entropy

MDPI

Article

A Rolling Bearing Fault Diagnosis Method Based on the WOA-VMD and the GAT

Yaping Wang [1,2], Sheng Zhang [2], Ruofan Cao [2], Di Xu [2] and Yuqi Fan [2,*]

[1] Key Laboratory of Advanced Manufacturing and Intelligent Technology, Harbin University of Science and Technology, Harbin 150080, China; wangyaping@hrbust.edu.cn
[2] School of Mechanical and Power Engineering, Harbin University of Science and Technology, Harbin 150080, China; cxzzxc147258@163.com (S.Z.); kk837910389@163.com (R.C.); dixu115@hrbust.edu.cn (D.X.)
* Correspondence: yuqifan@163.com

Abstract: In complex industrial environments, the vibration signal of the rolling bearing is covered by noise, which makes fault diagnosis inaccurate. In order to overcome the effect of noise on the signal, a rolling bearing fault diagnosis method based on the WOA-VMD (Whale Optimization Algorithm-Variational Mode Decomposition) and the GAT (Graph Attention network) is proposed to deal with end effect and mode mixing issues in signal decomposition. Firstly, the WOA is used to adaptively determine the penalty factor and decomposition layers in the VMD algorithm. Meanwhile, the optimal combination is determined and input into the VMD, which is used to decompose the original signal. Then, the Pearson correlation coefficient method is used to select IMF (Intrinsic Mode Function) components that have a high correlation with the original signal, and selected IMF components are reconstructed to remove the noise in the original signal. Finally, the KNN (K-Nearest Neighbor) method is used to construct the graph structure data. The multi-headed attention mechanism is used to construct the fault diagnosis model of the GAT rolling bearing in order to classify the signal. The results show an obvious noise reduction effect in the high-frequency part of the signal after the application of the proposed method, where a large amount of noise was removed. In the diagnosis of rolling bearing faults, the accuracy of the test set diagnosis in this study was 100%, which is higher than that of the four other compared methods, and the diagnosis accuracy rate of various faults reached 100%.

Keywords: fault diagnosis; signal decomposition; GAT; rolling bearing; WOA-VMD

check for updates

Citation: Wang, Y.; Zhang, S.; Cao, R.; Xu, D.; Fan, Y. A Rolling Bearing Fault Diagnosis Method Based on the WOA-VMD and the GAT. *Entropy* **2023**, *25*, 889. https://doi.org/10.3390/e25060889

Received: 21 April 2023
Revised: 7 May 2023
Accepted: 18 May 2023
Published: 1 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rolling bearings are widely used in various fields of the machinery industry due to their advantages of high speed, high efficiency and low noise. They are often used in harsh operating environments, which result in the large dispersion of life and a high failure rate. According to statistics, about 30% of mechanical failures in rotating machinery equipment that use rolling bearings are related to bearing damage [1]. The use of vibration signals generated during the working process for fault diagnosis may not only reduce the probability of accidents related to mechanical equipment, but it can also provide reliable decision support for the later maintenance of the equipment [2,3]. As fault detection is based on physical quantities (such as stator current [4], stray flux [5], thermal image [6], etc.), and because vibration-based measurement methods are lower in cost, are easier than direct observation, more sensitive to external interference, and often used in practical engineering, a vibration-based fault detection method is adopted in this paper.

In general, the vibration signal collected through the equipment is mixed with a considerable amount of noise. Due to the existence of noise, the performances of the mechanical equipment and equipment health prediction are very poor. To achieve optimal noise reduction, it is particularly important to select a suitable noise reduction method. The

accuracy of fault diagnosis depends to a certain extent on the noise reduction effect. For the problem of noise reduction, Chen applied EMD (Empirical Mode Decomposition) to process the original signal containing the interference signal and then further carried out a feature extraction process [7]. The extracted time-domain and frequency-domain features formed a new rolling bearing fault feature set. Zhou solved the strong noise in bearing vibration signals by combining the wavelet threshold with EMD, which achieved good results [8]. Although the above noise reduction methods are all effective, they also have certain drawbacks, such as mode mixing and end effect in the decomposition process [9,10]. Jia decomposed the collected rolling bearing vibration signals using EEMD (Ensemble Empirical Mode Decomposition) to effectively remove the noise [11]. Niu also used EEMD for noise reduction, which did not completely remove the interference signal in the vibration signal as compared to EMD and increased the calculation time [12]. Donoho first proposed the wavelet transform method, which adopted soft and hard thresholds to filter the wavelet coefficients in the decomposition process in order to achieve noise reduction [13]. Liang proposed a new method, the WT-IResNet (Wavelet Transform-Improved Residual Neural Network), for signal noise reduction [14]. Among the current signal noise reduction methods, it is wavelet transform that has a good effect [15]. Threshold setting is a key step in the wavelet decomposition process [16], which is usually selected by means of experimentation or manual experience. Since it cannot be adjusted according to different signals, the model is poorly adaptive. The VMD solves to a certain extent the problems of mode mixing and end effects, and it can effectively separate components and achieve the adaptive frequency-domain separation of the signal, mainly by obtaining the bandwidth and frequency center of each IMF [17]. Duo used the VMD method for the problem of bearings with high external environmental interference and a large noise component [18]. Wu proposed an independent component (ICA) algorithm based on the VMD because the fault signal generated in the gearbox was very weak and easily affected by external environmental noise and other factors [19]. It was proposed that two key parameters, the number of mode decompositions and the quadratic penalty factor, need to be set artificially before the VMD decomposition. However, this artificial setting is subjective and poorly generalized. Therefore, how to improve the choice of these two key parameters is very important for the subsequent diagnosis of rolling bearing faults.

After the appearance of new technology for rolling bearing condition detection and fault diagnosis, new ideas have constantly emerged with the progress of the times [20]. Rolling bearing fault diagnosis technology has also entered the peak of development [21]. Gao and Li used a convolutional neural network for the fault diagnosis of rolling bearings, and their results showed that it had high recognition accuracy for different types of faults [22,23]. Ma proposed a fault diagnosis method based on complementary integrated empirical mode decomposition combined with principal component analysis and limit gradient lifting; the method has been verified to be effective [24]. However, these fault diagnosis methods also have some limitations, and do not explore the relationships and interdependences of data. To solve this problem, some studies proposed to display data in the form of irregular plots [25–28]. Compared with traditional methods, it is necessary to establish graph data first, which increases the degree of complexity and poses severe challenges to standard neural network-based methods. It makes some important operations (e.g., convolution) easy to apply to the Euclidean domain, but it is difficult to model the graph data in non-Euclidean space [29]. A graph neural network (GNN) [30] is an artificial intelligence algorithm derived from graph theory, that can process graph data. Due to the ability of the GNN to model the interdependence between data and embed it into the extracted features, this method has gradually become a research hotspot in the field of rolling bearing fault diagnosis. Gao utilized all vibration samples to construct a undirected weighted K-nearest neighbor graph and also used a depth graph neural network for fault diagnosis. The effectiveness of this method was then verified with gear and bearing data [31]. Li proposed a graph convolutional network (GCN) combined with a weighted horizontal visibility graph (WHVG) for bearing fault diagnosis [32]. The WHVG is used to

convert a time series into graphical data from a geometric perspective, which improves the graph isomorphic network (GIN) to GIN+, while learning graph representation and fault classification. Finally, the effectiveness of the WHVG and GIN+ was verified with three actual bearing datasets. In another study, Zhang transformed acoustic signals into a graph and modeled the graph using a GCN to carry out the fault diagnosis of roller bearings [33]. Yu first constructed a graph dataset and then realized fault classification using a fast deep GCN [34]. Li transformed the vibration signals of rolling bearings into horizontal visibility graphs, and then modeled the graph data with a GNN to realize fault classification [35]. The above methods ignore the importance of input information and the interdependence of data. For different neighborhoods, the fault information is related to different degrees. Therefore, if the same weight is given, a certain degree of information will be lost, which will affect the outcome of the rolling bearing fault diagnosis.

Hence, a rolling bearing fault diagnosis method based on the WOA-VMD and GAT is proposed in this paper. The main contributions of this method are as follows:

1. By separating the signals with a fixed bandwidth, the problems of mode mixing and end effect are solved to some extent. Two key parameters in the VMD are determined using the WOA optimization algorithm, allowing the model to adaptively decompose the signal.
2. Node classification of graph structure data is carried out using the attention mechanism method, which assigns different attention weights to different neighborhoods so as to identify more important information.

2. Signal Decomposition and Reconstruction Based on the WOA-VMD

The variational mode decomposition can be estimated for each signal component by solving a frequency-domain variational optimization problem. It is assumed that the decomposed IMF components are all low-bandwidth signals, and that the fault feature frequency appears in the central frequency of the IMF components. The IMF components are then reconstructed to achieve the purpose, the VMD model is solved iteratively. The center frequency ω_k is used to decompose the decomposed k-mode components. Finally, the following problems are obtained:

$$\begin{gathered} \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \\ s.t \sum_k u_k = f \end{gathered} \tag{1}$$

where $\{u_k\} = \{u_1, \cdots, \mathrm{u}_k\}$, u_k is the original signal, $u_1, \ldots, u_k$ are the k mode components obtained after decomposition, and $\{\omega_k\} = \{\omega_1, \cdots, \omega_k\}$ is the frequency center of each mode component after decomposition. If Equation (1) is to be solved, it needs to be converted into the unconstrained problem of Equation (2). The conversion mode is mainly solved with the use of Lagrange multipliers and quadratic penalty terms.

$$\begin{gathered} L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \| \partial_t [(\delta(t) + \frac{j}{\pi t}) \times u_k(t)] e^{-j\omega_k t} \| + \| f(t) - \\ \sum_k u_k(t) \|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle, \end{gathered} \tag{2}$$

where α is the penalty factor, and $\lambda(t)$ is the Lagrange multiplier.

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |u_k(\omega)|^2 d\omega}{\int_0^\infty |u_k(\omega)|^2 d\omega}, \tag{3}$$

The two parameters of the VMD, the penalty factor α and the decomposition layers k, need to be considered in the decomposition process to create certain limitations for the VMD. Too small or too large a value will affect the algorithm's effect, so the optimal parameter combination [k, α] needs to be determined. At present, the center frequency observation method is widely used. This method mainly determines the value of k by observing the center frequency under different values of k, without an accurate basis, and

it can only determine the number of modes k, but not the penalty parameter α, which ultimately leads to poor noise reduction.

2.1. Whale Optimization Algorithm

In this paper, the whale optimization algorithm is used to adaptively determine the two parameters mentioned above in order to achieve a better noise reduction effect. In the process of hunting, humpback whales surround their prey in groups and move in a spiral, during which they constantly spit out bubbles, thus forming a spiral "bubble net" shown in Figure 1. This will make the space for the prey's movement smaller and smaller, until it can be swallowed in a single bite [36].

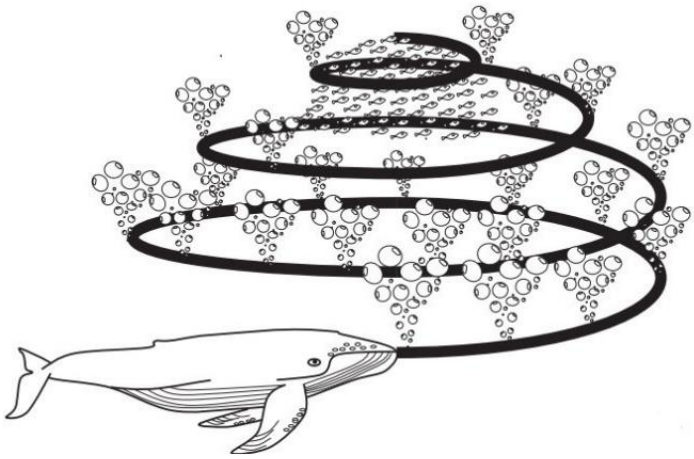


Figure 1. Humpback whales simulate "bubble net" feeding behavior.

However, the algorithm initially does not know the optimal location, and the WOA is goal-oriented to find the prey. When whales find the best prey, they also find the best whale location, thus allowing the rest of the whales to move closer to it. This behavior can be expressed as follows:

$$\vec{D} = \left| \vec{C} \cdot \vec{X^*}(t) - \vec{X}(t) \right|, \tag{4}$$

$$\vec{X}(t+1) = \vec{X^*}(t) - \vec{A} \cdot \vec{D}, \tag{5}$$

where $\vec{X}$ is the position vector, $\vec{X^*}$ represents the optimal position obtained currently, $\vec{A}$ and $\vec{C}$ are the coefficient vectors, t is the current number of iterations and $\vec{D}$ is the distance between the prey and the whales. If there is a better solution, $\vec{X^*}$ will be iteratively updated.

Vectors $\vec{A}$ and $\vec{C}$ can be calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a}, \tag{6}$$

$$\vec{C} = 2 \cdot \vec{r}, \tag{7}$$

where $\vec{a}$ is linear in its descent from 2 to 0, while $\vec{r}$ is a random vector within [0, 1].

Simulating the humpback whale's unique hunting method is achieved through a spiral motion that updates the position and a narrowing ring mechanism, which is often referred to as the bubble net strategy. Assuming that the probability of prey capture using these two methods is 50%, the total strategy is expressed as follows:

$$X(t+1) = \begin{cases} \vec{X^*}(t) - \vec{A} \cdot \vec{D}, & \{ 0 \le p < 0.5 \\ \vec{D'} \cdot e^{bl} \cos(2\pi l) + \vec{X^*}(t) & \{ 0.5 \le p \le 1 \end{cases}, \tag{8}$$

where $\vec{D'} = \left|\vec{X^*}(t) - \vec{X}(t)\right|$, b is the coefficient of the spiral search, and l is a random number within $[-1, 1]$.

In addition to the above strategy, it will also hunt according to its position, again through a change in vector $\vec{A}$. When $|\vec{A}| > 1$, it moves away from the prey; conversely, it moves closer to the prey. In this way, global optimization properties can be effectively improved. The mathematical model is as follows:

$$\vec{D} = \left|\vec{C} \cdot \vec{X}_{random}(t) - \vec{X}(t)\right|, \tag{9}$$

$$\vec{X}(t+1) = \vec{X}_{random}(t) - \vec{A} \cdot \vec{D}, \tag{10}$$

where $\vec{X}_{random}$ is the whale position randomly selected from the current group of whales.

Then, the minimum value of envelope entropy is selected as the fitness function; the envelope entropy represents the sparsity of the original signal. When there is more noise in the IMF and less effective information, the envelope entropy value is larger; otherwise, the envelope entropy value is smaller.

The envelope spectrum of the signal $X(i)(i = 1, 2, \ldots, N)$ is calculated with the following equation:

$$\left.\begin{array}{l} E_p = -\sum_{j=1}^{N} p_j lg p_j \\ p_j = a(j) / \sum_{j=1}^{N} a(j) \end{array}\right\}, \tag{11}$$

where $a(j)$ is the envelope signal of k mode components decomposed by the VMD after Hilbert demodulation, p_j is the sequence of probability distributions obtained by calculating the normalization of $a(j)$, N is the number of sampling points, and the entropy value of a calculated sequence of probability distributions p_j is the envelope entropy E_p.

2.2. WOA-VMD Parameter Optimization

Since the WOA is simple, and easy to implement and because few parameters were set, it was decided that the penalty factor and decomposition layers in the VMD decomposition would be optimized first. The steps are shown in Figure 2.

(1) Set the number of whales, the maximum number of iterations and the optimization dimension, and initialize the position information. Set the mode component and penalty factor as $k = [100,\ 3]$ and $\alpha = [2000,\ 7]$;
(2) Use the VMD algorithm to decompose the input signal and obtain each IMF function. Calculate the envelope entropy of each IMF according to Equation (11). The envelope entropy was used as a fitness function to find the optimal whale location and retain it;
(3) Start the iteration. Generate a random number p in the interval $(-1, 1)$. If $p < 0.5$, it is directly transferred to step 4; otherwise, Equation (8) is used for position update, namely for spiral contraction;
(4) Determine the value of $|A|$. If $|A| < 1$, update the position type of Equation (5), surrounded by the contraction; otherwise, update the position according to Equation (10), that is, change it to random exploration;
(5) Calculate the fitness of each whale and compare it with the previously reserved optimal position. If it is better, replace it with the new optimal solution;
(6) Determine whether the iteration is terminated. If $t \leq t_{max}$, then $t = t + 1$; return to Step 3. Otherwise, the iteration ends and the optimal parameter combination $[k, \alpha]$ is saved.

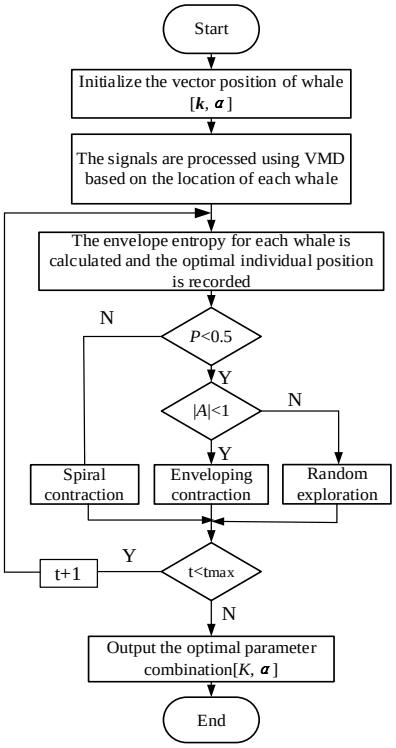


Figure 2. WOA Optimization Process.

2.3. Screening of IMF Component Coefficients

To filter the decomposed mode components, the Pearson correlation coefficient method is adopted in this paper.

2.3.1. Pearson Correlation Coefficient Analysis

The Pearson correlation coefficient method focuses on the correlation between signals to determine the similarity between two signals. The Pearson correlation coefficient method is as follows:

$$\rho_{x_t x_{IMF}} = \frac{cov(x_t, x_{IMF})}{\sigma_{x_t}\sigma_{x_{IMF}}} = \frac{E(x_t - E(x_t))(x_{IMF} - E(x_{IMF}))}{\sigma_{x_t}\sigma_{x_{IMF}}}, \tag{12}$$

$$\sigma_{x_t} = \sqrt{E({x_t}^2) - E^2(x_t)}, \tag{13}$$

$$\sigma_{x_{IMF}} = \sqrt{E({x_{IMF}}^2) - E^2(x_{IMF})}, \tag{14}$$

$$cov(x_t, x_{IMF}) = E(x_t x_{IMF}) - E(x_t)E(x_{IMF}), \tag{15}$$

where $\rho_{x_t x_{IMF}}$ represents the overall correlation coefficient, E represents the expected value, and $cov(x_t, x_{IMF})$ represents covariances of x_t and x_{IMF}.

The equation for the correlation coefficient value $r(x_t, x_{IMF})$ is as follows:

$$r(x_t, x_{IMF}) = \frac{\sum(x_t - \overline{x_t})(x_{IMF} - \overline{x_{IMF}})}{\sqrt{\sum(x_t - \overline{x_t})^2}\sqrt{\sum(x_{IMF} - \overline{x_{IMF}})^2}}, \tag{16}$$

2.3.2. Correlation Component Discrimination

The value range of $r(x_t, x_{IMF})$ is $[-1, 1]$; that is, $-1 \leq r(x_t, x_{IMF}) \leq 1$. Therefore, the classification of $r(x_t, x_{IMF})$ is shown in Table 1.

Table 1. Correlation discrimination table.

Interrelation	$r(x_t, x_{IMF})$ Coefficient Values
Very weakly correlated or uncorrelated	0.0–0.2
weakly correlated	0.2–0.4
Moderate correlation	0.4–0.6
Strongly related	0.6–0.8
Extremely strong correlation	0.8–1.0

In this section, reference [37] is used to select the mode components. In general, it is considered that the mode components with a strong correlation or above the original signal are those with less noise information. Mode components with correlation degrees below a strong correlation are removed, and then, mode components with a strong correlation or above are reconstructed to complete the noise reduction.

The signal decomposition and reconstruction method proposed in this paper uses the VMD to decompose signals, which can effectively avoid the phenomenon of mode mixing in noise reduction with the traditional method, and can retain the effective information of the original signal. Then, two key parameters, k and α, of the VMD are determined adaptively using the WOA optimization algorithm to improve the generalization ability of the model.

This method is used for noise reduction. The parameters are initialized for the WOA. The envelope entropy value corresponding to each whale is calculated and recorded as optimal. The optimal combination of parameters is output, and the optimal k and α are used to decompose the original signal. Pearson is then used to filter the obtained results. Specific steps are shown in Figure 3.

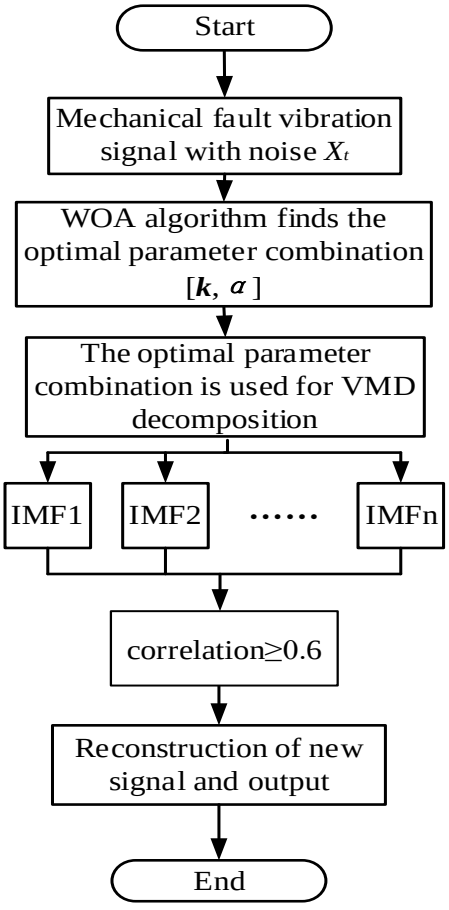


Figure 3. WOA-VMD-based signal noise reduction process.

3. Fault Diagnosis of Rolling Bearing Based on the GAT

If the fault can be found in time during the actual operation of rolling bearings and maintenance or remedial measures can be taken, which is of great significance for the reliability of the bearing operation. In this section, fault diagnosis methods of rolling bearings are studied based on Section 2. The GAT network model is established to diagnose rolling bearings. Meanwhile, a multi-head attention mechanism is used to perform node classification of graph structure data. To investigate ways of combining the two methods, increasing the weight of important information and improving diagnostic accuracy. At the same time, the parameter settings and data sets in the experiment are elaborated, and the experimental results are analyzed in detail.

3.1. KNN Graph Construction Method for Fault Diagnosis Data

Graph attention neural networks need to build graphs to represent the correlation between different faults as input data. Therefore, for the data set V after noise reduction using the WOA-VMD, the concept of Graph is introduced in order to represent the data itself and the relationship between them. Additionally, the KNN method is used to construct the graph model.

In the KNN figure, for each node (e.g., fault data point in V) to find the first K nearest neighbor points, the nearest neighbor of the obtained node x_i can be expressed as follows:

$$Ne(x_i) = KNN(K', x_i, \psi), \tag{17}$$

where NN returns the K nearest neighbors of the node x_i in set Ψ, $\Psi = [x_{i+1}, x_{i+2}, \ldots, x_{i+m}]$ denotes that there are m samples, and $Ne(x_i)$ represents the neighbors of the node x_i.

The edge weight between KNN nodes can be estimated using the Gaussian kernel weight function, and defined as follows:

$$e_{ij} = \exp\left(-\frac{\left(\|x_i, x_j\|\right)^2}{2\xi}\right), x_j \in Ne(x_i), \tag{18}$$

where e_{ij} is edge weight between nodes x_i and x_j, and ξ is the bandwidth of the Gaussian kernel. An example KNN composition is shown in Figure 4 [29].

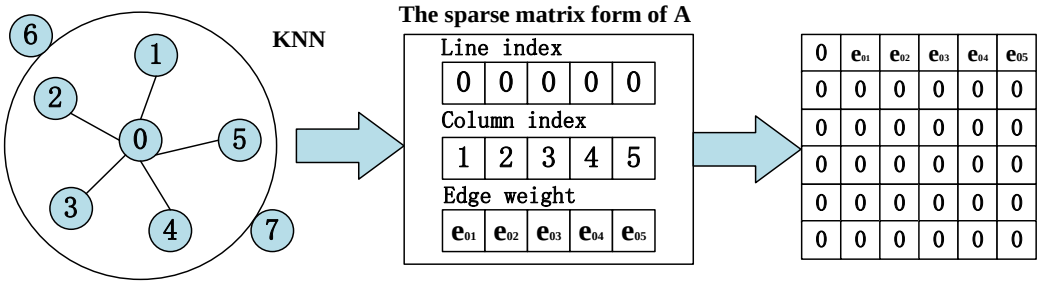


Figure 4. Construct an example of adjacency matrix based on KNN.

The rolling bearing fault diagnosis data in this paper are based on PyTorch Geometry (PYG) [38], a deep learning framework developed by PyTorch to construct undirected graphs. Firstly, nine different types of data from Case Western Reserve University are input, the number of labels is set, the time-domain vibration signal is loaded as the input, the graph model is constructed using *KNN*, and weights are assigned. Then, PYG is used for data encapsulation to complete the establishment of the graph model. The composition process is shown in Figure 5.

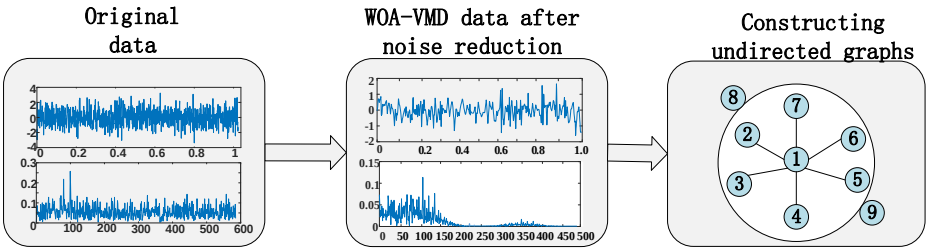


Figure 5. Composition process of the undirected graph of fault data.

3.2. Graph Attention Layer Construction Method

This section will start from the attention layer of a single graph. Firstly, node information in the constructed graph model is input, and then input features are transformed into higher-order features via a linear transformation. In addition, attention (weights) is allocated to each node through self-attention; a represents the shared attention mechanism, and $R^{F'} \times R^{F'} \to R$, which is used to calculate the attention coefficient e_{ij}, that is, the influence coefficient of node i on node j.

$$e_{ij} = a\left(W\vec{h}_i, W\vec{h}_j\right), \tag{19}$$

The attention calculation above only considers any two nodes in the graph, whereas in the general case, every node in the graph needs to be considered; thus, the overall graph information is lost. Therefore, this section only calculates the correlation between node i and node $j \in N_i$ in its neighborhood to the target node, where N_i is a domain of node i, and then normalizes it in all j options with the softmax function.

$$\alpha_{ij} = softmax_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k \in N_i} exp(e_{ik})}, \tag{20}$$

Using *LeakyReLU* as the activation function.

$$e_{ij} = LeakyReLU(\vec{a}^T[W\vec{h}_i \parallel W\vec{h}_j]), \tag{21}$$

where $\parallel$ indicates the splicing operation, and T indicates transposition. The complete equation for calculating the weight factor is shown below.

$$\alpha_{ij} = \frac{LeakyReLU(\vec{a}^T[W\vec{h}_i \parallel W\vec{h}_j])}{\sum_{k \in N_i} exp(LeakyReLU(\vec{a}^T[W\vec{h}_i \parallel W\vec{h}_k]))}, \tag{22}$$

The attention coefficient after normalization is calculated for the corresponding linear combination of features, and the final output feature vector of each node is obtained after calculation via the nonlinear activation function as shown below.

$$\vec{h}_1' = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W\vec{h}_j\right), \tag{23}$$

In addition, in this section, weights are mainly assigned through a multi-head attention mechanism, which is detailed in Figure 6. Each head attention mechanism will finally do a summation and averaging process to $\vec{h}_1'$ [39].

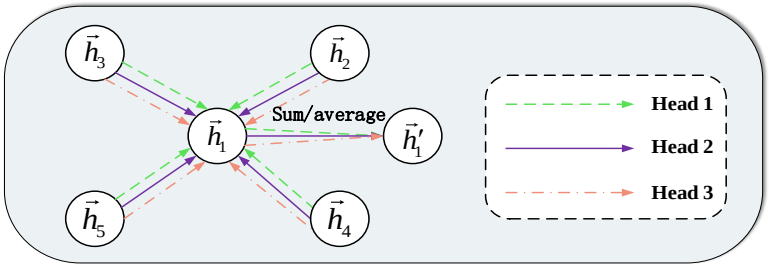


Figure 6. Multi-head attention mechanism schematic diagram.

The output of the above equation is stitched together using the following independent attention mechanisms:

$$\vec{h}_i' = \|_{\beta=1}^{K} \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h}_j\right), \tag{24}$$

where $\|$ represents the splicing operation, α_{ij}^k represents the normalized attention coefficient calculated by the β-th attention mechanism, and W^k is the weight matrix of the corresponding input linear transformation. It is worth noting here that, in this setup, the final output h' returned will consist of the K and F' features of each node, not just F'. The specific flow is shown in Figure 7.

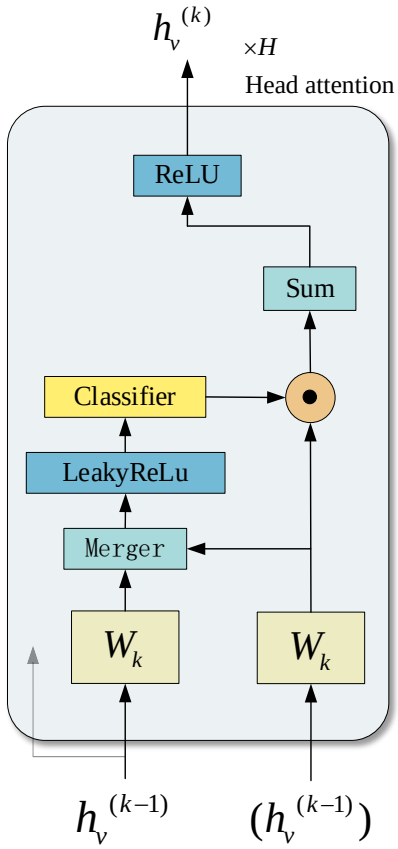


Figure 7. Diagram of GAT model.

3.3. Formatting of Mathematical Components

In this section, the graph attention neural network model is constructed, and some notations used are introduced. The graph is represented as $G = \{V, E\}$. The graph adjacency matrix is represented as A. The node feature matrix of the graph is represented as $F \in R^{N \times d}$. N is used to represent the number of nodes in the graph, and d is used to

represent the dimension of the node features. The features of a node in the graph are each a row of F.

The basic framework of the graph attention neural network is the combination of the graph filtering layer and nonlinear activation layer. Figure 8 shows two graph filtering layers and activation layers. The output of the i-th graph filter layer is denoted as $F^{(i)}$. In particular, $F^{(0)}$ is initialized to the nodal feature matrix F. The output dimension of the i-th graph filter layer is denoted as d_i. Since the structure of the graph does not change, it follows that $F^{(i)} \in R^{N \times d_i}$. The i-th layer of the graph filter layer can be described as follows:

$$F^{(i)} = h_i\Big(A, \alpha_{i-1}(F^{i-1})\Big), \tag{25}$$

where $\alpha_{i-1}()$ denotes the activation function applied element-by-element after the $(i-1)$-th graph filter layer. It is worth noting that α_0 denotes a constant function; as in practice, the input features are not usually activated.

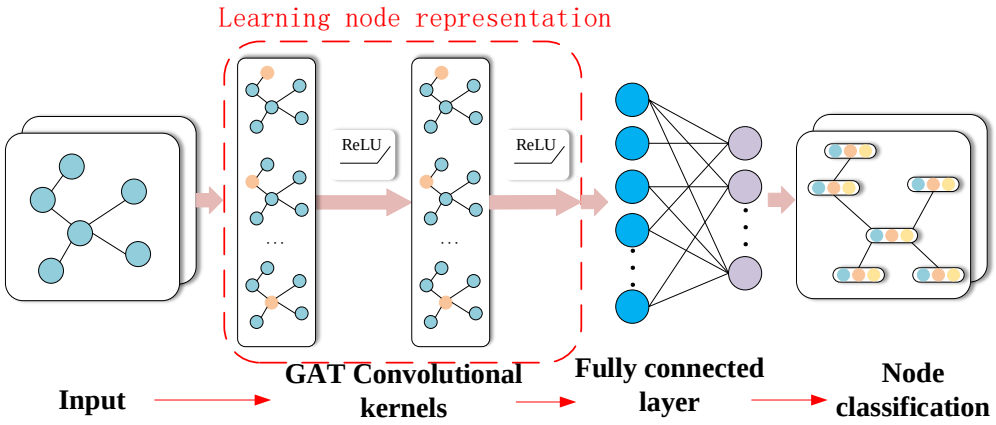


Figure 8. Schematic diagram of GAT layer structure.

Depending on the specific downstream task, the final output $F^{(L)}$ can be used as an input for a particular layer. The downstream task is parametric learning. The GAT model in this paper takes the entire graph as inputs to generate node representations, which are then used to train a node classifier. Specifically, let $GAT_{node}()$ represent a GAT model with multiple graph filter layers stacked. The function $GAT_{node}()$ takes the graph structure and node features as inputs and the learned node features as the output, and is expressed as follows:

$$F^{(out)} = GAT_{node}(A, F, \Theta_1), \tag{26}$$

where Θ_1 represents a model parameter, $A \in R^{N \times N}$ is the adjacency matrix, $F \in R^{N \times d_{in}}$ represents the node features of the input, and $F_{out} \in R^{N \times d_{out}}$ indicates the node features of the output.

The output node features are then used for node classification, as shown below:

$$Z = Soft\big(F^{out}; \Theta_2\big), \tag{27}$$

where $Z \in R^{N \times C}$ represents the output node category probability matrix, and $\Theta_2 \in R^{d_{out} \times C}$ is the parameter matrix that converts the feature F_{out} into a dimension equal to class number C.

The i-th line of Z represents the category distribution of the prediction node, and the prediction label is usually the label with the highest probability. The whole process can be summarized as follows:

$$Z = f_{GAT}\Big(A, F^{(ip)}; \Theta\Big), \tag{28}$$

where function $f_{GAT}()$ contains the processes of Equations (24) and (25), Θ contains Θ_1 and Θ_2. The parameter Θ can be learned by minimizing the following objective function:

$$L_{train} = \sum_{v_i \in V_l} l(f_{GAT}(A, F; \Theta)_i, y_i), \tag{29}$$

where $f_{GAT}(A, F; \Theta)_i$ represents the i-th line of the matrix, that is, the category probability distribution of node v_i; y_i represents its corresponding label; and $l(\cdot, \cdot)$ represents some kind of loss function, such as a cross-entropy loss function.

Figure 9 shows the fault diagnosis flow chart of the GAT.

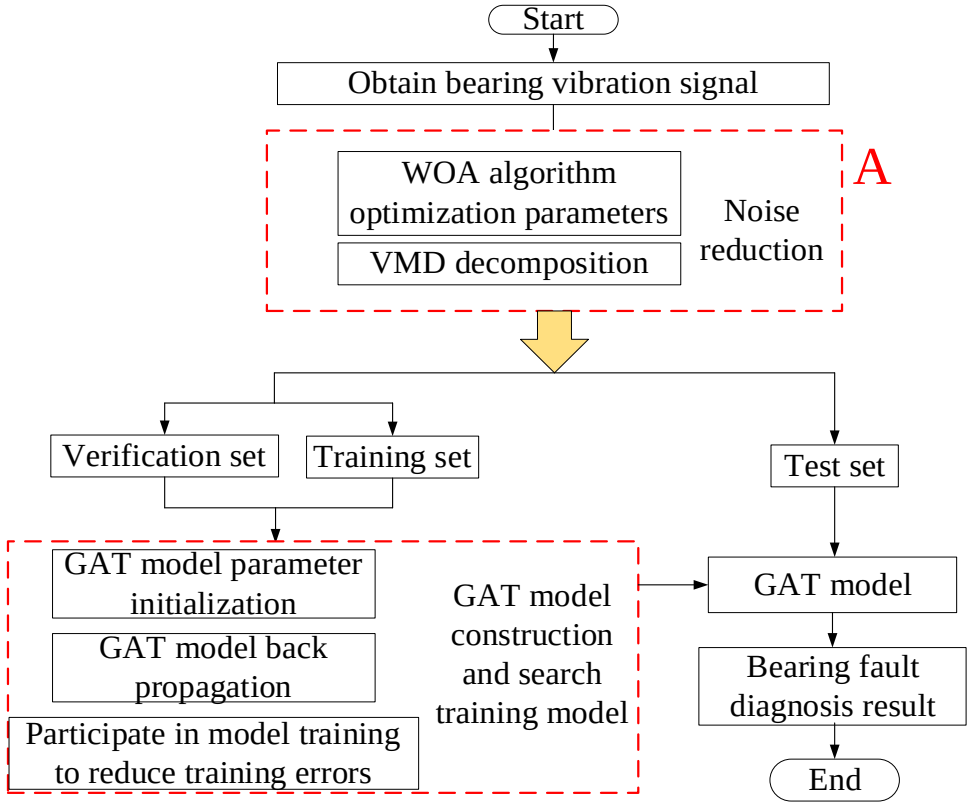


Figure 9. Fault diagnosis flow chart.

Step 1: The signal reconstruction and decomposition methods in Part A of the figure are described in Section 2.

Step 2: The GAT model is initialized and relevant model parameters are set.

Step 3: The data after noise reduction are initialized, and the graph model is constructed through the *KNN* method. The constructed graph model is divided into the test, training and validation sets.

Step 4: The training set data are input into the GAT model, the model is trained, the output error is obtained through the validation set, and the error is back-propagated to update the network model parameters.

4. Experimental Verification

4.1. WOA-VMD

In this section, the simulation fault signal and the test bench data in our laboratory are used to verify the superiority and effectiveness of the WOA-VMD noise reduction algorithm.

4.1.1. Simulation Verification

Firstly, the simulation fault signal is used for verification. The equation of the simulation vibration signal is as follows:

$$\begin{cases} s_1 = 0.2\cos(2\pi f_1 t + 10) \\ s_2 = 0.4\sin(2\pi f_2 t + 10) \\ s_3 = 0.2\sin(2\pi f_3 t) \end{cases}, \tag{30}$$

where $f_1 = 80$ Hz, $f_2 = 200$ Hz, $f_3 = 300$ Hz, sampling points $N = 1024$.

The time and frequency domain diagrams of the simulation signal s_1 are shown in Figure 10.

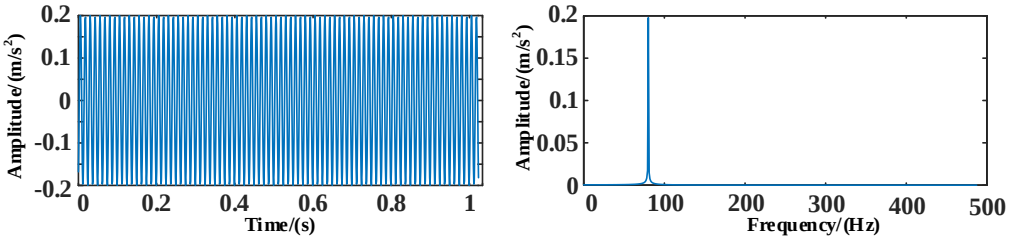


Figure 10. The time and frequency domain diagrams of signal s_1.

The time and frequency domain diagrams of the simulation signal s_2 are shown in Figure 11.

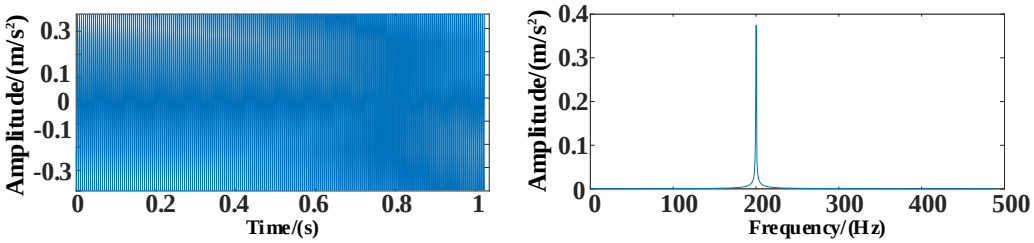


Figure 11. The time and frequency domain diagrams of signal s_2.

The time and frequency domain diagrams of the simulation signal s_3 are shown in Figure 12.

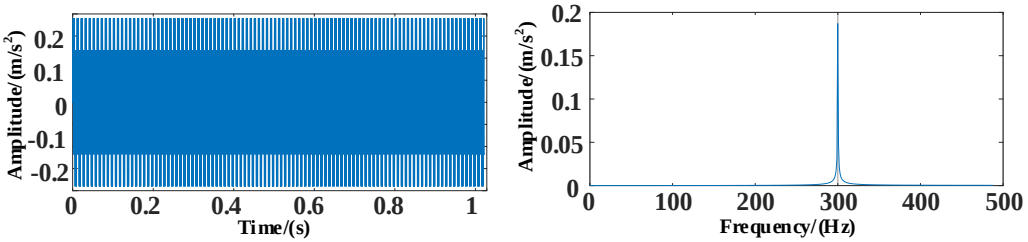


Figure 12. The time and frequency domain diagrams of signal s_3.

Figure 13 shows the time and frequency domain diagram of the simulation signal Z after mixing.

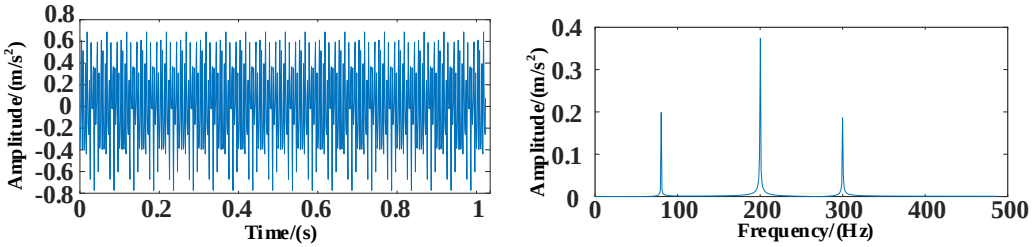


Figure 13. The time and frequency domain diagrams of signal Z.

The Gaussian white noise $n(t)$ of -10 dB [40] is added to signal Y. The time and frequency domain diagrams after adding noise are shown in Figure 14.

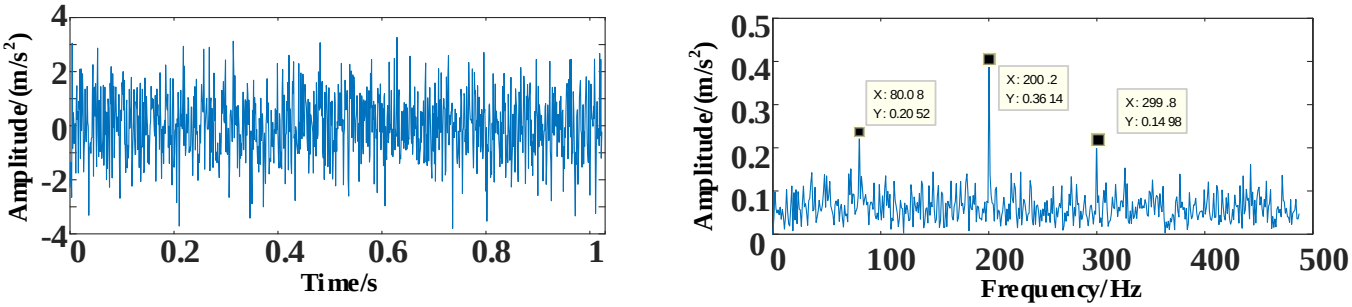


Figure 14. The time and frequency domain diagrams of signal after noise addition.

It can be seen in Figure 14 that the simulated signal after adding noise $n(t)$ is more real than the one before. The range of variation in the time-domain diagram increases, and the difference in amplitude is likewise increased. The amplitude interleaving in the frequency domain is due to the addition of simulated high-intensity noise.

In what follows, the mixed signal with noise is input into the WOA-VMD model for decomposition, and multiple mode components and frequency-domain diagrams are obtained. The number of decomposition layers and penalty factor in the VMD algorithm is optimized using the WOA algorithm, and the best results are obtained. The whale algorithm parameters are set, as in Table 2.

Table 2. Whale optimization algorithm parameter setting value.

Parameter	Values
Population size	10
Maximum number of iterations	20
Number of variables	2
Range of decomposition layers	[100,3]
Penalty factor range	[2000,7]

The final parameters are obtained after 20 iterations. The iteration curve of penalty factor α, shown in Figure 15a, enters convergence after the fifth time, with a final convergence value of 1652. The iteration curve of the optimal decomposition layer k is shown in Figure 15b. After the second iteration, the optimal solution is found to be eight layers. The iterative curve of the envelope entropy, shown in Figure 15c, enters convergence after the thirteenth iteration, with the final envelope entropy value of 9.7633.

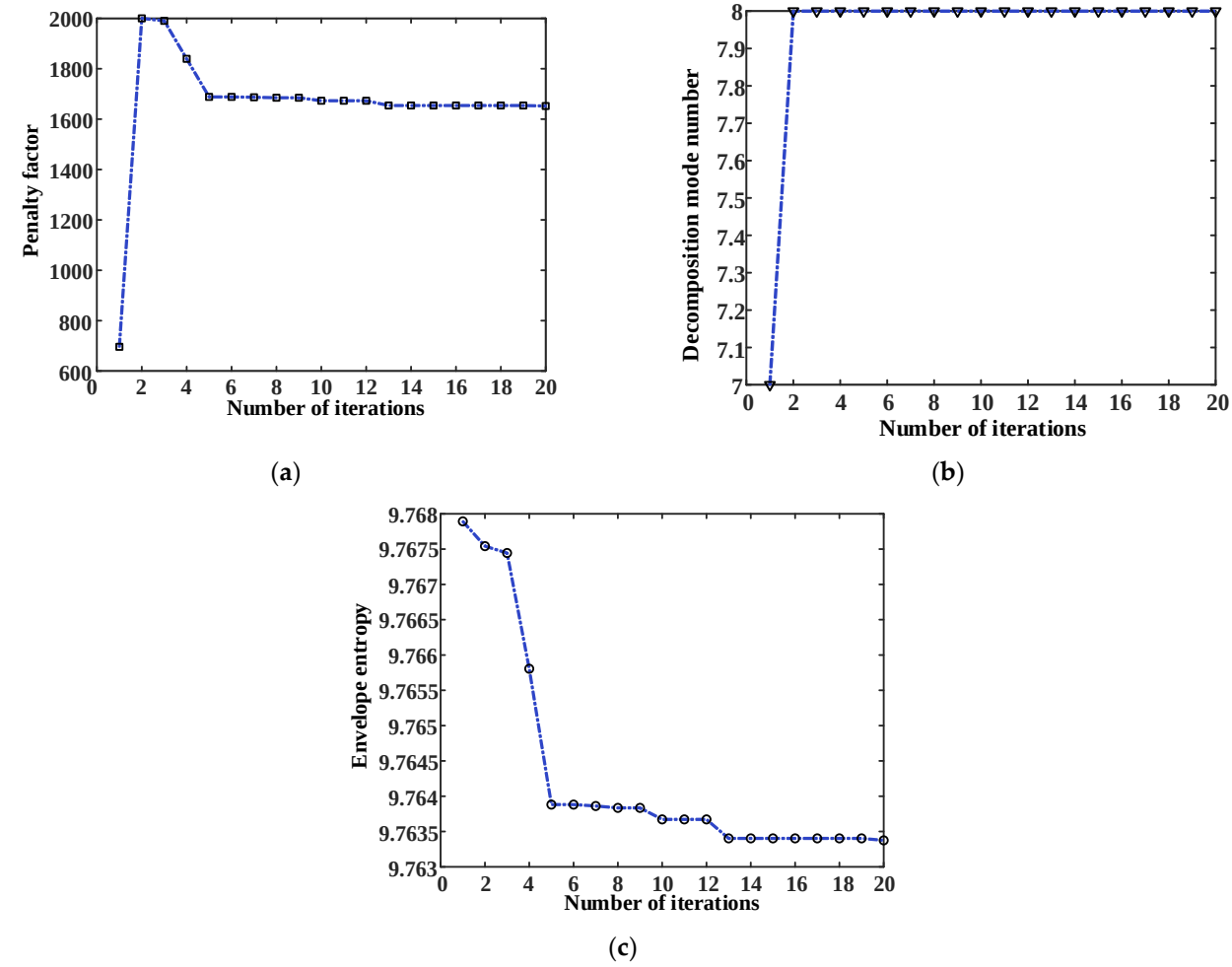


Figure 15. WOA-VMD iteration curve. (**a**) Optimization process curve of penalty factor; (**b**) Optimization process curve of decomposition mode number; (**c**) Envelope entropy iteration curve of WOA.

Figure 16 shows the time and frequency diagram of multiple mode components obtained from the decomposition of the optimal combination of parameters.

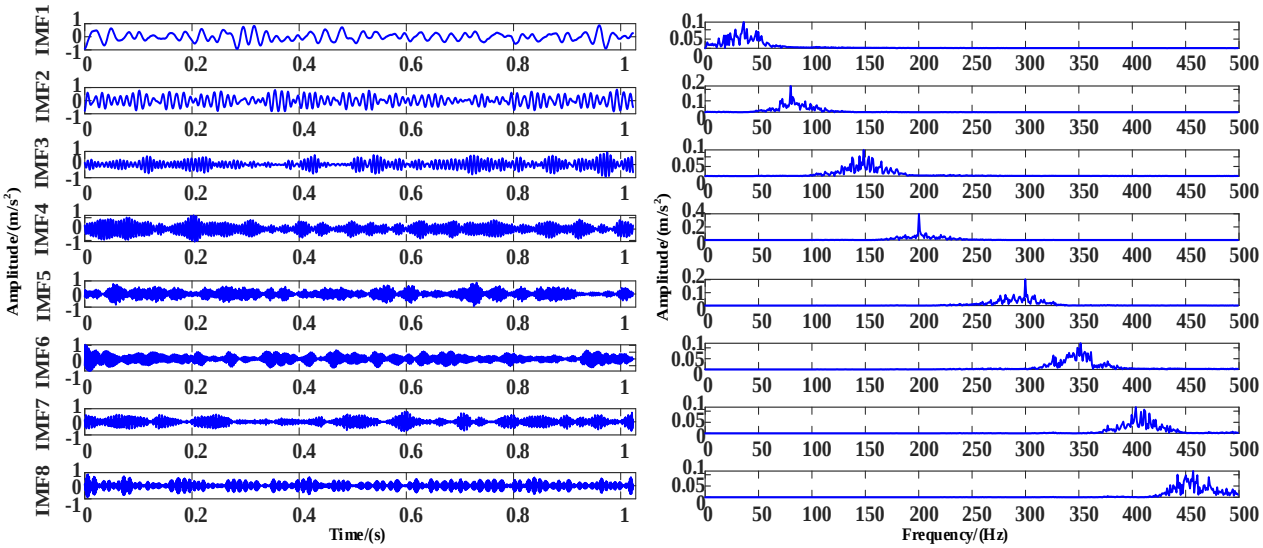


Figure 16. The time and frequency domain diagram of IMF after WOA-VMD decomposition.

Then, the mode components obtained from decomposition are selected using the Pearson correlation coefficient method, and those whose relationship value with the original signal is greater than or equal to 0.6 are reconstructed. The correlation values of the calculated mode components are shown in Table 3.

Table 3. Correlation coefficient value of each IMF.

Mode Component	Correlation Coefficient
IMF1	0.5167
IMF2	0.8005
IMF3	0.4942
IMF4	0.7966
IMF5	0.7854
IMF6	0.6258
IMF7	0.4135
IMF8	0.4638

The Pearson correlation coefficients of IMF2, 4, 5 and 6 exceed 0.6, discarding IMF1, 3, 7 and 8. IMF2, 4, 5 and 6 are reconstructed into signals after noise reduction. Then, the signals are compared after the noise reduction of EMD, EEMD, CEEMD and GA-VMD. The respective time and frequency domain diagrams are shown below.

As can be seen in Figures 17–21, the WOA-VMD provides a better noise reduction effect compared to the other four methods, and it has an obvious noise reduction effect on the whole frequency band.

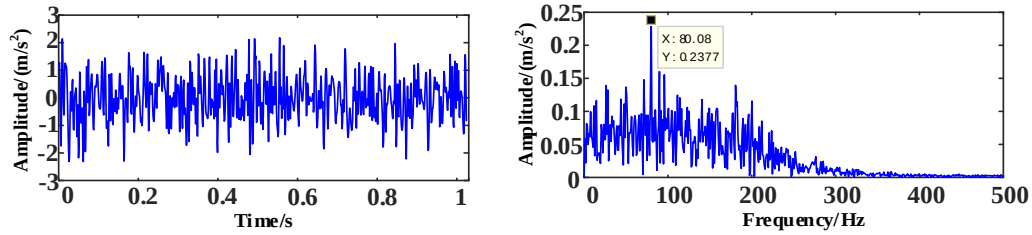


Figure 17. The time and frequency domain diagrams of EMD noise reduction.

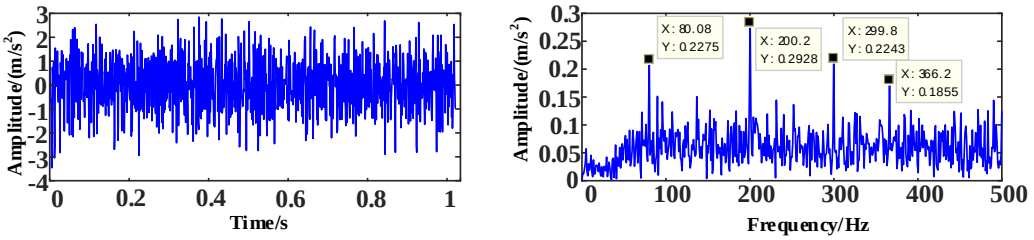


Figure 18. The time and frequency domain diagrams of EEMD noise reduction.

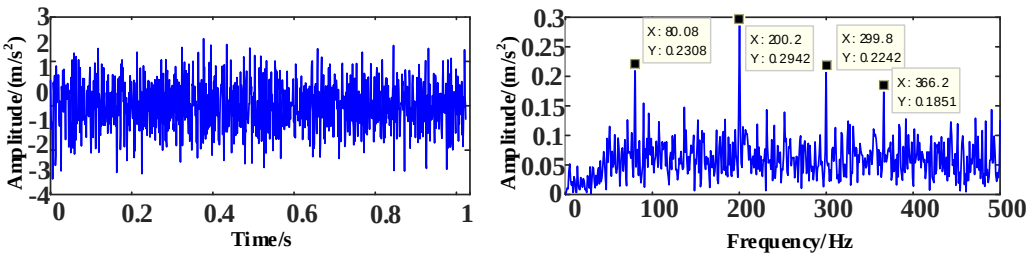


Figure 19. The time and frequency domain diagrams of CEEMD noise reduction.

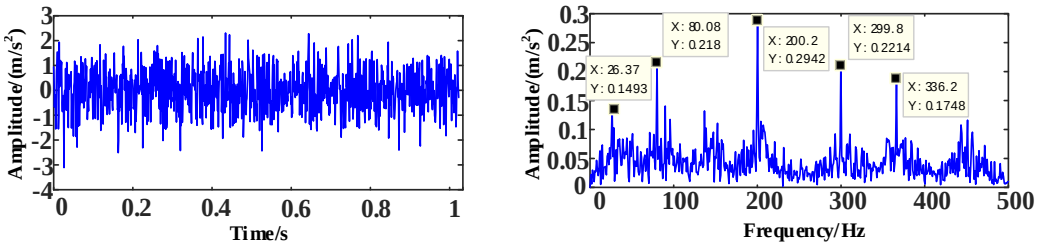


Figure 20. The time and frequency domain diagrams of GA-VMD noise reduction.

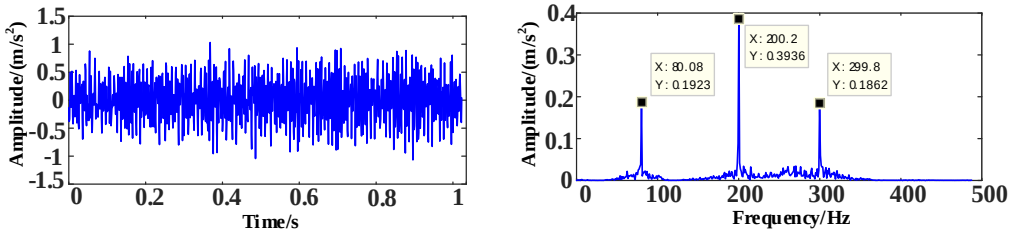


Figure 21. The time and frequency domain diagrams of WOA-VMD noise reduction.

As can be seen in Table 4, the root mean square error after the WOA-VMD signal decomposition and reconstruction is 0.213, the signal to noise ratio is 6.912 and the noise reduction effect is more obvious as compared with other algorithms.

Table 4. Comparison of different noise reduction methods under simulation data.

Noise Reduction Algorithms	Root Mean Square Error	Signal to Noise Ratio
EMD noise reduction	0.738	4.018
EEMD noise reduction	0.780	3.958
CEEMD noise reduction	0.785	3.804
GA-VMD noise reduction	0.702	4.112
WOA-VMD noise reduction	0.213	6.912

4.1.2. Validation of Laboratory Data

As can be seen in Figure 22, the bearing fault diagnosis test bench consists of a touch panel, motor speed controller, motor, radial loading hydraulic system, ADI150 uniaxial acceleration sensor, axial loading hydraulic system, main shaft, two support 6210 and 18,720 bearing, the ER-16K bearing to be measured and a force arm beam adjusting device. The bearing type is ER-16K, and detailed parameters are shown in Table 5. The acceleration sensor was used to obtain the vibration acceleration information of 13 bearing fault states, including 10 single point faults and 3 compound faults (CF). The experimental data were obtained at a sampling frequency of 25.6 kHz. A total of 10 groups were collected under each fault state, with each group comprising 32,768 sample points.

Table 5. Parameters of the test bench.

Bearing Parameters	Values	Bearing Parameters	Values
Outer ring diameter	51.99 mm	Inner ring diameter	25.40 mm
Weight	0.28 kg	Rolling diameter	7.92 mm
Number of rolling elements	9	Contact angle	0°
Maximum load (static)	7830 N	Maximum load (dynamic)	10,810 kN

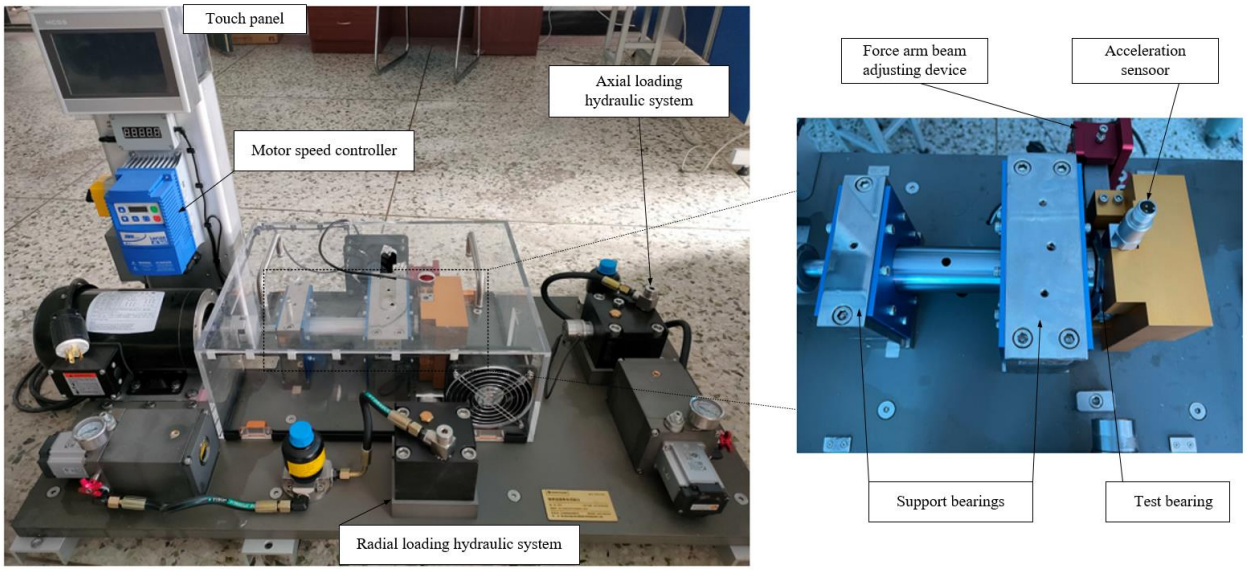


Figure 22. Test bench for rolling bearings.

The damage of the fault bearing is man-made, as shown in Figure 23. The inner and outer ring faults were caused by the use of a laser marking machine to create indentation in the groove of the outer ring ball. The red block represents the position of the laser groove. The rolling body fault represents the punching of holes in the rolling body.

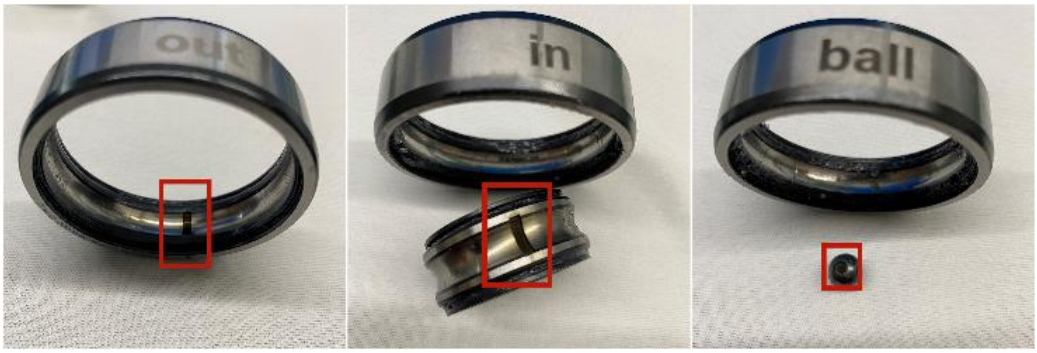


Figure 23. Damage diagrams of rolling bearings.

In the experiment, as shown in Table 6, three different loads were set, and different fault locations, damage degrees and experimental speeds were set under three different loads. In addition, there were three healthy sets of bearing data for each of the three loads.

The WOA-VMD method was used for the signal decomposition and reconstruction of the original data, and the feasibility of the proposed method was observed. The time and frequency domain diagrams of the vibration signals are shown in Figure 24.

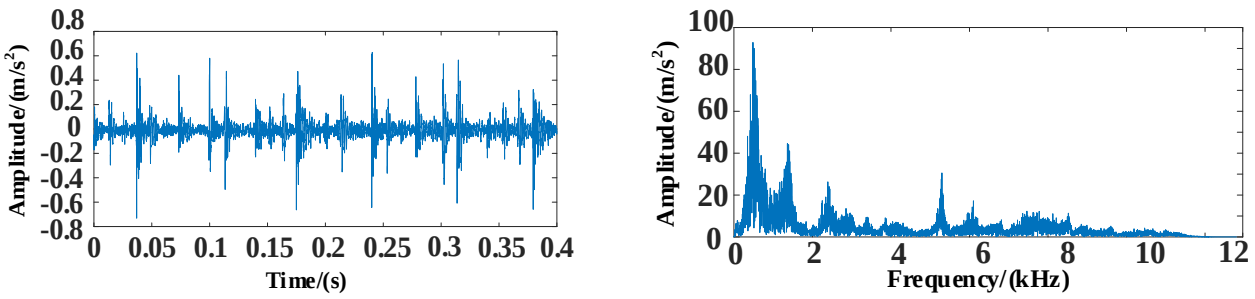


Figure 24. The time and frequency domain diagrams of vibration signals.

Table 6. A data set of bearing fault diagnosis experiment.

Radial Loading Force	Fault Location	Data Set	Degree of Damage
0 kg	Inner ring	I_L_0	mild
		I_M_0	moderate
		I_H_0	heavy
	Outer ring	O_L_0	mild
		O_M_0	moderate
		O_H_0	heavy
	Rolling ball	B_L_0	mild
		B_M_0	moderate
		B_H_0	heavy
100 kg	Inner ring	I_L_100	mild
		I_M_100	moderate
		I_H_100	heavy
	Outer ring	O_L_100	mild
		O_M_100	moderate
		O_H_100	heavy
	Rolling ball	B_L_100	mild
		B_M_100	moderate
		B_H_100	heavy
200 kg	Inner ring	I_L_200	mild
		I_M_200	moderate
		I_H_200	heavy
	Outer ring	O_L_200	mild
		O_M_200	moderate
		O_H_200	heavy
	Rolling ball	B_L_200	mild
		B_M_200	moderate
		B_H_200	heavy

In this experiment, the inner ring fault under a 100 N load is selected as experimental data, and the sampling frequency is 1024 Hz. It can be seen in Figure 24 that there is noise in the data. The WOA-VMD is applied for noise reduction processing. The decomposed IMF components and iterative curves are shown in Figures 25 and 26.

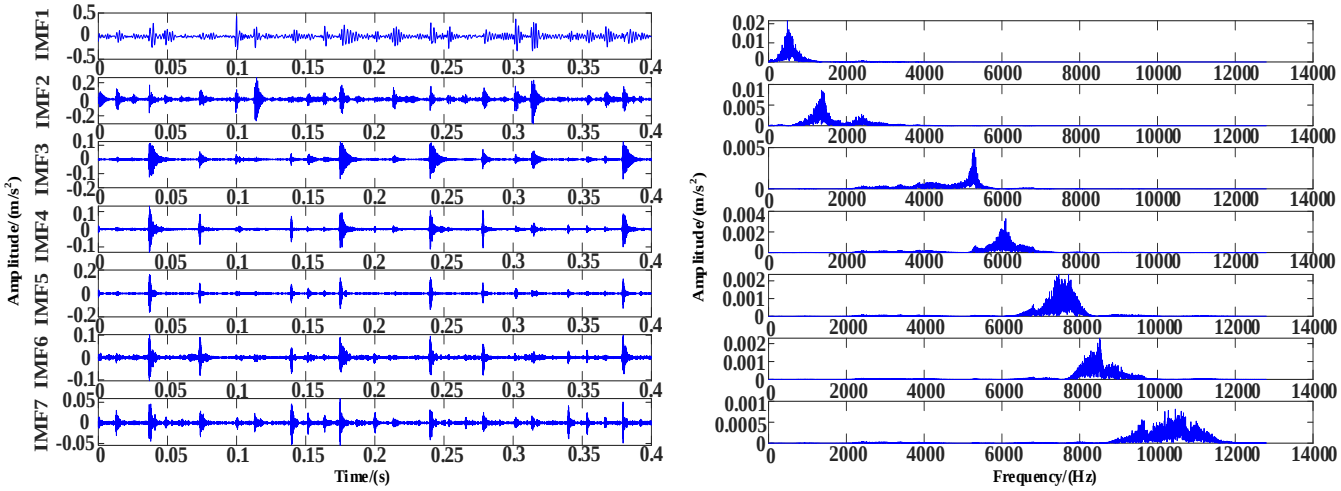


Figure 25. The time and frequency domain diagram of IMF after WOA-VMD noise decomposition.

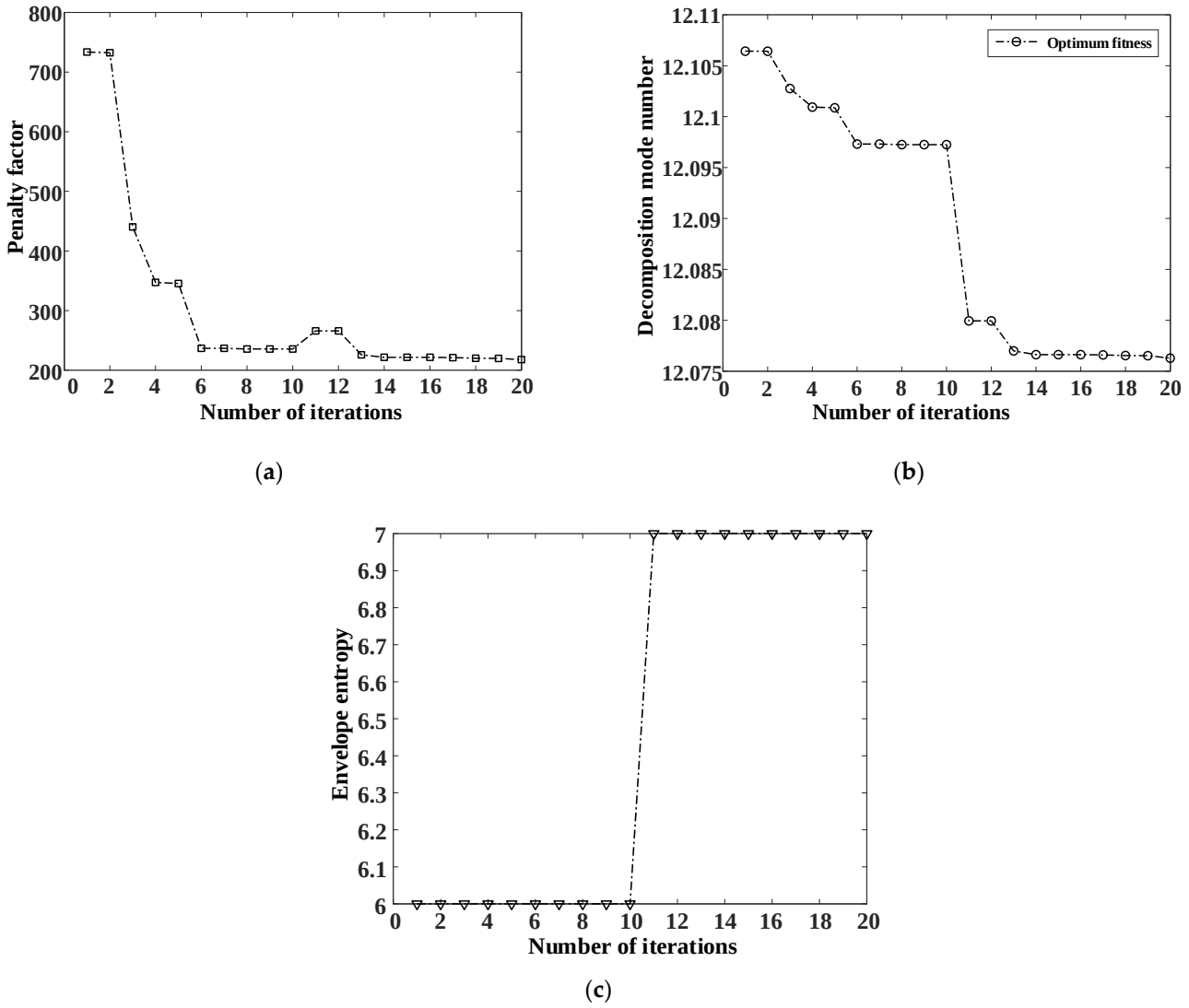


Figure 26. WOA-VMD iteration curve. (**a**) Optimization process curve of penalty factor; (**b**) Optimization process curve of decomposition mode number; (**c**) Envelope entropy iteration curve of WOA.

The correlation coefficient values between IMF components after the WOA-VMD decomposition and the original signal are calculated, as shown in Table 7.

Table 7. Correlation coefficient value of each IMF.

Mode Component	Correlation Coefficient
IMF1	0.8262
IMF2	0.6075
IMF3	0.2630
IMF4	0.2203
IMF5	0.2069
IMF6	0.1824
IMF7	0.1065

After the correlation analysis, the correlation coefficients of IMF1 and 2 were higher than 0.6, so IMF1 and 2 were selected for the reconstruction, and the noise reduction process

of the signal was completed. The time and frequency domain diagrams of the signal after the WOA-VMD noise reduction are shown in Figure 27.

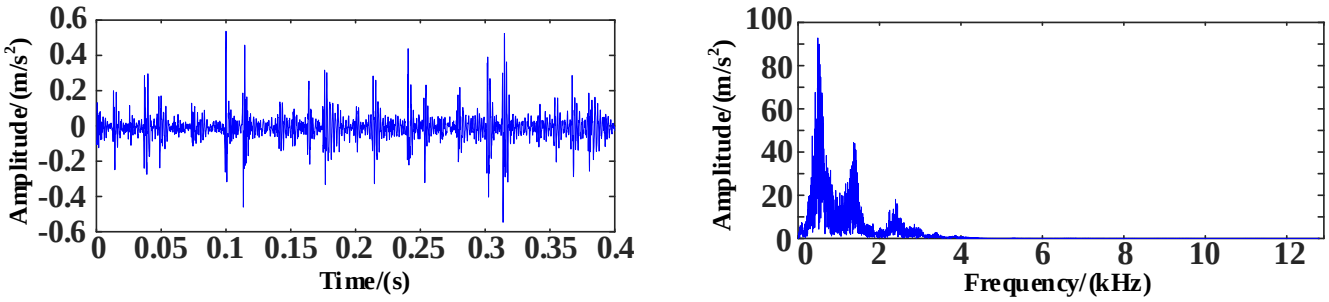


Figure 27. The time and frequency domain diagrams of the signal after WOA-VMD noise reduction.

As can be seen in Figure 27, the noise reduction effect in the high-frequency part of the signal is very obvious, with a large amount of noise having been removed. This allowed for effective information to be retained, the ideal effect to be achieved and the preparation for subsequent fault diagnosis.

4.2. GAT

In this section, two sets of data are chosen to verify the superiority and effectiveness of the GAT fault diagnosis algorithm. The data comprise the experimental bearing data from Case Western Reserve University [41] and the test bench data from this laboratory.

4.2.1. Case Western Reserve University Data Verification

The bearing test platform of Case Western Reserve University [42] is shown in Figure 28. The fault sizes were set to 0.007 inches, 0.014 inches and 0.021 inches, with 1 inch = 2.54 cm. In addition, the rotational speeds were set to 1797 r/min, 1772 r/min, 1750 r/min and 1730 r/min. The specific different fault states are shown in Table 8.

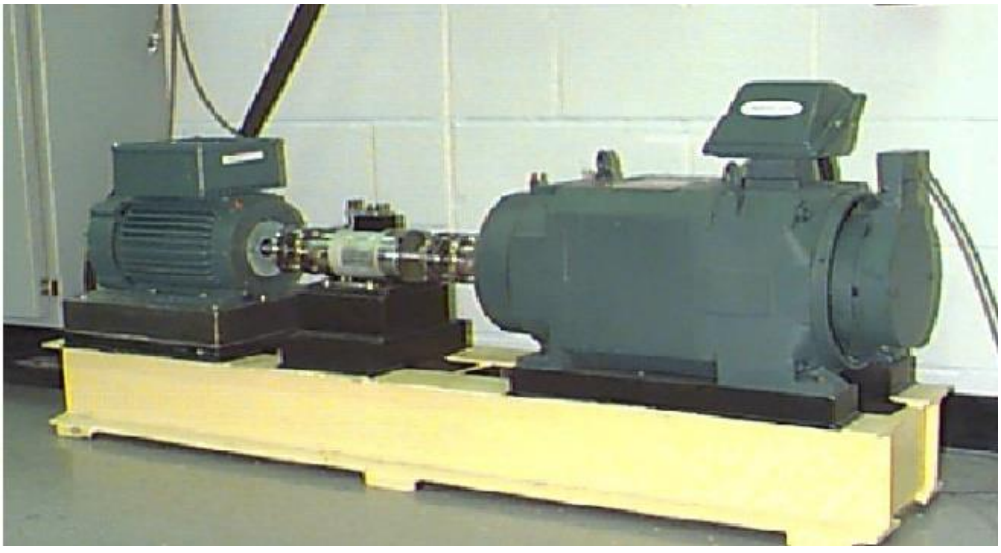


Figure 28. Case Western Reserve University bearing Experimental platform.

Table 8. Experimental parameter settings for different fault states.

Experiment Number	Fault Size	Fault Location	Location of Collection End
No. 1	0	Normal	
No. 2	0.007 inch	Inner ring fault	12 k Driving end
No. 3	0.007 inch	Outer ring fault	12 k Driving end
No. 4	0.007 inch	Rolling fault	12 k Driving end
No. 5	0.014 inch	Inner ring fault	12 k Fan end
No. 6	0.014 inch	Outer ring fault	12 k Fan end
No. 7	0.014 inch	Rolling fault	12 k Fan end
No. 8	0.021 inch	Inner ring fault	48 k Driving end
No. 9	0.021 inch	Outer ring fault	48 k Driving end
No. 10	0.021 inch	Rolling fault	48 k Driving end

In this section, the size of the convolutional kernel and the number of convolutional layers of the GAT will be discussed. A total of 60% of overall samples was randomly selected as the training set, 20% as the validation set and 20% as the test set. The specific GAT parameter settings are shown in Table 9.

Table 9. GAT model parameters.

Parameter Name	Value	Parameter Name	Value
Number of the training set sample groups	97	Node deactivation rate	0.2
Number of sample groups of verification set	25	Second fully connected layer	[1024,1024]
Number of test set sample groups	122	Batch normalization	1024
Convolutional kernel of the first layer	[2048,2048]	Loss function	Cross-entropy loss function
Convolutional kernel of the second layer	[2048,2048]	Optimizer	Stochastic gradient descent
Activation function of the first layer	Relu	Training times	100
Activation function of the second layer	Relu	Batch size	64
First fully connected layer	[2048,1024]	Learning rate	0.01

(1) Influence of the Convolutional Kernel Size of Different Graphs

The size of the convolutional kernel Θ, an important parameter of the network model, affects the accuracy of rolling bearing fault type identification. The impact on accuracy is discussed by comparing different sizes of convolutional kernels.

As shown in Table 10, when the convolution kernel size increases from $\Theta \in R^{1024\times1024}$ to $\Theta \in R^{2048\times2048}$, the accuracy becomes higher, and the calculation time increases accordingly. When the convolution kernel size increases from $\Theta \in R^{2048\times2048}$ to $\Theta \in R^{4096\times4096}$, the calculation time increases, but overfitting occurs. Therefore, the final convolutional kernel size selected for the graph in this paper is $\Theta \in R^{2048\times2048}$.

Table 10. Diagnostic accuracy of different graph convolution kernel size.

Size	$\boldsymbol{\Theta \in R^{1024\times1024}}$	$\boldsymbol{\Theta \in R^{2048\times2048}}$	$\boldsymbol{\Theta \in R^{4096\times4096}}$
Accuracy	90.40%	96.85%	92.81%
Time/s	42.25	61.58	120.54

(2) The Effect of Different Convolution Layers

In general, the more convolutional layers there are, the more filters are superimposed to solve the learning problem hierarchically. By deepening the layers, information can be transmitted at different levels. In this section, considering that other conditions are the same, the impact of two to six convolutional layers on the diagnostic accuracy is compared, as shown in Figure 29; the loss value is shown in Figure 30.

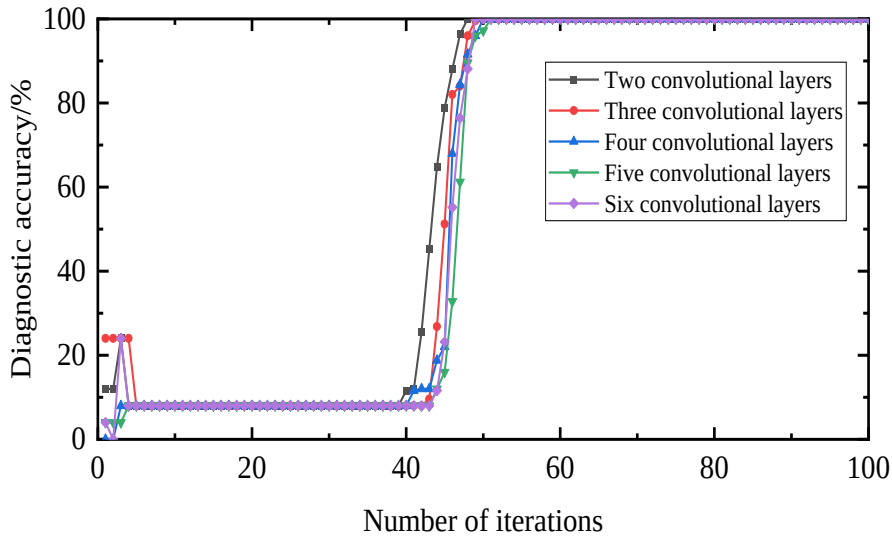


Figure 29. Diagnostic accuracy under different network layers.

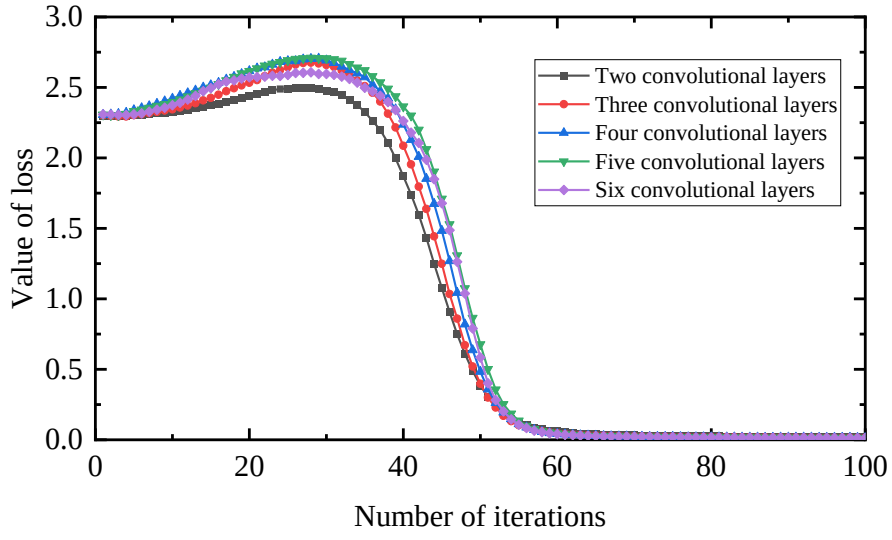


Figure 30. Loss value under different network layers.

The comparison shows that an increase in the number of layers has less impact on this experiment, but increases the time for iterations to converge. In contrast, for the loss value, the fewer the layers, the smaller the loss value, the faster the iteration speed and the more realistic the prediction. Therefore, the number of convolution kernel layers there are in this paper is two.

This time, the effectiveness of the proposed method is verified using TSNE dimensional reduction visualization, and the superiority of the GAT model is verified by comparing several fault diagnostic models.

Figure 31a shows the sample distribution of the initial data set. Before the model is trained, the label classification effect is not good. The labels of the same fault are not

aggregated and the labels of different categories are mixed. It can be seen in Figure 31b that the classification of all kinds of labels was completed. The aggregation of the same category is good, and there is no mixing of different category labels. This proves that the rolling bearing fault diagnosis method proposed in this paper is effective.

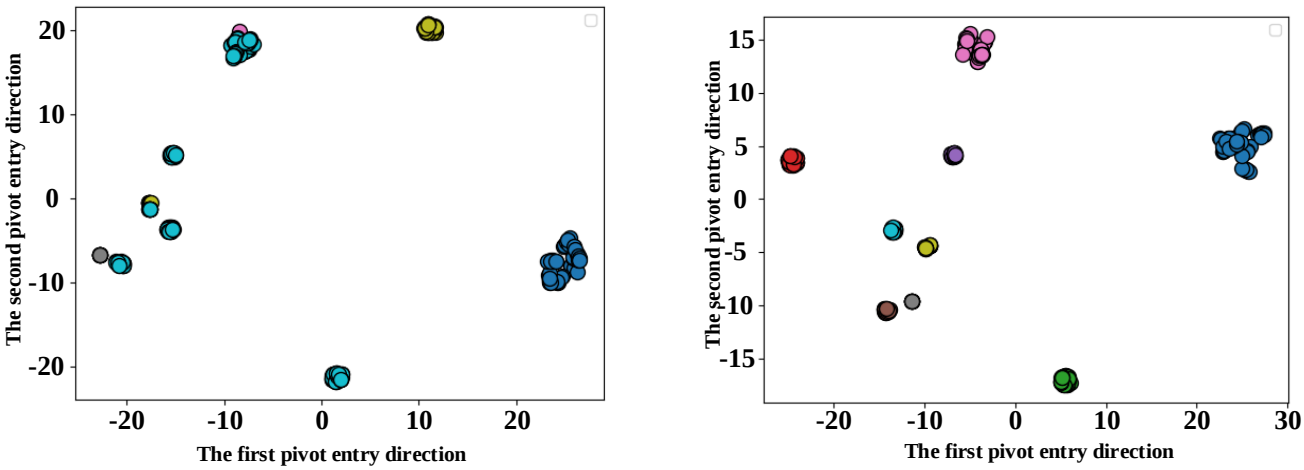


Figure 31. GAT rolling bearing fault diagnosis model visualization. (**a**) Initial data set visualization; (**b**) Visualization of GAT output results.

As can be seen in Figures 32 and 33, the accuracy of the MLP and Attention models reached about 80% after 100 training iterations, but they are not in a convergence state. Therefore, the graph neural network model converges faster and has higher diagnostic accuracy than the traditional neural network model. In the change curve of the loss value, it can be seen that the loss value of the graph neural network is lower than that of the traditional neural network.

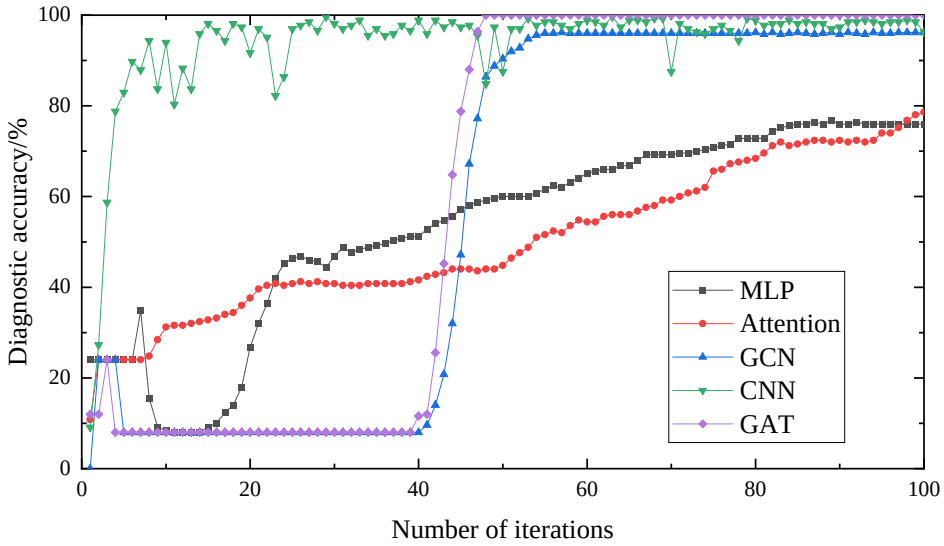


Figure 32. Diagnostic accuracy of different methods.

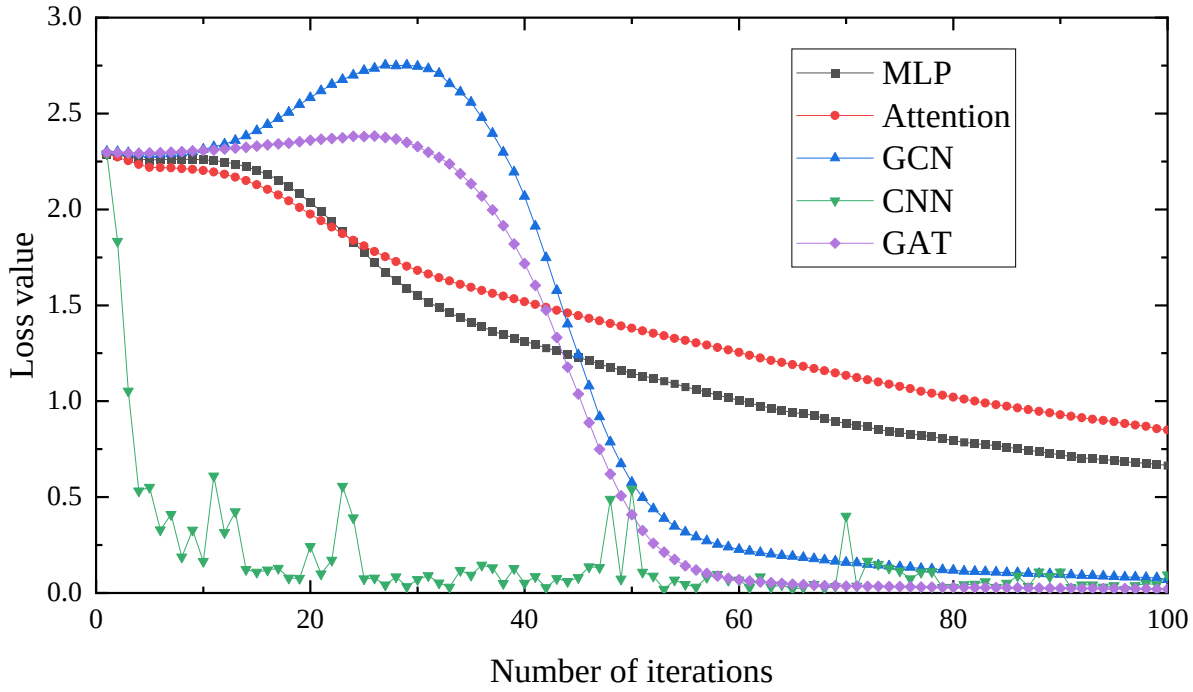


Figure 33. Loss value of different methods.

After 100 training iterations, the accuracy of the GCN, CNN and GAT models reached about 100%, while the iteration time and loss value of the GCN model are higher than those of the CNN and GAT, and the stability of the GCN and GAT is better than that of the CNN. This indicates that the method proposed in this paper is superior, with fast iteration speed, good model stability and strong generalization ability that could better solve the problem of rolling bearing fault diagnosis.

The comparison confusion matrix obtained by the classification of the test set V_{test} is shown in Figure 34. The accuracy of the experiment is calculated according to the comparison of the confusion matrix, as shown in Table 11.

Table 11. Diagnostic accuracy of test sets of different algorithms.

MLP	Attention	GCN	CNN	GAT
82.40%	70.8%	99.6%	98.32%	100%

The confusion matrix can analyze the classification results of samples in detail. In the figure above, the horizontal coordinates represent the predicted label of the samples, and the vertical coordinates represent the actual label of the samples. As can be seen in Figure 34, the Attention, MLP, GCN and CNN models all deviate in their diagnostic effects, while the GAT is accurate in diagnosing various faults, and its diagnostic effect is better than that of the other four methods.

It can be seen in Tables 11 and 12 that the test set accuracy of the GAT is 100%, higher than that of the MLP, Attention, GCN and CNN models, which proves that the method proposed in this paper can realize a more accurate diagnosis of fault data.

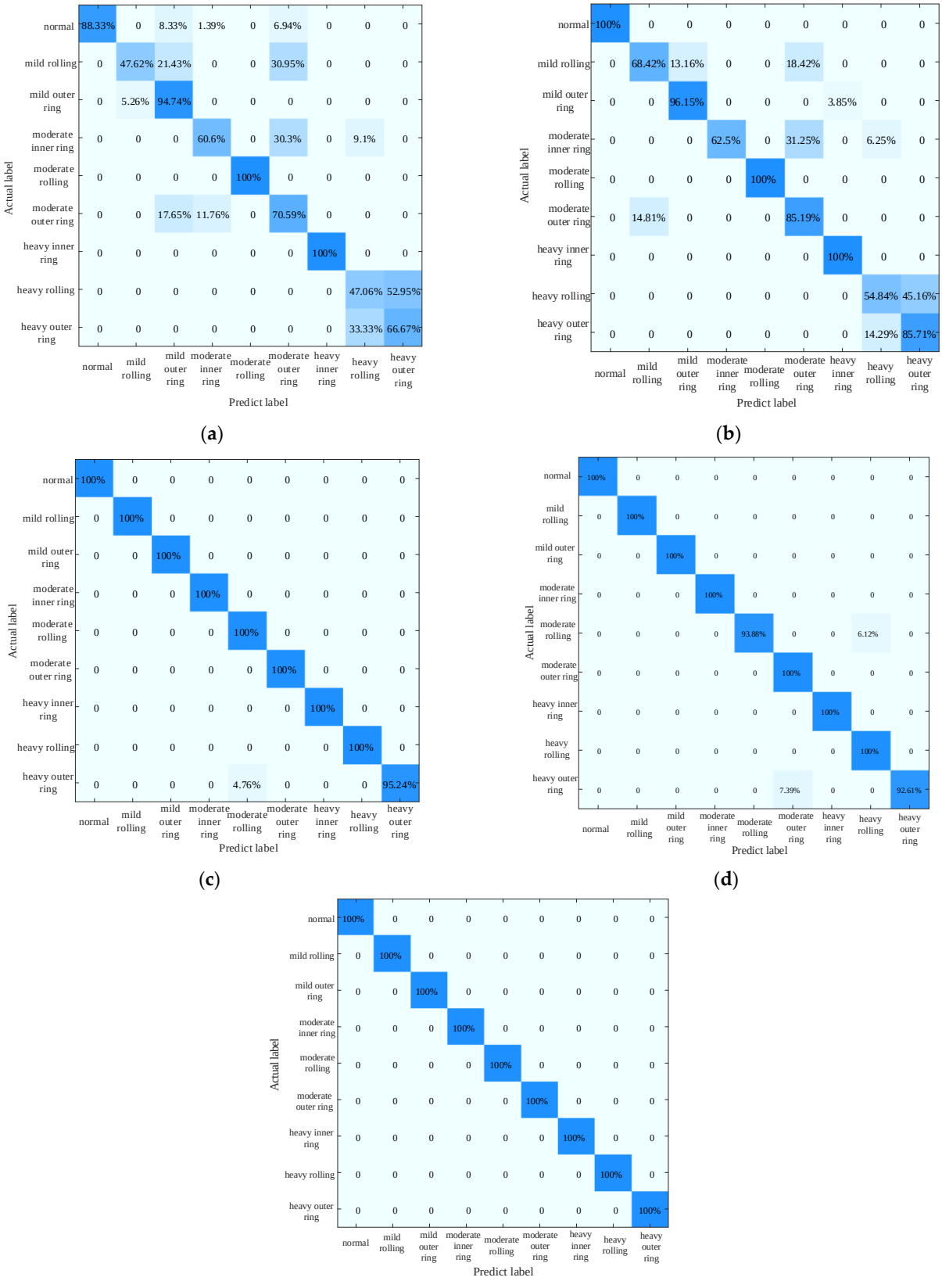


Figure 34. Confusion matrix of different methods. (**a**) Attention; (**b**) MLP; (**c**) GCN; (**d**) GCN; (**e**) GAT.

Table 12. The diagnostic precision of test sets of different algorithms.

	MLP	Attention	GCN	CNN	GAT
normal	100%	83.30%	100%	97.19%	100%
mild rolling	68.42%	47.62%	100%	96.16%	100%
mild outer ring	96.15%	94.74%	100%	100%	100%
moderate inner ring	62.50%	60.60%	100%	95.37%	100%
moderate rolling	100%	100%	100%	88.06%	100%
moderate outer ring	85.19%	70.59%	100%	96.42%	100%
heavy inner ring	100%	100%	100%	97.42%	100%
heavy rolling	54.84%	47.06%	100%	100%	100%
heavy outer ring	85.71%	66.67%	95.24%	98.25%	100%

4.2.2. Validation of Laboratory Data

The experimental data of the rolling bearings obtained in the experiment were decomposed and reconstructed using the WOA-VMD signal to eliminate noise components, and were subsequently used as input. The network model and structural parameters established above were used to divide the input data. The comparative experimental results are shown in Figures 35 and 36.

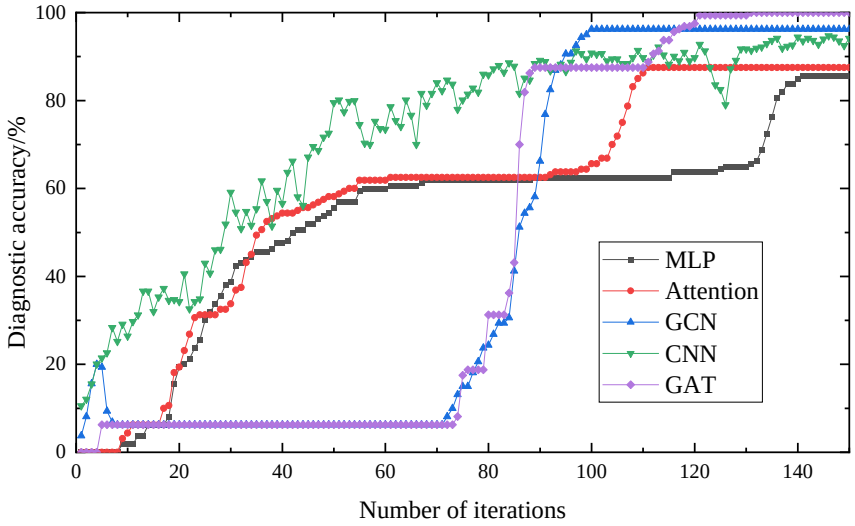


Figure 35. Diagnostic accuracy of different methods.

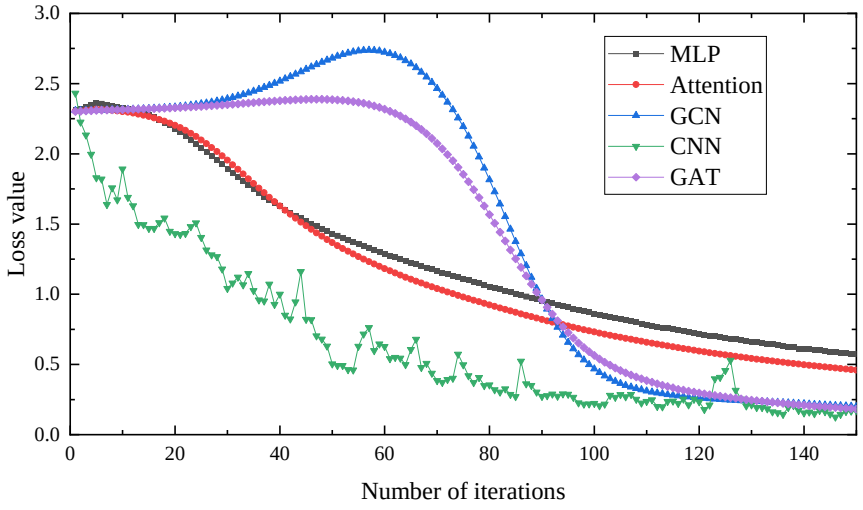


Figure 36. Loss value of different methods.

It can be seen in Figure 35 that the diagnostic accuracy of the MLP and Attention models are lower than that of the GCN, CNN and GAT models. It can be seen that the diagnostic accuracy of the graph structure used in this paper as input is higher, and the diagnostic accuracy of the GAT is higher than that of the GCN and CNN models. It can be seen in Figure 36 that the loss value of the GAT and CNN is lower than that of the GCN model, but the stability of the GAT is better. Therefore, the GAT has a certain stability and better accuracy.

According to Table 13, the test set accuracy of the GAT is 100%, which is higher than that of the MLP, Attention, GCN and CNN models.

Table 13. Diagnostic accuracy of test sets of different algorithms.

MLP	Attention	GCN	GCN	GAT
85.62%	87.25%	96.25%	94.12%	100%

As can be seen in Table 14, the diagnostic accuracy rate of the GAT for various fault signals can reach 100%, while the diagnosis results of the Attention, MLP, GCN and CNN models are all biased. It can be seen that the diagnostic effect of the graph neural network model is better than that of other models, and the GAT has better stability than the GCN. Therefore, the superiority of the GAT used in this paper is proved via the accuracy and precision rate indexes.

Table 14. Diagnostic accuracy of test sets of different algorithms.

	MLP	Attention	GCN	CNN	GAT
normal	100%	100%	100%	100%	100%
mild rolling	0%	0%	0%	100%	100%
mild outer ring	100%	100%	100%	80%	100%
moderate inner ring	100%	100%	100%	50%	100%
moderate rolling	100%	100%	100%	100%	100%
moderate outer ring	100%	100%	100%	90%	100%
heavy inner ring	100%	100%	100%	100%	100%
heavy rolling	100%	100%	100%	100%	100%
heavy outer ring	0%	0%	100%	100%	100%

5. Conclusions

In this paper, a fault diagnosis model based on the WOA-VMD and GAT was proposed for the identification of fault in rolling bearings with background noise.

1. The original signal was decomposed using the WOA-VMD, which effectively solved the phenomenon of mode mixing that occurs in traditional modal decomposition. After comparing the noise reduction effects of EMD, EEMD, CEEMD and GA-VMD, the experimental results showed that the root mean square error of the WOA-VMD is 0.213, and the signal-to-noise ratio is 6.912. Thus, the WOA-VMD has the best noise reduction effect.
2. The KNN method was used to construct the graph structure data, and a multi-headed attention mechanism was used to build the GAT rolling bearing fault diagnosis model, which assigned higher weights to the important neighborhoods and improved the sensitivity of the model to graph data containing faults. The diagnostic accuracy of the GAT method was 100%, which was 17.6%, 29.2%, 0.4% and 1.68% higher than that of the MLP, Attention, GCN and CNN models, respectively. This proves that the GAT can achieve more accurate diagnostic decisions for fault data sets.

Although the fault diagnosis algorithm proposed in this study has certain advantages in diagnosis, it is limited to rolling bearings with a constant speed. In the future, the application scope of the research will be extended to rolling bearings with variable speed.

In addition, measurements based on physical quantities (as stator current, stray fluxes, thermal images, etc.) will be considered.

Author Contributions: Conceptualization, data curation, writing—original draft preparation, Y.W. and S.Z.; validation, formal analysis, S.Z., R.C. and Y.F.; writing—review and editing, Y.W., S.Z. and R.C.; visualization, supervision, D.X. and Y.F.; funding acquisition, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52175502.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, N.; Lei, Y.; Lin, J.; Ding, S. An Improved Exponential Model for Predicting Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Ind. Electron.* **2015**, *62*, 7762–7773. [CrossRef]
2. Xia, T.; Zhuo, P.; Xiao, L.; Du, S.; Wang, D.; Xi, L. Multi-stage fault diagnosis framework for rolling bearing based on OHF Elman AdaBoost-Bagging algorithm. *Neurocomputing* **2021**, *433*, 237–251. [CrossRef]
3. Wang, Z.; Yao, L.; Chen, G.; Ding, J. Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals. *ISA Trans.* **2021**, *114*, 470–484. [CrossRef] [PubMed]
4. Kanemaru, M.; Tsukima, M.; Miyauchi, T.; Hayashi, K. Bearing Fault Detection in Induction Machine Based on Stator Current Spectrum Monitoring. *IEEJ J. Ind. Appl.* **2018**, *7*, 282–288. [CrossRef]
5. Harlişca, C.; Szabó, L.; Frosini, L.; Albini, A. Diagnosis of rolling bearings faults in electric machines through stray magnetic flux monitoring. In Proceedings of the 2013 8th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 23–25 May 2013.
6. Azeez, A.; Alkhedher, M.; Gadala, M. Thermal Imaging Fault Detection for Rolling Element Bearings. In Proceedings of the 2020 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 4 February–9 April 2020.
7. Chen, L.; Zhang, C. Fault Diagnosis of Rolling Bearing Based on EMD Envelop Spectrum Features and PCA-PNN. *Coal Mine Mach.* **2022**, *43*, 173–176.
8. Zhou, K.; Yan, Z.; Xin, Y.; Wu, X.; Zhang, C. Fault Diagnosis Method of Annealing Kiln Roller Bearings Based on Sensitive Feature Evaluation. *Noise Vib. Control* **2021**, *41*, 147–154, 160.
9. Jones, A. The mathematical theory of rolling-element bearings. In *Mechanical Design and Systems Handbook*; McGraw-Hill: New York, NY, USA, 1966.
10. Cao, H.; Niu, L.; He, Z. Method for Vibration Response Simulation and Sensor Placement Optimization of a Machine Tool Spindle System with a Bearing Defect. *Sensors* **2021**, *12*, 8732–8754. [CrossRef]
11. Jia, Y.; Li, G.; He, K.; Dong, X. Denoising method for vibration signal of hob based on grey criterion and EEMD. *Chin. J. Sci. Instrum.* **2019**, *40*, 187–194.
12. Ge, J.; Niu, T.; Xu, D.; Yin, G.; Wang, Y. A Rolling Bearing Fault Diagnosis Method Based on EEMD-WSST Signal Reconstruction and Multi-Scale Entropy. *Entropy* **2020**, *22*, 290. [CrossRef]
13. Donoho, D. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **1995**, *41*, 613–627. [CrossRef]
14. Liang, P.; Wang, W.; Yuan, X.; Liu, S.; Zhang, L.; Cheng, Y. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105269. [CrossRef]
15. Costa, C.; Kashiwagi, M.; Mathias, M. Rotor failure detection of induction motors by wavelet transform and Fourier transform in non-stationary condition. *Case Stud. Mech. Syst. Signal Process.* **2015**, *1*, 15–26. [CrossRef]
16. Jumah, A. Denoising of an Image Using Discrete Stationary Wavelet Transform and Various Thresholding Technique. *Signal Inf. Process.* **2013**, *4*, 33–41. [CrossRef]
17. Liu, C.; Wu, Y.; Zhen, C. Rolling Bearing Fault Diagnosis Based on Variational Mode Decomposition and Fuzzy C Means Clustering. *Proc. Chin. Soc. Electr. Eng.* **2015**, *35*, 3358–3365.
18. Duo, M.; Ji, G.; Zhu, H.; Yang, X. Bearing Fault Diagnosis Based on VMD Noise Reduction and CNN. *Noise Vib. Control* **2021**, *41*, 155–160.
19. Wu, L.; Hao, R.; Lu, Y. Application of VMD-FastICA in Fault Diagnosis of Gearbox. *J. Shijiazhuang Railw. Univ.* **2020**, *33*, 14–20.
20. Gustafsson, O.; Tallian, T. Detection of damage in assembled rolling element bearings. *Trans. ASAE* **1962**, *5*, 197–205. [CrossRef]
21. Hou, X. Feature Enhancement and Intelligent Recognition for Composite Faults of Rolling Bearings. Master's Thesis, Beijing University of Chemical Technology, Beijing, China, 2019.

22. Li, H.; Zhang, Q.; Qin, X.; Sun, Y. Fault diagnosis method for rolling bearings based on short-time Fourier transform and convolution neural network. *Vib. Shock* **2018**, *37*, 124–131.
23. Gao, S.; Pei, Z.; Zhang, Y.; Li, T. Bearing Fault Diagnosis Based on Adaptive Convolutional Neural Network with Nesterov Momentum. *IEEE Sens. J.* **2021**, *21*, 9268–9276. [CrossRef]
24. Ma, D.; He, Y.; Li, M.; Tang, Q.; Hu, M. Rolling bearing fault diagnosis method based on CEEMD-PCA-XGBoost. *Mech. Electr. Eng.* **2022**, *40*, 186–194.
25. Chen, K.; Hu, J.; Zhang, Y.; Yu, Z.; He, J. Fault Location in Power Distribution Systems via Deep Graph Convolutional Networks. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 119–131. [CrossRef]
26. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Multireceptive Field Graph Convolutional Networks for Machine Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2021**, *68*, 12739–12749. [CrossRef]
27. Zhao, X.; Jia, M.; Liu, Z. Semisupervised Graph Convolution Deep Belief Network for Fault Diagnosis of Electormechanical System with Limited Labeled Data. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5450–5460. [CrossRef]
28. Li, T.; Zhao, Z.; Sun, C.; Yan, R.; Chen, X. Domain Adversarial Graph Convolutional Network for Fault Diagnosis Under Variable Working Conditions. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]
29. Li, T.; Zhou, Z.; Li, S.; Sun, C.; Yan, R.; Chen, X. The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study. *Mech. Syst. Signal Process.* **2021**, *168*, 108653. [CrossRef]
30. Zhang, Z.; Cui, P.; Zhu, W. Deep Learning on Graphs: A Survey. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 249–270. [CrossRef]
31. Gao, Y.; Chen, M.; Yu, D. Semi-supervised graph convolutional network and its application in intelligent fault diagnosis of rotating machinery. *Measurement* **2021**, *186*, 110084. [CrossRef]
32. Li, C.; Mo, L.; Yan, R. Fault Diagnosis of Rolling Bearing Based on WHVG and GCN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–11. [CrossRef]
33. Zhang, D.; Stewart, E.; Entezami, M.; Roberts, C.; Yu, D. Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network. *Measurement* **2020**, *156*, 107585. [CrossRef]
34. Yu, X.; Tang, B.; Zhang, K. Fault Diagnosis of Wind Turbine Gearbox Using a Novel Method of Fast Deep Graph Convolutional Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–14. [CrossRef]
35. Li, C.; Mo, L.; Yan, R. Rolling Bearing Fault Diagnosis Based on Horizontal Visibility Graph and Graph Neural Networks. In Proceedings of the 2020 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), Xi'an, China, 10–15 October 2020.
36. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]
37. Yin, G. Research on Useful Life Prediction of Rolling Bearing Based on Pearson-KPCA Multi-feature Fusion. Master's Thesis, Harbin University of Science and Technology, Harbin, China, 2021.
38. Fey, M.; Lenssen, J. Fast graph representation learning with PyTorch Geometric. *arXiv* **2019**, arXiv:1903.02428.
39. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
40. Cui, H.; Guan, Y.; Chen, H. Rolling Element Fault Diagnosis Based on VMD and Sensitivity MCKD. *IEEE Access* **2021**, *9*, 120297–120308. [CrossRef]
41. Case Western Reserve University Bearing Data Center Website. Available online: https://engineering.case.edu/bearingdatacenter/download-data-file (accessed on 22 April 2023).
42. Smith, W.; Randall, R. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64–65*, 100–113. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.