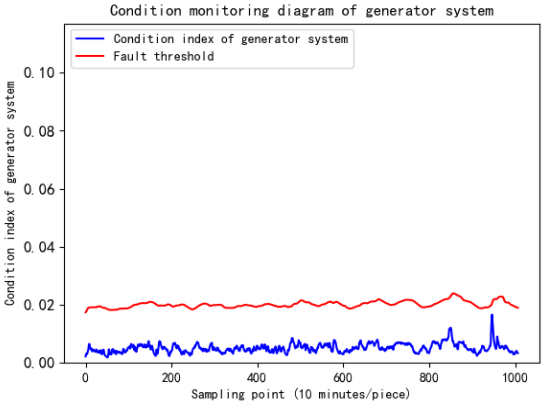
FIGURE 15. Status monitoring diagram of subsystems.



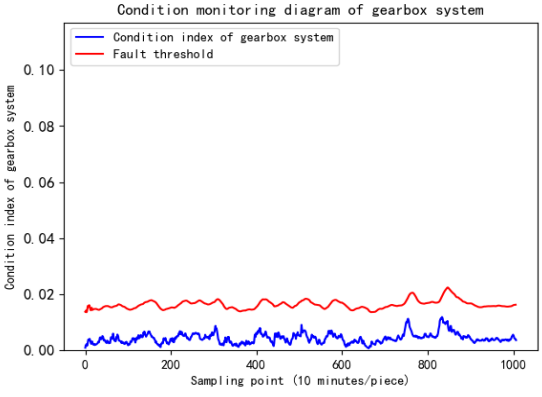
(a) Hydraulic system condition monitoring diagram



(b) Magnification of partial fault determination of hydraulic system



(c) Generator system condition monitoring diagram



(d) Gearbox system condition monitoring diagram

FIGURE 16. Status monitoring diagram of subsystems.

be seen from the figure, the hydraulic system was affected by the temperature rise of the converter and high-speed bearing of the gearbox before the near failure, leading to a slight pull up of the state index, but it did not exceed the fault threshold.

6) HYDRAULIC SYSTEM FAILURE

The SCADA system fault record shows that Unit 106 in the wind farm was shut down at 15:36 on April 26, 2019, after the SCADA system sent out the ''hydraulic motor temperature is too high'' fault alarm. The operation data of the unit 7 days before the time of failure were selected to do the hydraulic system fault warning experiment. The normal regression model was used to reconstruct each characteristic fault parameter and calculate the relative error. The state monitoring diagram of each subsystem is shown in Figure 16:

As shown in Figure 16(a) and (b), the state detection diagram of the hydraulic system and the locally enlarged diagram can be obtained: The state index of the hydraulic system shows a jitter rising trend from the 850th sampling point, and it starts to increase substantially from the 990th sampling point until the moment of failure, and its state index increases to around 0.11. In Figure 16(b), the red dotted line is the fault occurrence point (the 1008th sampling point), and the yellow dotted line is the fault warning point (the 994th sampling point), indicating that the condition monitoring system can send out the generator system fault alarm about 2.33 hours in advance (14 sampling points). By comparing Figure 16 (c) and (d), it can be seen that both the generator system and the gearbox system run smoothly, and no alarm occurs during the experimental period.

IV. SUMMARIZES

Reliable condition monitoring and fault diagnosis technology are of great significance to the operation and maintenance of wind turbines. It can not only monitor the state parameters of wind turbines in real-time, grasp the health information of wind turbines, but also find out the potential failure symptoms as early as possible, reduce the failure rate, ensure the safe and efficient operation of large wind turbines, and then promote the development of new energy industry. According to the characteristics of wind turbine operation data distribution, this paper designs an abnormal data processing scheme combining the DBSCAN clustering algorithm and normal power interval estimation. In this scheme, the DBSCAN clustering algorithm is first used to remove the outlier noise anomaly data with low density in the original data. Then the normal data interval is set based on the least square method and 3-Sigma criterion. Finally, the abnormal data outside the interval are eliminated to get the health data. The experimental results show that the anomalous data processing scheme proposed in this paper can effectively deal with all known abnormal data types and provide high-quality health data samples for subsequent studies. In this paper, a fault diagnosis model is built, and the characteristic parameters of different faults are analyzed. Aiming at the difficulty of super parameter tuning in the eXtreme Gradient Boosting (XGBoost) algorithm modeling, a fault diagnosis model of BOA-XGBoost based on the Bayesian optimization algorithm (BOA) is designed. The BOA algorithm is used to find the optimal super parameter combination of the XGBoost algorithm model, and the fault diagnosis of wind turbine generator system, gearbox system, and hydraulic system is realized efficiently. In order to further analyze the correlation between fault information and unit monitoring parameters, a feature weight measurement method based on a tree model was studied. The number of times that the monitoring parameters were used as splitting features in the construction of the classification tree model was used as the feature weight to complete the intuitive mapping from SCADA system monitoring data to fault features. Finally, according to the importance of fault feature monitoring parameters, a wind turbine condition monitoring scheme based on the information fusion of multi-feature monitoring parameters is designed. The early fault characteristics of wind turbines are identified by real-time monitoring whether the operating state index of the turbine exceeds the fault threshold. Firstly, the BOA-XGBOOST algorithm was used to build a normal regression model for multiple fault characteristic parameters, and the reconstruction error of each characteristic parameter was calculated in real-time. Then, the typical parameters from different sources and different scales were fused into the operating state index according to the characteristic weight, and the consistent expression of the functional state of the wind turbine was obtained from many monitoring parameters. Finally, a dynamic fault threshold design scheme based on the adaptive principle is designed. The fault threshold is set in sections by sliding window, which fully considers the operation situation of the unit in the previous period of time, and solves the problems of solid subjectivity, weak generalization ability, and easy false alarm in setting the fixed threshold artificially. The experimental results show that the designed condition monitoring scheme can warn generator system faults 3.67 hours in advance, gearbox system faults 5.17 hours in advance, and hydraulic system faults 2.33 hours in advance.

REFERENCES

[1] S. Chen, Y. Ma, L. Ma, F. Qiao, and H. Yang, ''Early warning of abnormal state of wind turbine based on principal component analysis and RBF neural network,'' in *Proc. 6th Asia Conf. Power Electr. Eng. (ACPEE)*, Apr. 2021, pp. 547–551, doi: 10.1109/ACPEE51499.2021.9437063.

[2] M. Kenisarin, V. M. Karslı, and M. Çağlar, ''Wind power engineering in the world and perspectives of its development in Turkey,'' *Renew. Sustain. Energy Rev.*, vol. 10, pp. 341–369, Aug. 2006.

[3] N. Golait, R. M. Moharil, and P. S. Kulkarni, ''Wind electric power in the world and perspectives of its development in India,'' *Renew. Sustain. Energy Rev.*, vol. 13, no. 1, pp. 233–247, Jan. 2009.

[4] S. Tao, Z. Qian, Y. Pei, A. Wang, and F. Zhang, ''Wind turbine failure detection based on SCADA data and data mining method,'' in *Proc. 8th Renew. Power Gener. Conf. (RPG)*, 2019, Art. no. 61573046.

[5] G. Xiong, Y. Zhang, P. Yang, Z. Zhu, and P. Hu, ''An analysis of the burning accident of the tower bottom cabinet of G58-850 KW wind turbine group,'' *IOP Conf. Ser., Earth Environ. Sci.*, vol. 508, Jul. 2020, Art. no. 012054.

[6] Y. Lin, L. Tu, H. Liu, and W. Li, ''Fault analysis of wind turbines in China,'' *Renew. Sustain. Energy Rev.*, vol. 55, pp. 482–490, Mar. 2016.

[7] J. Maldonado-Correa, S. Martín-Martínez, E. Artigao, and E. Gómez-Lázaro, "Using SCADA data for wind turbine condition monitoring: A systematic literature review," *Energies*, vol. 13, no. 12, p. 3132, Jun. 2020.
[8] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587.
[9] N. Li, L. Meng, B. Geng, and Z. Jing, "Power plant fan fault warning based on bidirectional feature compression and state estimation," in *Advances in Intelligent Information Hiding and Multimedia Signal Processing*. Singapore: Springer, 2020.
[10] K. Lu, S. Gao, W. Sun, Z. Jiang, X. Meng, Y. Zhai, Y. Han, and M. Sun, "Auto-encoder based fault early warning model for primary fan of power plant," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 358, Dec. 2019, Art. no. 042060.
[11] H. Hermawan and W. Caesarendra, "An investigation of shaft failure on induced draft fan in a steam power plant," *J. Phys., Conf. Ser.*, vol. 1845, no. 1, Mar. 2021, Art. no. 012080.
[12] Y. Yang, Y. Bai, C. Li, and Y.-N. Yang, "Application research of ARIMA model in wind turbine gearbox fault trend prediction," in *Proc. Int. Conf. Sens., Diagnostics, Prognostics, Control (SDPC)*, Aug. 2018, pp. 520–526.
[13] K. B. Abdusamad, D. W. Gao, and E. Muljadi, "A condition monitoring system for wind turbine generator temperature by applying multiple linear regression model," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2013, pp. 1–8.
[14] P. Bangalore and L. B. Tjernberg, "An artificial neural network approach for early fault detection of gearbox bearings," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 980–987, Mar. 2015.
[15] Y. Yang, A. Liu, H. Xin, and J. Wang, "Fault early warning of wind turbine gearbox based on multi-input support vector regression and improved ant lion optimization," *Wind Energy*, vol. 24, no. 8, pp. 812–832, Aug. 2021.
[16] M. Sawant, S. Thakare, A. P. Rao, A. E. Feijóo-Lorenzo, and N. D. Bokde, "A review on state-of-the-art reviews in wind-turbine-and wind-farm-related topics," *Energies*, vol. 14, no. 8, pp. 1–30, 2021, doi: 10.3390/en14082041.
[17] A. Adouni, D. Chariag, D. Diallo, M. B. Hamed, and L. Sbita, "FDI based on artificial neural network for low-voltage-ride-through in DFIG-based wind turbine," *ISA Trans.*, vol. 64, pp. 353–364, Sep. 2016.
[18] S. Qin, K. Wang, X. Ma, W. Wang, and M. Li, "An improved SVM based wind turbine multi-fault detection method," in *Proc. Int. Conf. Pioneering Comput. Sci., Eng. Educ.*, in Communications in Computer and Information Science, vol. 727, 2017, pp. 27–38.
[19] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017.
[20] R. Kumar, M. Ismail, W. Zhao, M. Noori, A. R. Yadav, S. Chen, V. Singh, W. A. Altabey, A. I. H. Silik, G. Kumar, J. Kumar, and A. Balodi, "Damage detection of wind turbine system based on signal processing approach: A critical review," *Clean Technol. Environ. Policy*, vol. 23, no. 2, pp. 561–580, Mar. 2021, doi: 10.1007/s10098-020-02003-w.
[21] Z. Hou, X. Lv, and S. Zhuang, "Monitoring and analysis of wind turbine working status based on LSSVR," in *Proc. 3rd World Conf. Mech. Eng. Intell. Manuf. (WCMEIM)*, Dec. 2020, pp. 18–21, doi: 10.1109/WCMEIM52463.2020.00011.
[22] T. H. Kim and J. P. Haldar, "Efficient iterative solutions to complex valued nonlinear least-squares problems with mixed linear and antilinear operators," *Optim. Eng.*, Feb. 2021, doi: 10.1007/s11081-021-09604-4.
[23] K. T. Abd-Elwahab and A. A. Hassan, "SCADA data as a powerful tool for early fault detection in wind turbine gearboxes," *Wind Eng.*, Dec. 2020, doi: 10.1177/0309524X20969418.

YILONG SHI received the B.E. degree in electronic information science and technology from Anhui Engineering University, in 2019. He is currently pursuing the postgraduate degree with the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China. His current research interests include wind turbine fault warning and diagnosis.

YIRONG LIU received the B.E. degree in engineering from Shandong University of Science and Technology, in 2018, and the M.E. degree in electronic and communication engineering from Shandong University, China, in 2021. with State Grid Shandong Electric Power Company. His research interests include microgrid and wind turbine fault warning and diagnosis.

XIANG GAO received the B.S., master's, and Ph.D. degrees in information and communication, in 2006, 2008, and 2012, respectively. He is currently the Director of the Department of Electronics, School of Mechatronics and Information Engineering, Shandong University. His current research interests include wireless communication and 5G technology.