J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 5233--5256

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Machine Learning-Based Data Mining and Fault Identification for Condition Monitoring of Multi-Dimensional Anthropomorphic Wind Turbines

Shuqiao Chen[1,✉], Peng Zhang[1], Hui Ma[1], Shuo Zhou[1]

[1] *Mengdong Concord Zalutqi Wind Power Co., Ltd., Tongliao, Inner Mongolia, 029100, China*

ABSTRACT
Due to the complex structure of multi-dimensional anthropomorphic wind turbine and the harsh operating environment, in order to reduce its maintenance cost, it has become a popular research hotspot to get fast and effective condition diagnosis and fault early warning through big data mining and analysis of wind turbine condition monitoring. The article clarifies the basic mechanism and typical faults of multi-dimensional anthropomorphic wind turbine, and after analyzing the characteristic frequency of faults on the transmission chain of multi-dimensional anthropomorphic wind turbine, it proposes the anomaly detection method of wind turbine condition monitoring data based on the auxiliary eigenvectors improved density clustering (DBSCAN), which realizes the accurate identification of different types of normal data, valid anomalous data containing fault information, and invalid anomalous data in the monitoring data. It realizes the accurate identification of different types of normal data, valid abnormal data with fault information, and invalid abnormal data in monitoring data. Subsequently, the actual historical data of the wind farm is used as the experimental data set to realize the identification of the operating status of the wind turbine. Finally, the DBN-Dropout wind turbine fault identification method is proposed by combining Deep Confidence Network and Dropout technique. The experimental results indicate that the recognition rate of this paper's model for nine faults is as high as 99.88%, and the superiority and accuracy of this paper's model in feature extraction and fault diagnosis are verified by comparing its performance with other fault detection models.
Keywords: dbscan; condition monitoring; fault identification; multidimensional anthropomorphic wind turbines

1. Introduction

With the transformation of the global energy structure and the enhancement of environmental protection awareness, wind energy, as a clean and renewable form of energy, has been widely noticed

✉ Corresponding author.
E-mail address: lyj005566@126.com (S. Chen).
Received 01 February 2024; Revised 20 April 2024; Accepted 10 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-295

and applied [1-2]. As the core equipment of wind power generation, the operation status of wind turbines directly affects the efficiency and reliability of wind power generation. Therefore, it is of great significance to conduct research on condition monitoring and fault diagnosis technology for wind turbines [3-5]. Wind turbine consists of complex mechanical and electrical systems, the operating environment is variable and harsh, long-term operation will inevitably occur a variety of faults. These failures will not only affect the normal operation of wind turbines and reduce power generation efficiency, but also may lead to equipment damage and even cause safety accidents [6-8]. Therefore, it is of great significance to carry out research on wind turbine condition monitoring and fault diagnosis technology to improve the reliability of wind turbines, reduce operation and maintenance costs, and ensure energy security [9-11].

By monitoring the operating conditions of wind power, better economic benefits and long-term development prospects can be obtained. The most direct embodiment is the economic benefits, real-time monitoring of the wind power generation system can help technicians to timely find out the freezing phenomenon of the wind power generation system due to the extreme external temperature environment, and also timely find out the resonance phenomenon of the system caused by the electric wave, and can take timely maintenance measures, thus effectively reducing the maintenance cost, preventing major accidents, and thus reducing the major economic losses [12-15]. In the wind power generation monitoring system, the use of computer data processing technology, the data uploaded to the computer receiving system, and then after a complex data analysis process, the data is organized, and finally draw the corresponding state diagram. The wind power monitoring system can monitor the system in real time and record its data to provide accurate data reference for future design and transformation, thus promoting the development of wind power system from the side [16-19].

Literature [20] conceptualized a WTG state detection and fault diagnosis method with monitoring and data acquisition covariance analysis as the core logic to achieve continuous anomaly detection of WTGs. Literature [21] relies on the deep learning method of deep self-coder network to train the WTG monitoring data to clarify the WTG health indicators, and then identify and diagnose the WTG faults. Literature [22] used a novel unsupervised learning method, i.e., denoising autoencoder (DAE), as the underlying architecture to construct a WTG fault detector, which can effectively capture the nonlinear correlation between multiple sensor variables and the time dependence of each sensor variable to improve the fault detection accuracy. Literature [23] envisioned a state monitoring algorithm based on Gaussian process, supplemented with standard IEC sub-tank power curves and individual sub-tank probability distributions to identify and evaluate the abnormalities of WTG operation, and revealed the effectiveness of this fault diagnosis strategy with feedback from actual cases. Literature [24] carried out comparative evaluation experiments, revealing that based on the decision number, the fault diagnosis rate of the random forest WTG fault diagnosis model is 92.68% and 91.98 respectively, which can provide early warning of faults for WTG operation and maintenance personnel. Literature [25] designed a data-driven strategy for generator bearing fault state detection based on time-series temperature data, and the feasibility was verified in a wind turbine data set. Literature [26] describes a wind turbine unit fault prediction model based on the Random Forest algorithm as the basic architecture, which can effectively detect the operational status of wind turbine units at low cost.

Firstly, the basic structure of the wind turbine is described, and the working modes and typical failures of each major component of the wind turbine including the generator, wind turbine wheel, gearbox, pitch system, yaw system and hydraulic system are comprehensively analyzed, while the failure frequency characteristics of different components after failure are analyzed. Subsequently, a transformer condition monitoring data anomaly detection method based on auxiliary feature vectors and density clustering is proposed. Auxiliary feature vectors are constructed for each class of state quantities and clustered using the DBSCAN model, which realizes the accurate identification of the conceptual drift, invalid and valid anomalies of the condition monitoring data. Based on the k-distance graph method, a heuristic parameter selection method based on the “category number-Eps” curve is proposed to optimize the parameters of the DBSCAN model under unsupervised conditions. Subsequently, we combine the historical data of multi-dimensional mimetic wind turbines in the wind farm to realize the monitoring and identification of the operating status of multi-dimensional mimetic wind turbines. Then the deep confidence network is introduced for the research of wind turbine fault identification algorithm, and the DBN-Dropout fusion method is titled, which comprehensively analyzes the performance of the model in terms of multiple dimensions such as identification accuracy, loss value, confusion matrix, and so on. At the same time, the model of this paper is compared with five other advanced fault diagnosis models in order to verify the effectiveness of the model of this paper.

2. Basic structure and typical failures of multidimensional mimetic wind turbines

2.1. Basic structure of multidimensional anthropomorphic wind turbines

The most widely used doubly-fed asynchronous wind turbine is composed of wind turbine, nacelle and tower. The nacelle includes the wind turbine, generator, variable speed gearbox, yaw system, hydraulic system, pitch system, SCADA system, and so on.

2.2. Major Components and Failures of Multi-Dimensional Mimetic WTGs

2.2.1. Generators. The generator in a wind turbine is usually of the doubly-fed induction type, which is one of the core components of the system and supplies power by converting rotating mechanical energy into electrical energy. On the rotor side of the wind turbine, the excitation current is generated through a connection between the converter and the grid, and on the stator side of the wind turbine, the transformer can be directly connected to the grid through a transmission line. Because generators work in harsh conditions and electromagnetic environments for long periods of time, the components are particularly susceptible to failure [27]. Common failure modes include excessive generator vibration, excessive generator noise, generator overheating, bearing overheating, or abnormal noise.

2.2.2. Wind turbines. Wind turbines are mainly composed of blades and hubs. Adopting aerodynamic principle design, it is the core component of wind turbine to capture wind energy. The blade set is fixed on the rotating shaft of the hub, and the torque generated by the wind turbine through rotation is transferred to the main drive shaft, and some mechanical and electrical equipment for pitch change are also installed inside the wind turbine. Due to the relatively complex and changeable working environment of wind turbines, stormy weather is more likely to erode the blades and hub.

2.2.3. Gearboxes. The gearbox, usually located inside the nacelle of the wind turbine, is the core component that connects the generator to the main shaft of the paddle. The gearbox usually consists of a one-stage planetary gear transmission and a two-stage parallel gear transmission, a feature that ensures that the generator can provide a stable power output despite fluctuations in speed. There are two common failures of gearboxes, including gear failure and bearing failure. Since the bearings are the core component of the gearbox, their failure can often cause devastating damage to the gearbox. Common gear failures include: broken teeth, tooth fatigue, and gumming; bearing failures include: wear, pitting, cracks, and surface spalling.

2.2.4. Pitch system. The main function of the pitch system is: by comparing the wind speed in the environment with the rated wind speed of the wind turbine, when the value of the wind speed in the external environment is higher than the rated value of the wind turbine, adjusting the pitch angle to make it smaller, to reduce the impact of the external environment wind on the wind turbine by reducing the resistance, when the value of the wind speed in the external environment is lower than the rated value of the wind turbine, adjusting the pitch angle to make it larger can increase the resistance, so that the blades can get a bigger Wind energy. Pitch system mainly includes two types: hydraulic pitch system and electric pitch system.

2.2.5. Yaw system. Yawing is the process by which the rotor of a wind turbine rotates around the wind axis. The main function of the yaw system is to control and change the wind direction inside the wind turbine compartment by applying the external ambient wind direction, so as to make its operation more stable and reliable and maximize the capture of wind energy. Yaw systems can be categorized into two types according to different yaw methods: passive yaw systems and active yaw systems. Typical faults of the yaw system include: yaw positioning deviation, broken teeth of yaw gear ring, ineffective yaw limit switch, lubricant leakage, abnormal noise, etc.

2.2.6. Hydraulic systems. The function of the hydraulic system is to increase the force by changing the liquid pressure and transferring the power to the mechanical unit. Due to the advantages of high liquid density, stepless speed regulation, smooth and reliable overload protection, and relatively simple component replacement, it is widely used in large wind turbines. This includes generator cooling, inverter temperature control and gearbox lubrication. In addition, it is also used in generator cooling, converter temperature control and gear box lubricant cooling, the hydraulic system plays an irreplaceable and important role. Typical faults of the hydraulic system are: high oil temperature, insufficient hydraulic pressure, abnormal noise, excessive vibration, frequent start and stop of the hydraulic pump.

2.2.7. SCADA system. Doubly-fed wind turbines are typical mechatronic energy power plants that require a particularly high degree of automation.A SCADA system is a computer-based DCS and power automation monitoring system, i.e. a data acquisition and monitoring system. During turbine operation, the SCADA system needs to monitor hundreds of operating parameters. Based on these operating data, the turbine operating status is observed and protected and monitored. These real-time tracking parameters are used to assess the overall performance of the turbine and to determine the presence of faults.

3. Multidimensional mimetic wind turbine condition monitoring

3.1. Wind turbine drive chain characteristic frequency

3.1.1. Rolling bearing characteristic frequencies. Bearings are very important mechanical components in the transmission system, it is responsible for transferring the load on the transmission shaft, when the transmission shaft is subjected to a certain weight of the load, and at the same time to a certain speed to drive the bearing to rotate, the bearing due to these loads with a certain weight will produce excitation, so that the bearing produces vibration. In the process of generating vibration, due to the different internal structure of the bearing, the vibration frequency of each component is also very different, so you can monitor the vibration frequency of each component to determine the operating status of the bearing.

Under normal working condition, the vibration frequency of each component of the bearing is related to its own parameters. To spherical ball bearings, for example, bearing pitch circle diameter with $D_m(mm)$, the diameter of the rolling body with $d(mm)$, the inner circle radius with $r_1(mm)$, the outer circle radius $r_2(mm)$, the rolling body and roll to the center of the contact angle with α, the number of rolling body with Z.

In order to analyze the various parts of the bearing motion parameters, the first to do the following assumptions: no relative sliding between the raceway and the rolling body, bear radial, axial load without deformation of each part. Then, the characteristic frequency of rolling bearings can be calculated according to the following formula.

Cage failure frequency:

$$f_{FTF} = \frac{1}{2}(1-\frac{d}{D_m})f_r \cos\alpha \tag{1}$$

Frequency of rolling body rotation failures:

$$f_{BSF} = \frac{1}{2}(1-\frac{d^2}{D_m^2}\cos^2\alpha)f_r \frac{D_m}{d} \tag{2}$$

Frequency of outer ring failures:

$$f_o = \frac{1}{2}(1-\frac{d}{D_m}\cos\alpha)f_r Z \tag{3}$$

Frequency of inner loop failures:

$$f_i = \frac{1}{2}(1+\frac{d}{D_m}\cos\alpha)f_r Z \tag{4}$$

In the formula:

D_m - the diameter of the circle where the center of the rolling body of the bearing is located.

d - The average diameter of the rolling body.

f_r - shaft rotation speed.

α - The angle between the direction of force on the rolling body and the perpendicular line of the inner and outer raceways.

z - The number of rolling bodies.

3.1.2. Gear meshing frequency. Gear box in the operation process, when the gear failure will produce two different frequency signals, one is due to the meshing frequency and the interaction of high harmonics generated by the carrier frequency, the other is the low-frequency amplitude phase changes generated by the modulation signal.

1) Frequency generated by gear and shaft rotation f_z:

$$f_z = \frac{z}{60} \tag{5}$$

Eq:

z - the speed of the gear, unit: revolutions 1 minute.

2) Gear mesh vibration arises from the gear mesh transmission when the gear stiffness is changed after a complete periodic shock acting on the gear. Gear mesh frequency f_n:

$$f_n = f_a \times n_a = f_b \times n_b \tag{6}$$

In Eq:

f_a, f_b - the frequency of the rotation speed of the master and follower wheels.

n_a, n_b - The number of teeth of the master and slave wheels.

When a full cycle of shock is applied to the gear, the harmonic $X_x(t)$ of the gear is written as:

$$X_x(t) = \sum_{r=0}^{R} A_r \cos(2\pi r f_x + \beta_r) \tag{7}$$

where:

r maximum harmonic number, $r = 0,1,\cdots,R$.

In the normal state, the vibration characteristics of the gear is manifested as a small amplitude of the meshing frequency harmonic wave, and once the failure occurs, the gear meshing stiffness is affected by the overload, gear fracture, meshing position change, its vibration generated by the magnitude of the meshing frequency harmonic wave will change significantly, and the more serious the tooth failure occurs, the amplitude of the harmonic part of the amplitude of the growth of the larger.

3.2. *Pre-processing of condition monitoring data*

Condition monitoring data are typically multivariate time series data and are usually represented by a design matrix as shown in equation (8).

$$[T \mid X \mid] = \begin{bmatrix} t_1 & x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ t_2 & x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_m & x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{bmatrix} \tag{8}$$

where matrix $X \in \mathbb{R}^{m\times n}$, consists of m samples (observations) and n types of state quantities (variables). The time vector $T = [t_1, t_2, \ldots, t_i, \ldots, t_m]^{\mathrm{T}}$ is the measurement time of each sample [28]. In practice, the time interval between neighboring observations is usually the same.

With the passage of time, the sample size in X will grow without limit and m will tend to infinity, which brings more difficulties and challenges to the detection of anomalous data.

3.2.1. Data segmentation.

If $K-1$ interruptions occur in dataset X, depending on the duration of the data interruption, X it can be reasonably divided into K segments as shown in equation (9).

$$X = [X(1), X_{inc}(1), \ldots, X(k), X_{inc}(k), \ldots, X_{inc}(K-1), X(K)]^T \tag{9}$$

Where: $X(k)$ is the k nd $(k = 1, \ldots, K)$ data segment. $X_{inc}(k)$ is the period of interruption of data segments $X(k)$ and $X(k+1)$, during which the vast majority of the data has missing measurements.

3.2.2. Data standardization.

Different types of state quantities usually have different scales and orders of magnitude, in order to avoid state quantities of larger orders of magnitude dominating the calculation results, it is necessary to use the z-score method to standardize each type of state quantities into a sequence of 0 mean and unit variance, taking the type j state quantities in data segment $X(k)$ as an example, the standardization method is shown in equation (10).

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_j}{\sigma_j}, i = 1, \ldots, mk, j = 1, \ldots, n \tag{10}$$

where $\tilde{x}_{i,j}$ is the standardized measure of the i nd sample, mk is the number of samples in $X(k)$. $\bar{x}_j$ and σ_j are the mean and standard deviation of the state measure in category j, respectively.

3.2.3. Auxiliary eigenvector construction.

In order to establish pattern recognition rules between different types of data and accurately identify valid and invalid anomalous data, this section constructs two new feature variables for each class of state quantities, which are defined as the backward differential rate of change γ and the over-warning value markers w, respectively, which in turn extends each class of state quantities into a three-dimensional degree of auxiliary feature vector. As an example, the auxiliary eigenvector for the ith sample of the state quantity of class j in $X(k)$ is shown in Eq. (11).

$$e_{i,j} = \left[\tilde{x}_{i,j} \quad \gamma_{i,j} \quad w_{i,j} \right] \tag{11}$$

Then mk sample consists of the eigenvector matrix $e_{i,j} \in \mathbb{R}^{mk\times 3}$.

The two eigenvariables are presented separately as follows.

1) Backward differential rate of change γ: In normal state, the deviation between neighboring measurements is small, while the deviation between abnormal values and normal values is large, thus the differential rate of change is an important indicator of whether the data are abnormal or not. Calculate the γ between the $i-1$th and ith samples of the jth type of state measure using the formula.

$$\gamma_{i,j} = \begin{cases} \tilde{x}_{i,j} - \tilde{x}_{i,1,j} & 2 \le i \le mk \\ 0 & i = 1 \end{cases} \tag{12}$$

2) Excessive warning value mark w: when the measured value of the state quantity exceeds its warning value will trigger a fault warning, this kind of abnormal data is usually the most concerned about the equipment management personnel, the calculation method of w is shown in equation (13).

$$w_{i,j} = \begin{cases} \tilde{x}_{i,j} & x_{i,j} \ge x_{w,j} \\ 0 & x_{i,j} < x_{w,j} \end{cases} \tag{13}$$

where $x_{w,j}$ is the early warning value of the Type j state quantity.

3.3. Density-based clustering model and parameter selection methods

3.3.1. DBSCAN clustering model. The density-based clustering model (DBSCAN) relies on the assumption that normal data are concentrated in densely populated spatial regions and belong to a certain category, while outliers are located in sparsely populated spatial regions and do not belong to any category. Taking $X(k)$ clustering as an example, several important concepts in the DBSCAN model are introduced as follows.

1) Eps Neighborhood: The Eps neighborhood of the prd sample x_p, in $X(k)$ denotes the region centered at x_p, with a radius of Eps, as shown in equation (14).

$$N_{Eps}(x_{p,\cdot}) = \{x_{q,\cdot} \in X(k) \mid dist(x_{p,\cdot}, x_{q,\cdot}) \le Eps\} \tag{14}$$

where x_q is the qrd sample in $X(k)$, $dist(x_{p,\cdot}, x_{q,\cdot})$ is the distance between $x_{p,\cdot}$ and $x_{q,\cdot}$, and when $dist(x_{p,\cdot}, x_{q,\cdot}) \le Eps$, it indicates that $x_{q,}$ is located within the radius of the Eps neighborhood of $x_{p,}$. In this paper, the Euclidean distance function is chosen to quantify the degree of proximity between samples, which is calculated as shown in Eq. (15).

$$\mathrm{dist}(\mathbf{x}_{p,\cdot}, \mathbf{x}_{q,\cdot}) = \sqrt{\sum_{j=1}^{n} (\tilde{x}_{p,j} - \tilde{x}_{q,j})^2} \tag{15}$$

2) Core point: if the Eps neighborhood of $x_{p,}$ contains at least MinPts of samples, then $x_{p,\cdot}$ is said to be a core point.

3) Boundary point: If the number of samples in the Eps neighborhood of x_q is less than MinPts, but it lies within the neighborhood of some core point, then $x_{q,\cdot}$ is said to be a boundary point.

4) Cluster: cluster c is a non-empty subset of $X(k)$ and satisfies the following conditions:

(1) $\forall x_{p,}, x_{q,}$, if $x_p, \in c$ and x_q are points accessible to the density of x_p (i.e., x_p, is a core point and x_q, is within the radius of the Eps of $x_{p,\cdot}$), then $x_q, \in c$.

(2) $x_{q,\cdot}$ is connected to the density of $x_{p,\cdot}$ if $\forall x_{p,\cdot}, x_{q,\cdot} \in c$.

According to the above definition, the condition for forming a clustered cluster c is that c contains at least one core point.

5) Noise points: let the set of clustering clusters $C = \{1, 2, \ldots, c\}$, the clustering result of noise points is 0, indicating that they do not belong to any of the clusters in c.

3.3.2. DBSCAN model parameter selection methods. The Eps and MinPts parameter selection method based on the k-distance graph is proposed by the pioneers of the DBSCAN model, which realizes the selection of MinPts and Eps respectively through two stages.
1) MinPts selection: such that the minimum value of $k = MinPts$, k needs to satisfy the requirement of $(k \geq 3) \cap (k \geq n+1)$.
2) Selection of Eps: Find the position of the most drastic change of the curve, i.e., the "inflection point", on the k-distance graph, and its corresponding vertical coordinate value is the optimal value of Ep.

3.3.3. Neighborhood radius parameter selection methods. A neighborhood radius parameter selection method based on the "category number-Eps" curve is proposed, and the specific procedure is as follows:
1) Define $Eps(s_0)$ as the lower threshold of Eps. When $Eps < Eps(s_0)$, the majority of samples cannot contain no less than MinPts observations within a very small Eps radius, and thus are regarded as anomalous data by the DBSCAN model. As Eps increases, many split clusters are created, resulting in a sharp increase in the number of categories N in the clustering result. Let N_0 be the number of categories at $Eps = 0$. When the number of categories N starts increasing to $N_1 (N_1 > N_0)$, its corresponding Eps value is $Eps(s_0)$.
2) $Eps_{\max}$ is defined as the upper threshold value of Eps, which is determined by k distance from an "inflection point" in the graph with a large Eps value.
3) A logarithmic lattice vector $\{Eps(s_0)|_{i=1}, \ldots, Eps(i), \ldots, Eps_{max}|_{i=10}\}$ of 10 values is generated between $Eps(s_0)$ and $Eps_{\max}$.

4. Multi-dimensional anthropomorphic wind turbine operating condition monitoring experiment

4.1. Performance analysis of the improved algorithm

The algorithmic experiments are tested under Windows 10 system using matlab environment with hardware configuration of Intel Corei5 processor, 8G RAM and 1T hard disk. 2 classic artificial datasets were selected for the quiz: Shapeset and Jain. Shapeset dataset is 2D data contains 788 data points. Jain dataset is 2D dataset contains 373 data points. A comparison of the clustering effect before and after the improvement of the DBSCAN algorithm is shown in Figure 1. A comparison of the improved results of the DBSCAN clustering algorithm is shown in Table 1. As can be seen from the clustering results graph, the improved DBSCAN algorithm is more accurate in the number of clusters classified, due to the ε parameter by the optimal value of the iterative results of the selection, so the number of noise points is controlled at a lower level. Therefore, the improvement method proposed in this paper not only ensures the effectiveness of the algorithm to improve the accuracy of the algorithm, and there is no large gap in the time-efficient complexity of the algorithm are kept on the order of O(n2).

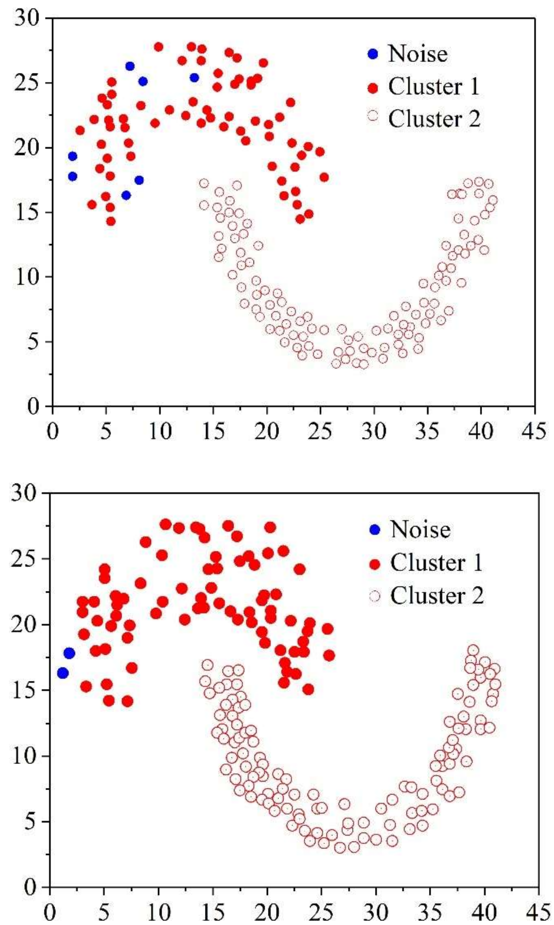


Fig. 1. The DBSCAN algorithm improved the combination of the clustering

Table 1. Improved effect of DBSCAN clustering algorithm

Data set	dimension	Data points	Clustering algorithm	ε	MinPts	Running time	F-Measure
Jain	2	368	Classic DBSCAN	2.32	4	0.452s	0.9855
			Improved DBSCAN	2.42	6	0.503s	0.9963
Shapeset	2	773	Classic DBSCAN	1.15	4	0.642s	0.8125
			Improved DBSCAN	1.28	9	0.755s	0.9752

4.2. Identification of wind turbine operating conditions

For the actual situation of wind turbine operation, the parameter that can intuitively reflect whether the unit is running normally or not is mainly the real-time power of the unit, but there is a certain deviation in considering all the operating states of the unit from the perspective of active power alone, and it is not possible to effectively differentiate between normal wind turbines and faulty wind turbines under the state of waiting for the wind. Therefore, we introduce the two characteristic parameters of the gearbox oil temperature and the main shaft temperature, which are of high relevance to the active power, for identifying the state of wind turbines under the state of waiting for the wind. Therefore, the two characteristic parameters of gearbox oil temperature and main shaft temperature, which have a high correlation with active power, are introduced to recognize the state of wind turbine under the wind condition. In the data preprocessing stage, the extracted data are divided into two categories according to the size of the real-time wind speed: one is the low wind speed area where the real-time wind speed is smaller than the starting wind speed, and the other is the medium-high wind speed area where the real-time wind speed is larger than the starting wind speed. The data of medium-high wind speed zone and low wind speed zone after screening at a certain time point are shown in Table 2 and Table 3.

Table 2. Medium high wind speed sample number

Numbering	Real-time wind speed (m/s)	Active power (kW)	Gear box oil temperature (℃)	Spindle temperature (℃)
1	4.23	129.32	54.98	61.86
2	5.4	168.55	55.26	61.84
3	5.87	202.04	55	62.33
4	5.43	160.47	55.47	63.01
5	4.91	129.1	54.91	61.73
6	5.26	294.06	61.37	72.73
7	8.83	1196.12	61.64	73.34
8	10.55	1528.18	61.78	73.95
...	...	...	...	...
299	12.06	1621.34	63.68	76.27
300	12.49	1639.29	63.67	76.28

Table 3. Sample size of low anematic area

Numbering	Real-time wind speed (m/s)	Active power (kW)	Gear box oil temperature (℃)	Spindle temperature (℃)
1	2.14	-3.3	44.23	33.49
2	1.92	-3.11	43.03	33.43
3	1.5	-3.33	50.99	46.4
4	1.51	-3.28	56.81	52.39
5	1.26	-3.23	48.74	27.24
6	1.4	-3.31	55.62	50.25
7	1.45	-3.31	52.51	47.12
8	2.11	-3.24	58.49	53.07
...	...	...	...	...
16	2.7	-3.28	62.98	56.63
17	2.73	-3.55	38.13	36.41
18	2.99	-3.32	50.15	45.36

The experimental analysis is carried out using the continuous data of a wind farm from December 1, 2023 to March 31, 2024 as the historical data set, which contains 350 wind turbines. All the wind turbine data of 10 time nodes are randomly selected as the sample set for state identification, and the experimental results of the historical data are analyzed as shown in Table 4.

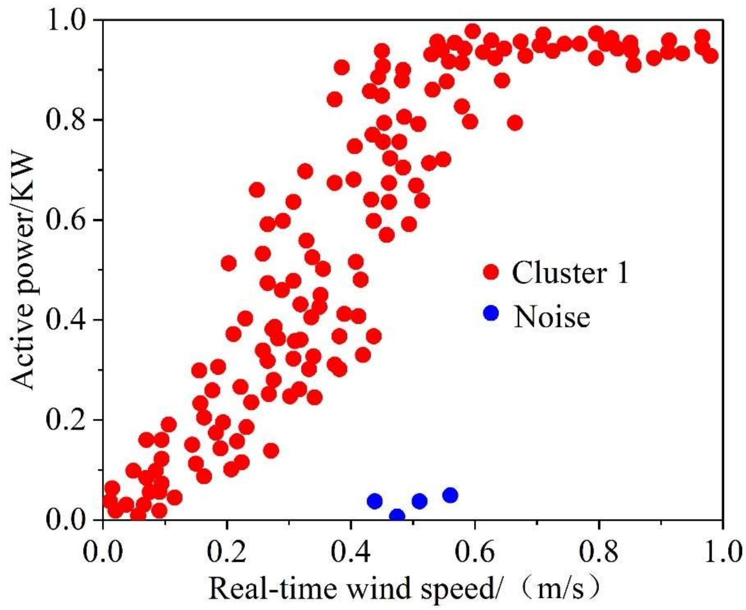
Table 4. Analysis of historical data experiment results

Numbering	Sample size	Correct number	Error number	Running time	F-Measure
1	350	350	0	0.505s	1

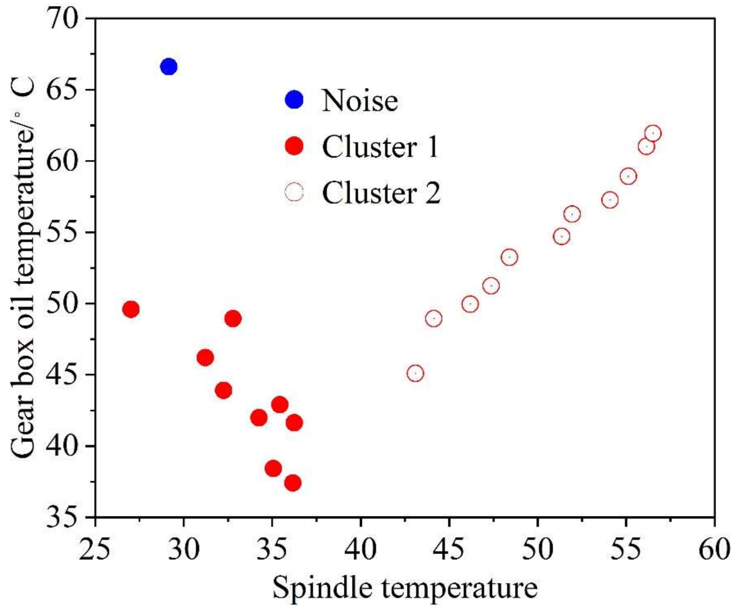
2	350	349	1	0.502s	0.9971
3	350	350	0	0.485s	1
4	350	350	0	0.476s	1
5	350	350	0	0.498s	1
6	350	350	0	0.492s	1
7	350	349	1	0.515s	0.9971
8	350	350	0	0.493s	1
9	350	350	0	0.503s	1
10	350	350	0	0.466s	1

The results of the 10th experiment of state identification are shown in Fig. 2. Figure a shows the results of the clustering operation on real-time wind speed and active power in the medium and high wind speed zone, and the results show the existence of 4 noise points, i.e., the existence of 3 turbines in faulty downtime status. Figure b shows the results of clustering operation for gearbox oil temperature and main shaft temperature in low wind speed zone, the results show the existence of 1 noise point, i.e., the existence of 1 wind turbine in faulty shutdown condition. Comparing the experimental data with the actual data, it is found that the experimental results coincide with the actual situation at this point in time.

The clustering results of the improved DBSCAN algorithm have a higher precision through the test of F-Measure evaluation criteria, and the quality of the clustering results is better. And in the processing of the actual data, the running time of the algorithm is shorter, and it can more accurately determine the current operating status of the wind turbine.



(a) Real-time wind speed



(b) Spindle temperature

Fig. 2. Status recognition of the 10th test results

5. Fault recognition model based on deep confidence network

5.1. Restricted Boltzmann Machine (RBM)

RBM is a special kind of energy-based generative probabilistic model that can learn to get the inherent laws in the data.The basic structure of RBM is shown in Fig. 3.

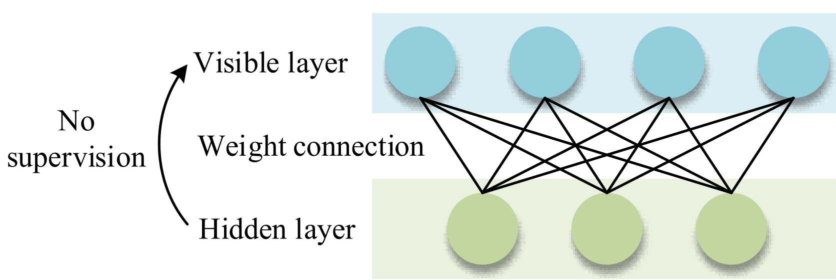


Fig. 3. Basic structure of the RBM

Each independent RBM contains two layers of neurons, the visible and the hidden layers, and there are bidirectional weight connections only between the visible and the hidden layers, while there are no connections within the visible layer and within the hidden layer, and this connection restriction greatly reduces the network topology complexity. The visual layer is denoted as $v = \{v_1, v_2, \cdots, v_m\}$ in the figure and its state space is {0, 1} or real-valued, and the implicit layer is denoted as $h = \{h_1, h_2, \cdots, h_n\}$ and its state space is {0, 1}. The weight connectivity matrix is denoted as W, and the bias of the visible and hidden layers are defined as b and c, respectively [29].

Each neuron node can have a different state. When the state of each node is determined, the state of the model is also determined. When the state space is {0, 1} i.e. discrete values, the following energy function is obtained for a particular state (v,h) of the RBM:

$$E(v,h) = -\sum_{m}^{i=1} b_i v_i - \sum_{n}^{j=1} c_j h_j - \sum_{m,n}^{i,j=1} w_{i,j} v_i h_j \tag{16}$$

where m,n represents the number of units in the visual and hidden layers, respectively. Definition $\theta = (W;b;c)$ is the set of model parameters, when the value of each parameter is determined, the energy distribution of the RBM is also determined, and then the joint probability distribution between the neurons of the visual layer and the hidden layer can be obtained as:

$$P(v,h;\theta)=\frac{1}{Z(\theta)}e^{-E(v,h;\theta)} \tag{17}$$

$$Z(\theta)=\sum_{v,h}e^{-E(v,h;\theta)} \tag{18}$$

$Z(\theta)$ is the normalization factor, also called the distribution function, is the bridge between energy and probability transformation. When the RBM conforms to the Boltzmann distribution and the energy is minimized, the above equation transforms the energy model into a probabilistic model, which in turn transforms the energy minimization problem into a probabilistic maximization problem to be solved, and from this the maximum likelihood estimation can be utilized to solve the model parameter θ. However, for the RBM with m explicit and n hidden layer neurons, there are a total of 2^{m+n} combinations, and calculating $Z(\theta)$ would require utilizing all possible combinations of cases for (v,h), which is clearly unrealistic.

Therefore, consider calculating the marginal distribution of the joint distribution formula (19):

$$p(v;\theta)=\sum_{h}p(v,h)=\frac{\sum_{h}e^{-E(v,h;\theta)}}{\sum_{v}\sum_{h}e^{-E(v,h;\theta)}} \tag{19}$$

At this point, maximizing the likelihood function of $p(v;\theta)$ is equivalent to maximizing the likelihood function of $P(v,h;\theta)$. That is, RBM training is to determine parameter θ such that $p(v;\theta)$ is maximized on a defined training set. Notice that RBM has inter-layer restrictions, so given a random input visible layer vector v conditions, all hidden layer neuron states are independent of each other, according to the formula to get the conditional probability formula corresponding to the hidden layer vector h. Similarly given a hidden layer vector h, the conditional probability of the corresponding visual layer vector v is as in Eq.

$$p(h\mid v)=\prod_{j}^{n}p(h_j\mid v) \tag{20}$$

$$p(v\mid h)=\prod_{i}^{m}p(v_i\mid h) \tag{21}$$

For the classification task, the structural units of the RBM must be binary states, so the activation probabilities of the neurons in each layer are obtained based on the sigmoid activation function $sigm(x)=\frac{1}{1+e^{-x}}$, which is common in classification problems:

$$p(h_j=1\mid v)=sigm(c_j+\sum_{i=1}v_iw_{i,j}) \tag{22}$$

$$p(v_i=1\mid h)=sigm(b_i+\sum_{j=1}w_{i,j}h_j) \tag{23}$$

The essence of RBM training is to better fit the training data by maximizing the probability that the learned model conforms to the distribution of the input data. That is, the probability value $p(v;\theta)$ in Eq. is maximized by continuously adjusting the internal parameters of the model $\theta=(W;b;c)$. According to the great likelihood estimation, taking logarithms on both sides of the equation and performing stochastic gradient descent (SGD) yields the result of the following equation:

$$\frac{\partial \log p(v;\theta)}{\partial \theta} = \frac{\partial \log\left(\frac{\sum_h e^{-E(v,h;\theta)}}{\sum_v \sum_h e^{-E(v,h;\theta)}}\right)}{\partial \theta}$$
$$= \frac{\partial \log\left(\sum_h e^{-E(v,h;\theta)}\right)}{\partial \theta} - \frac{\partial \log\left(\sum_v \sum_h e^{-E(v,h;\theta)}\right)}{\partial \theta} \tag{24}$$
$$= -\sum_h p(h \mid v) \frac{\partial E(v,h;\theta)}{\partial \theta} + \sum_v \sum_h p(v \mid h) \frac{\partial E(v,h;\theta)}{\partial \theta}$$

When calculating the above equation, $\frac{\partial E(v,h;\theta)}{\partial \theta}$ can be easily obtained, while $p(h \mid v)$ and $p(v \mid h)$ are directly calculated with high time complexity, so they are generally obtained using alternating Gibbs sampling:

$$\begin{array}{ll} h_0 \sim p(h \mid v_0) & v_1 \sim p(v \mid h_0) \\ h_1 \sim p(h \mid v_1) & v_2 \sim p(v \mid h_1) \\ \cdots\cdots & v_{k+1} \sim p(v \mid h_k) \end{array} \tag{25}$$

The training sample x_0 is used as input v_0 to the visual layer, and a number of alternating samples are performed to obtain sample v_{k+1} with joint probability approximation $p(v \mid h_k)$.

5.2. DBN Training Algorithm

Each layer of neurons of DBN implements a nonlinear transformation of the input data of that layer, and the essence of this transformation is the feature transformation of the input data, and when enough RBMs are stacked, the intrinsic features of the data can be well mined through many iterations of abstraction. The basic structure of deep confidence network is shown in Fig. 4.

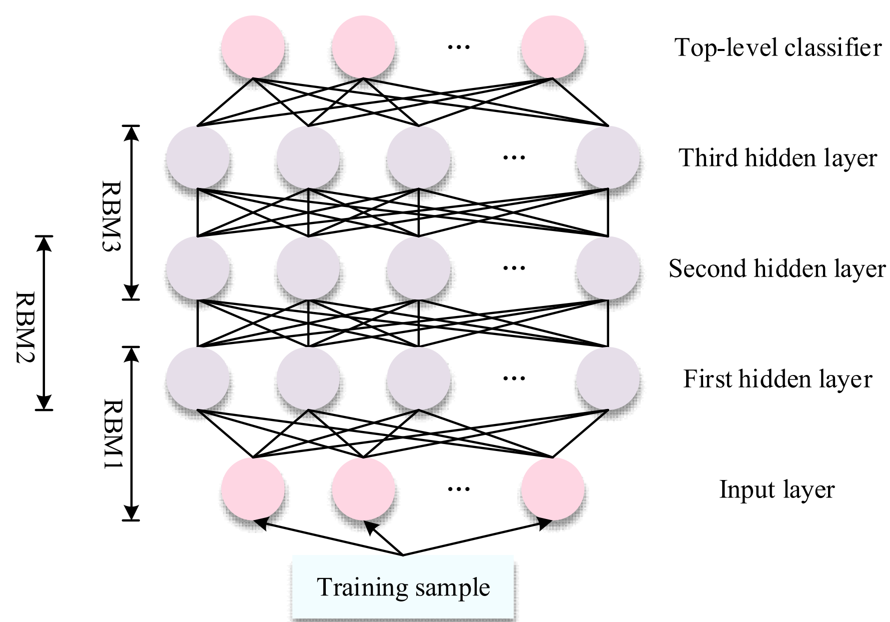


Fig. 4. Basic structure of the deep confidence network

The number of nodes in the input layer of the DBN model depends on the input data dimensions, and the number of nodes in the output layer depends on the number of categories. The training process consists of two phases: unsupervised pre-training and supervised fine-tuning [62,66]: the unsupervised RBM training rule is first utilized to train the DBN model sequentially layer by layer from the bottom up to obtain better initial values of the parameters, and then the overall network is fine-tuned from the top down using the (BP) algorithm [30]. Unsupervised pre-training is a distinguishing feature of DBN from other deep models. The process mainly utilizes the RBM training rule to map the input data into high-level abstract features through layer-by-layer greedy feature learning, which is the main reason for DBN's powerful feature extraction capability. In the pre-training process, only one RBM is considered at a time, and after the implied layer output is obtained, it is used as the visible layer input of the latter RBM, thus converting the deep network training into the training of multiple shallow networks, avoiding the complexity during the overall DBN training, and improving the training efficiency. The deep confidence network training process is shown in Fig. 5.

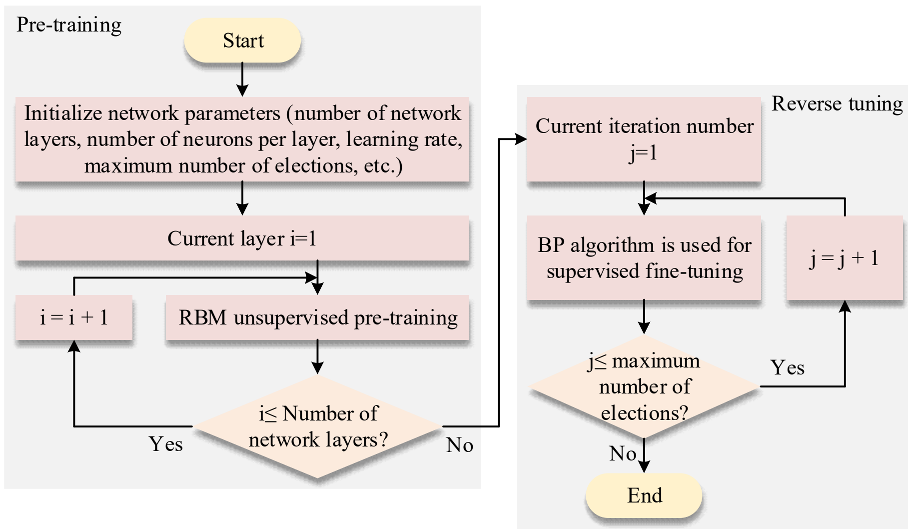


Fig. 5. Deep confidence network training process

5.3. DBN-Dropout Fault Identification Model

In the specific classification or regression task, the desired result is not to let the deep model fit the training data nearly perfectly, but hope that the trained model can make accurate and efficient judgments on the unknown data, and at this time, the error generated by the model acting on the unknown data is the generalization error. To ensure that the model fits these unknown data satisfactorily, it is necessary to make the generalization error as small as possible, which usually requires a large amount of labeled data for model training, however, the vast majority of the data in reality is unlabeled, which requires manual labeling, increasing the workload. In addition, as the depth of the network and the number of neuron nodes in the layer gradually increase, the parameters in the network also increase substantially, which leads to the generation of overfitting. Overfitting makes the network overlearn certain features that are specific to the training set but not to the unknown samples, and an intuitive manifestation of this is that there is a large gap between the model's effect on the training set and the test set, where the effect on the training set is ideal while the effect on the test set is far from being able to meet expectations.

6. Multi-dimensional mimetic wind turbine fault detection experiments

The evaluation metrics MSE, RMSE, MAE, and r2_score are selected in this section.

6.1. Experimental results and analysis

Ten experiments are conducted using the wind turbine fault diagnosis model proposed in this paper, with 110 iterations per experiment. The experimental results of the model in this paper are shown in

Table 5. The model has the highest recognition accuracy of 99.93% in the wind turbine fault recognition problem, and the average recognition accuracy reaches 99.903%.

Table 5. Experimental results of this paper

Experimental serial number	Training set loss value	Training set accuracy(%)	Test set loss value	Test set accuracy(%)
1	0.0044	99.84	0.0051	99.93
2	0.0032	99.89	0.0046	99.92
3	0.0037	99.9	0.0058	99.89
4	0.002	99.89	0.0063	99.88
5	0.0036	99.92	0.0057	99.9
6	0.0039	99.87	0.0069	99.92
7	0.0028	99.94	0.0067	99.93
8	0.0013	99.96	0.0072	99.87
9	0.0014	99.89	0.0067	99.89
10	0.0038	99.91	0.0044	99.9
Mean value	0.0030	99.901	0.0059	99.903

The trend of the recognition accuracy and Loss value of the model in this paper is shown in Fig. 6. As the number of iterations increases, the loss value gradually decreases and the accuracy gradually increases, indicating that the model performs well.

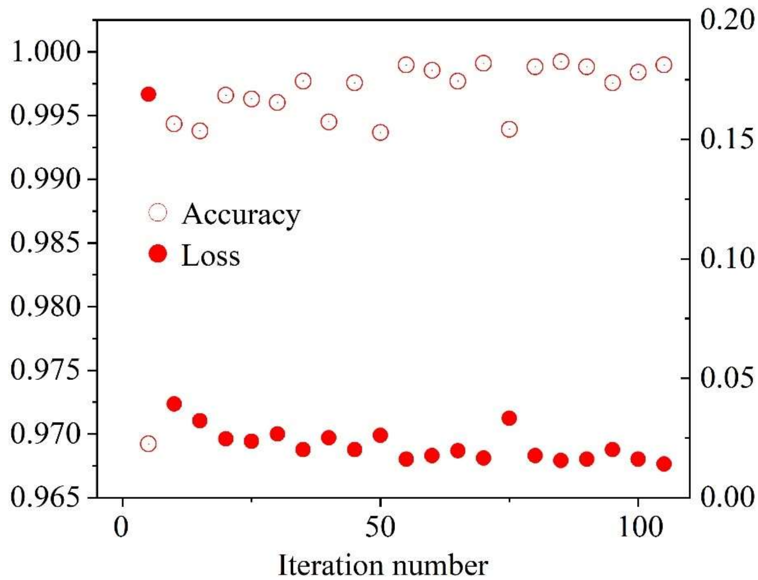


Fig. 6. The identification accuracy and loss value of this paper are changing

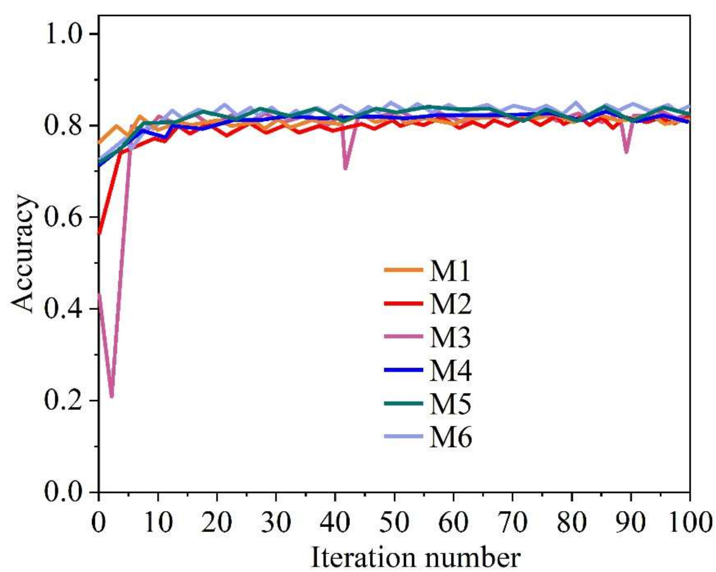
6.2. *Comparative Experiments*

In order to further verify the validity of the models in this paper, the models are compared and experimented. Five groups of advanced fault diagnosis models are set up to compare with the models in this paper, namely, the 1DCNN model, the LSTM model with added SENet attention (SE-LSTM), the Resnet18 network model, the dual-channel multiscale convolutional neural network model (DM-CNN), and the dual-channel multiscale convolutional-long- and short-term memory network model with no added attention (DM-CNN- LSTM). For the convenience of comparative description, each model adopts the designation M1-M6, and the parameters of different models are shown in Table 6.M1 represents the 1DCNN model.M2 represents the SE-LSTM model.M3 represents the Resnet18 model.M4 represents the DM-CNN model.M5 represents the DM-CNN-LSTM model.M6 represents the model of this paper.M6 represents the model of this paper.M6 represents the model of this paper.

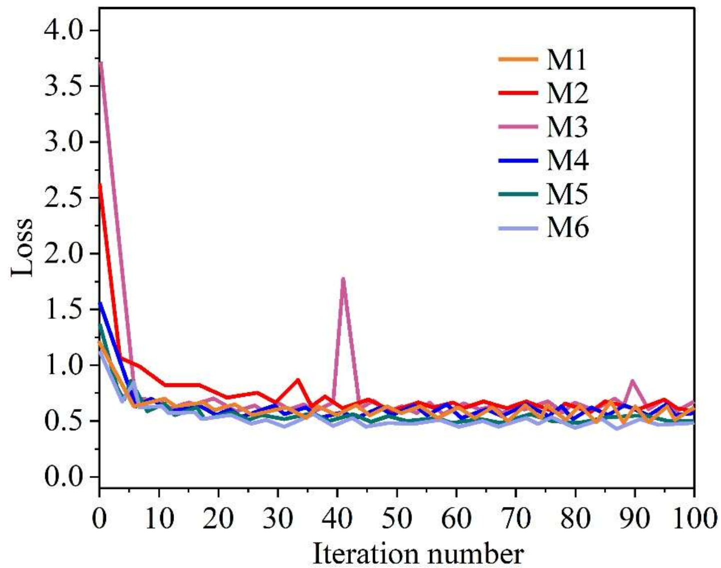
Table 6. Different model parameters

Code number	model	Layer number	Parameter quantity
M1	1DCNN	12	514563
M2	SE-LSTM	12	21623
M3	Resnet18	19	11281992
M4	DM-CNN	15	339782
M5	DM-CNN-LSTM	16	552355
M6	Ours	17	513936

A comparison of the different models' accuracy and Loss change trend is shown in Fig. 7. Figure (a) shows in detail the trend of different models' recognition accuracy during the iteration process.The M3 model has relatively poor recognition effect in the initial stage, which is much lower than the other five models. In contrast, the accuracy of the other five models in the initial stage has exceeded 90%, among which the M1 model has the best performance with an accuracy of 96.72%. As the number of iterations increases, all six models begin to converge near the 17th iteration and gradually reach a stable state. Among them, the M3 model shows two large fluctuations after convergence, which may be due to the Dropout parameter adjustment. When the number of iterations reaches 100, the M6 model performs significantly better than the other five models, and its recognition accuracy reaches 99.8%, while the M1 model has the lowest recognition accuracy of 99.25%, and the other models' recognition accuracies are all between the M1 and M6 models. Figure b demonstrates the trend of the loss values of different models during the iteration process. It is observed that the recognition accuracy of the model shows high consistency with the trend of the loss value. Specifically, the M3 model has the largest initial loss value and the largest volatility. The M1 model is stabilized with the largest value of loss. The M2 model loss values begin to converge at 40 iterations, and the rate of convergence is the slowest of the six models. The M6 model has the most stable loss profile and reaches steady state with the smallest loss value of 0.045. The rest of the models have stable loss values between the M1 and M6 models.



(a) Accuracy



(b) Loss

Fig. 7. Different model precision and loss change trend

The performance of different models under different evaluation metrics is shown in Table 7.The M6 model exhibits the lowest values under both MSE and RMSE evaluation criteria, 0.0002 and 0.01392, respectively, which reflects that the model has a very small deviation between the predicted and actual values, and its prediction accuracy is significantly higher than that of other models. The M1 model, on the other hand, has the highest MSE and RMSE values of 0.0017 and 0.0415, respectively, and performs the worst. The evaluation results in MAE and r^2_score show a slightly different ordering. In the MAE metric, lower values represent lower prediction errors and better model performance. On the contrary, a higher r^2_score indicates that the model explains the data better and has a higher prediction accuracy. The M6 model has the best MAE performance with a value of 0.0003 and also has the highest r2_score score of 0.9959. The M2 model has the largest MAE value of 0.0028 and the M1 model has the lowest r2_score score of 0.9681. Combining these four evaluation indicators, it can be concluded that the M6 model performs the best of all models, while the M1 model performs the worst. The models are ranked from best to worst as M6, M5, M3, M2, M4, and M1.

Table 7. Performance of different models under different evaluation indicators

	M1	M2	M3	M4	M5	M6
MSE	0.0017	0.0013	0.0011	0.0015	0.00062	0.0002
RMSE	0.0415	0.014	0.015	0.0142	0.01585	0.01392
MAE	0.0027	0.0028	0.0016	0.0023	0.0008	0.0003
r2_score	0.9681	0.9711	0.9715	0.9807	0.9918	0.9959

The different model accuracy and time comparisons are shown in Fig. 8. In terms of accuracy, the M6 model holds the first place with a significant lead over the other models with 99.88% recognition rate. However, in terms of computing speed, the M6 model is higher than the training time of the other four models, even though the time taken by the M6 model is 340.53 seconds lower than that of the M3 model, which is 1322.34 seconds. The training duration is closely related to the number of parameters that the models need to be trained.The M3 and M6 models require more training parameters than the other models, which leads to their relatively longer training duration. Once the model is trained, the time required to process a test data is only at the microsecond level, and the difference between several models is negligible, all of which have met the requirements of real-world applications. Therefore, the importance of model accuracy far exceeds the training time.

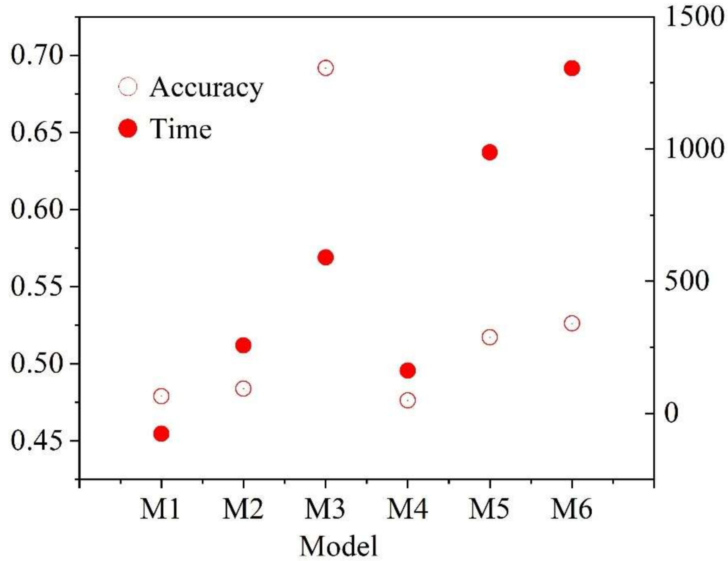
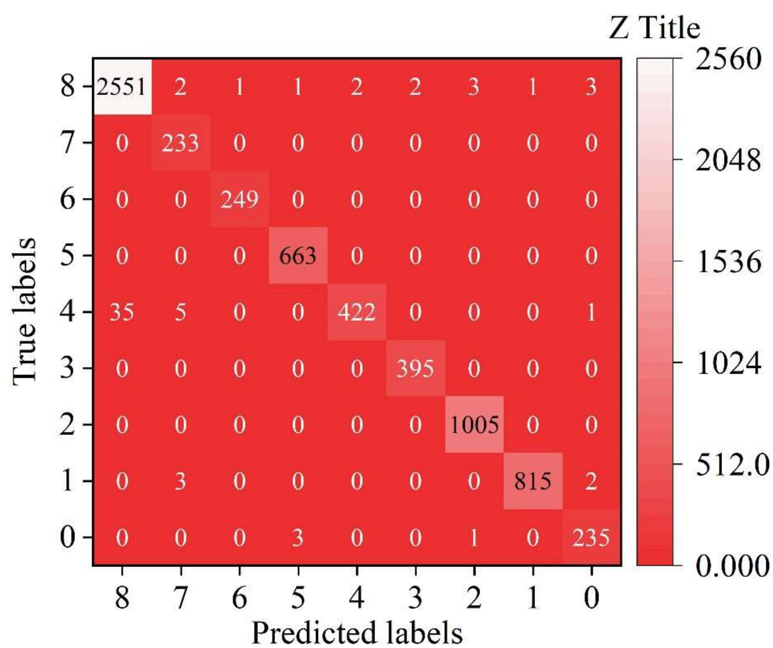
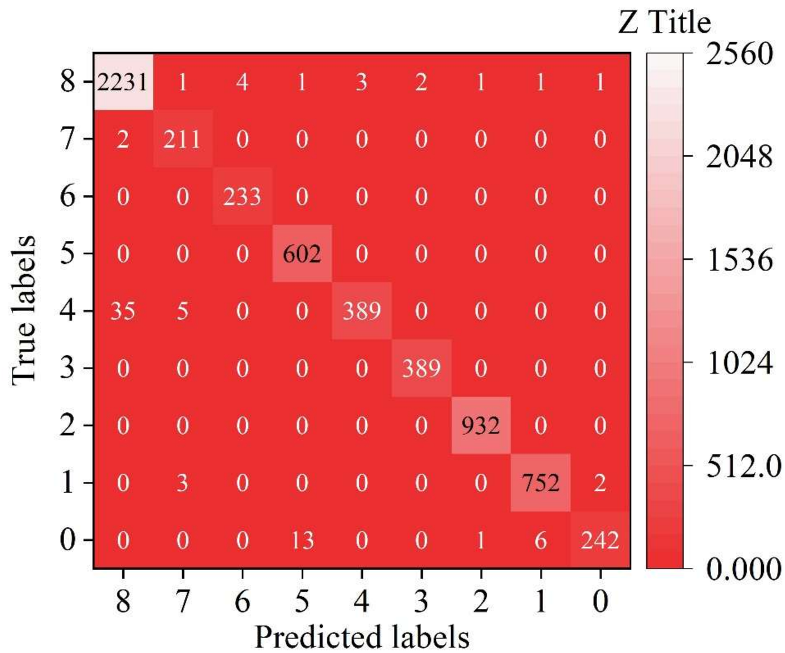


Fig. 8. Different model precision and time contrast

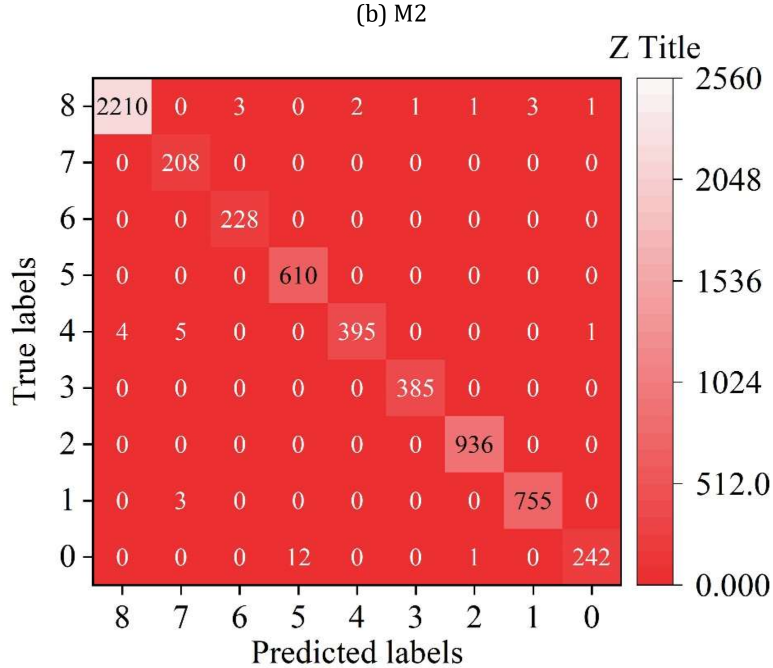
The confusion matrices of the different models are shown in Fig. 9 (Figs. a~f are M1, M2, M3, M4, M5, and M6, respectively). As can be seen from the figure, in contrast, the performance of model M6 is significantly superior, achieving 100% accuracy in the identification of fault types 3, 4, 5 and 6, which is extremely outstanding. This result reaffirms the excellent fault recognition capability of the model in this paper and shows that the model can be effectively applied to the fault diagnosis process of wind turbines.



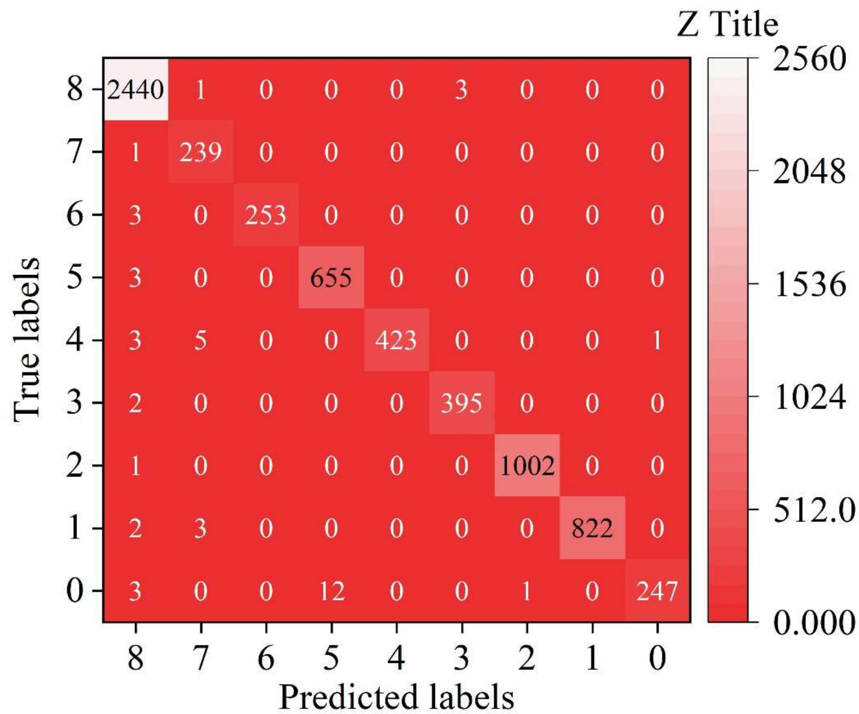
(a) M1


Z Title
2560
2048
1536
1024
512.0
0.000
True labels
Predicted labels
8 7 6 5 4 3 2 1 0
2231 1 4 1 3 2 1 1 1
2 211 0 0 0 0 0 0 0
0 0 233 0 0 0 0 0 0
0 0 0 602 0 0 0 0 0
35 5 0 0 389 0 0 0 0
0 0 0 0 0 389 0 0 0
0 0 0 0 0 0 932 0 0
0 3 0 0 0 0 0 752 2
0 0 0 13 0 0 1 6 242

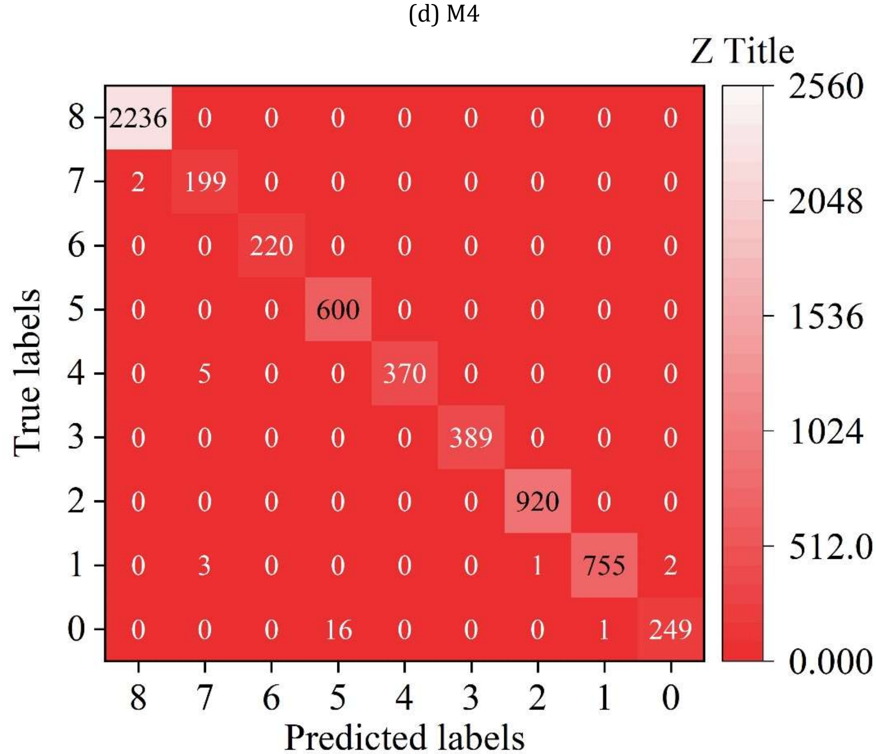
(b) M2


Z Title
2560
2048
1536
1024
512.0
0.000
True labels
Predicted labels
8 7 6 5 4 3 2 1 0
2210 0 3 0 2 1 1 3 1
0 208 0 0 0 0 0 0 0
0 0 228 0 0 0 0 0 0
0 0 0 610 0 0 0 0 0
4 5 0 0 395 0 0 0 1
0 0 0 0 0 385 0 0 0
0 0 0 0 0 0 936 0 0
0 3 0 0 0 0 0 755 0
0 0 0 12 0 0 1 0 242

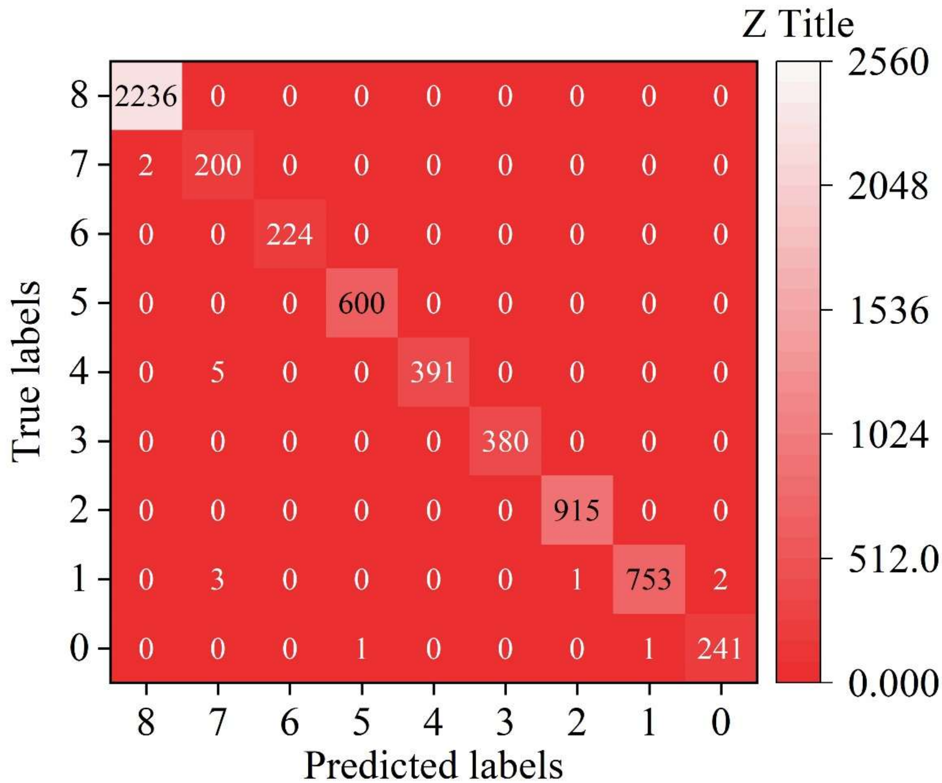
(c) M3


Z Title
2560
2048
1536
1024
512.0
0.000
True labels
Predicted labels
8 7 6 5 4 3 2 1 0
8 2440 1 0 0 0 3 0 0 0
7 1 239 0 0 0 0 0 0 0
6 3 0 253 0 0 0 0 0 0
5 3 0 0 655 0 0 0 0 0
4 3 5 0 0 423 0 0 0 1
3 2 0 0 0 0 395 0 0 0
2 1 0 0 0 0 0 1002 0 0
1 2 3 0 0 0 0 0 822 0
0 3 0 0 12 0 0 1 0 247

(d) M4


Z Title
2560
2048
1536
1024
512.0
0.000
True labels
Predicted labels
8 7 6 5 4 3 2 1 0
8 2236 0 0 0 0 0 0 0 0
7 2 199 0 0 0 0 0 0 0
6 0 0 220 0 0 0 0 0 0
5 0 0 0 600 0 0 0 0 0
4 0 5 0 0 370 0 0 0 0
3 0 0 0 0 0 389 0 0 0
2 0 0 0 0 0 0 920 0 0
1 0 3 0 0 0 0 1 755 2
0 0 0 0 16 0 0 0 1 249

(e) M5



(f) M6

Fig. 9. Confusion matrix of different models

7. Conclusion

In this study, a multi-dimensional anthropomorphic wind turbine state monitoring and fault detection model based on machine learning is proposed, and its effectiveness and feasibility in practical applications are verified through experiments.

In the 10th experiment of multi-dimensional anthropomorphic generator set state identification, the results of clustering operation on real-time wind speed and active power in the middle and high wind speed zone show the existence of four noise points. The results of clustering operation on gearbox oil temperature and main shaft temperature in low wind speed region show the existence of 1 noise point. Comparing the experimental data with the actual data it is found that the experimental results coincide with the actual situation. The clustering results of the improved DBSCAN algorithm can more accurately determine the current operating status of the wind turbine.

The DBN-Dropout fault recognition model of this paper is compared with other five advanced fault diagnosis models in the comparison experiment, and the results show that: the fault recognition accuracy of the model in this paper is as high as 99.903%, and the highest recognition accuracy is 99.93%, which has a significant advantage compared with other comparison models.

References

[1] Hossain, M. L., Abu-Siada, A., & Muyeen, S. M. (2018). Methods for advanced wind turbine condition monitoring and early diagnosis: A literature review. Energies, 11(5), 1309.

[2] Kong, Z., Tang, B., Deng, L., Liu, W., & Han, Y. (2020). Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units. Renewable Energy, 146, 760-768.

[3] Liu, Z., & Zhang, L. (2020). A review of failure modes, condition monitoring and fault diagnosis methods

for large-scale wind turbine bearings. Measurement, 149, 107002.

[4] Salameh, J. P., Cauet, S., Etien, E., Sakout, A., & Rambault, L. (2018). Gearbox condition monitoring in wind turbines: A review. Mechanical Systems and Signal Processing, 111, 251-264.

[5] Du, Y., Zhou, S., Jing, X., Peng, Y., Wu, H., & Kwok, N. (2020). Damage detection techniques for wind turbine blades: A review. Mechanical Systems and Signal Processing, 141, 106445.

[6] Chen, H., Liu, H., Chu, X., Liu, Q., & Xue, D. (2021). Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network. Renewable Energy, 172, 829-840.

[7] Cho, S., Choi, M., Gao, Z., & Moan, T. (2021). Fault detection and diagnosis of a blade pitch system in a floating wind turbine based on Kalman filters and artificial neural networks. Renewable Energy, 169, 1-13.

[8] Liu, Z., Zhang, L., & Carrasco, J. (2020). Vibration analysis for large-scale wind turbine blade bearing fault detection with an empirical wavelet thresholding method. Renewable Energy, 146, 99-110.

[9] Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., ... & Nenadic, G. (2019). Machine learning methods for wind turbine condition monitoring: A review. Renewable energy, 133, 620-635.

[10] Peeters, C., Guillaume, P., & Helsen, J. (2018). Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy. Renewable Energy, 116, 74-87.

[11] Tautz - Weinert, J., & Watson, S. J. (2017). Using SCADA data for wind turbine condition monitoring–a review. IET Renewable Power Generation, 11(4), 382-394.

[12] Cheng, F., Wang, J., Qu, L., & Qiao, W. (2017). Rotor-current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier. IEEE transactions on industry applications, 54(2), 1062-1071.

[13] Xiang, L., Wang, P., Yang, X., Hu, A., & Su, H. (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement, 175, 109094.

[14] Maheswari, R. U., & Umamaheswari, R. (2017). Trends in non-stationary signal processing techniques applied to vibration analysis of wind turbine drive train–A contemporary survey. Mechanical Systems and Signal Processing, 85, 296-311.

[15] Artigao, E., Martín-Martínez, S., Honrubia-Escribano, A., & Gómez-Lázaro, E. (2018). Wind turbine reliability: A comprehensive review towards effective condition monitoring development. Applied energy, 228, 1569-1583.

[16] Chen, L., Xu, G., Zhang, Q., & Zhang, X. (2019). Learning deep representation of imbalanced SCADA data for fault detection of wind turbines. Measurement, 139, 370-379.

[17] Gao, Z., & Liu, X. (2021). An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. Processes, 9(2), 300.

[18] Dhiman, H. S., Deb, D., Muyeen, S. M., & Kamwa, I. (2021). Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. IEEE Transactions on Energy Conversion, 36(4), 3462-3469.

[19] Wang, T., Han, Q., Chu, F., & Feng, Z. (2019). Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review. Mechanical Systems and Signal Processing, 126, 662-685.

[20] Dao, P. B., Staszewski, W. J., Barszcz, T., & Uhl, T. (2018). Condition monitoring and fault detection in wind turbines based on cointegration analysis of SCADA data. Renewable Energy, 116, 107-122.

[21] Zhao, H., Liu, H., Hu, W., & Yan, X. (2018). Anomaly detection and fault analysis of wind turbine components based on deep learning network. Renewable energy, 127, 825-834.

[22] Jiang, G., Xie, P., He, H., & Yan, J. (2017). Wind turbine fault detection using a denoising autoencoder with temporal information. IEEE/Asme transactions on mechatronics, 23(1), 89-100.

[23] Pandit, R. K., & Infield, D. (2018). SCADA - based wind turbine anomaly detection using Gaussian process models for wind turbine condition monitoring purposes. IET Renewable Power Generation, 12(11), 1249-1255.

[24] Hsu, J. Y., Wang, Y. F., Lin, K. C., Chen, M. Y., & Hsu, J. H. Y. (2020). Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning. Ieee Access, 8, 23427-23439.

[25] Li, H., Deng, J., Yuan, S., Feng, P., & Arachchige, D. D. (2021). Monitoring and identifying wind turbine generator bearing faults using deep belief network and EWMA control charts. Frontiers in Energy Research, 9, 799039.

[26] Canizo, M., Onieva, E., Conde, A., Charramendieta, S., & Trujillo, S. (2017, June). Real-time predictive maintenance for wind turbines using Big Data frameworks. In 2017 ieee international conference on prognostics and health management (icphm) (pp. 70-77). IEEE.

[27] Junshuai Yan,Yongqian Liu,Hang Meng,Li Li & Xiaoying Ren. (2024). Wind turbine generator early fault diagnosis using LSTM-based stacked denoising autoencoder network and stacking algorithm. International Journal of Green Energy(11),2477-2492.

[28] Paweł Knes & Phong B. Dao. (2024). Machine Learning and Cointegration for Wind Turbine Monitoring and Fault Detection: From a Comparative Study to a Combined Approach. Energies(20),5055-5055.

[29] Yunbin Sun,Zaohong Zhou,Qiang Li & Hongjun He. (2025). A novel risk assessment method for advanced and environmentally friendly construction technologies integrating RBM and I-OPA. Alexandria Engineering Journal648-660.

[30] Yongtao Xi,Xiang Zhang,Bing Han,Yu Zhu,Cunlong Fan & Eunwoo Kim. (2024). Advanced Human Reliability Analysis Approach for Ship Convoy Operations via a Model of IDAC and DBN: A Case from Ice-Covered Waters. Journal of Marine Science and Engineering(9),1536-1536.