Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 2091382, 10 pages
https://doi.org/10.1155/2020/2091382

Research Article

Fault Detection of Wind Turbine Pitch System Based on Multiclass Optimal Margin Distribution Machine

Mingzhu Tang,[1,2] **Zijie Kuang**,[1] **Qi Zhao**,[1] **Huawei Wu**,[2] **and Xu Yang**[3]

[1]*School of Energy and Power Engineering, Changsha University of Science & Technology, Changsha 410114, China*
[2]*Hubei Key Laboratory of Power System Design and Test for Electrical Vehicle, Hubei University of Arts and Science, Xiangyang 441053, China*
[3]*State Grid (Beijing) Integrated Energy Service Company Limited, Beijing 100176, China*

Correspondence should be addressed to Huawei Wu; whw_xy@hbuas.edu.cn

Received 11 June 2020; Accepted 19 June 2020; Published 4 August 2020

Guest Editor: Yong Chen

In response to the unbalanced sample categories and complex sample distribution of the operating data of the pitch system of the wind turbine generator system, this paper proposes a method for fault detection of the pitch system of the wind turbine generator system based on the multiclass optimal margin distribution machine. In this method, the power output of the wind turbine generator system is used as the main status parameter, and the operating data history of the wind turbine generator system in the wind power supervisory control and data acquisition (SCADA) system is subject to correlation analysis with the Pearson correlation coefficient, to eliminate the features that have low correlation with the power output status parameter. Secondary analysis is performed to the remaining features, thus reducing the number and complexity of samples. Datasets are divided into the training set for training of the multiclass optimal margin distribution machine fault detection model and test set for testing. Experimental verification was carried out with the operating data of one wind farm in China. Experimental results show that, compared with other support vector machines, the proposed method has higher fault detection accuracy and precision and lower false-negative rate and false-positive rate.

1. Introduction

Wind turbine generator systems are usually used in complex and unstable natural environments and eroded by sunlight, rain, wind, and sand all the year round. In addition, the wind turbine generator systems work at high altitudes, and their main parts are in high-altitude nacelles, which may lead to faults during operation. Long downtime of the wind turbine generator system arising from failure will result in a lot of operation and maintenance costs and part replacement costs, low power generation efficiency of the wind farm, and huge economic losses [1].

The pitch system is a critical part of the wind turbine generator system, mainly consisting of the blades, hubs, and other parts. These parts account for a large proportion in terms of the average maintenance time, material costs, and corresponding technical personnel [2]. Hence, it is of particular importance to guarantee the safe and stable operation of the pitch system of the wind turbine generator system. Timely and efficient status monitoring and fault detection of the pitch system has excellent economic benefits and engineering application values for the wind power industry [3].

The current fault detection of wind turbine generator systems is mainly based on the data analysis of the wind power supervisory control and data acquisition (SCADA) system. A correlation model is built by analyzing the data (e.g., power, vibration, and temperature) generated in the operation of the wind turbine generator system, to obtain the operating status, fault, and other information of the wind turbine generator system and thus detecting faults [4].

Fault detection mainly involves two aspects: feature selection and detection model [5–7]. The status parameters reflecting the faults of the wind turbine generator system are

selected from the SCADA system, and the detection model is built through training, which is used for status monitoring and fault detection of the wind turbine generator system. Pandit and Infield proposed a method of status monitoring of the wind turbine generator system based on the Gaussian process [8]. The wind power curve is predicted according to the data of the SCADA system and used for yaw fault detection of the wind turbine generator system. Liu et al. presented a method of wavelet transform fault detection based on the generative adversarial network, in which the normal operating data of the wind turbine generator system is converted into rough fault data based on the prior knowledge, and a generative adversarial network model is built for fault detection [9]. Ruiming et al. proposed a method based on SCADA data and dynamical network marker, which is constructed as a fault warning signal of a wind turbine [10]. The techniques including the multinode complex network and the correlation and cross-correlation analysis of the denosed method, and the experiment verifies its convenience and robustness. However, the excessive use of the feature parameters based on artificial experience will introduce human influencing factors into fault detection, resulting in interference in the detection process. Due to the particularity of the SCADA system, the operating data of the wind turbine generator system may be missing or abnormal, and it is difficult to extract effective features from a large amount of raw data, which may lead to low efficiency [11]. Furthermore, the current SCADA system is not yet mature, which may involve strong coupling of status parameters. The use of these parameters will lead to redundancy and ultimately model overfitting. Hence, more potential of SCADA data needs to be explored [12].

The support vector machine (SVM), as a machine learning method based on the statistical theory, has good learning performance. It has been successfully applied in many fields such as multiclass recognition and regression forecasting [13–15] and favored by a large number of scholars in the field of fault research on wind turbine generator systems, including fault diagnosis and prediction of wind turbine generator systems via the SVM [16–18]. Liu et al. presented the diagonal spectrum and clustering binary tree are combined with the SVM for fault detection of the gearboxes of wind turbine generator systems [19]. Hang et al. proposed a method of fault diagnosis of wind turbine generator systems based on the multilevel fuzzy SVM classifier, in which the fault feature vectors are extracted from vibration signals utilizing empirical mode decomposition, and the kernel function parameters of the fuzzy clustering algorithm are optimized, and the faults of wind turbine generator systems are diagnosed via the multilevel fuzzy SVM [20]. Saari et al. were to detect and identify wind turbine bearing faults by using fault-specific features extracted from vibration signals. Automatic identification was achieved by training models by using these features as an input for a one-class support vector machine [21]. In the SVM, however, classification is based on the identification of the hyperplane with the minimum margin, leading to low generalization performance and, in the case of complex nonlinear multiclassification, final optimization may become a nondifferentiable nonconvex process [22]. In order to solve this problem, Zhang and Zhou put forward the multiclass optimal distribution machine (mcODM), in which a distribution model is built based on the sample distribution features during fault detection. The sample mean and sample variance are taken into account for higher classification performance [23]. The experiments of multiple datasets have verified the accuracy and generalization performance of this model and the model complexity is relatively low during the optimization.

To resolve the unbalanced samples and complex distribution in fault detection of the pitch system of the wind turbine generator system, a method for fault detection of the pitch system of the wind turbine generator system based on mcODM is proposed. This method mainly consists of three parts. First, the SCADA data of the wind turbine generator system are preprocessed, including data cleaning and normalization. Secondly, the correlation of parameters is analyzed according to the operation mechanism of the wind turbine generator system and the Pearson correlation coefficient, followed by feature selection. Finally, sample sets are built, including the training set for training of the detection model and the test set for testing of this model, using the actual operating data of one wind farm in China as experimental data. Experimental results show that this method has higher accuracy and precision of fault detection and lower false-negative rate and false-positive rate.

2. Pitch System of Wind Turbine Generator System

The pitch system of the wind turbine generator system is used to change the upwind area of the blades when the rotor is facing the wind, thus controlling the rotation torque of the rotor. In combination with the yaw system, the wind turbine generator system can maintain the stable efficiency of power generation under different wind conditions [24]. At present, the pitch systems of wind turbine generator systems are mainly divided into the hydraulic pitch system and electric pitch system.

The hydraulic pitch system is equipped with a set of crank sliding structure to drive all blades for synchronous pitching. This system has a fast response to pitch signals and large pitch torque, which is conducive to the centralized layout and integration. It is mostly used in large-sized wind turbine generator systems. However, it is a nonlinear system that has a relatively complex structure and may be subject to hydraulic oil leakage, jamming, etc. [25].

The electric pitch system is equipped with an independent control mechanism for each blade and composed of the pitch controller, servo driver, and standby power supply, in which the pitch of each blade is controlled separately. Its transmission features a relatively simple structure, stable operation, and high reliability, but has large inertia due to its poor dynamic features. Where the wind speed changes rapidly, frequent pitching may lead to controller overheat and damage to the body [26].

Once the pitch system of the wind turbine generator system fails, the blade pitch will be abnormal and the

rotation torque of the rotor will not be the expected value. If the speed is too low, the wind energy capture rate will be affected. The mechanical energy generated in the rotation will be transferred to the generator through the gearbox transmission chain, resulting in the abnormal speed of the generator and ultimately affecting the power output of the generator. Accordingly, the safe and stable operation of the pitch system is essential for the stable and efficient power generation of the wind turbine generator system.

During fault detection of the pitch system of the wind turbine generator system, an important step is to acquire the status parameters that effectively reflect the features of the pitch system from a lot of SCADA data. Due to the particularity of the SCADA system that involves the complex and diverse parameters of the pitch system, including strong coupling parameters, it is necessary for feature selection to optimize the model complexity to reduce the calculation time and the amount and select the effective status parameters and also to take redundant items into account to delete excess parameters and avoid model overfitting [27, 28].

The method proposed in this paper is for fault detection of the electric pitch systems of large-sized wind turbine generator systems. The experimental data are the actual operating data of a wind farm, and various categories of samples are used. The method involves the typical data category imbalance, complex distribution, and the like.

3. Fault Detection of Pitch System of Wind Turbine Generator System

Fault detection of the pitch system of the wind turbine generator system consists of the preprocessing of the operating data acquired, selection of effective features, and building of the sample sets, including training sets for the training of the detection model and the test set for testing. Figure 1 shows the mcODM-based process for fault detection of the pitch system of the wind turbine generator system.

3.1. Data Cleaning and Preprocessing. In order to obtain the fault samples of the pitch system of the wind turbine generator system, the actual operating data of the wind turbine generator system of one wind farm are used, including the sensor monitoring data during normal operation and at the failure time of the pitch system. Unstable environmental factors and sensor abnormalities under the actual operating conditions will cause information processing errors, data losses, data abnormalities, and other problems. Thus, the obtained raw data are cleaned and preprocessed as follows:

Step 1: delete the "no data" variable in the dataset

Step 2: delete all status variables with a value of "0"

Step 3: according to the fault record of the wind turbine generator system, select the data from 30 min before a fault to 30 min after the fault

Step 4: normalize the sample data by the following formula:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{1}$$

where X is a status parameter, $X_{\min}$ and $X_{\max}$ represent the minimum and maximum value of the status variable, respectively, and X' represents the normalized value.

Normalization makes the model smoother and more convergent to find the optimal solution.

3.2. Feature Selection. According to the mechanism analysis of the pitch system, when the pitch system fails, the main status parameter that is ultimately affected is the power output of the wind turbine generator system. Hence, the correlation between the power output and other operating parameters of the wind turbine generator system is analyzed based on the Pearson correlation coefficient during feature selection, to delete the parameters that are little correlated to the pitch system.

The Pearson correlation coefficient was proposed by the British statistician Karl Pearson in the 20th century. It reflects the degree of correlation between two variables and calculated by the following formula:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left((X - \mu_X)(Y - \mu_Y)\right)}{\sigma_X \sigma_Y}, \tag{2}$$

where $\operatorname{cov}(X,Y)$ represents the covariance of the two variables, and μ_X/μ_Y and σ_X/σ_Y represent the mean and standard deviation of the two variables, respectively.

The aforesaid formula defines the population correlation coefficient. When the sample size of the variables X and Y is n, the Pearson correlation coefficient is given by

$$r = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}, \tag{3}$$

where r represents the degree of linear correlation between the two variables. It ranges from -1 to 1, i.e., $-1 \le r \le 1$, as described as follows: $0 < r < 1$: the two variables are positively correlated. The closer r is to 1, the greater the positive correlation of the variables; $-1 < r < 0$: the two variables are negatively correlated. The closer r is to -1, the greater the negative correlation of the variables; $|r| = 1$: the two variables are linearly correlated; and $r = 0$: the two variables are linearly independent of each other.

In order to further reduce the sample size as well as the computational complexity of the model, and avoid model overfitting, the status variables selected in the first step are subject to a secondary Pearson correlation analysis, to delete some highly correlated parameters and resolve the redundancy.

Following the feature selection of the datasets based on the Pearson correlation coefficient, the normal and fault samples are divided into the training set for model training and test set for model performance testing.

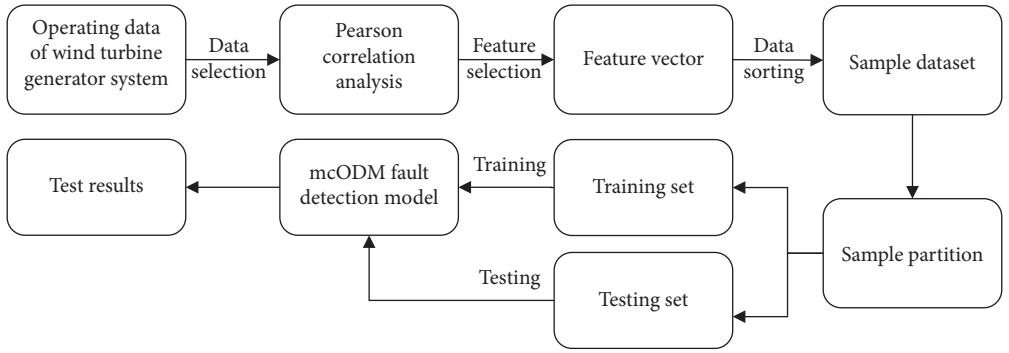

Figure 1: mcODM-based process for fault detection of the pitch system of wind turbine generator system.

3.3. mcODM Algorithm. A feature set $X = [x_1, \ldots, x_k]$ is assumed, corresponding to the category label set $Y = [K]$, where $[K] = \{1, \ldots, k\}$. The training set is $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$. The mapping function φ is defined, and the sample set is mapped by the kernel function κ to the high-dimensional space φ: $X \longrightarrow \mathrm{H}$. The corresponding weight vectors are $\omega_1, \ldots, \omega_k$. A scoring function is $\omega_l^{\mathrm{T}}\varphi(x)$ defined for each weight vector ω_l. The feature value of each sample and the corresponding label will maximize the value of the scoring function of the samples, i.e., $h(x) = \mathrm{argmax}_{l\in Y}\omega_l^{\mathrm{T}}\varphi(x)$, thereby leading to a margin definition:

$$\gamma_h(x, y) = \omega_y^{\mathrm{T}}\varphi(x) - \max_{l\neq y}\omega_l^{\mathrm{T}}\varphi(x). \tag{4}$$

When a negative margin is generated in calculation, the category provided by the classifier will be incorrect.

Let $\overline{\gamma}$ represent the mean of margins, the optimal margin distribution machine can be expressed as follows:

$$\begin{aligned} \min_{\omega,\overline{\gamma},\xi_j,\varepsilon_j} \quad & \Omega(\omega) - \eta\overline{\gamma} + \frac{\lambda}{m}\sum_{j=1}^{m}\left(\xi_j^2 + \varepsilon_j^2\right), \\ \text{s.t.} \quad & \gamma_h\left(x_j, y_j\right) \geq \overline{\gamma} - \xi_j, \\ & \gamma_h\left(x_j, y_j\right) \leq \overline{\gamma} + \varepsilon_j, \forall j, \end{aligned} \tag{5}$$

where $\Omega(\omega)$ is a regular term, η and λ are balance parameters, ξ_j and ε_j are the positive and negative deviations of the margin $\gamma_h(x_j, y_j)$ and its mean $\overline{\gamma}$, respectively, and $(1/m)\sum_{j=1}^{m}(\xi_j^2 + \varepsilon_j^2)$ is the variance.

The margin mean can be fixed at 1 by scaling of ω. The deviation of the sample (x_j, y_j) and margin mean will be $|\gamma_h(x_j, y_j) - 1|$. Then, the optimal margin distribution machine can be expressed as follows:

$$\begin{aligned} \min_{\omega,\xi_j,\varepsilon_j} \quad & \Omega(\omega) + \frac{\lambda}{m}\sum_{j=1}^{m}\frac{\xi_j^2 + \tau\varepsilon_j^2}{(1-\theta)^2}, \\ \text{s.t.} \quad & \gamma_h\left(x_j, y_j\right) \geq 1 - \theta - \xi_j, \\ & \gamma_h\left(x_j, y_j\right) \leq 1 + \theta + \varepsilon_j, \forall j, \end{aligned} \tag{6}$$

where $\tau \in [0, 1)$ is a parameter balancing two different deviations (greater or less than the margin means). $\theta \in [0, 1)$ is a zero-loss parameter which can control the number of support vectors, i.e., the sparseness of the solution. $(1-\theta)^2$ is a substitution loss used to change the aforesaid second item into a 0-1 loss function.

The regular term is $\Omega(\omega) = \sum_{l=1}^{k}(\|\omega_l\|_{\mathrm{H}}^2/2)$, and the mcODM is ultimately expressed as follows:

$$\begin{aligned} \min_{\omega_l,\xi_j,\varepsilon_j} \quad & \frac{1}{2}\sum_{l=1}^{k}\|\omega_l\|_{\mathrm{H}}^2 + \frac{\lambda}{m}\sum_{j=1}^{m}\frac{\xi_j^2 + \tau\varepsilon_j^2}{(1-\theta)^2}, \\ \text{s.t.} \quad & \omega_{y_j}^{\mathrm{T}}\varphi\left(x_j\right) - \max_{l\neq y_j}\omega_l^{\mathrm{T}}\varphi\left(x_j\right) \geq 1 - \theta - \xi_j, \\ & \omega_{y_j}^{\mathrm{T}}\varphi\left(x_j\right) - \max_{l\neq y_j}\omega_l^{\mathrm{T}}\varphi\left(x_j\right) \leq 1 + \theta + \varepsilon_j, \forall j, \end{aligned} \tag{7}$$

where λ, τ, and θ are the aforementioned balance parameters.

The parameters are selected by the grid search method, among which λ is determined in the sequence $[2^0, 2^2, 2^4, \ldots, 2^{20}]$ and τ and θ in $[0.2, 0.4, 0.6, 0.8]$.

3.4. Evaluation Criteria for Fault Detection Performance. To evaluate the fault detection performance of the model, a confusion matrix [29] is introduced, as defined in Table 1.

The following five evaluation indicators are obtained via the confusion matrix:

$$\begin{aligned} \text{accuracy} &= \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP} + \mathrm{TN}}, \\ \text{precision} &= \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \\ F1 - \text{score} &= \frac{2}{(1/\text{precision}) + (1/\text{recall})}, \\ \mathrm{FPR} &= \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}, \\ \mathrm{FNR} &= \frac{\mathrm{FN}}{\mathrm{TP} + \mathrm{FN}}. \end{aligned} \tag{8}$$

Table 1: Confusion matrix.

	Number of predicted faulty samples	Number of predicted normal samples
Number of actual faulty samples (P)	TP	FN
Number of actual normal samples (N)	FP	TN

Note. TP: the number of the samples classified as faulty samples and predicted to be faulty in the sample set; FP: the number of the samples classified as faulty samples and predicted to be normal in the sample set; TN: the number of the samples classified as normal samples and predicted to be normal in the sample set; FN: the number of the samples classified as normal samples and predicted to be faulty in the sample set.

Table 2: Some data of faulty fan on July 23, 2016.

Status parameter	Time									
	0:38:02	0:40:14	0:46:18	0:55:46	1:01:42	1:05:16	1:16:22	1:21:08	1:32:34	1:40:22
Rotor speed (r/m)	17.38	16.97	9.36	1.23	0	0	0	0	0.21	0.14
Generator speed (r/m)	1763.6	1702.8	1214.6	27.3	8.9	7.3	9.0	6.1	953.4	1651.6
Temperature of bearing A (°C)	43.8	42.6	44.0	47.7	48.5	48.4	47.1	46.7	454	45.2
Temperature of bearing B (°C)	45.1	45.8	46.2	44.5	48.8	47.3	46.4	45.5	46.8	47.1
Current of pitch motor 1 (A)	120	80	0	0	0	0	0	0	0	40
Current of pitch motor 2 (A)	70	50	20	0	0	0	0	0	30	20
Current of pitch motor 3 (A)	50	40	40	0	0	0	0	0	10	50
Brake pressure (N)	0	27.53	110.14	124.47	160.29	143.21	120.45	150.41	140.12	134.47
1 min average wind speed (m/s)	8.32	9.95	10.39	8.87	9.81	10.32	11.14	9.48	8.22	9.13
U1 phase winding current (A)	890	832	32	6	6	6	6	6	4	4
U2 phase winding current (A)	883	838	37	6	6	6	6	6	4	4
U3 phase winding current (A)	880	828	35	6	6	6	6	6	4	4
Lubricant filter inlet pressure (N)	−3.96	−3.78	−3.67	−3.54	−3.66	−3.92	−3.96	−3.88	−3.68	−3.86
Lubricant filter outlet pressure (N)	5.77	6.04	3.03	2.78	4.66	3.45	6.20	2.97	3.42	2.64
Pressure of main hydraulic system (N)	143.12	151.24	160.14	152.13	130.24	163.48	153.46	143.34	132.04	130.19
Wind angle (°)	4.21	3.24	5.25	3.54	2.14	1.67	3.45	2.87	2.21	4.15
Grid voltage (kV)	407.2	406.4	405.7	407.2	403.1	401.8	406.7	405.3	404.1	405.9
Pitch angle 1 (°)	0.25	0.33	89.04	89.04	89.04	89.04	89.04	89.04	89.04	89.04
Pitch angle 2 (°)	0.27	0.34	89.02	89.02	89.02	89.02	89.02	89.02	89.02	89.02
Pitch angle 3 (°)	0.27	0.33	89.04	89.04	89.04	89.04	89.04	89.04	89.04	89.04
Pitch controller location (°)	56.21	52.47	0	0	0	0	0	0	0	0
Voltage of pitch capacitor (V)	59.23	59.17	59.10	59.11	59.14	59.18	59.17	59.21	59.33	59.12
Yaw speed (°/s)	0	0.12	0	0	0	0	0	0	0	0
Nacelle vibration (mm)	0.06	0.04	0.01	0.01	0.01	0.02	0.02	0.01	0.02	0.03
Gearbox inlet oil temperature (°C)	46.5	44.1	45.6	43.9	43.2	44.5	44.7	43.5	42.3	44.9
Generator torque deviation (N·m)	56.63	44.12	23.14	17.19	0.12	0.11	0	0	0	0

4. Experimental Analysis

4.1. Data Description. In order to verify the effectiveness of the proposed fault detection method, the actual operating data of one wind farm in Shandong in one year was used in the experiment. This wind farm includes 33 variable-speed and variable-pitch wind turbine generator systems in total, which are separately connected to the monitoring center through sensors. The data were sampled at intervals of 2 s and stored in the database.

Among them, the main power supply of the pitch system of the #11 wind turbine generator system failed on March 14, 2016. The failure lasted from 0 : 43 to 1 : 29. The data from 30 min before the fault to 30 min after the fault were selected as the experimental data, to effectively classify samples and fully reflect fault features. Accordingly, the status parameters were selected from 0 : 13 to 1 : 59 on March 14. Part of the original data is given in Table 2.

4.2. Selection of Sample Features. According to the operation mechanism of the wind turbine generator system, when the pitch system fails, the status parameter affected directly is the power output of the wind turbine generator system. The correlation between the power output and each variable was analyzed based on the Pearson correlation coefficient, to select effective variables.

The raw data of the aforesaid status parameters were first subject to data cleaning, to eliminate "no data" and the data corresponding to the value "0" of all status variables. After the data were normalized, the correlation with the output power was calculated. Some calculation results are given in Table 3.

As can be seen from the correlation results in Table 3, some variables of the status parameters have a low correlation with the output power. Based on the nature of the Pearson correlation coefficient, the variables with the absolute value of the correlation coefficient less than 0.55 were deleted, and those with the absolute value of the correlation coefficient greater than 0.55 were taken as the main influencing factors of the fault, as indicated by the bold part in Table 3. In order to prevent model overfitting due to the interference of redundant variables in model training, these status variables were subject to a secondary calculation with the Pearson correlation coefficient to identify the redundant parameters that have a high correlation and simplify the sample size. Some secondary Pearson calculation results are given in Table 4.

Table 3: Some correlation calculation results 1.

Status parameter	Pearson correlation coefficient
Rotor speed (r/m)	**0.97753**
Generator speed (r/m)	**−0.99738**
Temperature of bearing A (°C)	0.33852
Temperature of bearing B (°C)	−0.26418
Current of pitch motor 1 (A)	**0.79989**
1 min average wind speed (m/s)	−0.20292
U1 phase winding current (A)	**0.99992**
U2 phase winding current (A)	**0.99988**
Pitch angle 2 (°)	**−0.94072**
Pitch angle 3 (°)	**−0.95908**
Yaw speed (°/s)	0.03103
Voltage of pitch capacitor (V)	**0.56784**
Pressure of main hydraulic system (N)	0.02426
U3 phase winding current (A)	**0.99985**
Current of pitch motor 2 (A)	**0.80794**
Lubricant filter inlet pressure (N)	−0.02909
Lubricant filter outlet pressure (N)	**−0.81853**
Current of pitch motor 3 (A)	**0.80471**
Wind angle (°)	0.21073
Grid voltage (kV)	**−0.88644**
Pitch angle 1 (°)	**−0.94581**
Nacelle vibration (mm)	**0.58414**
Brake pressure (N)	0.12353
Gearbox inlet oil temperature (°C)	−0.25683
Generator torque deviation (N·m)	**−0.85781**
Pitch controller location (°)	**−0.85769**

Table 4: Some correlation calculation results 2.

Pearson correlation coefficient	Pitch angle 1	U1 phase winding current (A)	Rotor speed	Current of pitch motor 1
Yaw angle 1	**0.99998**	−0.60992	**−0.97406**	−0.83843
U2 phase winding current (A)	−0.58954	**0.99863**	0.64357	0.51048
Pitch angle 2	**0.99981**	−0.60992	**−0.97406**	−0.83843
Current of pitch motor 2	−0.85505	0.53974	0.83172	**0.91606**

As can be seen from some calculation results in Table 4, the correlation coefficient of the yaw angle 1 and pitch angle 1 of the blade was close to 1, and that of the pitch angle 2 and rotor speed was also close to 1. The same status parameters of different parts also had a high correlation. They essentially had the same effect during the operation of the pitch system. If these status parameters are considered simultaneously in a model, redundant variables will be introduced, which will increase the complexity and calculation of the model and may lead to overfitting and other problems. Therefore, the redundant parameters were eliminated in conjunction with the correlation results in Tables 3 and 4. The sample feature set was built with the remaining status parameters.

4.3. Experimental Results. The sample set corresponding to the normal operation of the wind turbine generator systems was classified as a normal category and that corresponding to the failure of the main power supply of the pitch system as a fault category. The entire sample set was divided into two parts: training set and testing set, including normal and fault data, respectively. The training set was used to train the mcODM model, while the testing set to test the model. The one-versus-rest SVM (ovrSVM) and one-versus-one SVM (ovoSVM) were compared in the experiment on the Matlab platform.

According to the performance evaluation indicators of the model, five indicators were compared, i.e., the accuracy, precision, $F1$-score, false-negative rate, and false-positive rate. The test set was subject to tenfold cross-validation, using the average of results.

The comparison results of accuracy and precision are given in Table 5, the box chart of accuracy is presented in Figure 2, and the box chart of precision is presented in Figure 3. The comparison results of $F1$-score and FPR and FNR are given in Table 6.

As shown above, the accuracy, precision, and $F1$-score of the mcODM model were higher than those of the other two models, while its false-negative rate and false-positive rate were the lowest.

In order to verify the universality of the method proposed in this paper, the operating data of multiple wind turbine generator systems with failure in their pitch systems in this wind farm were used in the experiment. The number of 23 wind turbine generator system occurred the overtemperature of the servo drive of the pitch blade 1 on July 23, 2016, in this wind farm, the comparison results of accuracy and precision are given in Table 7, the box chart of accuracy

Table 5: Comparison of fault detection performance for the main power supply of pitch system 1.

Model	Accuracy	Precision
mcODM	**96.11% (±0.0345)**	**95.09% (±0.0185)**
ovrSVM	93.84% (±0.0237)	92.62% (±0.0095)
ovoSVM	91.32% (±0.0514)	90.77% (±0.0126)

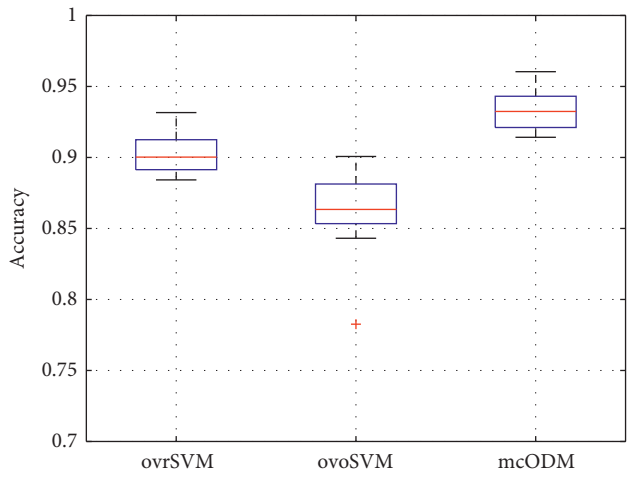

Figure 2: Box chart of fault detection accuracy for the main power supply of the pitch system.

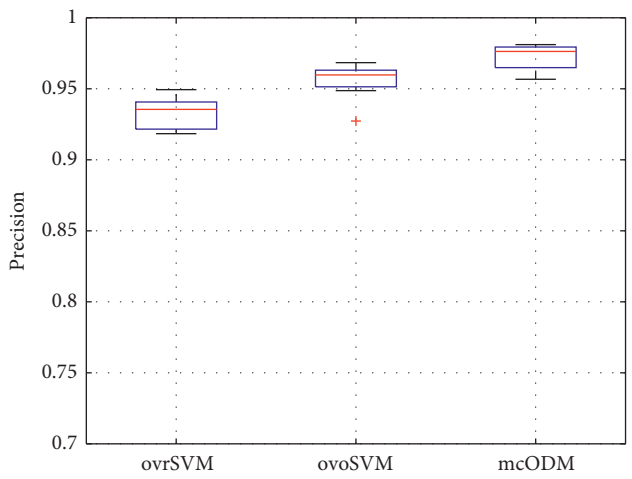

Figure 3: Box chart of fault detection precision for the main power supply of the pitch system.

Table 6: Comparison of fault detection performance for the main power supply of pitch system 2.

Model	$F1$-score	FPR	FNR
mcODM	**96.03% (±0.0027)**	**5.01% (±0.0127)**	**3.01% (±0.0116)**
ovrSVM	93.75% (±0.0013)	8.26% (±0.0084)	6.43% (±0.0237)
ovoSVM	90.14% (±0.0019)	10.64% (±0.0092)	8.18% (±0.0134)

is presented in Figure 4, and the box chart of precision is presented in Figure 5. The comparison results of $F1$-score and FPR and FNR are given in Table 8. The number of 28 wind turbine generator system occurred the emergency stop of the pitch system on June 8, 2016, in this wind farm, the comparison results of accuracy and precision are given in Table 9, the box chart of accuracy is presented in Figure 6, and the box chart of precision is presented in Figure 7. The comparison results of $F1$-score and FPR and FNR are given in Table 10.

In the fault detection at the overtemperature of the servo drive of the pitch blade 1 and the emergency stop of the pitch system, the mcODM model has the highest accuracy, precision, and $F1$-score and lowest false-negative rate and false-positive rate.

It can be seen from the aforesaid comparison results that, in terms of the faults of the pitch systems of different wind turbine generator systems, the mcODM model has high efficiency in sample classification and capabilities in generalization, since the distribution model is built based on the features of the sample distribution. In conjunction with the aforesaid method of feature selection, the status parameters of low correlation can be eliminated, thus reducing the sample size and model training burden and avoiding overfitting. When the mcODM algorithm is combined with the proposed feature selection method in fault detection of pitch systems of wind turbine generator systems, higher capabilities can be achieved in fault detection.

5. Conclusions

This paper proposes the mcODM-based method for fault detection of pitch systems of wind turbine generator systems. The features are extracted according to the operating features of the pitch system and the Pearson correlation coefficient of the wind turbine generator system. The correlation of status parameters is fully considered, and the model complexity is subject to secondary Pearson analysis, which can eliminate the redundant parameters and avoid model overfitting while ensuring the detection rate. This also solves the problem of selecting the feature parameters reflecting the faults of pitch systems from a large amount of SCADA data. Considering the detection model, the mcODM model has been successfully applied in fault detection of the pitch systems of wind turbine generator systems. Due to the combination of the margin mean and variance and full consideration to the sample distribution features, this model solves the problem of inefficient classification arising from the sample category unbalance and complex distribution of pitch system fault samples.

In order to verify the universality of this method, the SCADA data of wind turbine generator systems with different pitch system faults were used in the fault detection experiment. At the same time, the ovrSVM and ovoSVM models were introduced for comparison. The experimental results show that the proposed method has good performance in generalization and high accuracy and precision as well as low false-negative rate and false-positive rate in fault detection of the pitch systems of wind turbine generator systems.

Since wind turbine generator systems are affected by multiple factors (e.g., operating environment and load) and

Table 7: Comparison of fault detection performance at overtemperature of servo drive of pitch blade 1.

Model	Accuracy	Precision
mcODM	**92.07% (±0.0214)**	**90.26% (±0.0426)**
ovrSVM	89.94% (±0.0186)	88.75% (±0.0329)
ovoSVM	86.44% (±0.0621)	86.16% (±0.0084)

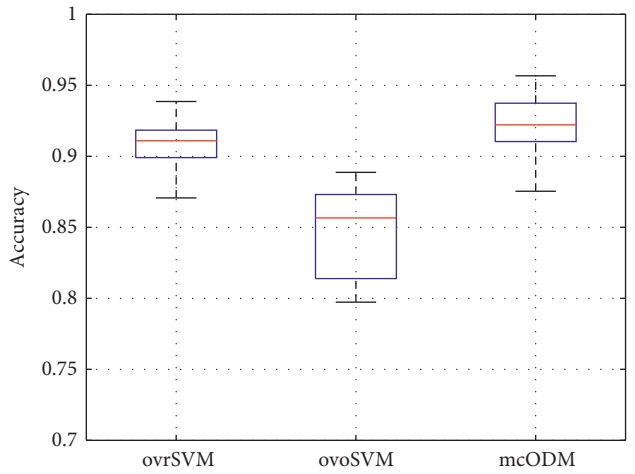

Figure 4: Box chart of fault detection accuracy at overtemperature of servo drive of pitch blade 1.

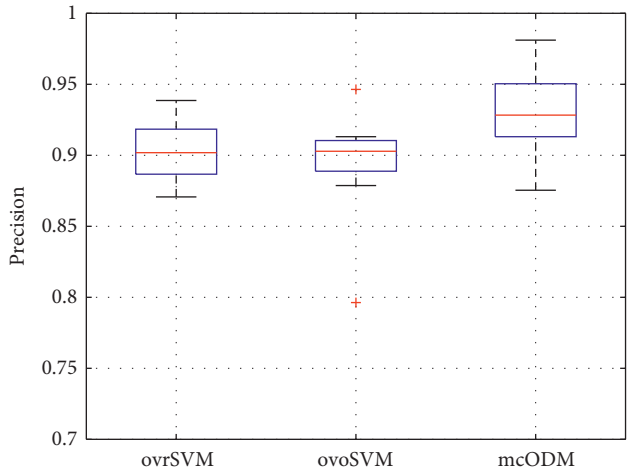

Figure 5: Box chart of fault detection precision at overtemperature of servo drive of pitch blade 1.

Table 8: Comparison of fault detection performance at overtemperature of servo drive of pitch blade 2.

Model	$F1$-score	FPR	FNR
mcODM	**91.93% (±0.0023)**	**8.24% (±0.0134)**	**6.07% (±0.0112)**
ovrSVM	89.81% (±0.0018)	11.07% (±0.0121)	8.82% (±0.0082)
ovoSVM	85.78% (±0.0012)	13.39% (±0.0219)	10.48% (±0.0148)

their operating conditions are changing during fault detection, it is difficult to meet the fault detection requirements for the entire wind turbine generator systems in most cases.

Table 9: Comparison of fault detection performance at the emergency stop of pitch system 1.

Model	Accuracy	Precision
mcODM	**94.73% (±0.0219)**	**94.41% (±0.0427)**
ovrSVM	89.14% (±0.0271)	88.83% (±0.0316)
ovoSVM	90.81% (±0.0184)	89.76% (±0.0251)

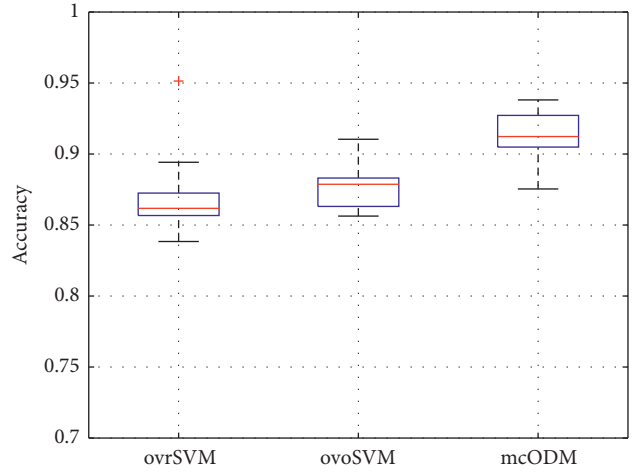

Figure 6: Box chart of fault detection accuracy at the emergency stop of the pitch system.

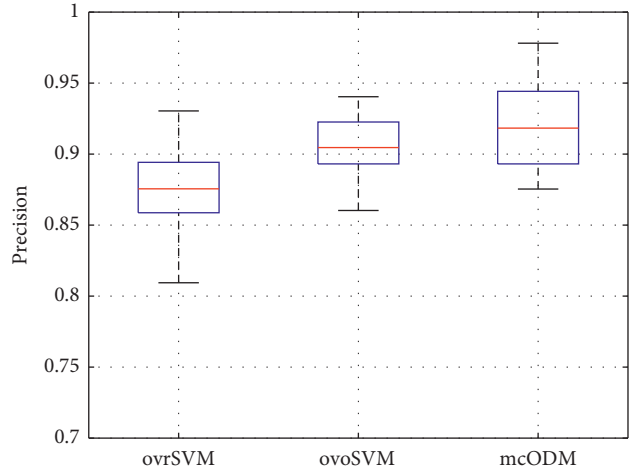

Figure 7: Box chart of fault detection precision at the emergency stop of the pitch system.

Table 10: Comparison of fault detection performance at the emergency stop of pitch system 2.

Model	$F1$-score	FPR	FNR
mcODM	**93.16% (±0.0024)**	**6.25% (±0.0121)**	**3.52% (±0.0057)**
ovrSVM	88.15% (±0.0019)	11.70% (±0.0079)	8.43% (±0.0081)
ovoSVM	90.08% (±0.0021)	9.07% (±0.0237)	7.51% (±0.0048)

Therefore, the research on status monitoring and fault detection of the entire wind turbine generator systems under changing conditions can help effectively reduce the fault rate and improve operating stability.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

All authors contributed equally to this work.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 61403046), the Natural Science Foundation of Hunan Province, China (Grant no. 2019JJ40304), Changsha University of Science and Technology "The Double First Class University Plan" International Cooperation and Development Project in Scientific Research in 2018 (Grant no. 2018IC14), the Research Foundation of Education Bureau of Hunan Province (Grant no. 19K007), Hunan Provincial Department of Transportation 2018 Science and Technology Progress and Innovation Plan Project (Grant no. 201843), the Key Laboratory of Renewable Energy Electric-Technology of Hunan Province, the Key Laboratory of Efficient and Clean Energy Utilization of Hunan Province, Innovative Team of Key Technologies of Energy Conservation, Emission Reduction and Intelligent Control for Power-Generating Equipment and System, CSUST, Hubei Superior and Distinctive Discipline Group of Mechatronics and Automobiles (Grant no. XKQ2020009), and Major Fund Project of Technical Innovation in Hubei (Grant no. 2017AAA133).

References

[1] J. Chen, F. Wang, and K. A. Stelson, "A mathematical approach to minimizing the cost of energy for large utility wind turbines," *Applied Energy*, vol. 228, pp. 1413–1422, 2018.

[2] J. Tautz-Weinert and S. J. Watson, "Using SCADA data for wind turbine condition monitoring-a review," *IET Renewable Power Generation*, vol. 11, no. 4, pp. 382–394, 2017.

[3] M. D. Hossain, A. Abu-Siada, and S. Muyeen, "Methods for advanced wind turbine condition monitoring and early diagnosis: a literature review," *Energies*, vol. 11, no. 5, p. 1309, 2018.

[4] M. Tang, S. X. Ding, C. Yang et al., "Cost-sensitive large margin distribution machine for fault detection of wind turbines," *Cluster Computing*, vol. 22, no. 3, pp. 7525–7537, 2019.

[5] L. Li, H. Luo, S. X. Ding, Y. Yang, and K. Peng, "Performance-based fault detection and fault-tolerant control for automatic control systems," *Automatica*, vol. 99, pp. 308–316, 2019.

[6] L. Li and S. X. Ding, "Optimal detection schemes for multiplicative faults in uncertain systems with application to rolling mill processes," *IEEE Transactions on Control Systems Technology*, 2020.

[7] W. Long, T. Wu, J. Jiao, M. Tang, and M. Xu, "Refraction-learning-based whale optimization algorithm for high-dimensional problems and parameter estimation of PV model," *Engineering Applications of Artificial Intelligence*, vol. 89, Article ID 103457, 14 pages, 2020.

[8] R. V. Pandit and D. Infield, "SCADA-based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes," *IET Renewable Power Generation*, vol. 12, no. 11, pp. 1249–1255, 2018.

[9] J. Liu, F. Qu, X. Hong, and H. Zhang, "A small-sample wind turbine fault detection method with synthetic fault data using generative adversarial nets," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 3877–3888, 2018.

[10] F. Ruiming, W. Minling, G. Xinhua, S. Rongyan, and S. Pengfei, "Identifying early defects of wind turbine based on SCADA data and dynamical network marker," *Renewable Energy*, vol. 154, pp. 625–635, 2020.

[11] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems & Signal Processing*, vol. 107, pp. 241–265, 2018.

[12] J. Dai, W. Yang, J. Cao, D. Liu, and L. Xing, "Ageing assessment of a wind turbine over time by interpreting wind farm SCADA data," *Renewable Energy*, vol. 116, no. 2, pp. 199–208, 2018.

[13] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information," *Renewable Energy*, vol. 118, no. 4, pp. 357–367, 2018.

[14] H. Zheng, Y. Zhang, J. Liu, H. Wei, J. Zhao, and R. Liao, "A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers," *Electric Power Systems Research*, vol. 155, pp. 196–205, 2018.

[15] Z. Wang, L. Yao, Y. Cai, and J. Zhang, "Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis," *Renewable Energy*, vol. 155, pp. 1312–1327, 2020.

[16] Y. Li, S. Liu, and L. Shu, "Wind turbine fault diagnosis based on Gaussian process classifiers applied to operational data," *Renewable Energy*, vol. 134, pp. 357–366, 2019.

[17] A. Soussa, M. D. Mouss, S. Aitouche, H. Melakhessou, and M. Titah, "The MAED and SVM for fault diagnosis of wind turbine system," *International Journal of Renewable Energy Research*, vol. 7, no. 2, pp. 758–769, 2017.

[18] Q. W. Gao, W. Y. Liu, B. P. Tang, and G. J. Li, "A novel wind turbine fault diagnosis method based on intergral extension load mean decomposition multiscale entropy and least squares support vector machine," *Renewable Energy*, vol. 116, pp. 169–175, 2018.

[19] W. Liu, Z. Wang, J. Han, and G. Wang, "Wind turbine fault diagnosis method based on diagonal spectrum and clustering binary tree SVM," *Renewable Energy*, vol. 50, no. 2, pp. 1–6, 2013.

[20] J. Hang, J. Zhang, and M. Cheng, "Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine," *Fuzzy Sets and Systems*, vol. 297, pp. 128–140, 2016.

[21] J. Saari, D. Strömbergsson, J. Lundberg, and A. Thomson, "Detection and identification of windmill bearing faults using a one-class support vector machine (SVM)," *Measurement*, vol. 137, pp. 287–301, 2019.

[22] M. Tanveer, A. Sharma, and P. N. Suganthan, "General twin support vector machine with pinball loss function," *Information Sciences*, vol. 494, pp. 311–327, 2019.

[23] T. Zhang and Z. Zhou, "Multi-class optimal margin distribution machine," in *Proceedings of the International*

Conference on Machine Learning, pp. 4063–4071, Sydney, Australia, 2017.

[24] E. Muljadi and C. P. Butterfield, "Pitch-controlled variable-speed wind turbine generation," *IEEE Transactions on Industry Applications*, vol. 37, no. 1, pp. 240–246, 2001.

[25] X. Yin, W. Zhang, Z. Jiang, and L. Pan, "Adaptive robust integral sliding mode pitch angle control of an electro-hydraulic servo pitch system for wind turbine," *Mechanical Systems and Signal Processing*, vol. 133, Article ID 105704, 2019.

[26] H. Li, C. Yang, Y. Hu, X. Liao, Z. Zeng, and C. Zhe, "An improved reduced-order model of an electric pitch drive system for wind turbine control system design and simulation," *Renewable Energy*, vol. 93, no. 8, pp. 188–200, 2016.

[27] Q. Jiang, S. Yan, H. Cheng, and X. Yan, "Local-global modeling and distributed computing framework for nonlinear plant-wide process monitoring with industrial big data," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[28] Q. Jiang, S. Yan, X. Yan, H. Yi, and F. Gao, "Data-driven two-dimensional deep correlated representation learning for nonlinear batch process monitoring," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2839–2848, 2020.

[29] M. Tang, Q. Zhao, S. X. Ding et al., "An improved lightGBM algorithm for online fault detection of wind turbine gearboxes," *Energies*, vol. 13, no. 4, Article ID 807, 2020.