

Article

Rolling Bearing Fault Diagnosis based on Multiscale Permutation Entropy and SOA-SVM

Xi Zhang¹, Hongju Wang^{1*}, Mingming Ren¹, Mengyun He¹ and Lei Jin¹

¹ School of Mechanical Electronic & Information Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China

* Correspondence: qishangyinyin@163.com

Abstract: The service conditions of underground coal mine equipment are poor, and it is difficult to accurately extract the fault characteristics of rolling bearings. In order to better improve the accuracy of fault identification of rolling bearings, a fault detection method based on multiscale permutation entropy and SOA-SVM is proposed. First, the whale optimization algorithm is used to select the modal analysis number K and the penalty factor α of the variational mode decomposition algorithm. Then, the vibration signal of rolling bearings is dissolved according to the optimized variational mode decomposition algorithm, and the multi-scale permutation entropy of the main intrinsic mode function is calculated. Finally, the feature values of the matrix are entered into the SVM algorithm optimized by the seagull optimization algorithm to obtain the classification result. The experimental results based on the published rolling bearing datasets of Western Reserve University show that the identification success rate of the proposed method can reach 98.75%. The fault detection of the rolling bearings can be completed accurately and efficiently.

Keywords: whale optimization algorithm; variational mode decomposition; seagull optimization algorithm; support vector machine; multi-scale permutation entropy; fault diagnosis

1. Introduction

As a key component of rotating machinery and equipment, the operating conditions of rolling bearings immediately impact the working characteristics of mining fans. When there is a problem with a rolling bearing, the damage point is constantly colliding with other parts that it touches, resulting in shock oscillation and unstable, nonlinear, multi-frequency data signals [1]. Sudden faults such as loose or damaged rolling bearings will cause uneven bearing capacity, expansion of frictional resistance or shutdown, leading to faults such as displacement, unbalance, and surge of the mining fan. The problems caused by rolling bearings account for about 50% of the common failures of mining fans, and the shutdown time caused by rolling bearings also accounts for about 45%. Therefore, the accurate identification of faults of rolling bearings is of key practical significance to the safety and stability of mining fans.

Vibration analysis method is widely used in rolling bearing fault diagnosis because it reveals the inherent characteristics of bearing fault [2]. Generally speaking, the reliability analysis method mainly includes three levels: data preprocessing, fault feature extraction and failure mode classification [3]. Because the evaluation of vibration signal usually shows characteristics of optimal control and instability, the research in recent years is mainly concentrated on time-frequency analysis technology [4]. At present, there are two types of time-frequency analysis technology. The first methods

do not need to establish the primary parameters before examining the vibration signals. A very typical example is empirical mode decomposition (EMD) [5]. EMD is a responsive reliability analysis technology, which can dissolve all complicated data signals into several characteristic modal analysis function formulas according to the original vibration. Although several applications have proved the efficiency of EMD in detecting rolling bearing faults [6], it still has issues with the terminal effect and modal aliasing. The second methods need to set some main parameters before they are used to analyze vibration signals, such as wavelet transform (WT). However, this method must define the wavelet basis function and threshold in advance [7], and the choice of wavelet basis function has a considerable influence on the final output. Therefore, wavelet transform does not have adaptive characteristics.

Dragomiretschi [8] introduced variational mode decomposition as a method for determining the frequency center and bandwidth of a variational model. Compared with empirical mode decomposition and wavelet transform, variational mode decomposition has a rigorous mathematical theoretical foundation, and can separate vibration signals efficiently and accurately. Although the frequencies of the vibration signals can be adaptively divided by the VMD method, the attenuation results are still limited by the choice of the modal number K and the penalty parameter α . Z. Zhang [9] determined the selection of K value by observing the center frequency of intrinsic mode function (IMF). Z. Guo [10] selected the appropriate number of decomposition layers by setting the threshold of multi-scale permutation entropy. With the increasing applications of intelligent algorithms, researchers tend to combine intelligent algorithms with parameter optimization of VMD. G. A. Ran [11] introduced grey wolf algorithm to optimize K . J. Li [12] introduced genetic algorithm to optimize K and α at the same time. Although it takes a very long time to optimize the parameters of variational mode decomposition with intelligent algorithm, it has become a research hotspot because it considers the coupling impact of the two factors on the decomposition effect.

Following the dissolution of the vibration data signal into a sequence of IMFs via VMD, the next task is how to obtain the fault information from the obtained IMF weights. Richman explicitly proposed sample entropy [13]. Because sample entropy is less sensitive to data length and noise, it is of general concern. Permutation entropy (PE) was suggested by Bandt [14] to analyze the plurality of mechanical systems and assess their conditions. Since PE considers complexity in terms of relatively close proximity, it is simple and not compromised by noise. However, sample entropy and permutation entropy estimate complexity only on a single scale, which will produce adverse results when applied to the analysis of data on multiple time scales. In view of this shortcoming, Costa [15] developed a method for assessing the complexity of unprocessed time series at different scales using a multi-scale sample entropy approach. However, the complexity estimation of the actually measured bearing fault vibration signal by multi-scale sample entropy is poor, and the processing of long-time series is particularly time-consuming. To assess the complexity of time-series data, Aziz and ARIF [16] introduced the multiscale permutation entropy (MPE). In addition, the stability and robustness of MPE were verified. J. Zheng [17] employed MPE and SVM to identify rolling bearing defects, proving the superiority of MPE in feature extraction of rolling bearing faults. Therefore, MPE is selected as a special tool for SVM algorithm in this paper.

At this stage, the specific methods used for rolling bearing fault classification include SVM [18], extreme learning machine [19], BP neural network [20], etc. In small samples, SVM has strong generalization ability and relatively simple structure. The SVM solid model has two key main parameters C and g , where C is the penalty index, the tolerance for deviation. If the C value is too large, it is easy to multicollinearity; and if the C value is too small, it is easy to underfit. If C is too large or too small, it will lead to poor generalization ability of SVM [21]. g is the main parameter after the RBF function

formula is evaluated as a kernel function. It categorizes the data after projecting it explicitly to the interior space with new features. The larger the g value is, the less applicable space vectors are, and the smaller the g value is, the more applicable space vectors are. The number of applicable space vectors can compromise the rate of training. Intelligent optimization algorithm is often used to select C and g of support vector machine. J. Zheng [22] optimized SVM for rolling bearing defect type detection using the cuckoo search method, and its overall recognition rate reached 98.03%.

Inspired by previous scientific research, a combined model based on multi-scale permutation entropy and SOA-SVM is pointed out in this paper. First, the envelope entropy is adopted as the fitness function of the whale optimization algorithm to obtain the preset parameter pair of the variational mode decomposition algorithm $[K, \alpha]$. Then, the bearing vibration signal is decomposed by using the variational mode decomposition algorithm optimized the parameters to generate a set of intrinsic mode functions. The multi-scale permutation entropy of the main intrinsic mode functions is calculated on the basis of the kurtosis and correlation coefficient to form the feature vector. Finally, the SOA-SVM method is employed to identify four statuses of rolling bearing (normal, inner ring fault, outer ring fault and rolling element fault).

2. The Proposed Method

2.1. WOA-VMD

VMD has a high signal attenuation efficiency as a prominent time-frequency analysis approach.[23]. For the original signal $x(t)$, it can be decomposed into a series of intrinsic mode functions IMFs u_k in automation mode.

$$x(t) = \sum_k u_k \quad (1)$$

To guarantee the minimum sum of the bandwidth of each center frequency, the process can be expressed as:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (2)$$

where $\{u_k\} = \{u_1, \dots, u_k\}$ is a series of decomposed intrinsic mode functions, $\{\omega_k\} = \{\omega_1, \dots, \omega_k\}$ is the center frequency corresponding to each intrinsic mode function. In order to arrive at the best solution in equation (2), Lagrange penalty factor L and secondary penalty factor α are introduced.

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 \\ + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (3)$$

The combination of decomposition levels K and penalty factor α has a significant impact on the decomposition result of the variational mode decomposition method [24]. Over decomposition and erroneous components will result if the K value is too high; under decomposition will result if the K value is too low. The bandwidth surrounding each center frequency will be too narrow if the value is too high. The bandwidth surrounding each center frequency will be too wide if the value is too low. Improper parameter selection will lead to the difficulty of subsequent feature extraction, which will affect the final accuracy of fault recognition. Therefore, reasonable parameter setting is very important to get satisfactory decomposition results.

WOA was explicitly proposed by Mirjalili [25] based on scientific research on the hunting behavior of whales. WOA is selected because this method has the advantages of fast improvement speed, strong global convergence and a few parameters. The specific steps of the WOA optimization are as follows:

(1) Initialization of parameters such as whale individual population, location and iteration times. The i -th individual location is :

$$X_i = r \cdot (ub - lb) + lb \quad (4)$$

where r is a random number within the range of $[0,1]$. X_i is in the range of $[lb, ub]$. lb is the minimum value of the parameter boundary and ub is the maximum value of the parameter boundary.

(2) When $p < 0.5$ and $|A| < 1$, shrink and surround according to the best search agent, as shown in equation (5):

$$\begin{cases} \vec{X}(i+1) = \vec{X}^*(i) - \vec{A} \cdot \vec{D} \\ \vec{D} = |\vec{C} \cdot \vec{X}^*(i) - \vec{X}(i)| \\ \vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} = 2 \cdot \vec{r}_2 \\ \vec{a} = 2 - 2(i/i_{\max}) \end{cases} \quad (5)$$

where \vec{r}_1 , \vec{r}_2 and p are random numbers, and the value range is $[0,1]$. i is the current number of iterations; i_{\max} is the maximum number of iterations.

When $p < 0.5$ and $|A| \geq 1$, a random search agent is selected to iterate and update the expression (6). \vec{X}_{rand} is the whale position vector selected randomly.

$$\begin{cases} \vec{X}(i+1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \vec{D} \\ \vec{D} = |\vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X}(i)| \\ \vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \\ \vec{C} = 2 \cdot \vec{r}_2 \\ \vec{a} = 2 - 2(i/i_{\max}) \end{cases} \quad (6)$$

When $p \geq 0.5$, the spiral contraction method is adopted for iteration, as shown in formula (7):

$$\begin{cases} \vec{X}(i+1) = \vec{D}' \cdot e^{bm} \cdot \cos(2\pi m) + \vec{X}^*(i) \\ |\vec{D}'| = |\vec{X}^*(i) - \vec{X}(i)| \end{cases} \quad (7)$$

where \vec{D}' is the distance between simulated whales and prey; b is the defined helix constant; m is a random number between $(-1,1)$.

(3) Check if the termination requirements have been satisfied or if the maximum number of repetitions has been reached. If not, return to step (2). If yes, output the best search agent.

Using whale optimization algorithm, it is also necessary to select the appropriate fitness function[26]. In this paper, the envelope entropy proposed by Tang Guiji [27] is used as the fitness function, and its expression is as follows:

$$\begin{cases} Q_j = a(j) / \sum_{j=1}^N a(j) \\ E_e = - \sum_{j=1}^N Q_j \lg Q_j \end{cases} \quad (8)$$

where Q_j is a sequence of probability distribution processed by the envelope signal; $a(j)$ is the envelope signal got by Hilbert Demodulation of the original signal[28]; E_e is envelope entropy which can quantitatively measure the sparsity of vibration signals [29].

When the signal contains a large number of interference components, the fault impact and modulation phenomenon caused by the fault will be hidden in the signal, resulting in the weakening of the sparsity of the signal, and the envelope entropy value is large at this time. The sparsity of the signal is high and the envelope entropy value is low when it contains clear fault impact and modulation events. The envelope entropy is used as the fitness function for the parameter optimization of VMD, and its minimum is taken as the search goal of the algorithm to complete the optimization of relevant parameters.

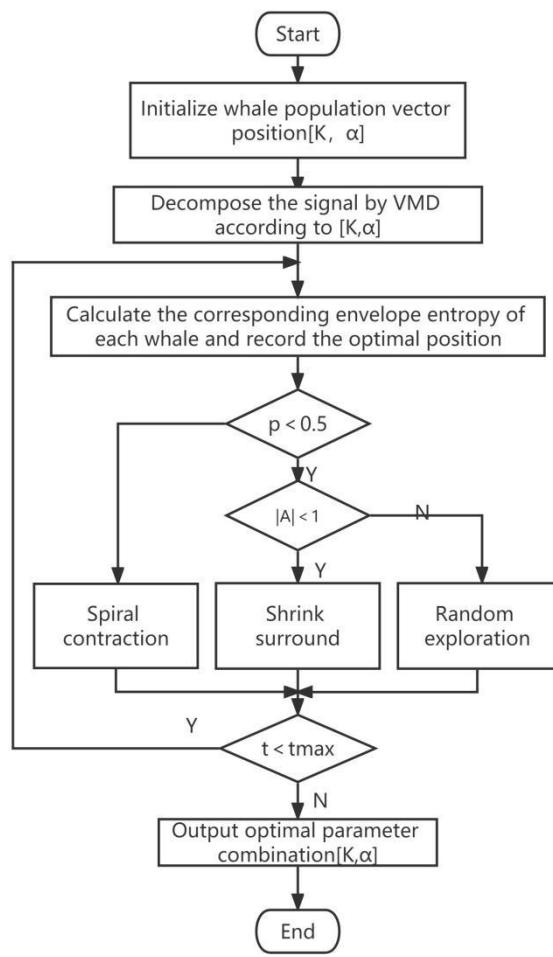


Figure 1. Flowchart of IWOA optimizing VMD parameters.

The process of optimizing VMD parameters with WOA is shown in Figure 1. First, initialize the whale group position vector $[K, \alpha]$. The fitness function is the envelope entropy, and each whale's fitness is evaluated. Then, by judging the size of the convergence factor, the iterative formula is selected for iterative update until the termination conditions are met, and the optimal VMD parameters are output. The upper boundary of the whale group position is set to $[10, 3000]$; The lower boundary is set to $[3, 300]$. The convergence criterion is 10 and the population number is set at 20.

2.2. Multiscale Permutation Entropy and Its Parameter Setting

The related concepts of multi-scale permutation entropy are shown in reference [30], and its theory is described as follows:

The original time series with length n is coarsened to obtain a new time series:

$$y_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} x_i, j = 1, 2, \dots, [N/s] \quad (9)$$

where s is the scale factor, $s = 1, 2, \dots, [N/s]$ indicates rounding. Each scale sequence's time reconstruction of each scale sequence $y_j^{(s)}$ are as follows:

$$Y_l^{(s)} = \{y_l^{(s)}, y_{l+\tau}^{(s)}, \dots, y_{l+(m-1)\tau}^{(s)}\} \quad (10)$$

The reconstructed sequences are placed in order if their values are same. A collection of symbol sequences can be produced for every scale sequence, where $r = 1, 2, 3, \dots, R$ and

$R \leq m!$. One of the permutations is the symbol sequence $S(r)$, and the chance of each symbol sequence occurring is determined ($r = 1, 2, \dots, R$). Information entropy is used to define the permutation entropy of various symbol sequences.

$$H_p(m) = - \sum_{r=1}^R P_r \ln P_r \quad (11)$$

When $P_r = 1/m!$, the maximum value is reached. For convenience, normalization is usually performed.

$$H_p = H_p(m) / \ln(m!) \quad (12)$$

Four parameters must be established before MPE can be used: time series length N , encapsulation dimension m , scale factor s , and time delay τ . Because m is the number that specifies the maximum number of permutations $m!$, Permutation entropy depends largely on the choice of encapsulation dimension m . In addition, the length of time series N should be more than $5 * m$ factorial [31] to obtain reliable statistics. Bandt [14] pointed out that this method is suitable for the case where the encapsulation dimension is $3 \leq m \leq 7$. The approach will not function if the encapsulation dimension m is too small since there are too few different states. When the encapsulation dimension m is too large, on the other hand, it will be too time-consuming. Typically, the encapsulation dimension m is chosen based on balancing information content impairment and measurement complexity. m is set to 5 in this article. We put $\tau = 1$ here since the time delay τ has no significant impact on the outcome. Calculation efficiency is jeopardized when N is too high. The criterion of $N \geq 5m!$ cannot be met if N is too small. Taking this control into account, the data length of 2048 points is sufficient to get a stable permutation entropy. Therefore, N is set to 2048. The scale factor s is set to 15 to obtain permutation entropy of each scales. Finally, we put $\tau = 1$ here since the time delay τ has no significant impact on the outcome.

3.3. SOA-SVM

Support Vector Machine (SVM) was proposed in the early 1990s. It is based on the statistical learning theory's VC measure idea and the structural risk reduction principle. It can balance the amount of computation and the ability of computation on the basis of limited sample information. Its superior data classification and recognition ability makes it very effective in rolling bearing fault diagnosis. The population classification in the case of linear separability is shown in Figure 2. It can be seen from the figure that two different types of samples are divided by the optimal hyperplane H , and the purpose of SVM classifier is to find the hyperplane.

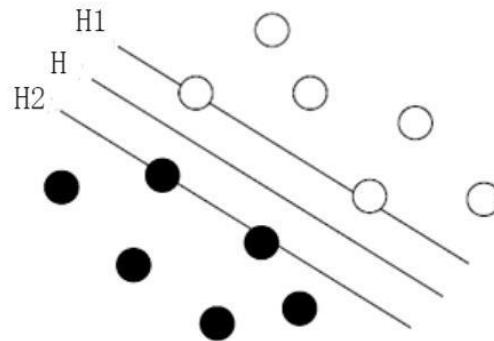


Figure 2. Optimal hyperplane diagram.

Let the two types of sample sets in the graph be $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$; n is the number of samples and x_i is the i th input value of the sample feature space. In the linearly separable state, the optimal hyperplane solved by the support vector classifier can be transformed into the following constraint problem:

$$\begin{cases} \min \varphi(\omega) = \frac{1}{2} \|\omega\|^2 \\ \text{s.t. } y_i(\omega \cdot x_i + m) \geq 1, i = 1, 2, \dots, l \end{cases} \quad (13)$$

where ω is the normal vector of the hyperplane; m is the offset; $\omega \cdot x_i + m = 0$ is the hyperplane to be solved.

If the nature of the sample is linearly inseparable, the support vector machine will map the sample from the current space to the high-dimensional space Λ using nonlinear mapping Ψ . In this way, the problem of linear inseparability can be transformed into linear separability. Therefore, the optimal hyperplane can be obtained on the high-dimensional space Λ , but the kernel function $k(x_i, x_j)$ must meet the positive definite matrix condition, that is:

$$k(x_i, x_j) = \Psi(x_i) \cdot \Psi(x_j) \quad (14)$$

By selecting the appropriate kernel function $k(x_i, x_j)$, the nonlinear samples can be linearized and classified. After the relaxation variable ξ_i is introduced, the expression of the original classification hyperplane is:

$$\begin{cases} \min \varphi(\omega) = \frac{1}{2} \|\omega\|^2 + C(\sum_{i=1}^n \xi_i) \\ y_i(\omega \cdot \Psi(x_i) + b) \geq 1 - \xi_i, i = 1, 2, \dots, n \end{cases} \quad (15)$$

where C is the penalty factor. After introducing the Laplace multiplication operator, the optimal classification hyperplane problem is transformed into a dual quadratic programming problem. At the same time, it is substituted into the inner product transformation of the kernel function, and formula (15) becomes:

$$\begin{cases} \max L(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j (k(x_i, x_j)) \\ \text{s.t. } 0 \leq \alpha_i \leq C, \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0 \end{cases} \quad (16)$$

The final classification hyperplane can be expressed as:

$$f(x) = \operatorname{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i k(x_i * x) + a^* \right\} \quad (17)$$

The parameter setting of support vector machine algorithm affects its learning and generalization ability, so how to select the optimal parameters has great research value. When $k(x_i, x_j)$ is the radial basis function, The debugging of penalty factor C and kernel width g is the major focus of SVM parameter adjustment.

Gaurav Dhiman [32] introduced Seagull optimization algorithm (SOA) as a new swarm intelligence optimization technique in 2018. The algorithm mainly simulates the migration behavior and attack behavior of seagull groups in nature. Migration refers to the movement of seagulls from one position to another while seagulls should meet three conditions: avoiding collision, direction of the best position direction and approaching the best position[33].

(1) Avoid collision. In order to prevent the occurrence of a collision between adjacent seagulls, add a new variable a . The formula is :

$$c(t) = A \times p(t) \quad (18)$$

where $c(t)$ represents the new position of seagulls after collision avoidance; t indicates the current number of iterations; $p(t)$ indicates the initial position of the seagull; A represents the motion behavior of seagulls in a given search space. The calculation formula of A is :

$$A = f - t \times (f/n_{\max}) \quad (19)$$

where the value of A is adjusted by f linearly, and it decreases linearly from f to 0; n_{\max} is the maximum number of iterations.

(2) Direction of the best position . On the premise of not colliding with other individuals, seagulls will move in the direction of the best position. The formula is :

$$m(t) = B \times [Z_{\text{best}}(t) - p(t)] \quad (20)$$

where $m(t)$ indicates the direction of the best position of the seagull; B is mainly responsible for balancing global search and local search. In order to obtain an appropriate balance number, the calculation formula of B is :

$$B = 2 \times A \times A \times r_d \quad (21)$$

where r_d is a random number between $[0, 1]$.

(3) Approaching the best position. The seagull will soar in the route of the best position to achieve a new one after landing in a safe location away from other seagulls.

$$d(t) = |c(t) + m(t)| \quad (22)$$

where $d(t)$ indicates the new position where the seagull meets three conditions.

During migration, seagulls can constantly change the angle and speed of attack. With the help of gravity and wings, they can maintain a certain height in the air. When seagulls attack their prey, they move spirally in the air, and their motion behavior is represented by a , b and c components respectively. The formulas of motion behavior are as follows:

$$a = r \times \theta \quad (23)$$

$$b = r \times \sin \theta \quad (24)$$

$$c = r \times \cos \theta \quad (25)$$

$$r = u \times e^{\theta v} \quad (26)$$

where r is the spiral radius in the movement of seagulls; u and v are the correlation constants of spiral shape; θ Represents the angle, which is a random number between $[0, 2\pi]$; The formula of seagull's attack behavior obtained from the movement behavior is :

$$p(t) = d(t) \times a \times b \times c + Z_{best}(t) \quad (27)$$

where $Z_{best}(t)$ indicates the best seagull position. The steps of optimizing support vector machine with seagull algorithm are as follows:

(1) Initialize the population parameters of seagull optimization algorithm, the number of iterations, and the value range of C and g .

(2) Determine the fitness function of seagull optimization algorithm, and evaluate the adaptability of seagull individuals on the basis of the value of fitness function. According to the principle of seagull optimization algorithm, find the optimal fitness value and the optimal position obtained by seagull.

(3) According to the best individual position of seagull, the optimal values of parameters C and g are obtained.

(4) The optimal parameters C and g are assigned to the support vector machine for training, and the optimized support vector machine classification model is obtained.

(5) Input the test samples, then the optimized SVM classification model will output the predicted labels of the test samples and compare the predicted labels with the actual labels to obtain the classification accuracy.

3. Experiment and Results

3.1. Experimental System

For analysis, select Case Western Reserve University (CWRU) rolling bearing data information. The selected data material is the mechanical vibration data signal of the SKF6205 rolling bearing on the motor drive side. Rolling bearing bore, outside and turning body are wire by EDM 0.007 diameter. The push motor is rated at 1797 rpm. The sampling rate is 12kHz. Figure 3 depicts the structure of the test service platform. When a roller bearing is damaged, the roller bearing can come into contact with failure locations, causing significant shock vibrations. The cycle time for the flip body to touch the common fault location varies by common fault type. Figure 4 shows the frequency domain waveforms of the vibration data signals for the four rolling bearing cases.

There are four different sorts of flaws that are investigated. Each type of fault is represented by 50 data groups, 30 of which are training samples in a known state and the remaining 20 are diagnostic test samples. Each batch of vibration data has a sample length of 2048. The sample data set is shown in Table 1.

Table 1. Description of bearing dataset.

Type	Fault Size/mm	Category Label
Normal	—	1
Inner Race Fault	0.1778	2
Outer Race Fault	0.1778	3
Rolling Element fault	0.1778	4

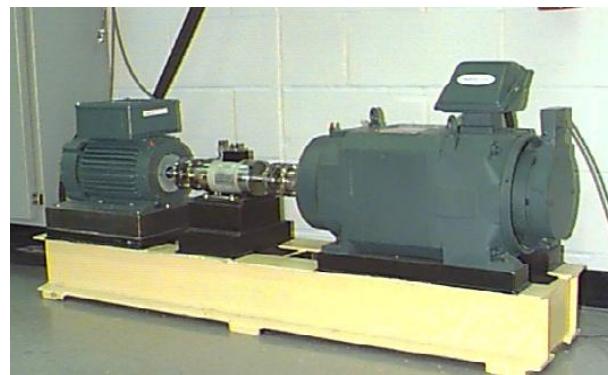


Figure 3. Data set experiment platform of CWRU.

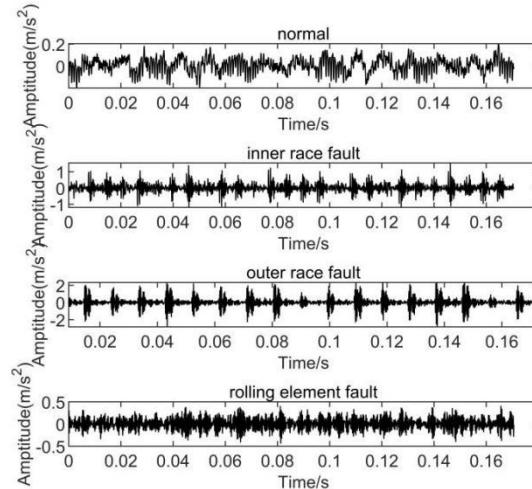


Figure 4. The time domain waveforms of vibration signals under four bearing states.

3.2. Results and Discussion

The vibration signal is decomposed using the enhanced VMD method. Taking the outer ring defect as an example, the whale algorithm is utilized to improve the parameters of the VMD algorithm. Figure 5 shows the minimal envelope entropy fluctuation as the number of generations in the WOA optimization process grows. The minimum envelope entropy in the fourth iteration is 3.6576 As seen in the figure. The optimization procedure is complete when the number of iterations hits 10, and the optimization parameters $[K, \alpha]$ is [10, 1469]. The best parameters in Table 2 are used to set the VMD algorithm parameters. Then the optimized VMD algorithm is used to

decompose the vibration signals of different damage positions and degrees of rolling bearing.

In the VMD approach, the measurement index is a crucial factor that decides if the decomposition result is satisfactory. In mechanical fault identification, kurtosis is an important index of vibration signal analysis. Kurtosis is a numerical statistic that depicts the features of random variables' distribution. It is a dimensionless parameter. Because kurtosis has nothing to do with factors such as bearing rotation speed, size and mechanical load, and is very sensitive to impact signal, it is particularly suitable to study surface damage faults [34].

The main IMFs after vibration signal decomposition are selected according to kurtosis. According to table 3, it can be seen that the kurtosis value when modes $n = 3, 9$ and 10 are always the larger of the ten modes when different outer ring fault samples are analyzed and calculated; By analyzing the kurtosis values of different modal components in the inner ring fault with the same method, it can be known that the kurtosis values of the corresponding modal components in this state when $n = 3, 6$ and 10 are the larger three; When $n = 3, 4$ and 10 , the kurtosis value of the corresponding modal component is the larger in the rolling element fault. However, through the analysis of bearing signals under normal conditions, the results shown in Table 6 are obtained, and the kurtosis value of each intrinsic mode function has no obvious law. The reason for this result may be that the definition of kurtosis criterion makes the vibration data in normal state not suitable for kurtosis criterion analysis. According to the correlation coefficient [35] between each intrinsic mode function and the original signal, the modal component n equal to $1, 2$ and 4 are selected as the main intrinsic mode function in normal state for subsequent analysis.

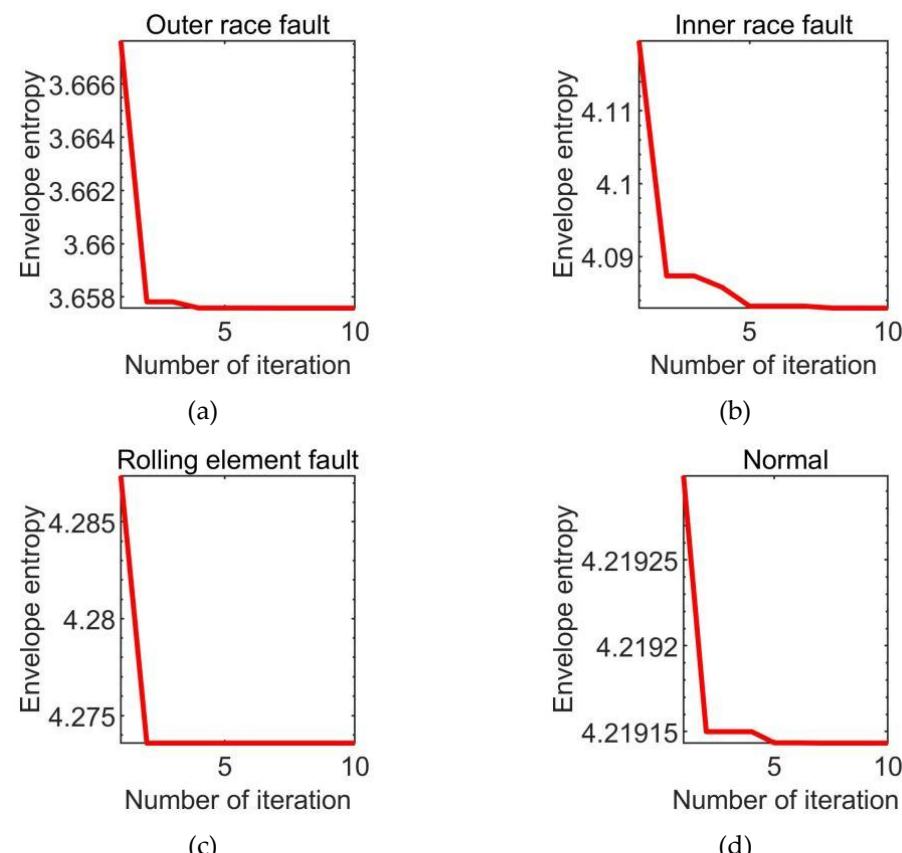


Figure 5. Optimized VMD curve with IWOA.

Table 2. Optimization parameters obtained by using the IWOA.

	Normal	Inner race	Outer race	Rolling element
K	6	10	10	10
α	1108	1219	1469	2000

Table 3. Kurtosis value of each mode of outer ring fault.

Imf1	Imf2	Imf3	Imf4	Imf5	Imf6	Imf7	Imf8	Imf9	Imf10
3.110	3.699	4.460	2.159	3.583	2.249	2.695	2.687	4.751	5.099
2.975	2.649	4.488	2.281	3.570	2.358	2.897	2.915	5.241	5.067
3.177	2.410	4.561	2.205	3.620	2.374	2.763	2.740	5.163	5.429

Table 4. Kurtosis value of each mode of inner ring fault.

Imf1	Imf2	Imf3	Imf4	Imf5	Imf6	Imf7	Imf8	Imf9	Imf10
2.859	2.128	3.565	2.534	2.951	3.150	2.997	2.109	3.106	3.625
2.972	2.020	3.545	2.480	2.782	3.138	2.573	1.930	2.712	3.218
2.699	2.017	3.417	2.475	2.920	3.149	2.619	2.335	2.452	3.500

Table 5. Kurtosis value of each mode of rolling element fault.

Imf1	Imf2	Imf3	Imf4	Imf5	Imf6	Imf7	Imf8	Imf9	Imf10
2.493	2.676	3.403	2.990	2.654	2.363	2.714	2.482	2.699	3.018
2.885	2.867	3.324	3.400	3.324	2.974	2.630	2.723	2.941	3.340
2.450	2.891	3.367	3.013	2.992	2.925	3.032	2.514	2.689	3.333

Table 6. Kurtosis value of each mode of normal.

Imf1	Imf2	Imf3	Imf4	Imf5	Imf6
2.825	3.110	1.545	2.559	3.956	3.014
2.970	3.261	1.583	2.706	2.946	2.854
2.329	3.030	1.600	2.996	2.911	3.316

Table 7. Correlation coefficient of each mode of normal.

Imf1	Imf2	Imf3	Imf4	Imf5	Imf6
0.649	0.644	0.304	0.382	0.090	0.048
0.631	0.658	0.300	0.383	0.097	0.049
0.632	0.654	0.289	0.384	0.092	0.047

The MPE of three main intrinsic mode functions in four states is calculated respectively. According to the results shown in Figure 6, when $s = 1, 2$, the permutation entropy calculated by rolling element fault and inner ring fault is very close; When $s = 3$, the permutation entropy calculated by inner ring fault and outer ring fault is very close. If the feature vector is constructed based on these, it may cause the disorder of later state classification. Considering the average distance and minimum distance of the entropy of imf1 multi-scale arrangement in four states, the value of the optimal scale factor s of imf1 is chosen to be 7. Using the same method, the optimal scale factor s of imf2 is 9 and the optimal scale factor s of imf3 is 4.

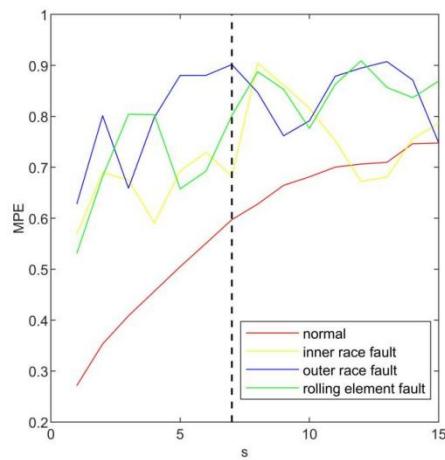


Figure 6. MPE of imf1.

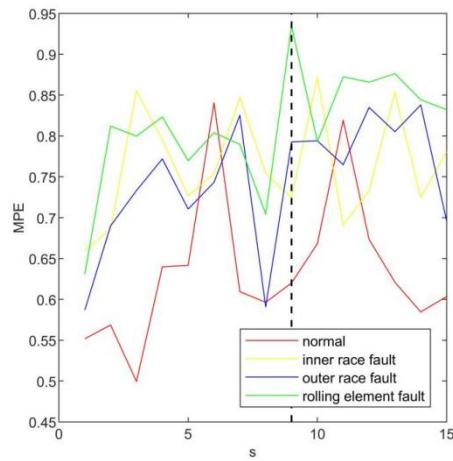


Figure 7. MPE of imf2.

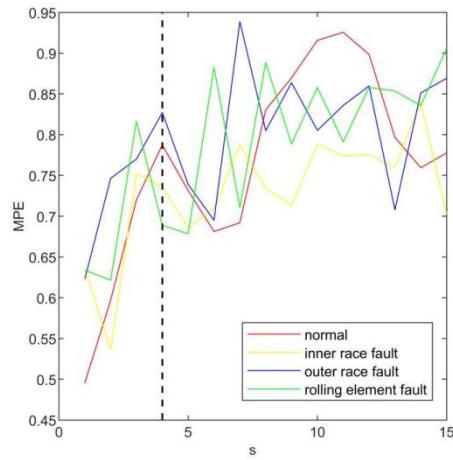


Figure 8. MPE of imf3.

Using the feature vector construction method proposed in this paper, the corresponding optimal multi-scale permutation entropy of all samples are calculated to form the feature vector. There are $50 \times 4 \times 3$ permutation entropy, 50×4 feature vectors. 30×4 feature vectors of the training samples are used to train the model of support vector machine and optimize its parameters. 20×4 feature vectors of test samples are used as unknowns for the final classification test. Four sets of feature vectors are

given here, as shown in Table 8. The digital labels 1 ~ 4 in the table represent the normal state, inner ring fault, outer ring fault and rolling element fault respectively.

Table 8. Feature vectors and labels.

Feature value 1	Feature value 2	Feature value 3	Label
0.597	0.742	0.482	1
0.682	0.621	0.832	2
0.902	0.834	0.750	3
0.802	0.846	0.811	4

120 feature vectors of training sets similar to table 8 are input into SOA-SVM for training. As can be seen from Figure 9, after two iterations, the fitness value can reach 100%. The optimum parameters C and g are 35.609 and 1.991, respectively.

The learned detection entity model is used to identify rolling bearing faults. Figure 10 shows the confusion matrix results obtained by applying the WOA-VMD-SOA-SVM approach in four different common scenarios. There is an incorrectly classified sample, which identifies the rolling element fault as the inner ring fault, and the detection set's ultimate identification accuracy is 98.75%. The findings show that the fault detection approach can correctly identify common rolling bearing defects in a variety of conditions.

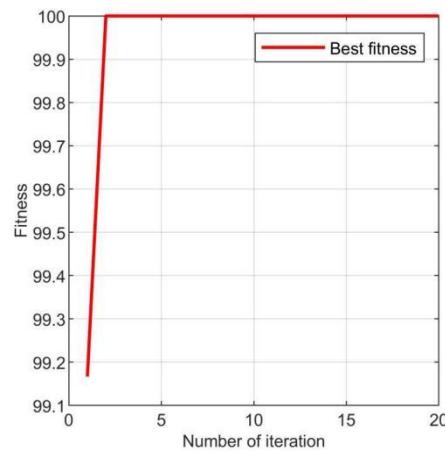


Figure 9. Convergence curve of the SOA-SVM method.

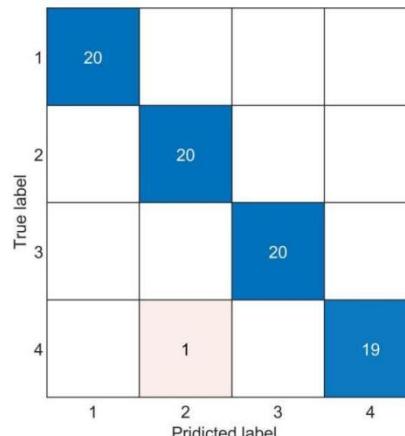


Figure 10. Confusion matrix of the WOA-VMD-SOA-SVM method.

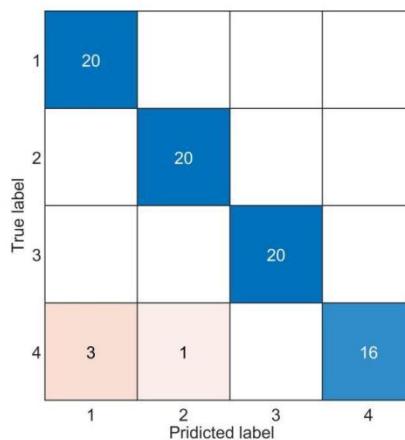


Figure 11. Confusion matrix of the VMD-SOA-SVM method.

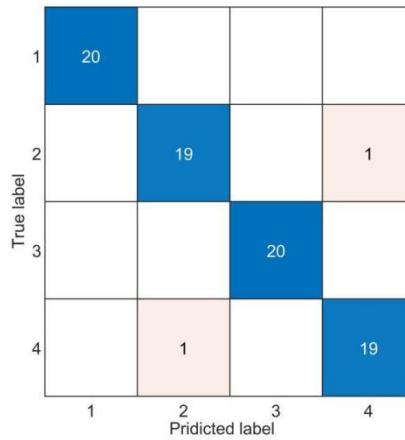


Figure 12. Confusion matrix of the WOA-VMD-PSO-SVM method.

To better verify the effectiveness of the improved VMD optimization algorithm, WOA-VMD and non-boosted VMD are compared in this paper. The VMD primary parameter K is set to 8 and α is set to 2000 in this case. Figure 11 shows the results of the confusion matrix applying VMD-SOA-SVM fault detection way. The accuracy on the detection set is 95.00%. Figure 12 shows the results of the confusion matrix applying the WOA-VMD-PSO-SVM fault detection approach. The accuracy on the detection set is 97.50%. According to the comparison of Figures 10 and 11, it can be shown that the WOA-VMD method's actual effect is stronger than that of the non-improved VMD method, indicating that the WOA-VMD method can more precisely collect the information content of common rolling bearings fault characteristics. In addition, according to the comparison of Figures 10 and 12, it can be shown that the actual effect of using WOA-VMD-SOA-SVM is better than that of applying WOA-VMD-PSO-SVM, indicating that SOA is more powerful than PSO.

4. Conclusions

In this article, we mentioned a fault detection method for rolling bearings that integrated WOA-VMD, multi-scale permutation entropy and SOA-SVM algorithm. Rolling bearing fault detection and analysis were carried out from the fields of data processing, fault feature extraction and fault feature recognition.

The key parameters of VMD were obtained using the whale optimization algorithm, and then the information of fault characteristics was obtained using the improved VMD method. According to the results, WOA-VMD may reasonably retrieve the fault information content of rolling bearings. In feature extraction, we found that the scale

factors s were 7, 9 and 4 respectively in order to obtain the optimal multi-scale permutation entropy of three imfs. The SOA approach was used to optimize the parameters of the penalty factor C and the kernel function g in the SVM fault detection entity model. The results showed that the SOA-SVM method had good classification characteristics, and the mean diagnosis accuracy can reach 98.75%. Compared with the results of other methods, it can be seen that this method can reasonably diagnose different damage types of the rolling bearings.

In the future work, we will focus on building a test service platform for mining fans, collecting mechanical vibration data signals of rolling bearings, and certifying the feasibility analysis of applying the methods mentioned in the article to fault detection of mining fans.

6. Patents

Author Contributions: Conceptualization, X.Z. and H.W.; methodology, H.W.; software, H.W and M.R.; validation, M.H. and L.J.; formal analysis, M.H.; investigation, L.J.; resources, M.H.; data curation, M.R.; writing—original draft preparation, H.W.; writing—review and editing, X.Z.; visualization, H.W.; supervision, M.R.; project administration, M.H.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by funds by the National Natural Science Foundation of China No.51774293.

Data Availability Statement: This research used the published rolling bearing datasets of Western Reserve University. The access of the datasets is <https://engineering.case.edu/bearingdatacenter/welcome>.

Acknowledgments: The authors appreciate the editor's helpful remarks and ideas, as well as those of the anonymous reviewers. The authors also would like to thank the National Natural Science Foundation of China No.51774293 for its support.

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

References

- [1] C. He, T. Wu, R. Gu, Z. Jin, R. Ma, and H. Qu, "Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm – Extreme learning machine," *Measurement*, vol. 173, 2021.
- [2] H. Liu and M. Han, "A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings," *Mechanism and Machine Theory*, vol. 75, pp. 67-78, 2014.
- [3] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism and Machine Theory*, vol. 78, pp. 187-200, 2014.
- [4] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4919-4939, 2012.
- [5] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Mathematical Physical & Engineering Sciences*, vol. 454, no. 1971, pp. 903-995, 1998.
- [6] Q. Du and S. Yang, "Improvement of the EMD method and applications in defect diagnosis of ball bearings," *Measurement Science & Technology*, vol. 17, no. 8, p. 2355, 2006.
- [7] H. Wang, J. Chen, and G. Dong, "Feature extraction of rolling bearing's early weak fault based on EEMD and tunable Q-factor wavelet transform," *Mechanical Systems and Signal Processing*, vol. 48, no. 1-2, 2014.
- [8] Dominique, Zosso, Konstantin, and Dragomiretskiy, "Variational Mode Decomposition," *IEEE Transactions on Signal Processing A Publication of the IEEE Signal Processing Society*, 2014.
- [9] Z. Zhang, X. Zhang, P. Wu, and X. Li, "Compound fault extraction method via self-adaptively determining the number of decomposition layers of the variational mode decomposition," *Rev Sci Instrum*, vol. 89, no. 8, p. 085110, Aug 2018.
- [10] Z. Guo, M. Liu, Y. Wang, and H. Qin, "A New Fault Diagnosis Classifier for Rolling Bearing United Multi-Scale Permutation Entropy Optimize VMD and Cuckoo Search SVM," *IEEE Access*, vol. 8, pp. 153610-153629, 2020.
- [11] G. A. Ran, C. A. Jie, A. Rh, W. A. Hua, and B. Ww, "Incipient fault diagnosis of rolling bearings based on adaptive variational mode decomposition and Teager energy operator," *Measurement*, vol. 149, pp. 106941-106941.
- [12] J. Li, W. Chen, K. Han, and Q. Wang, "Fault Diagnosis of Rolling Bearing Based on GA-VMD and Improved WOA-LSSVM," *IEEE Access*, vol. 8, pp. 166753-166767, 2020.
- [13] Richman, Joshua, S., Moorman, J., and Randall, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology*, 2000.
- [14] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [15] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale Entropy Analysis of Complex Physiologic Time Series," *Physical Review Letters*, vol. 89, no. 6, pp. 705 - 708, 2007.
- [16] W. Aziz and M. Arif, "Multiscale Permutation Entropy of Physiological Time Series," in *International Multitopic Conference*, 2005.
- [17] J. Zheng, J. Cheng, and Y. Yu, "Multiscale Permutation Entropy Based Rolling Bearing Fault Diagnosis," *Shock and Vibration*, 2014,(2014-3-2), vol. 2014, no. pt.1, pp. 1-8, 2014.
- [18] L. Shuang and L. Meng, "Bearing Fault Diagnosis Based on PCA and SVM," in *2007 International Conference on Mechatronics and Automation*, 2007.
- [19] Hou and Z. Ran, "Rolling Bearing Fault Diagnosis Based on Wavelet Packet and Improved BP Neural Network for Wind Turbines," *Applied Mechanics & Materials*, vol. 347-350, pp. 117-120, 2013.
- [20] B. Qin, G. D. Sun, L. Y. Zhang, J. G. Wang, and J. Hu, "Fault Features Extraction and Identification based Rolling Bearing Fault Diagnosis," *Journal of Physics: Conference Series*, vol. 842, p. 012055, 2017.

[21] Z. Wang, L. Yao, and Y. Cai, "Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine," *Measurement*, vol. 156, 2020.

[22] R. Wang, Z. Zhang, Z. Xia, J. Miao, and Y. Guo, "A new approach for rolling bearing fault diagnosis based on EEMD hierarchical entropy and improved CS-SVM," in *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, 2019.

[23] H. Liu and J. Xiang, "A Strategy Using Variational Mode Decomposition, L-Kurtosis and Minimum Entropy Deconvolution to Detect Mechanical Faults," *IEEE Access*, vol. 7, pp. 70564-70573, 2019.

[24] H. Ma, Q. Tong, and Y. Zhang, "Applications of Optimization Parameters VMD to Fault Diagnosis of Rolling Bearings," *Zhongguo Jixie Gongcheng/China Mechanical Engineering*, vol. 29, no. 4, pp. 390-397, 2018.

[25] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51-67, 2016.

[26] W. Yang, Z. Yang, Y. Chen, and Z. Peng, "Modified Whale Optimization Algorithm for Multi-Type Combine Harvesters Scheduling," *Machines*, vol. 10, no. 1, 2022.

[27] Tang, G.J.; Wang, X.L. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing. *J. Xi'an Jiaotong Univ.* 2015, 49, 73–81.

[28] P. Ma, H. Zhang, W. Fan, and C. Wang, "Fault diagnosis using an improved fusion feature based on manifold learning for wind turbine transmission system," *Journal of Vibroengineering*, vol. 21, no. 7, pp. 1859-1874, 2019.

[29] T. Liang, H. Lu, and H. Sun, "Application of Parameter Optimized Variational Mode Decomposition Method in Fault Feature Extraction of Rolling Bearing," *Entropy (Basel)*, vol. 23, no. 5, Apr 24 2021.

[30] Y. Li, M. Xu, Y. Wei, and W. Huang, "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree," *Measurement*, vol. 77, pp. 80-94, 2016.

[31] M. Matilla-Garcia, "A non-parametric test for independence based on symbolic dynamics," *Journal of Economic Dynamics and Control*, vol. 31, no. 12, pp. 3889-3903, 2007.

[32] G. Dhiman and V. Kumar, "Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems," *Knowledge-Based Systems*, vol. 165, pp. 169-196, 2019.

[33] J. Wang, Y. Li, and G. Hu, "Hybrid seagull optimization algorithm and its engineering application integrating Yin-Yang Pair idea," *Engineering with Computers*, pp. 1-37.

[34] J. Tan, X. Chen, and Z. He, "Impact Signal Detection Method with Adaptive Stochastic Resonance," *Journal of Mechanical Engineering*.

[35] J. Li, X. Chen, and Z. He, "Adaptive stochastic resonance method for impact signal detection based on sliding window," *Mechanical Systems & Signal Processing*, vol. 36, no. 2, pp. 240-255, 2013.