Research Paper

Rolling Bearing Fault Diagnosis Method Based on Improved Variational Mode Decomposition and Information Entropy

Liang GE[1),2)], Wen FAN[1)], Xiaoting XIAO[3)*], Fangji GAN[4)],
Xin LAI[1)], Hongxia DENG[1)], Qi HUANG[1)]

[1)] *School of Mechanical and Electrical Engineering, Southwest Petroleum University*
Chengdu 610500, China

[2)] *National Engineering and Research Center for Mountainous Highways*
Chongqing 400067, China

[3)] *School of Electrical Engineering and Information, Southwest Petroleum University*
Chengdu 610500, China
*Corresponding Author e-mail: xt_xiao@foxmail.com

[4)] *School of Manufacturing Industry and Engineering Sciences, Sichuan University*
Chengdu 610500, China

Due to the complex randomness and nonlinearity of rolling bearing vibration signal, it is challenging to extract fault features effectively. By analyzing the vibration mechanism of rolling bearing, it is found that the vibration signal of local damage defects of rolling bearing has the characteristics of periodic impact and amplitude modulation. The variational mode decomposition (VMD) algorithm has a good advantage in dealing with nonlinear and non-stationary signals and decomposing a signal into different modes. However, VMD has the problem of parameter selection, which directly affects the performance of VMD processing, and causes mode aliasing. Therefore, a rolling bearing fault diagnosis method based on improved VMD is proposed. A new fitness function combining differential evolution (DE) algorithm with gray wolf optimization (GWO) algorithm is proposed to form a new hybrid optimization algorithm, named DEGWO. The simulation results show that the improved VMD method based on DEGWO can adaptively remove the noise according to the characteristics of the signal and restore the original characteristics of the vibration signal. Finally, in order to verify the advantages of the research, the information entropy is extracted from the data of 1000 samples in the bearing database of Case Western Reserve University as the feature set, which is input into support vector machine (SVM) for fault diagnosis test. The results show that the diagnostic accuracy of this method is 96.5%, which effectively improved the accuracy of rolling bearing fault diagnosis.

Key words: rolling bearing; gray wolf optimization; fault diagnosis; variable mode decomposition.

1. Introduction

As an important part of mechanical equipment, rolling bearing seriously affects the normal operation of the whole mechanical equipment. According to the incomplete statistics, more than 30% of the rotating machinery faults are caused by rolling bearings, which may affect the rotating machinery operation and even cause large casualties [1]. To ensure the normal and safe operation of mechanical equipment, early fault diagnosis of rolling bearing is particularly important. With strong adaptation, high accuracy, easy and intuitive signal testing and processing, the vibration signal analysis method is widely used in this field. However, the rolling bear vibration signal generally shows non-stability and non-linearity. So the complexity of the time sequence cannot be measured and the internal information of the original signal cannot be separated for extracting the features [2]. The extracted signal features cannot reflect the fault information well, leading to poor diagnosis performance and dissatisfactory rolling bear fault diagnosis. Therefore, proper preprocessing of the vibration signal before feature extraction becomes necessary to improve the accuracy of rolling bearing fault diagnosis.

Currently, there are many nonlinear dynamic methods based on a statistical parameter estimation to extract the fault features [3]. However, due to their limitations, the preprocessing effect is not the best, which affects the accuracy of feature extraction. In 2014, Dragomiretskiy and Zosso [6] proposed a new adaptive signal processing method, called the variational mode decomposition (VMD), widely used in mechanical, medical, financial, biological, and other fields. It is also a popular data preprocessing algorithm in rolling bearing fault diagnosis [7, 8]. Existing researches prove that the VMD algorithm is superior to some traditional signal processing algorithms [9–12]. However, it lacks adaptability and needs to set the parameter group in advance. The choice of parameter groups directly affects the processing effect of the algorithm.

To solve the problem of parameter selection in the VMD algorithm, the gray wolf optimization (GWO) algorithm is introduced for intelligent optimization of parameter group. The above mentioned algorithm was first proposed in 2014 [13]. Its principle is to select the optimal parameter array according to the characteristics of the wolf social group. However, it may fall into the local optimal solution, so the DE algorithm is introduced to optimize it. The DE algorithm was first proposed in 1995, and it is based on global optimization [14]. The algorithm has strong global search ability and search efficiency. The initial population of the gray wolf algorithm can be optimized by using its differential mutation characteristics to enlarge the difference between the populations. Therefore, the search speed of the optimal solution is accelerated, and the classification performance of GWO is improved. The optimized algorithm is called DEGWO. In the

process of optimization, the fitness function is modified to avoid the influence of human factors. The optimal combination of preset modal scale and balance factor is automatically selected. Thus, the improved DEGWO-VMD algorithm is established.

The rest of this paper is arranged as follows. In the second section, the dynamic model of rolling bearing vibration signal is analyzed first. Then a novel preprocessing algorithm, an improved DEGWO-VMD algorithm, is proposed and described. In the third section, the proposed algorithm is validated by comparing it to other algorithms through simulation data. In the fourth section, the feasibility of this method is verified by the experiment of bearing database of Case Western Reserve University. The fifth section draws some conclusions and prospects.

2. Dynamic modeling and theoretical analysis of rolling bearing fault

By analyzing the bearing dynamic model, the vibration response characteristics of the fault bearing can be more clearly understood, and the internal relationship between the dynamic parameters and the response signals under fault conditions can be revealed, which provides a theoretical basis for the use of signal analysis method for bearing fault diagnosis [15]. The dynamic vibration model of rolling bearing is shown in Fig. 1, where ω is the angular velocity, k is stiffness, and m is mass.
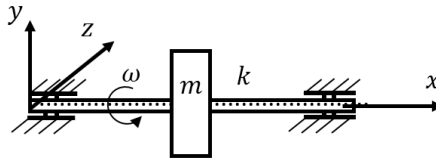

Fig. 1. The dynamic model of rolling bearing.

The system kinematics equation of single cycle impulse force and single-degree of freedom model is shown in Eq. (2.1):

$$(2.1) \qquad m\ddot{A} + c\dot{A} + kA = F_m,$$

where c is damping, A is the radial vibration amplitude of the system, and F_m is the non-inertia force of the rotor system. According to Newton's law, it can be seen that

$$(2.2) \qquad F_m = me\omega^2,$$

where e is the rotor radius. The stress analysis diagram of the rolling bearing is shown in Fig. 2, where G is the gravity of the system. Therefore, the expression of force F_s supported by the bearing is determined as follows:
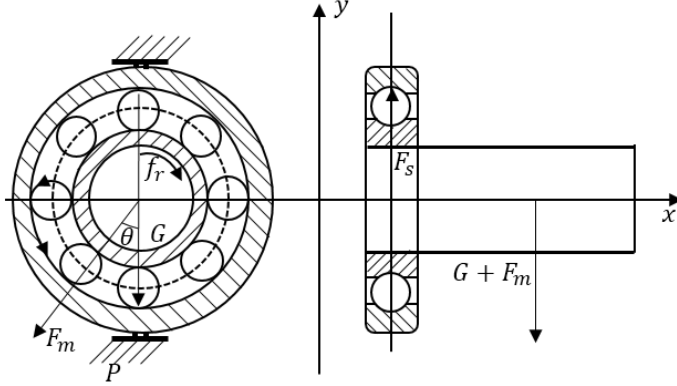
$$(2.3) \qquad \mathbf{F}_s = \mathbf{G} + \mathbf{F}_m.$$

FIG. 2. The stress analysis diagram.

2.1. *Modeling of fault impact energy and impact characteristics*

By solving Eq. (2.1), the amplitude A_i of the i-th impact can be obtained, as shown in Eq. (2.4)

$$(2.4) \qquad A_i = A_0 \cos\left(2\pi f_A t + \varphi_A\right) + C_A + randn(t),$$

where f_A is the modulation frequency, φ_A and C_A are arbitrary constants set in the simulation to $\varphi_A = 0$ and $C_A = 7$, and $randn(t)$ is a Gaussian white noise with a variance of 10^{-6} and mean value of 0. Pitting corrosion occurs at different bearing positions and f_A values are different, as shown in Table 1.

Table 1. The f_A value of different faults and the formula for calculating the characteristic frequency of faults.

Different fault conditions	Value of f_A	Calculation formula for fault characteristic frequency
Outer ring fault	$f_A = 0$	$f_o = \dfrac{N_b}{2}\left(1 - \dfrac{d}{D}\cos\theta\right)f_r$
Inner ring fault	$f_A = f_r$	$f_i = \dfrac{N_b}{2}\left(1 + \dfrac{d}{D}\cos\theta\right)f_r$
Rolling element failure	$f_A = f_c$	$f_b = \dfrac{D}{2d}\left(1 - \left(\dfrac{d}{D}\right)^2\cos^2\theta\right)f_r$

Additionally, f_r is the rotation frequency, N_b is the number of rollers, θ is the pressure angle, f_b is the roller failure frequency, f_o is the outer ring failure frequency, f_i is the inner ring failure frequency, d is the roller diameter, D is the bearing pitch diameter, and f_c is the cage rotation frequency, as shown in Eq. (2.5):

$$(2.5) \qquad f_c = \frac{1}{2} \left(1 - \frac{d}{D} \cos\theta \right) f_r.$$

The impact oscillation $h(t)$ is an oscillation attenuation signal with the system's natural frequency as the oscillation frequency, as shown in Eq. (2.6):

$$(2.6) \qquad h(t) = \exp(-Bt) \cos\left(2\pi f_n t + \phi_\omega\right),$$

where f_n is the system's natural frequency and B is the attenuation constant.

According to Eqs. (2.4) and (2.6), the simulation model of rolling bearing fault vibration signal can be obtained, as shown in Eq. (2.7):

$$(2.7) \qquad x(t) = \sum_i A_i h(t - iT - \tau_i) + n(t),$$

where $x(t)$ on the left side of the equation is the vibration signal of rolling bearing pitting fault. On the right side of the equation, $n(t)$ is a Gaussian white noise with a mean value of 0 and variance of 1, t is time, $h(t)$ is the impact oscillation caused by pitting fault, $T = 1/f_g$ is the time difference between two shock oscillations, f_g is the fault characteristic frequency of different faults, A_i is the amplitude of the i-th impact, and τ_i is the fluctuation of the period T caused by the rolling element slipping during the i-th fault impact. Therefore, the mathematical model of the rolling bearing pitting failure is established.

The modeling parameters of the rolling bearing fault simulation signal are shown in Table 2. The parameters of bearing inner diameter, outer diameter and roller number are consistent with those of rolling bearing number-12k. According to Table 2, the fault impulse frequency of three kinds of fault signals can be obtained. Then, a group of rolling bearing fault simulation signals with 40 Hz rotating frequency can be obtained by using Eqs. (2.4), (2.6), (2.7) and parameters in Table 2, and the normalized signals can be processed.

Table 2. Fault signal simulation modeling parameters.

Parameter	D [mm]	d [mm]	N_b	f_r [Hz]	θ [°]	f_n [Hz]	B
Value	47.00	7.94	8	40	0	3000	2500

2.2. *Variational mode decomposition algorithm*

The objective of the VMD method for solving variational problems is to make the sum of the estimated bandwidth of each eigenmode function optimal. The process can be expressed as searching for the minimum value of the constrained variational model, which is shown in Eq. (2.8):

$$
(2.8) \quad
\begin{cases}
\displaystyle \max_{\{u_k\},\{w_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j w_k t} \right\|_2^2 \right\}, \\
\text{s.t.} \ \displaystyle \sum_k u_k = f,
\end{cases}
$$

where $\{u_k\} = \{u_1, u_2, ..., u_k\}$ is the set of k intrinsic mode function (IMF) components obtained by decomposition, $\{w_k\} = \{w_1, w_2, ..., w_k\}$ is the set of k IMF central frequencies, $\otimes$ is the convolution operator, and $\delta(t)$ is the unit impulse function.

The constrained variational problem is transformed into an unconstrained one to solve the above problem. Lagrange multiplier and quadratic penalty are added to Eq. (2.8) to obtain the augmented Lagrangian function:

$$
(2.9) \quad L\left(\{u_k\}, \{w_k\}, \lambda \right) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \otimes u_k(t) \right] e^{-j w_k t} \right\|_2^2
$$

$$
+ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle,
$$

where α is the balance factor determining the accuracy of vibration signal reconstruction and $\lambda(t)$ is a Lagrange multiplier determining the strictness of the constraints.

The VMD algorithm lacks adaptability when dealing with the rolling bearing vibration signal, so it is necessary to set the decomposition mode number k and balance factor α in advance. According to the simulation signal x_2, different combinations of K and α are selected to compare their effects on the decomposition results. Three cosine components with 5 Hz, 50 Hz, and 200 Hz are selected for the simulation signal, and a Gaussian white noise with a mean value of 0 and variance of 1 is selected:

$$
X_2 = x_1(t) + x_2(t) + x_3(t) + \xi, \quad t \in [0, 1],
$$

$$
x_1(t) = \cos(2\pi \cdot 5t),
$$

$$
(2.10) \quad x_2(t) = 0.25 \cos(2\pi \cdot 50t),
$$

$$
x_3(t) = 0.5 \cos(2\pi \cdot 200t).
$$

Figure 3 shows the decomposition results of selecting different combinations of K and α. By comparing Figs. 3a and 3b, it can be concluded that the balance factor remains unchanged. When the value of K is small, i.e., $K = 2$ (under decomposition), it means that the number of decomposition is less than the ideal value, and this can easily cause the problem of modal loss. When the value of K is large, i.e., $K = 8$ (over decomposition), it means that the number of decomposition is more than the ideal value, and it can easily cause false components. By comparing Figs. 3c and 3d, the preset scale remains unchanged. At this time, if the balance factor is too small, one component will be decomposed into other adjacent components, resulting in mode aliasing. If the balance factor is larger, the bandwidth will be smaller and false components will easily appear.
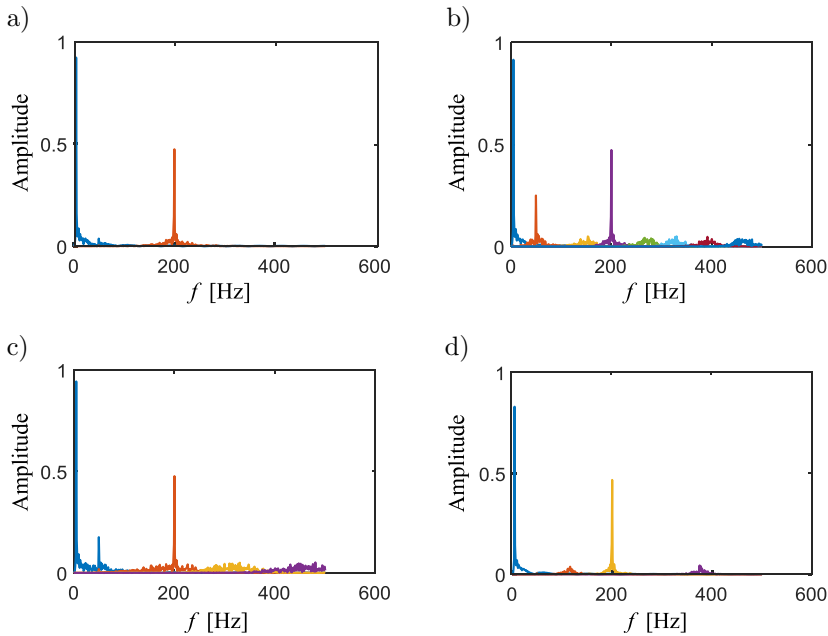

FIG. 3. Decomposition results of different combinations of K and α: a) $K = 2$, $\alpha = 2000$, b) $K = 8$, $\alpha = 2000$, c) $K = 4$, $\alpha = 100$, d) $K = 4$, $\alpha = 10\,000$.

Therefore, it is necessary to solve the parameter optimization problem of the VMD algorithm.

2.3. Gray wolf optimization algorithm

The GWO algorithm was used in this paper to optimize the parameters of the VMD algorithm. This is an optimization search method inspired by the predatory activity of the gray wolf. It has the characteristics of strong conver-

gence, few parameters and easy implementation. The specific modeling process
of GWO is illustrated as follows [13]:

2.3.1. Social hierarchy. According to the social class of the gray wolf population, the final optimal solution is defined as α. The second and third optimal solutions are defined as β and δ, respectively. The remaining candidate solutions are defined as ω. The core of the gray wolf algorithm is the search and hunting behavior of the wolf. By comparing the different positions randomly searched by the wolf with the three leadership positions, the better position is given to the level wolf group to achieve the purpose of optimization.

2.3.2. Surround your prey. The mathematical model of the behavior of wolf pack encirclement is as follows (see Fig. 4):

$$(2.11) \qquad\qquad D = |\mathbf{C} \cdot \mathbf{X}_p(t) - \mathbf{X}(t)|,$$

$$(2.12) \qquad\qquad \mathbf{X}(t+1) = \mathbf{X}_p(t) - \mathbf{A} \cdot D,$$

where $\mathbf{X}(t)$ and $\mathbf{X}_p(t)$ are the position vectors of gray wolf and prey, respectively, and t is the current iteration number. The expressions of vectors $\mathbf{A}_l$ and $\mathbf{C}_l$ are as follows:

$$(2.13) \qquad\qquad \mathbf{A}_i = 2\boldsymbol{\alpha} \cdot \mathbf{r}_1 - \boldsymbol{\alpha}, \quad i = 1,2,3,$$

$$(2.14) \qquad\qquad \mathbf{C}_i = 2\mathbf{r}_2,$$

$$(2.15) \qquad\qquad \boldsymbol{\alpha} = 2 \cdot \left(1 - \frac{i}{\max T}\right),$$

where i is the number of leaders, $\boldsymbol{\alpha}$ is a dynamic vector, which will decrease linearly from 2 to 0 in the whole iteration process, $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors between $[0, 1]$, and $\max T$ is the maximum number of iterations of the algorithm.
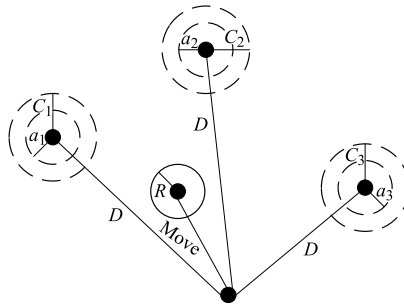

Fig. 4. Schematic diagram of gray wolf position update of GWO algorithm.

2.3.3. Hunting for prey. Usually, the exact location of the prey (the optimal solution) cannot be obtained in reality. Therefore, we need to make a hypothesis: α, β, δ can more accurately identify the specific location of the prey. Based on this, the positions of the rest gray wolves (ω) are updated according to the positions of the best three wolves in each iteration. Therefore, the mathematical model of hunting for a prey is as follows:

$$\begin{aligned} D_\alpha &= |\mathbf{C}_\alpha \cdot \mathbf{X}_\alpha - \mathbf{X}(t)|, \\ D_\beta &= |\mathbf{C}_\beta \cdot \mathbf{X}_\beta - \mathbf{X}(t)|, \\ D_\delta &= |\mathbf{C}_\delta \cdot \mathbf{X}_\delta - \mathbf{X}(t)|, \end{aligned}$$

(2.16)

$$\begin{aligned} \mathbf{X}_1(t+1) &= \mathbf{X}_\alpha(t) - \mathbf{A}_1 \cdot D_\alpha, \\ \mathbf{X}_2(t+1) &= \mathbf{X}_\beta(t) - \mathbf{A}_2 \cdot D_\beta, \\ \mathbf{X}_3(t+1) &= \mathbf{X}_\delta(t) - \mathbf{A}_3 \cdot D_\delta, \end{aligned}$$

(2.17)

$$(2.18) \qquad \mathbf{X}(t+1) = \frac{\mathbf{X}_1(t+1) + \mathbf{X}_2(t+1) + \mathbf{X}_3(t+1)}{3},$$

where D_α, D_β, and D_δ represent the distance from the individual of the grass-roots wolf pack to the prey, $\mathbf{X}_\alpha$, $\mathbf{X}_\beta$, and $\mathbf{X}_\delta$ represent the position vectors of the leadership, the wolf and the wolf, respectively, $\mathbf{X}(t)$ represents the position vector of the grass-roots wolf, $\mathbf{X}(t+1)$ is the position vector of the wolf after updating the position, $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ are swing factors, representing the search distance weight of the wolf around the leadership, the wolf and the wolf, respectively, and $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{A}_3$ are convergence factors representing the position distance weights from ω wolf to α wolf, β wolf, and δ wolf, respectively.

2.3.4. Attack or search for prey. It can be seen in Eq. (2.13) that $|\mathbf{A}|$ will decrease linearly with the change of $\mathbf{a}$ in the iterative process. In fact, $|\mathbf{A}|$ is a random vector with the value interval of $[-a, a]$. In order to improve the global search ability of the gray wolf optimization algorithm, when $|\mathbf{A}| < 1$, the gray wolf attacks the prey, and when $|\mathbf{A}| > 1$, the gray wolf leaves to find the next more suitable prey.

The steps of the GWO algorithm are shown in Fig. 5.

2.4. Differential evolution algorithm

In this paper, the DE algorithm is introduced into the gray wolf algorithm to optimize the searching performance. It is an algorithm based on global optimization with strong global search ability and search efficiency, and requiring fewer parameters.
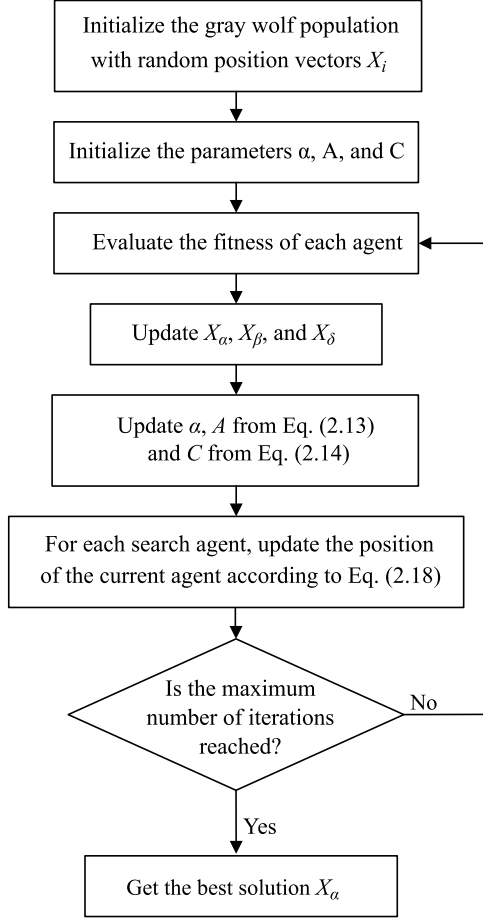
FIG. 5. The gray wolf optimization algorithm flow chart.

Defined in d-dimensional search space with the population size N, $X_k(g)$ is the k-th individual of g generation, where $k = 1, 2, ..., N$, $g = 1, 2, ..., t_{\max}$, and $t_{\max}$ is the maximum number of iterations. The calculation steps [14] are given as follows:

2.4.1. Population initialization. The initialization of each individual in the population can be expressed by Eq. (2.19):

$$(2.19) \qquad X_{k,p}(0) = X_p^L + \text{rand}(0, 1) \left(X_p^U - X_p^L \right),$$

where $\text{rand}(0, 1)$ is a random number whose value range is between $[0, 1]$, $X_{k,p}(0)$ is the value of the p-th dimension of the k-th individual in the initial population, and X_P^U and X_P^L are the upper and lower bounds of the p-th dimension, respectively.

2.4.2. Variation. A classical differential strategy is used to generate a mutant:

$$(2.20) \quad V_{k,p}(t+1) = X_{r_1,p}(t) + F \times (X_{r_2,p}(t) - X_{r_3,p}(t)), \quad r_1 \neq r_2 \neq r_3 \neq k,$$

where r_1, r_2, and r_3 represent three different randomly selected individuals, t is the current iteration number, and F is a parameter that can be selected and generated randomly.

2.4.3. Crossover operation. The crossover operation of $X(t)$ and its corresponding variant $V(t+1)$ are expressed as follows:

$$(2.21) \qquad U_{k,p}(t+1) = \begin{cases} V_{k,p}(t+1) & \text{if rand } (0,1) \leq P_{CR} or p = P_{\text{rand}}, \\ X_{k,p}(t) & \text{if rand } (0,1) > P_{CR} or p \neq P_{\text{rand}}, \end{cases}$$

where $U(t+1)$ is a new variant after crossover operation, the constant P_{CR} is a specific crossover probability, and P_{rand} is a randomly selected dimension.

2.4.4. Choice. To choose whether a new variant $U(t+1)$ is introduced into the next generation, DE uses the greedy standard to measure.

$$(2.22) \qquad X(t+1) = \begin{cases} U(t+1), & \text{if } f\left(U(t+1)\right) \leq f(X(t)), \\ X(t), & \text{if } f\left(U(t+1)\right) > f(X(t)), \end{cases}$$

where f is the loss function.

2.5. Improved fitness function

A proper adaptation function is of top priority in terms of the optimization process. At present, many algorithms take the envelope entropy E_p proposed by TANG [16] as the fitness function. After the signal $x(j)$ $(j = 1, 2, ..., m)$ is demodulated by the Hilbert transform, the envelope entropy E_p of the envelope signal $a(j)$ is obtained. The expression is shown in Eq. (2.23):

$$E_p = -\sum_{j=1}^{m} p_j \lg p_j,$$

$$(2.23)$$

$$p_j = \frac{a(j)}{\sum_{j=1}^{m} a(j)},$$

where p_j is the normalized expression of the envelope signal $a(j)$, E_p can quantitatively represent the variability of the original signal. E_p, also known as envelope variance, can reflect the size and degree of signal envelope fluctuation. The greater the envelope fluctuation of the signal, the greater the E_p of the signal, so the greater the noise contained in the signal.

However, during the experiment in Subsec. 2.2, it is found that if the E_p is too small, it is easy to lead to "over decomposition" of the optimal K value, resulting in mode aliasing and false components, which will ultimately affect the processing effect of the rolling bearing vibration signal. To solve the above problems, this study improved it. For IMF decomposed by VMD, if the center frequency ω distance of adjacent IMF is too close, it is easy to produce mode aliasing and false component. Based on this, a new parameter W_j is defined to measure the distance between two adjacent IMF's ω:

$$(2.24) \qquad W_j = \text{average}\,(\omega(i)) - \text{average}\,(\omega(j)), \quad j = i - 1, \quad i = 2, ..., K,$$

where K is the number of modal decomposition and average is the mean value of ω. To avoid mode aliasing and spurious components, we anticipate that the value W_j is as large as possible. But if it is too large, it will lead to insufficient decomposition of VMD, i.e., "under decomposition". To solve the problems existing in the VMD decomposition, a new fitness function is defined in this study as follows:

$$(2.25) \qquad \text{fitness} = \min \frac{E_p * \rho}{W_j},$$

where ρ is the adjusting parameter making the result more balanced. It is calculated that $\rho = 100$ makes the result more balanced. When a certain $[K, \alpha]$ is selected as the VMD algorithm parameter, a VMD decomposition is performed on the signal at a certain alpha wolf group position to obtain K IMF components. The fitness value of each IMF component is calculated and the smallest fitness value as the local minimum is obtained. The local value is used as the fitness value in the whole search process, i.e., fitness is taken as the fitness function and the local minimum value is taken as the search target.

2.6. Improved DEGWO-VMD algorithm

The preset modal scale K and balance factor α play a decisive role in the effect of variational mode decomposition of a rolling bearing vibration signal. GWO is used to optimize the parameters $[K, \alpha]$ of VMD. DE is introduced into GWO to classify the initial population. In the process of optimization, the improved fitness function is used, and the local minimum is taken as the search

target. The best $[K, \alpha]$ searching process of improved DEGWO-VMD algorithm is as follows (Fig. 6):

Step 1: Initialize the DEGWO parameters, including population size N, maximum iteration times $t_{\max}$, crossover probability P_{CR}, search dimension D, search range ub, lb, and scale factor F range.

Step 2: Initialize parameters α, A, and C. Then, use DE to process individual wolves according to Eq. (2.20) and produce intermediate. According to Eq. (2.22), the initial individuals are generated by the competitive selection and set $t = 1$.

Step 3: According to the objective function value of gray wolf, the best X_α, X_β, and X_δ are selected.

Step 4: According to Eq. (2.16), the distances between other individuals and the optimal X_α, X_β, and X_δ are calculated. The specific positions of each individual are updated according to Eqs. (2.17) and (2.18).

Step 5: Update the values of α, A, and C, and perform the DE crossover operation on the position of the individual population according to Eq. (2.21) to
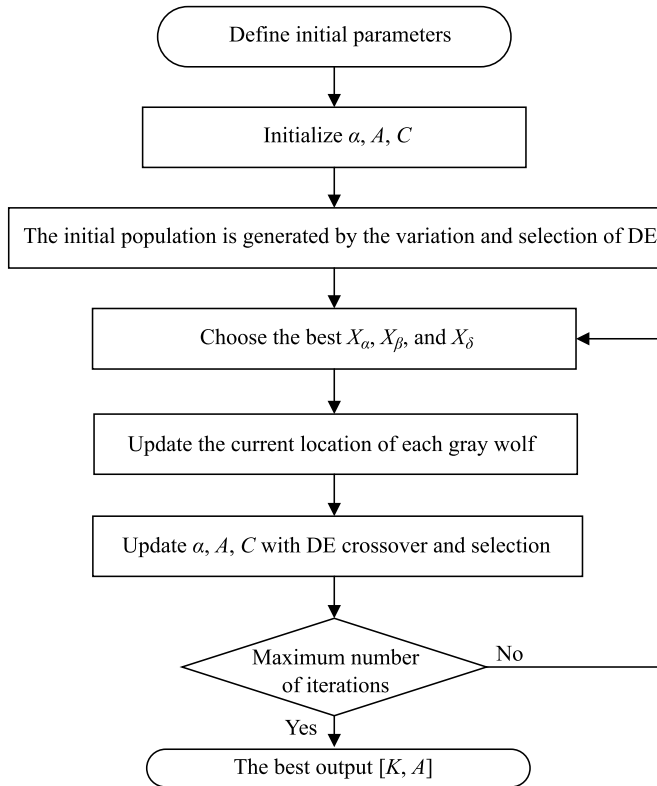

FIG. 6. Flow chart of the improved DEGWO-VMD algorithm.

retain better components. Then select new individuals according to Eq. (2.22). Calculate the objective function value of all gray wolf individuals, that is, the position of prey.

Step 6: Update the position of the first three wolves X_α, X_β, and X_δ according to the calculated objective function value.

Step 7: If the maximum number of iterations $t_{\max}$ is satisfied, the iteration will exit and output the objective function value of global optimal X_α and the optimal $[K, \alpha]$ value after rounding; otherwise, make $t = t + 1$ and return to step 3 to continue execution.

3. Signal simulation and algorithm verification

To verify the accuracy and effectiveness of the improved algorithm, the simulation signal is used for analysis. The simulation signal of a single-point fault signal on the outer ring of the rolling bearing is:

(3.1)
$$x(t) = \sum_i h\left(t - \frac{i}{f_g} - \tau_i\right) + n(t),$$

$$H(t) = Ae^{-kt}\sin\left(2\pi f_n t\right),$$

where $A = 1$ is the amplitude constant, $k = 800$ is the attenuation coefficient, $f_n = 1000$ Hz is the system resonance frequency $f_g = 128$ H, and $n(t)$ is a Gaussian white noise with a signal-to-noise ratio of 10.

It can be seen in Fig. 7 that the time domain waveform of the original signal is clear. In the corresponding spectrum diagram, the resonance frequency is 1000 Hz. After the time-domain waveform and spectrum diagram of the signal
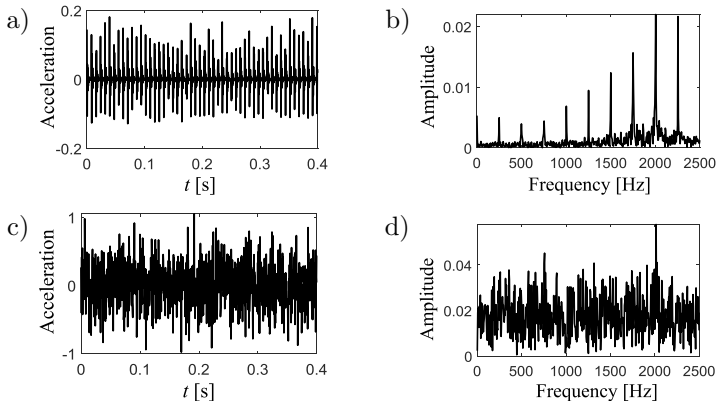

Fig. 7. Time-domain waveform and spectrum of the original and noisy signals: a) fault simulation signal, b) original signal spectrum, c) noisy signal, d) noise signal spectrum.

are mixed with noise, as shown in Figs. 7c and 7d, the fault feature becomes difficult to be observed. Figure 8 shows the envelope spectrum of the original and noisy signals. It can be seen in Fig. 8a, that the fault characteristic frequency (128 Hz) and its multiple frequencies of the simulated outer ring fault can be obtained in the envelope spectrum obtained by the Hilbert demodulation of the original signal. However, due to the influence of the noise signal, the fault characteristic frequency is completely submerged in the noise, as shown in Fig. 8b.
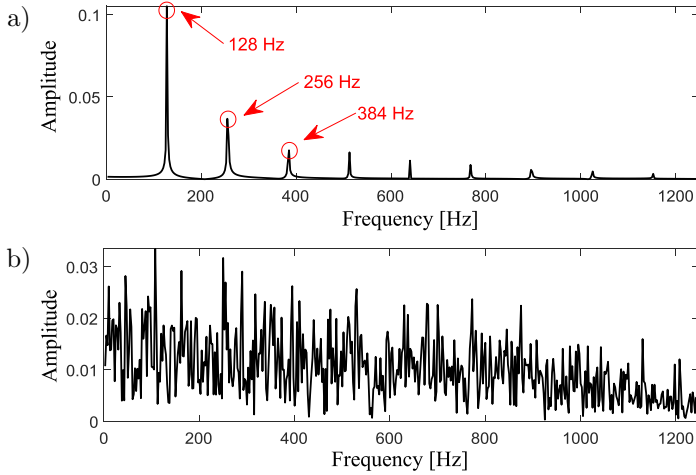
FIG. 8. a) Envelope spectrum of the original signal,
b) noise signal envelope spectrum (SNR = 10 dB).

The DEGWO-VMD algorithm is used to preprocess the signal. The relevant parameters of the DEGWO algorithm are set as follows: population size $N = 30$, maximum iteration times $t_{\max} = 20$, $[K, \alpha]$ search range $lb = [1, 1]$, $ub = [10, 4f_s]$, scale factor f upper bound is 0.8, lower bound is 0.2, crossover probability $P_{CR} = 0.2$, and convergence factor C_f decreases dynamically from 2 to 0 with the increase of iteration. By solving the DEGWO optimization algorithm, the optimal combination parameter of preset modal scale K and balance factor α is $[4, 500]$. According to this parameter, the VMD decomposition result is shown in Fig. 9. The normalized center frequencies of the decomposed IMFs are shown in Fig. 10. It can be seen in Fig. 10 that the center frequency spacing of each mode is moderate, and there is no mode aliasing.

As shown in Fig. 11, through the Hellinger value between two adjacent IMFs, the probability distance between IMF1 and IMF2 is the largest. Furthermore the distance between IMF2 and IMF3 is slightly different from the former. In contrast, the distance between IMF3 and IMF4 is the smallest, indicating that IMF2 and IMF3 are the boundary between noise signal and useful signal. The calculated correlation coefficient starts from IMF3. The correlation coefficient of the
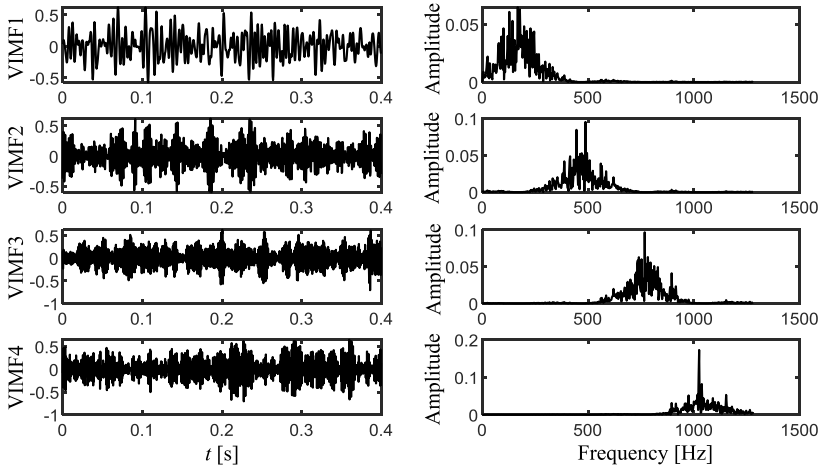
FIG. 9. Mode components of the VMD decomposition and their corresponding spectrum.
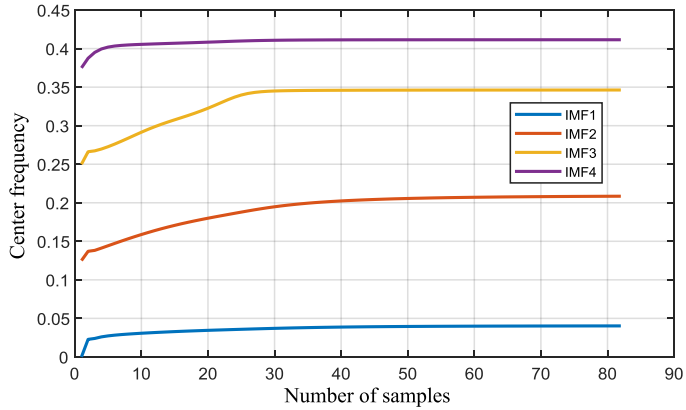

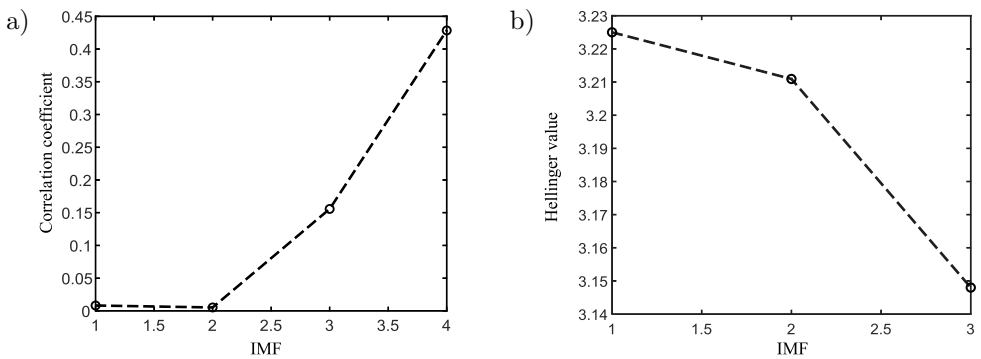FIG. 10. The normalized central frequency of each IMF.

FIG. 11. a) Correlation coefficient between each mode and the original signal, b) the Hellinger value between adjacent natural modes.

latter is large. The sub IMF3 and sub IMF4 are then used to reconstruct the signal. The envelope spectrum of the reconstructed signal is shown in Fig. 12a. The fault characteristic frequency of 128 Hz and the frequency doubling of 256 Hz and 384 Hz can be obtained from the reconstructed signal. The results show that the algorithm can effectively remove the noise signal and obtain the fault characteristic frequency of the signal.
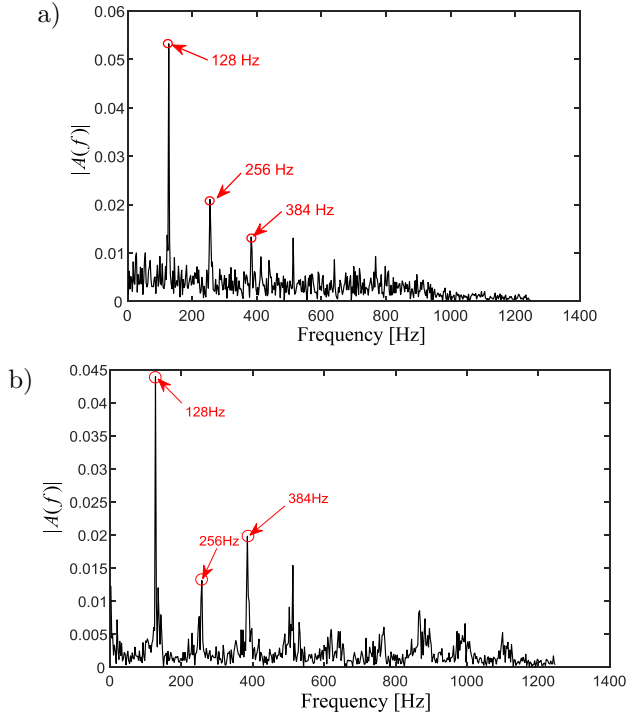

FIG. 12. a) Envelope spectrum of the improved DEGWO-VMD reconstructed signal, b) envelope spectrum of the original DEGWO-VMD reconstructed signal.

Figure 12b shows the reconstructed signal envelope spectrum of the original DEGWO-VMD algorithm (using the envelope entropy as the fitness function). The optimal combination parameter of the preset modal scale and balance factor is [8, 10240]. It can be seen in Fig. 12b that the pursuit of too small E_p leads to the excessive "over decomposition" of the optimal K value, resulting in false components.

For comparative analysis, the genetic algorithm (GA) and the particle swarm optimization (PSO) are used to optimize VMD parameters. The optimal combination parameters of the preset modal scale and balance factor obtained by GA are [4, 574]. The optimal combination parameters of the preset modal scale and balance factor obtained by PSO are [2, 103].

Figure 13 shows the envelope spectrum of the reconstructed vibration signal of the rolling bearing after being processed by the PSO-VMD and GA-VMD algorithms, respectively. Compared with Fig. 13, clear fault features of 128 Hz and its frequency doubling of 256 Hz and 384 Hz are obtained by this research method after de-noising. Only 128 Hz and 256 Hz characteristic frequencies are obtained by GA and PSO methods. The PSO optimization algorithm has a poor ability to suppress high-frequency noise. The GA optimization algorithm has a high noise component at high frequency, and the amplitude of characteristic frequency obtained by this research method is better than that of GA and PSO. The optimized method shows that this method can better preprocess the signal.
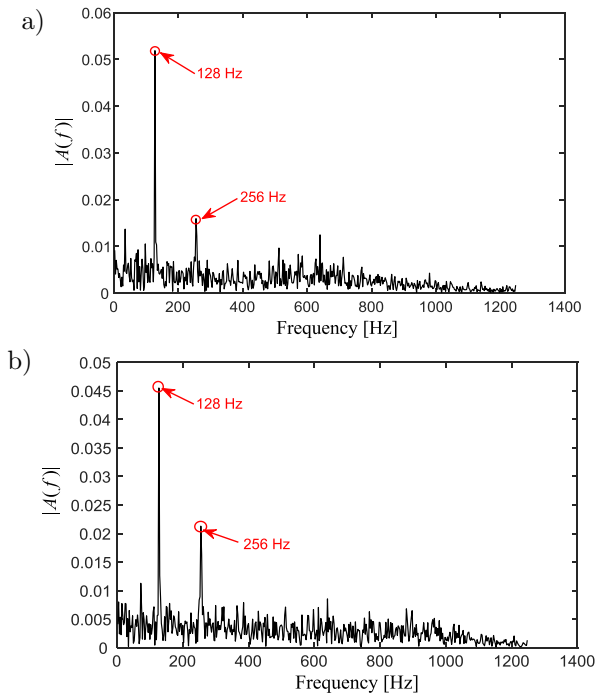

FIG. 13. a) Envelope spectrum of the PSO-VMD reconstructed signal, b) envelope spectrum of the GA-VMD reconstructed signal.

The evaluation indexes of different algorithms are listed in Table 3. They are mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and signal noise ratio (SNR):

$$(3.2) \qquad \mathrm{MAE} = \frac{1}{N} \sum_{t=1}^{N} |x(n) - y(n)|,$$

(3.3)
$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{x(n) - y(n)}{x(n)} \right| \cdot 100\%,$$

(3.4)
$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (x(n) - y(n))^2}.$$

Table 3. Comparison of evaluation indexes of different algorithms.

Algorithm		MAE	MAPE	RMSE	SNR [dB]
DEGWO-VMD	Improved	0.0721	1.5154	0.0900	9.1424
	Original	0.1325	10.4255	0.1662	3.6284
GA-VMD		0.1051	2.1900	0.1322	5.4303
PSO-VMD		0.0761	2.6395	0.0951	8.6785

It can be seen in Table 3 that the first four evaluation parameters of the improved DEGWO-VMD are smaller than the other three. SNR of the original fitness function is only 3.6284 dB, that of GA-VMD is only 5.4303 dB, and that of PSO-VMD is 8.6785 dB. The results show that the method of this study is better for the vibration signal preprocessing of the rolling bearing.

4. Test and analysis

4.1. Preprocessing and feature extraction of experimental data

Using the bearing database of Case Western Reserve University (CWRU) as the experimental data set, the effectiveness of the algorithm is verified by experiments [17, 18]. The experimental equipment is shown in Fig. 14. In this experiment, SKF-6205-2RS deep groove ball bearing is used. An electrical discharge machine (EDM) is used to arrange a single-point fault. The fault diameter
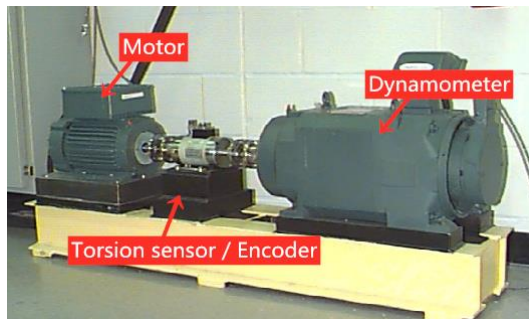

Fig. 14. Experimental equipment for the bearing fault signal acquisition.

is 0.007, 0.014, 0.021 inches and so on. Vibration signals are collected by 12 kHz and 48 kHz acceleration sensors.

To further prove the advantage of the proposed method, a Gaussian white noise with a signal-to-noise ratio of 18 dB is added to the sample data of the CWRU bearing database for experimental analysis.

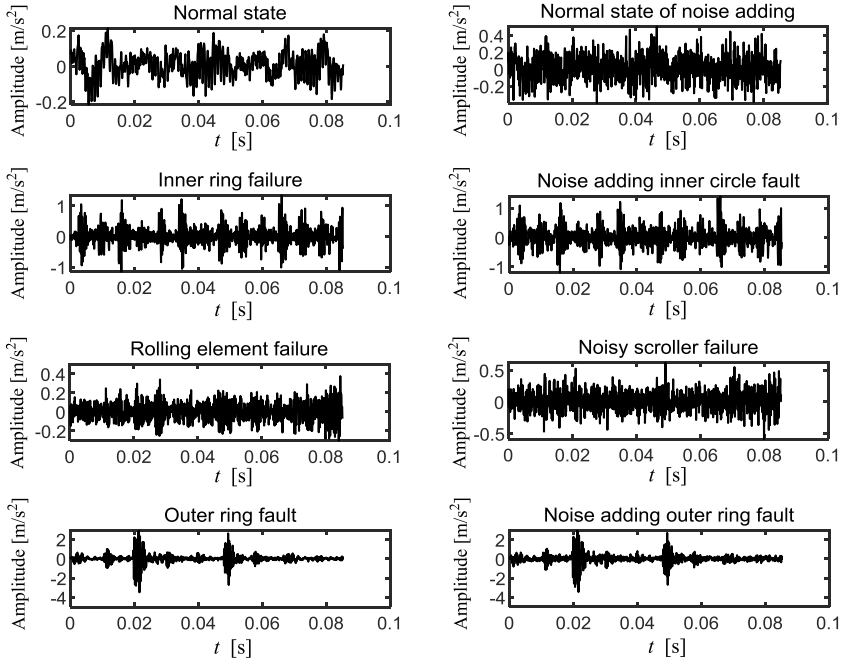The time-domain waveform of the data is shown in Fig. 15.

Fig. 15. The time-domain waveform of the signal.

4.1.1. CWRU signal preprocessing based on DEGWO-VMD. One thousand samples in the database were extracted as experimental data. Among them, groups 1–100 are in a normal state, 101–400 are inner ring fault data, 401–700 are rolling element fault, and 701–1000 are outer ring fault. The sample data is preprocessed by DEGWO-VMD, and the optimal parameter array is $[4, 1139]$. The normalized center frequency of each IMF is shown in Fig. 16. The correlation coefficient between each IMF and the original signal, and the Hellinger value between the two adjacent IMFs are shown in Fig. 17. Through the Hellinger value between two adjacent IMFs, it can be seen that there is an obvious inflection point in the probability distance of IMF2 and IMF3, indicating that this is the boundary between the noise signal and the useful signal. The calculated correlation coefficient starts from the IMF3. The latter IMF has a larger correlation coefficient. Then, sub IMF3 and sub IMF4 are used to reconstruct the signal.
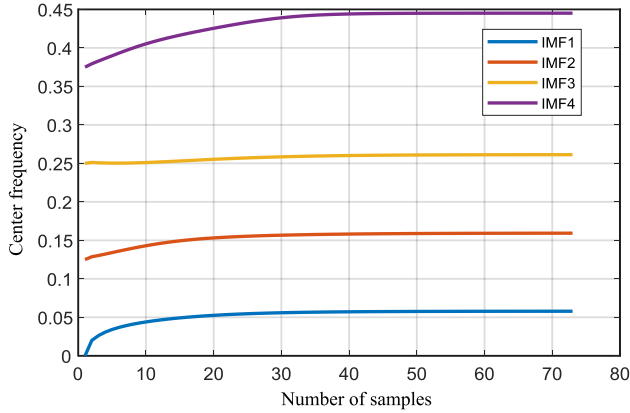
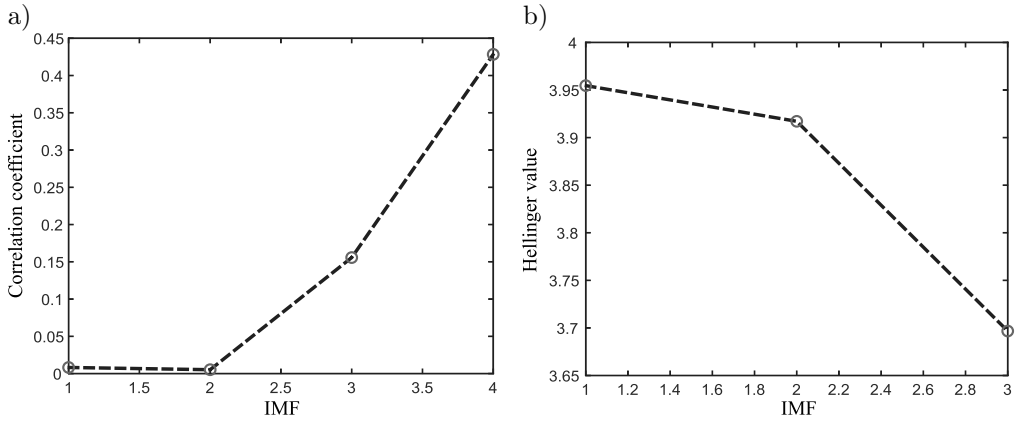Fig. 16. The normalized central frequency of each IMF.

Fig. 17. a) Correlation coefficient between each mode and the original signal, b) the Hellinger value between adjacent natural modes.

4.1.2. Feature extraction experiment of CWRU signal based on information entropy. Information entropy is introduced into the bearing fault diagnosis as it reflects the chaos of information inside the system, and to obtain effective entropy to reflect the bearing situation. In this paper, the multiscale permutation entropy, sample entropy and multiscale scatter entropy are selected to reflect the fault characteristic information.

(1) *Feature extraction of the rolling bearing vibration signal based on MPE:* Multiscale permutation entropy (MPE) is obtained by improving permutation entropy, which can quantify the fault characteristics of bearings by evaluating the randomness and complexity of sequences at multiple scales. Its core idea is: first, different time series are processed by multi-

scale coarse-grained; then, the permutation entropy of coarse-grained series is calculated respectively [19]. The specific steps are as follows:

(a) The time series $Y = \{y_i, \ i = 1, 2, ..., N\}$ with length n is coarse-grained, and the coarse-grained sequence is obtained as shown in Eq. (4.1):

$$(4.1) \qquad x_j^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} y_{i,j} = 1, 2, ... \left[\frac{N}{s}\right],$$

s is the scale factor, $[N/s]$ denotes rounding N/s.

(b) The phase space of each coarse-grained sequence of x_j^s is reconstructed.

$$(4.2) \qquad X_L^{(s)} = \left\{x_L^{(s)}, x_{L+\tau}^{(s)}, ..., x_{L+(m-1)\tau}^{(s)}\right\},$$

where L is the reconstructed component, $L = 1, 2, ..., N - (m-1)\tau$, τ is the delay response time, and m is the embedding dimension.

(c) The symbol sequence $S(r) = (j_1, j_2, ..., j_m)$ can be obtained by arranging the time reconstruction components in ascending order, where $r = 1, 2, ..., R$, and $m! \geq R$ The probability p_r of $S(r)$ is calculated.

(d) According to Eq. (4.3), the permutation entropy of each coarse-grained sequence is normalized to obtain the MPE value

$$(4.3) \qquad H_p = -\sum_{r=1}^{R} p_r \ln p_r.$$

(2) *Feature extraction of the rolling bearing vibration signal based on sample entropy:* Sample entropy measures the probability of multiple categories in the signal by analyzing the complexity of the time series. It has a strong anti-noise ability while mining the essential information of samples, and can get stable entropy value from small samples. The larger the value, the more complex the signal [20]. The specific process is as follows:

(a) First, an m-dimensional vector sequence with length n is constructed, as shown in Eq. (4.4):

$$(4.4) \qquad X(i) = [x_i, x_{i+1}, ..., x_{i+m-1}], \quad 1 \leq i \leq N - m.$$

(b) The maximum absolute difference between $x(i)$ and $x(j)$ elements is defined as the distance d_{ij}, as shown in Eq. (4.5):

$$(4.5) \qquad d_{ij} = \max\left[\|x(i+k) - x(j+k)\|\right], \quad i \neq j, \quad k \in (0, m-1).$$

(c) Given the threshold value of similarity tolerance parameter $r(r > 0)$, count the number A_i^m of $d_{ij} < r$ in (b), as shown in Eq. (4.6):

(4.6) $$A_i^m(r) = \frac{1}{N - m - 1} A_i^m.$$

(d) Calculate the average value $B_i^m(r)$ of $A_i^m(r)$:

(4.7) $$B_i^m(r) = \frac{1}{N - m - 1} \sum_{i=1}^{N-m} A(r)_i^m.$$

(e) Set the dimension to $m+1$ and recalculate Eq. (4.7) to get $B_i^{m+1}(r)$.

(f) The SampEn value can be obtained:

(4.8) $$\mathrm{SampEn}(m, r) = \left[-\ln \frac{B_i^{m+1}(r)}{B_i^m(r)} \right].$$

(3) *Feature extraction of the rolling bearing vibration signal based on composite multiscale dispersion entropy.* To analyze the complexity of the time series on more scales, DE is improved to obtain the composite multiscale dispersion entropy (CMDE) [21]. CMDE combines the information of multiple coarse-grained sequences to get more reliable entropy. The steps to calculate the CMDE value of the signal are as follows:

(a) The acquisition signal $x(t)$ with length N is mapped to $y(t)$ by normal distribution function processing:

(4.9) $$y(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x(t)} e^{\frac{-(t-mean)^2}{2\sigma^2}} \, \mathrm{d}t,$$

where *mean* is the average of $x(t)$, σ is the standard deviation.

(b) $y(t)$ is mapped to the set $z(t)$ from 1 to c by Eq. (4.10), where c is the number of classes:

(4.10) $$z^c(t) = \mathrm{round}\left[cy(t) + 0.5 \right].$$

(c) Given the embedding dimension m and delay d, the embedding vector $z_i^{m,c}$ is defined, as shown in Eq. (4.11):

(4.11) $$\begin{cases} z_i^{m,c} = \left\{ z_i^c, z_{i+d}^c, ..., z_{i+(m-1)d}^c \right\}, & i = 1, 2, ..., N - (m-1)d, \\ z_i^c = v_0, \quad z_{i+d}^c = v_1, ..., \quad z_{i+(m-1)d}^c = v_{m-1}. \end{cases}$$

The embedding vector $z_i^{m,c}$ is the scatter pattern $\pi_{v_1 v_2 ... v_{m-1}}$. Because each element in each $\pi_{v_1 v_2 ... v_{m-1}}$ has c values, there are a total of c^m dispersion patterns.

(d) Equation (4.12) is the probability expression of each dispersion mode:

$$(4.12) \qquad p\left(\pi_{v_1 v_2 \ldots v_{m-1}}\right) = \frac{\text{Number}\left(\pi_{v_1 v_2 \ldots v_{m-1}}\right)}{N - (m-1)\,d}.$$

Number $\left(\pi_{v_1 v_2 \ldots v_{m-1}}\right)$ is the number of $\pi_{v_1 v_2 \ldots v_{m-1}}$ mapped by $z_i^{m,c}$.

(e) DE is calculated:

$$(4.13) \quad \text{DE}(x, m, c, d) = -\sum_{\pi=1}^{c^m} p\left(\pi_{v_1 v_2 \ldots v_{m-1}}\right) \ln\left[p\left(\pi_{v_1 v_2 \ldots v_{m-1}}\right)\right].$$

(f) The value of CMDE, as shown in Eqs. (4.14) and (4.15), is:

$$(4.14) \qquad M_{K,j}^{\tau} = \frac{1}{\tau} \sum_{i=(j-1)\tau+K}^{j\tau+K-1} x_i, \quad 1 \le j \le \frac{N}{\tau}, \quad 1 \le K \le \tau,$$

$$(4.15) \qquad \text{CMDE} = \frac{1}{\tau} \sum_{K=1}^{\tau} \text{DE}\left[M_K^{\tau}, m, c, d\right].$$

A total of 1000 sample data constitute a 1000×23 matrix, the multiscale permutation entropy is 10 dimensions, the composite multiscale dispersion entropy is 12 dimensions, and the sample entropy is 1 dimension. The specific description is shown in Table 4.

Table 4. Description of sample data of the fault vibration signal.

Element	Fault type	Training sample number	Test sample size	Label
Inner ring	Normal state	1–100	23	1
	Slight failure	101–200	23	2
	Moderate failure	201–300	23	3
	Serious failure	301–400	23	4
Rolling element	Slight failure	401–500	23	5
	Medium failure	501–600	23	6
	Serious failure	601–700	23	7
Outer ring	Slight failure	701–800	23	8
	Medium failure	801–900	23	9
	Serious failure	901–1000	23	10

The principal component analysis (PCA) is used to reduce the dimension of the information entropy data set. After analysis, when it is reduced to five dimensions, the diagnosis effect is the best. The optimal characteristic parameters are composed of a 1000×5 matrix. The extracted parameter set is shown in Table 5.

Table 5. Description of sample data of fault vibration signal.

State	Sample	S1	S2	S3	S4	S5
Normal state	1	0.9958	0.9792	4.2201	4.0012	3.8575
	2	0.9935	0.9883	4.1673	4.0160	3.8811
	...					
	99	0.9968	0.9869	4.1447	3.9959	3.8556
	100	0.9945	0.9913	4.1847	4.0700	3.9436
Slight inner ring fault	101	0.9636	0.9814	3.5898	3.4762	3.2529
	102	0.9590	0.9688	3.5457	3.4068	3.1767
	...					
	199	0.9525	0.9823	3.4824	3.3898	3.0540
	200	0.9612	0.9777	3.5512	3.3994	3.0941
...						
Serious fault of rolling element	601	0.9794	0.9777	3.7292	3.6414	3.4873
	602	0.9867	0.9852	3.7217	3.6251	3.4587
	...					
	699	0.9829	0.9846	3.8138	3.6934	3.4800
	700	0.9765	0.9911	3.8002	3.6772	3.4863
...						
Serious outer ring fault	901	0.9723	0.9855	3.2202	3.1371	2.9869
	902	0.9769	0.9762	2.9817	3.0009	2.8123
	...					
	999	0.9718	0.9881	2.9345	2.9300	2.7402
	1000	0.9607	0.9752	2.8913	2.8653	2.6349

4.2. Fault diagnosis experiment based on nonlinear SVM

Support vector machine (SVM) has great advantages in small sample diagnosis [22]. Its principle is to find the optimal hyperplane and correctly divide different types of data samples. Because the fault signal of the rolling bearing is nonlinear, a nonlinear support vector machine is used. The kernel function is radial basis function (RBF): $F(x, x_i) = \exp\left(-\frac{\|x - x_i\|}{\sigma^2}\right)$. Therefore, the decision function of nonlinear support vector machine can be expressed as [23]:

$$(4.16) \qquad f(x) = \text{Sgn}\left(\sum_{k=1}^{n} a_i^* y_i F(x, x_i) + b^*\right).$$

The data in Table 5 is input into the SVM model for fault diagnosis, and 800×23 groups are selected as the training set, and the remaining 200×23 groups are selected as the test set. The diagnosis results and confusion matrix are shown in Fig. 18.
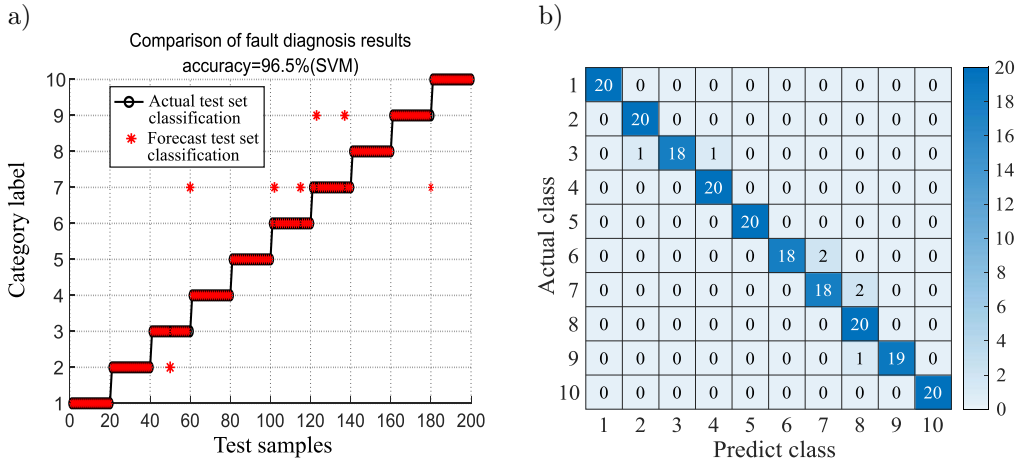
a) b)

Fig. 18. a) Model training results, b) model run confusion matrix.

The average test accuracy and computation time of the rolling bearing fault diagnosis method based on this paper are shown in Table 6. The diagnostic accuracy is 96.5%. The standard deviation is 2.8932. The running time is 35 s. The fault diagnosis accuracy of the rolling bearing is effectively improved, providing a better accuracy and effectiveness.

Table 6. Average testing accuracy and computation time comparison of the method in the experiment.

Methods	Average testing accuracy [%]	Standard deviation of testing accuracy	Average computation time [s]
Information entropy	96.5	2.8932	35

5. Conclusion and future work

In this paper, an improved DEGWO-VMD method was proposed to find the optimal parameter combination of mode size k and balance factor α to solve the problem of mode mixing and virtual component in the VMD algorithm. To verify the advantage of this method, the information entropy was extracted as the feature set. The feature matrix was input into SVM for fault diagnosis.

Using the simulation data, the preprocessing effects of PSO-VMD, GA-VMD and DEGWO-VMD before and after the improved fitness function were compared qualitatively and quantitatively. The results showed that the improved DEGWO-VMD algorithm is the most effective.

Through the fault diagnosis experiment analysis of the bearing database sample data of CWRU, the fault diagnosis method proposed in this paper achieved

96.5% accuracy, which effectively improved the rolling bearing fault diagnosis accuracy.

In future work, the improved DEGWO-VMD fault diagnosis model will be optimized to evaluate the fault degree.

Acknowledgments

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

References

1. Meng Z., Li J., Yin N., Pan Z., Remaining useful life prediction of rolling bearing using fractal theory, *Measurement*, **156**: 107572, 2020, doi: 10.1016/j.measurement.2020.107572.

2. Zhao M., Tang B., Tan Q., Bearing remaining useful life estimation based on time–frequency representation and supervised dimensionality reduction, *Measurement*, **86**: 41–55, 2016, doi: 10.1016/j.measurement.2015.11.047.

3. Wen-Hsiang H., Braun S.G., Editorial Statement: Prognostics and system health management for electromechanical systems, *Mechanical Systems and Signal Processing*, **113**: 1–4, 2018, doi: 10.1016/j.ymssp.2018.05.065.

4. Mishra C., Samantaray A.K., Chakraborty G., Rolling element bearing defect diagnosis under variable speed operation through angle synchronous averaging of wavelet denoised estimate, *Mechanical Systems and Signal Processing*, **72–73**: 206–222, 2016, doi: 10.1016/j.ymssp.2015.10.019.

5. Cheng J., Yang Y., Yang Y., A rotating machinery fault diagnosis method based on local mean decomposition, *Digital Signal Processing*, **22**(2): 356–366, 2012, doi: 10.1016/j.dsp.2011.09.008.

6. Dragomiretskiy K., Zosso D., Variational mode decomposition, *IEEE Transactions on Signal Processing*, **62**(3): 531–544, 2014, doi: 10.1109/TSP.2013.2288675.

7. Li X., Ma Z., Kang D., Li X., Fault diagnosis for rolling bearing based on VMD-FRFT, *Measurement*, **155**: 107554, 2020, doi: 10.1016/j.measurement.2020.107554.

8. He Z., Chen G., Hao T., Teng C., Hou M., Cheng Z., Weak fault detection method of rolling bearing based on testing signal far away from fault source, *Journal of Mechanical Science and Technology*, **34**(3): 1035–1048, 2020, doi: 10.1007/s12206-020-0206-4.

9. ZHANG M., JIANG Z., FENG K., Research on variational mode decomposition in rolling bearings fault diagnosis of the multistage centrifugal pump, *Mechanical Systems and Signal Processing*, **93**: 460–493, 2017, doi: 10.1016/j.ymssp.2017.02.013.

10. ANEESH C., KUMAR S., HISHAM P.M., SOMAN K.P., Performance comparison of variational mode decomposition over empirical wavelet transform for the classification of power quality disturbances using support vector machine, *Procedia Computer Science*, **46**: 372–380, 2015, doi: 10.1016/j.procs.2015.02.033.

11. BI F., LI X., LIU C., TIAN C., MA T., YANG X., Knock detection based on the optimized variational mode decomposition, *Measurement*, **140**: 1–13, 2019, doi: 10.1016/j.measurement.2019.03.042.

12. LI F., LI R., TIAN L., CHEN L., LIU J., Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions, *Mechanical Systems and Signal Processing*, **116**: 462–479, 2019, doi: 10.1016/j.ymssp.2018.06.055.

13. SHAHEEN M.A.M., HASANIEN H.M., ALKUHAYLI A., A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution, *Ain Shams Engineering Journal*, **12**(1): 621–630, 2021, doi: 10.1016/j.asej.2020.07.011.

14. STORN R., PRICE K., Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, **11**(4): 341–359, 1997, doi: 10.1023/A:1008202821328.

15. NAN G., TANG M., CHEN E., YANG A., Nonlinear dynamic mechanism of rolling element bearings with an internal clearance in a rotor-bearing system, *Advances in Mechanical Engineering*, **8**(11): 1687814016679588, 2016, doi: 10.1177/1687814016679588.

16. TANG G.J., WANG X.L., Application of parameter optimization variational modal decomposition method in early fault diagnosis of rolling bearing, *Journal of Xi'an JiaoTong University*, **49**(5): 73–81, 2015.

17. *Case Western Reserve University. Bearing data center seeded fault test data*, https://engineering.case.edu/bearingdatacenter/download-data-file.

18. 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 Transactions*, **114**: 470–484, 2021, doi: 10.1016/j.isatra.2020.12.054.

19. LI Y., GAO Q., MIAO B., ZHANG W., LIU J., ZHU Y., Application of the refined multiscale permutation entropy method to fault detection of rolling bearing, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **43**(5): 1–14, 2021, doi: 10.1007/s40430-021-02986-7.

20. GUO X., SHEN Y., YANG S., Application of sample entropy and Fractional Fourier transform in the fault diagnosis of rolling bearings [in Chinese], *Journal of Vibration and Shock*, **36**(18): 65–69, 2017, doi: 10.13465/j.cnki.jvs.2017.18.010.

21. ZHANG W., ZHOU J., A comprehensive fault diagnosis method for rolling bearings based on refined composite multiscale dispersion entropy and fast ensemble empirical mode decomposition, *Entropy*, **21**(7): 680, 2019, doi: 10.3390/e21070680.

22. LEE Y.E., KIM B.-K., BAE J.-H., KIM K.C., Misalignment detection of a rotating machine shaft using a support vector machine learning algorithm, *International Journal of*

Precision Engineering and Manufacturing, **22**(3): 409–416, 2021, doi: 10.1007/s12541-020-00462-1.

23. DASH Ch.S.K., SAHOO P., DEHURI S., CHO S.-B., An empirical analysis of evolved radial basis function networks and support vector machines with mixture of kernel, *International Journal on Artificial Intelligence Tools*, **24**(4): 1550013, 2015, doi: 10.1142/s021821301550013x.