



A novel intelligent approach based on WOAGWO-VMD and MPA-LSSVM for diagnosis of bearing faults

Ahmed Taibi¹ · Nabil Ikhlef¹ · Said Touati²

Received: 12 October 2021 / Accepted: 28 January 2022 / Published online: 10 March 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

The bearing fault diagnosis plays an important role to reduce catastrophic failures and ensure the continuity of running machines to avoid heavy economic loss. The vibration signals of rolling bearings are often nonlinear and nonstationary; it is difficult to extract sensitive features and diagnose faults by traditional signal processing methods. To solve this problem, a novel intelligent fault-diagnosis approach based on whale optimization algorithm grey wolf optimization-variational mode decomposition (WOAGWO-VMD) algorithm and the marine predators algorithm optimization-least squares support vector machine (MPA-LSSVM) is proposed in this paper. Firstly, hybrid algorithm WOAGWO is used to optimize the parameters of VMD and obtain the optimal combination ($K; \alpha$). Then, the optimized VMD algorithm is utilized to decompose the vibration signal of the rolling bearing into several intrinsic mode functions, and a new sensitive indicator is created to select the components containing the most information. For these components, the dispersion entropy feature, permutation entropy feature, and singular value feature are extracted to form the multi-feature vector. Finally, the feature vectors obtained are input to the MPA-LSSVM for diagnosis and identification. The validity and strength of the proposed method is verified by experimental data under different bearing conditions. The results have shown that the proposed method can extract the fault feature information of 16 bearing signals of different fault types effectively and identify them accurately.

Keywords Fault diagnosis · Feature extraction · WOAGWO · Variational mode decomposition · MPA · LSSVM

1 Introduction

Rotating machines are one of the necessary equipment widely used in modern industry as they represent the mainstay of various vital industries such as nuclear power plants, petrochemical plants, and others, and due to their great importance, the continuity of the work of

these machines is necessary, and despite their reliability, their work regularly under unstable conditions, such as variable rotational speed, overload, and unsteady thermal condition, makes it vulnerable to several failures. It can lead to massive financial and economic losses or even human life safety problems if left undetected [1, 2]. One of the main causes responsible for stopping rotating machines is bearings defects, because they are often exposed to rigorous environments during long-term operation. According to many studies, more than 45% of rotating machinery failures are due to bearing failures [3]. Therefore, it is crucial to identify the fault state, timely, necessarily and accurately to ensure the safe and continuous operation of machinery. The diagnosis of bearing faults has become the concern of many researchers from various fields, where, at the present time, it has formed a multi-disciplinary research structure. Its detection and diagnostic technologies mainly include vibration diagnostic technology, acoustic diagnostic technology, and temperature diagnostic

Nabil Ikhlef and Said Touati contributed equally to this work.

✉ Ahmed Taibi
taibi.ahmed@univ-jijel.dz

Nabil Ikhlef
ikhlefnaabil@yahoo.fr

Said Touati
saidtouati@yahoo.fr

¹ L2EI Laboratory, Jijel University, BP 98 Ouled Aïssa 18000, Jijel, Algeria

² Electrical Engineering Department (DGE), Nuclear Research Center of Birine, CRNB) Bp 180, Ain Oussera, Algeria

technology [4]. The combination of these technologies can improve diagnosis accuracy. Currently, vibration signal analysis (vibration diagnosis) is one of the most widely used techniques. According to Mohd Ghazali and Rahiman [5], the vibrational diagnosis technique represents more than 82% of the overall other techniques used in fault diagnosis. Due to the complex working conditions that bearings are subjected to, the vibration signal will inevitably be nonlinear and non-stationary and contain noise, which considerably increases the difficulties in diagnosing bearing faults [6, 7]. In the field of rolling bearing fault vibration signal processing, there are some popular signal processing methods for example wavelet transform (WT) [8], Hilbert-Huang transform (HHT) [9], empirical mode decomposition (EMD) [10], ensemble empirical mode decomposition (EEMD) [11, 12], complete ensemble empirical mode decomposition adaptive noise (CEEMDAN) [13], and local mean decomposition (LMD) [14]. Although these methods gave some results, there are some problems, where WT needs to select the wavelet basis and decomposition level, which partially limits their application, while HHT's limitation is energy leakage produced by the endpoint effect and unexpected negative frequency [15]. As for EMD, EEMD, and LMD methods, they have some adverse effects such as mode mixing and endpoint effect [16]. To solve the problems found in the aforementioned methods such as mode mixing and endpoint effect, Dragomiretskiy [17] proposed a non-iterative and adaptive signal processing method called variational mode decomposition (VMD) that has a solid mathematical basis, where it can adaptively decompose the vibration signal into several narrow-band signals of different frequencies. Marco Civera derived an effective comparison between the VMD, EMD, and CEEMDAN for the vibration-based structural health monitoring purposes and denoted the feasibility of VMD over the later methods [18]. Ye et al. [19] used VMD to decompose the original bearing vibration signal into several intrinsic mode functions (IMFs) and used the feature energy ratio to reconstruct the bearing vibration signal, then the multiscale permutation entropy is calculated to construct multidimensional feature vectors. And then the PSO-SVM model optimized is used in classification and identification of different faults of the rolling bearing. Fu et al. [20] proposed a new hybrid approach, coupling VMD algorithm, composite multiscale fine-sorted dispersion entropy (CMFSDE), and support vector machine (SVM) for fault diagnosis of rolling bearings. The original signal was decomposed into several IMFs by VMD,

and then calculated the CMFSDE value of each IMF to form fault features. Finally, these features are used as input to SVM classifier optimized by MSCAHHO to identify the fault types. The previous studies have proved the effectiveness of the VMD algorithm in signal decomposition. However, the selection of the main two parameters, the decomposition level K and the penalty factor α , has always limited the improvement of the performance of VMD. Nowadays, with the development of smart algorithms, many researchers have proposed algorithms to solve this problem. Li et al. [21] utilized the genetic algorithm (GA) to optimize VMD parameters and used the sample entropy as the fitness function. Yao et al. [22] proposed a hybrid gearbox fault diagnosis method based on GWO-VMD and DE-KELM. Where they used the grey wolf optimizer (GWO) to optimize the parameters of the VMD in order to eliminate effectively the noise present in the vibration signals. Zhang et al. [23] proposed a method to select the optimal parameters $[k, \alpha]$ in the VMD algorithm based on improved particle swarm optimization and the envelope entropy value as the fitness function. Feng et al. [24] applied the whale optimization algorithm (WOA) to optimize the parameters of VMD in order to reduce noise and obtain adaptive decomposition of vibration signals.

After decomposing the signal into several IMFs, the next procedure is to extract fault features from IMFs. With the evolution of artificial intelligence, bearing fault diagnosis is highly treated as a category of pattern recognition. Its feasibility and reliability largely related to the choice of feature vector [25]. Zhang et al. [26] decomposed the experiment gear fault data into several IMFs by the adaptive local iterative filtering, and calculated the permutation entropy of each IMF as the fault feature vector. Jin et al. [27] proposed a fault diagnosis method of train rolling bearing based on VMD to decompose the original signal into several intrinsic mode components (IMF) and calculated the distribution entropy of each component as the feature vector. Cheng et al. [28] proposed a bearing fault diagnosis method based on VMD and singular value feature (SVD) for feature extraction and G-G fuzzy clustering for fault identification. Therefore, we consider in this paper the multi-features based on dispersion entropy feature (DE), permutation entropy feature (PE), and SVD as the fault feature vector of rolling bearing vibration signal.

Obviously, the key step after feature extraction is to use an effective classifier for fault identification. Currently and with the continuous development of machine learning and deep learning, there are several classifiers

that have been presented and used for fault diagnosis in rotating machines, such as artificial neural network (ANN) [29], extreme learning machines (ELMs) [30], SVM [31, 32], and deep learning [33]. In addition to these methods, we mention least-square support vector machines (LSSVM). The LSSVM can successfully process non-linear data. Hence, it was effectively investigated by many literatures in diagnosing and fault classification caused by bearings [34], and gears [35]. However, the accuracy of the LSSVM classification is highly sensitive to the selection of the penalty parameter C and kernel parameter g . To determine the values of the optimal parameters, a combination of intelligent optimization algorithms and LSSVM can be used. There are many smart optimization algorithms, including GA, particle swarm optimization (PSO), WOA, and GWO. Gao et al. [36] used the PSO to optimize LSSVM and applied it to fault classification of rolling bearings. Hanoi University of Industry et al. [37] proposed an automatic fault diagnosis approach for rolling bearing based on EEMD-LaS and Optimized Classifier BSO-LSSVM. Although the previously mentioned intelligent algorithms have been successful in improving the LSSVM classifier parameter, they still have some limitations, where it is not accurate at finding the global optimum parameter optimization, and this is confirmed by the theory of no free lunch [38]. To solve this defect, novel meta-heuristic optimization algorithm called MPA has been applied. In this paper an, intelligent model was developed for fault diagnosis of rolling bearing. This model was based firstly on hybrid algorithm WOAGWO to optimize the parameters of VMD. Then, the optimized VMD algorithm is used to decompose the vibration signal into several IMFs. Secondly, a new sensitive indicator (SI) is created to define the components containing the most information and calculate DE, PE, and SVD of each IMF to form the multi-feature vector. Finally, we used the marine predator algorithm MPA to optimize the parameters of classifier LSSVM; hence, the optimized model is used to classify and identify the different fault types of rolling bearing.

This paper is organized as follows. Section 2 presents the techniques used for fault feature extraction, including WOAGWO-VMD algorithm, and SI. Section 3 elaborates the fault feature classification using MPA-LSSVM. Section 4 describes the proposed intelligent diagnosis method based on WOAGWO-VMD and MPA-LSSVM. Section 5 presents the experimental analysis to check the validity of the proposed method for rolling bearing fault diagnosis; finally, the conclusion of this paper is given in Sect. 6.

2 Fault feature extraction

2.1 Variational mode decomposition

VMD is a novel signal processing method, which is adopted to decompose the original signal into several IMFs. In VMD algorithm, the vibration signal is adaptively decomposed into several IMFs in order to find the variational problem. The variational problem can be expressed as follows:

$$\begin{cases} \min \left\{ \sum_k \left\| \partial_t \left[\delta(t) + \frac{j}{\pi t} * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k u_k(t) = f(t) \quad k = 1, 2, \dots, K \end{cases} \quad (1)$$

where u_k is the k th IMF and ω_k is the bandwidth center frequency. K denotes the decomposition number. $f(t)$ is the original signal. ∂_t is the partial derivative function and $\delta(t)$ is the Dirichlet function. In order to solve the variational problem, the parameters of penalty factor α and the Lagrange multiplication operator $\lambda(t)$ are introduced to transform the constrained variational problem into an unconstrained problem as follows:

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

where α is the penalty parameter; λ is the Lagrange multiplier.

To find the optimal solution of Eq. (2), the alternating direction method of multipliers is utilized [16]. The implementation steps for the algorithm are as follows:

- Step 1. Perform an iterative loop $n = n + 1$;
- Step 2. Update $\{\hat{u}_k(\omega)\}$ for all $\omega \geq 0$;

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

- Step 3. Update the modal center frequency $\{\hat{\omega}_k\}$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \quad (4)$$

- Step 4. Update Lagrange multiplication operator $\hat{\lambda}(\omega)$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_K \hat{u}_K^{n+1}(\omega) \right) \quad (5)$$

Step 5. Repeat steps 1–4 until the iteration stop condition is satisfied

$$\sum_K \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \varepsilon \quad (6)$$

Step 6. End.

where ε represents the tolerance of convergence criterion.

2.2 WOAGWO algorithm

2.2.1 Whale optimization algorithm

WOA is a new optimization algorithm introduced in metaheuristic algorithms by Mirjalili and Lewis. WOA simulates humpback whale hunting behavior [39]. These whales generally prefer to hunt small fish near the surface of the sea, where they use a special fishing technique called the bubble net feeding technique. WOA includes three types of mathematical behavioral simulations, which are as follows:

(a) Encircling prey

In order to begin the hunt, humpback whales locate the prey and then surround it, and this behavior can be represented by the following formulas:

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (7)$$

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (8)$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r} + \vec{a} \quad (9)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (10)$$

where $\vec{X}^*(t)$ is the best position of search agent (whale) obtained so far, t is the current iteration, and $\vec{X}(t)$ is the present position of search agent at iteration t . \vec{A} and \vec{C} are coefficient vectors. \vec{r} is a random number in $[0, 1]$. \vec{a} is linearly decreased from 2 to 0.

(b) Bubble-net attacking

This hunting method includes the following two behaviors: reducing the ring and continuing to surround the prey. These two behaviors are represented by the following equation:

$$\vec{X}(t+1) = \vec{D}^* \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) \quad (11)$$

$$\vec{D}^* = |\vec{X}^*(t) - \vec{X}(t)| \quad (12)$$

where b is a constant value that identifies the logarithmic spiral shape, l is a random number in the range $[-1, 1]$, and \vec{D}^* represents the distance between the whale and prey. Humpback whales rotate around their prey during predation and shrink their range, so each behavior has a 50% chance. It is expressed mathematically as follows:

$$\vec{X}(t+1) = \begin{cases} \vec{X}^* - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}^* \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (13)$$

where p is an arbitrary number between $[0$ and $1]$.

(iii) Search for prey

Exploration step: At this point, the humpback whales search each other's positions at random. They are represented mathematically as follows:

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \quad (14)$$

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (15)$$

\vec{X}_{rand} is the random whales in current iteration.

2.2.2 Grey wolf optimization

GWO was proposed by Mirjalili [40], an optimization algorithm that simulates the hunting of grey wolves in wildlife. This algorithm was inspired by the social hierarchy of grey wolves as these wolves are categorized into four classes: (α) wolf leader, (β) helping the leader, (δ) follows both previous wolves, and omega (ω).

(a) Social hierarchy

To emulate the social hierarchy of grey wolves, the fittest solution is regarded as α then β , and the third best solution is δ , and the remaining candidate solutions are considered ω .

(b) Encircling prey

The grey wolves encircle the prey in order to hunt, according to the following equation:

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (16)$$

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (17)$$

where t is the current number of iterations, $\vec{X}_p(t)$ represents the position vector of the prey, and $\vec{X}(t)$ is the position vector of a grey wolf. \vec{A} and \vec{C} are coefficient vectors that are calculated from the following equations:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{18}$$

$$\vec{C} = 2 \cdot \vec{r}_2 \tag{19}$$

where \vec{a} decreases linearly from 2 to 0 over the course of iterations, and \vec{r}_1 and \vec{r}_2 are random vectors between [0, 1].

(iii) Hunting

After the encircling process, a grey wolf begins to hunt for the best solution. Despite the reality that the best solution must be optimized, alpha wolf saves the best solution in each iteration and updates it if the solution is improved. Beta and delta can be used to identify the location of the prey. Thus, the best solutions are stored by each type of grey wolf and employed to update the position of grey wolves using the equations below.

$$\begin{cases} \vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right| \\ \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right| \\ \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \end{cases} \tag{20}$$

$$\begin{cases} \vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \\ \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \\ \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \end{cases} \tag{21}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{22}$$

(iv) Attacking prey (exploitation)

In this step, a grey wolf can perform hunting mechanism to try to stop the movement of the prey for attack them. This mechanism works by decreasing the value of \vec{a} , where the value of \vec{A} is also decreased by the value \vec{a} , resulting in a value in the range [-1, 1]. A grey wolf can be attacking the prey, if \vec{A} is less than -1 or greater than 1. However, the GWO is prone to stagnation in local solutions. Therefore, the researchers are attempting to discover various mechanisms for resolving this issue.

(e) Search for prey (exploration)

The searching mechanism is influenced by alpha, beta, and delta. These three categories are distinct from one

another. As a result, a mathematical equation is required for them to converge and attack the prey. Thus, the value of \vec{A} must be between 1 and -1; if its value is greater than 1 or less than -1, this forces the search agents to diverge from the prey. Furthermore, if \vec{A} is greater than 1, the search agent will try to find a better prey. \vec{C} is another component factor that affects the exploration phase of GWO. Overall, the GWO algorithm generates a random population. The prey’s location is assumed by alpha, beta, and delta. The distance between candidate solutions is then updated. After that, a is reduced from 2 to 0 in order to achieve a balance between the two phases. After that, if $\vec{A} > 1$, the search agents then move away from attacking the prey and if $\vec{A} < 1$ they go forward the prey. Finally, the GWO reached a satisfactory result and was terminated.

2.2.3 WOAGWO

WOA is a recently implemented optimization algorithm to solve many optimization problems. Standard WOA may work well for the best solution. However, refining the optimal solution with each iteration is not enough. To solve this limitation, we introduce an algorithm called WOAGWO, proposed by Mohammed H [41]; this algorithm is a hybrid between the WOA and GWO algorithm in order to improve the performances of WOA and to obtain better solutions. The WOA algorithm is hybridized by adding two sections. Firstly, to improve the hunting mechanism, a condition has been added in the exploitation phase in WOA according to Eq. (22). The effects of A1, A2, and A3 on exploitation performance are greater. As a result, a new condition has been added to WOA’s standard exploitation phase to bypass local optima where each A is greater than -1 or less than 1. Secondly, Eqs. (20), (21) and (22) have been adapted, and used within the conditions that have been added to the exploitation phase which contains A1, A2, and A3. Finally, to make the current solution progress toward the best solution and prevent the whale from changing to a position that is not better than the previous position as well, for these purposes, another new condition has been added to the exploration phase. The pseudocodes of WOAGWO are presented in Algorithm 2.

2.2.4 Hybrid WOA with GWO

Pseudocodes of WOAGWO algorithm:

Algorithm 2 WOAGWO

```

Initialize WOAGWO population  $X_i$  where  $(i=1,2,3,\dots,n_0)$ 
Evaluate the fitness function for each search agent,  $X^*$ = the best search agent
While (iteration < Max iterations )
  For each search agent: update  $a$ ,  $A$ ,  $C$ ,  $L$ , and  $p$  for WOA, and update  $A_1$ ,  $C_1$ ,  $A_2$ ,  $C_2$ ,  $A_3$ ,  $C_3$  for GWO
    If1 ( $p < 0.5$ )
      If2 ( $|A| < 1$ ) calculate new position of the present search agent by Eq.7
        If3 ( $f_{\text{current}} < f_{\text{previous}}$ )
          Position=new-position
        End if3
      Else if2 ( $|A| > 1$ )
        Randomly choose search agent ( $X_{\text{rand}}$ ) Change the position of the present search agent by Eq.(14)
        If4 ( $f_{\text{current}} < f_{\text{previous}}$ )
          Position = new-position
        End if4
      End if2
    Else if1 ( $p \geq 0.5$ )
      If5 ( $(|A_1| > 1 \parallel |A_1| < 1) \&\& (|A_2| > 1 \parallel |A_2| < 1) \&\& (|A_3| > 1 \parallel |A_3| < 1)$ )
        Position of the current search agent update by Eq.(22)
      End if5
    End if1
  End for
  Return search agents to inside the search space if it goes beyond the search space fitness value for each search agent is calculated.
  Update  $X^*$  if there is a better solution.
  Iteration =iteration +1
End while
Return  $X^*$ .

```

2.3 Parameter adaptive optimization of VMD method based on WOAGWO

It is noteworthy that the parameter settings $[k, \alpha]$ of VMD need to be set in advance when decomposing signals due to their high impact on results [42], where alpha is directly proportional to bandwidth since smaller alpha gives a smaller bandwidth and vice versa; on the other hand, smaller K will give raise to mode aliasing and if K is too large it will lead to useless component generation [46]. In this paper, after tremendous trials, and investigating multiple manuscripts [21, 22, 42–44], we finally chose a reasonable range of α which is [200, 4000], and the selection range of K is [2, 12]. Therefore, in this section, we are introducing the hybrid algorithm WOAGWO to optimize the parameters of VMD. The WOAGWO-VMD algorithm needs to define the fitness function. According to reference [43], information entropy is an eminent indicator for judging signal sparseness, where the higher the entropy value, the higher the noise content of the signal, while the smaller indicates that the signal contains more fault information [44]. In this paper, we consider the average of the weighted permutation entropy of modes obtained by VMD as an objective function of the WOAGWO optimization algorithm, as shown in Fig. 1. The purpose

of the parameter optimization process of VMD using WOAGWO is clearly interpreted as an efficient search for the minimum value of the objective function, as illustrated in Eq. (23):

$$\begin{cases} \text{fitness} = \min \left(\frac{1}{k} \sum_{i=1}^k \text{wpe}(i) \right) \\ \text{s.t } K = [2, 12], \alpha = [200, 4000] \end{cases} \quad (23)$$

where $\text{wpe}(i)$ is the weighted permutation entropy of the i th IMF component. K and α denote the mode number and the penalty factor respectively. The details of $\text{wpe}(i)$ is shown in [45], and its parameters used in this paper are embedding dimension $m=4$ and time delay $\tau=1$, and the parameters of WOAGWO are population size = 20 and maximum iteration = 15, and its initial parameters are shown in Table 1. The flowchart of WOAGWO-VMD algorithm is shown in Fig. 2, and its steps are as follows:

- Step 1:** Set the parameters of the WOAGWO algorithm, including the maximum number of iterations, the number of search agents, and the iteration range of K and α , and the parameters of the VMD algorithm also should be set.
- Step 2:** Initialize the search agents.

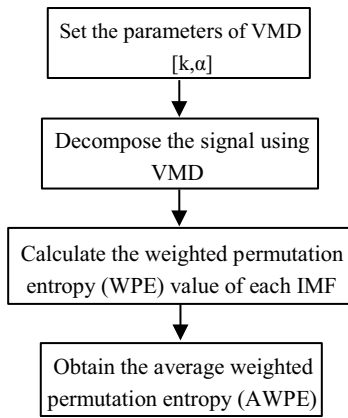


Fig. 1 Objective function of WOAGWO algorithm

Step 3: Calculate the objective function of each search agent and choose the search agent with minimum objective function as the initial best search

Step 4: To update the best search agent execute an iteration loop within the maximum number of iterations

- For each search agent update, the values of a , A , C , l , and p of WOA and $A1$, $C1$, $A2$, $C2$, $A3$, and $C3$ of GWO.
- Select the corresponding update formula to update the appropriate search agent to the different values of p and A .
- Calculate the updated objective function of each agent and select the best one to retain.

Step 5: At the end of iterations, the best search agent is the best parameter combination (K , α).

2.4 Simulation signal analysis

In order to validate the effectiveness of the WOAGWO-VMD algorithm proposed in this paper, it is applied to decompose the simulation signal mentioned in the literature [43]. The simulation signal is shown as follows:

The time domain waveform of the simulated signal $y(t)$ and its four components ($y1(t)$, $y2(t)$, $y3(t)$, and $y4(t)$) are shown as follows:

Table 1 Parameter setting

Algorithm	Parameter	Value	Reference
WOAGWO	a	2 to 0	[24]
	b	1	
	l	$[-1, 1]$	
GWO	r	$[0, 1]$	[41]
	a	2 to 0	
	r_1, r_2	$[0, 1]$	
MPA	P	0.5	[50]
	FADs	0.2	

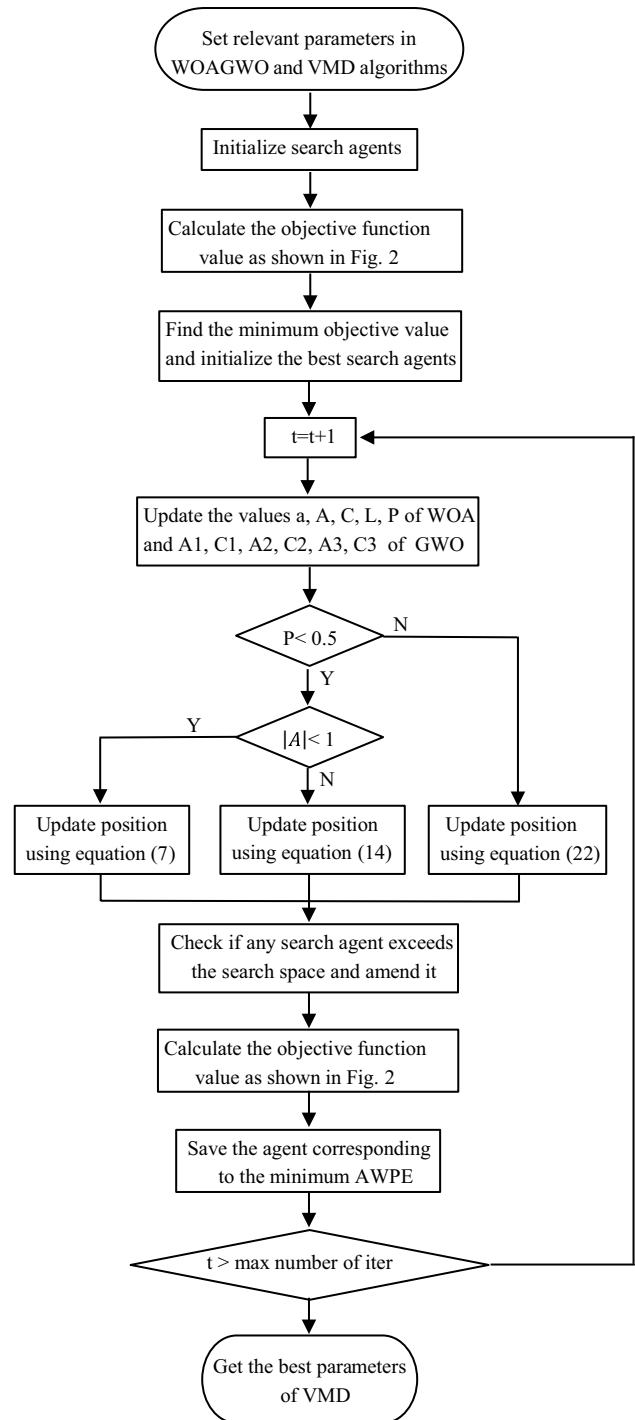


Fig. 2 Flowchart of WOAGWO-VMD

$$\begin{cases}
 y1(t) = 5 \times \sin(2\pi \times 50 \times t) \\
 y2(t) = 4 \times \cos(2\pi \times 100 \times t) \\
 y3(t) = \begin{cases} 3 \times \sin(2\pi \times 300 \times t) & t \in [[0,0.1], [0.3,0.4], [0.6,0.7], [0.9,1]] \\ 0 & t \in [[0.1,0.3], [0.4,0.6], [0.7,0.9]] \end{cases} \\
 y(t) = y1(t) + y2(t) + y3(t) + y4(t)
 \end{cases} \tag{24}$$

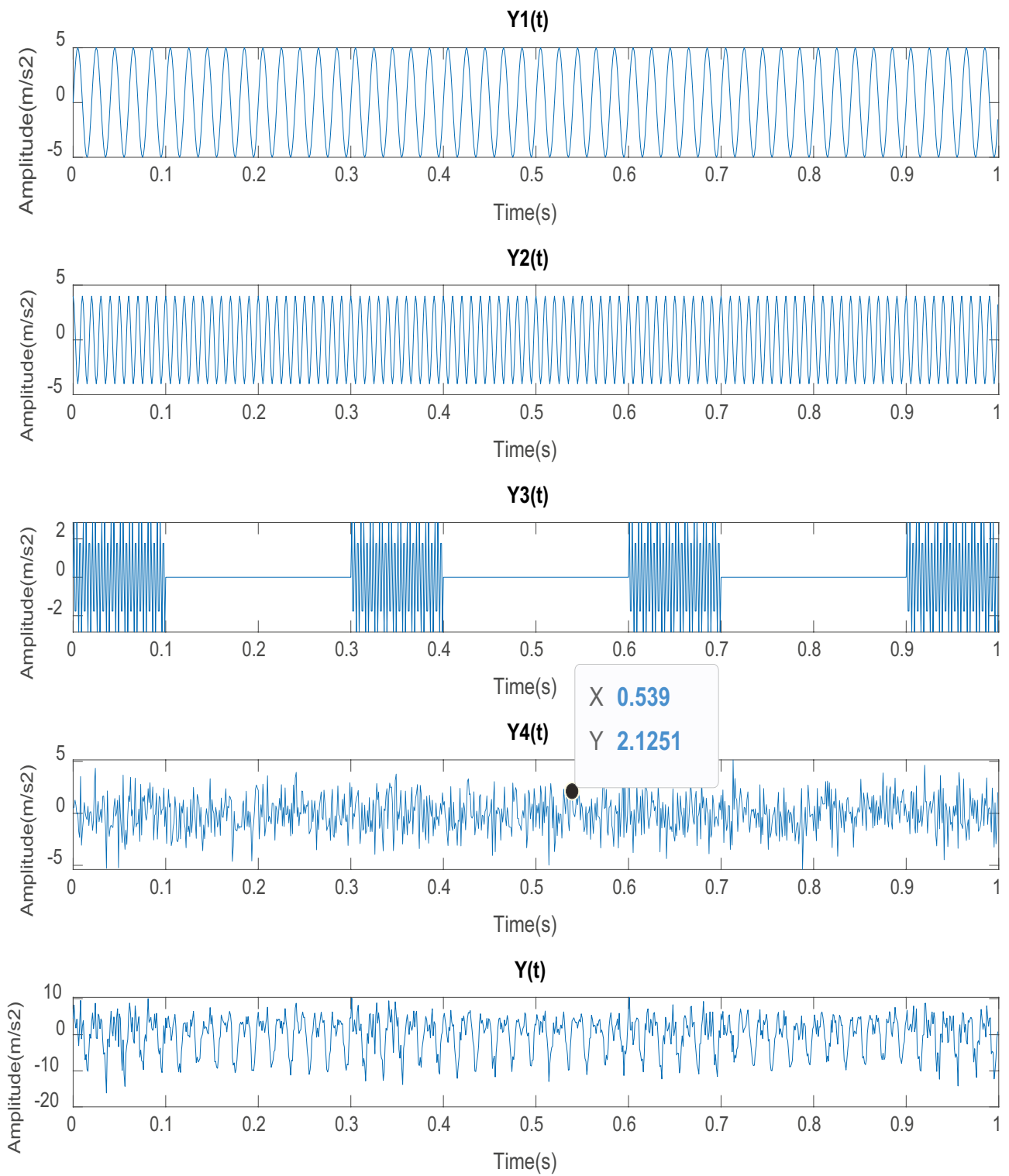


Fig. 3 The time domain waveform of each component signal and simulation signal

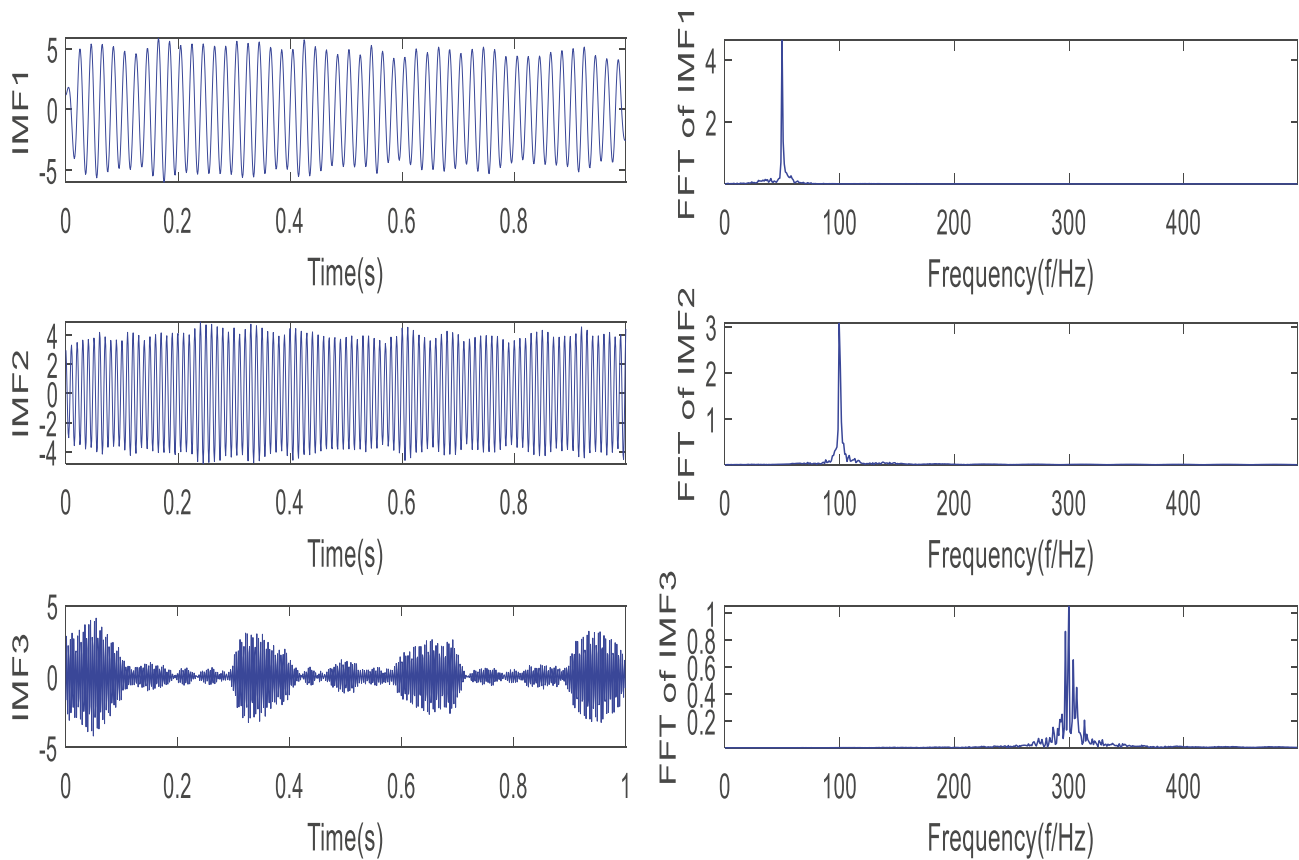


Fig. 4 WOAGWO-VMD decomposes the simulation signal

where $y_1(t)$ is the sine signal with amplitude 5 and frequency 50; $y_2(t)$ is the cosine signal with amplitude 4 and frequency 100; $y_3(t)$ is high-frequency intermittent signal; $y_4(t)$ is Gaussian white noise and $y(t)$ is the simulation signal. The time domain waveform of all these signals is shown in Fig. 3.

Firstly, WOAGWO is used to optimize the parameter combination $[K, \alpha]$ in the VMD algorithm. After that, the optimized VMD algorithm with the optimal combination $[K, \alpha]$ is used to decompose the simulation signal. The decomposition results of WOAGWO-VMD are shown in Fig. 4. Secondly, the EMD algorithm is used to decompose the same simulation signal. The decomposition results of EMD are shown in Fig. 5.

According to Fig. 4, it can be seen that the proposed method WOAGWO-VMD decomposed the simulation signal into 3 IMF components with 50 Hz, 100 Hz, and 300 Hz successfully. Therefore, the effectiveness of WOAGWO-VMD algorithm can be obtained. As shown in Fig. 5, it can be seen that IMF1 and IMF3 extracted 300-Hz and 50-Hz signals successfully, but mode-mixing still occurs in IMF2 and IMF3. From the results obtained, it can be concluded that the proposed WOAGWO-VMD

algorithm can decompose the frequency components in the simulation signal successfully, and overcome the mode mixing phenomenon in EMD algorithm.

2.5 Selection of sensitive components

The selection of components containing the most information is an important step in fault diagnosis, so for fault feature extraction a new method must be used to determine the sensitive components that have the largest contribution. Pearson correlation coefficient (Pcc) is an effective tool in the selection of IMF which contains the fault characteristics [47], where the bigger value of Pcc denotes the greater impact features. The sparseness is a statistical index that effectively reflects the amplitude distribution of the vibration signal [42]. The largest value of sparseness denotes the stronger data sparsity. Given the superiority of both Pearson correlation coefficient and sparseness, a new index called sensitive indicator (SI) was formulated based on the product of Pearson correlation coefficient and sparseness to select adaptively the sensitive components. The mathematical expression of SI is defined as:

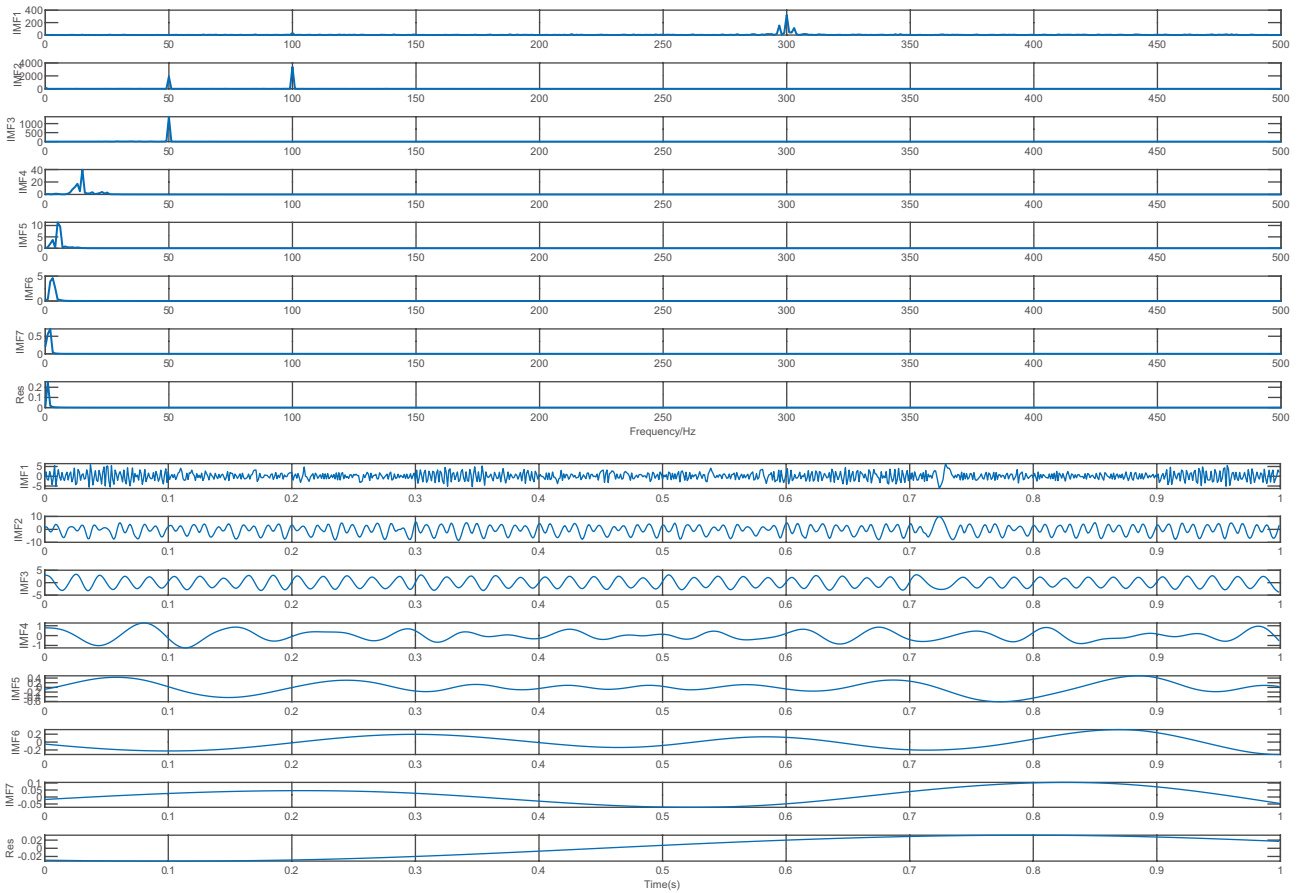


Fig. 5 EMD decomposes the simulation signal

$$S = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N x(n)^2}}{N \sum_{n=1}^N |x(n)|} \tag{25}$$

$$Pcc = \frac{E[(x - \bar{x})(y - \bar{y})]}{E[(x - \bar{x})^2]E[(y - \bar{y})^2]} \tag{26}$$

$$SI = Pcc \times S \tag{27}$$

where S is the sparseness of the signal $x(n)$, and N is the length of the signal $x(n)$; Pcc denotes the Pearson correlation coefficient between two signals (x and y), and $E[.]$ represents the expectation operator.

2.6 Multi-features

To extract bearing fault feature vectors, we propose a multi-feature feature extraction method based on DE, PE, and SVD.

2.6.1 Dispersion entropy

Dispersion entropy results from the integration of symbolic dynamics with Shannon entropy for the development of an algorithm capable of characterizing the irregularity of time series with a low computation time [48]. The calculation steps of dispersion entropy for the time series $x = \{x_i, i = 1, 2, \dots, N\}$ are as follows:

1. Use normal cumulative distribution function to map the time series \mathbf{x} to $y = \{y_1, y_2, \dots, y_N\}$ from 0 to 1

$$y_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \tag{28}$$

where σ and μ represent mean and standard deviations of time series x respectively.

2. Employ the linear algorithm to map y_i into integer from 1 to c and obtain a sequence z_j^c

$$Z_j^c = \text{round}(c \cdot y_j + 0.5) \tag{29}$$

where c represents the number of classes after mapping, Z_j^c is the j th member of the classified time series, and round is the rounding function.

3. Calculate the embedding vector $z_i^{m,c}$ by exploiting the following formula:

$$Z_i^{m,c} = \{Z_i^c, Z_{i+d}^c, \dots, Z_{i+(m-1)d}^c\}, i = 1, 2, \dots, N - (m - 1)d \tag{30}$$

where m and d represent the embedded dimension and the delay time respectively.

4. Calculate the dispersion entropy patterns $\pi_{v_0, v_1, \dots, v_{m-1}}$ ($v = 1, 2, \dots, c$) of each vector $Z_i^{m,c}$. $Z_i^c = v_0, Z_{i+d}^c = v_1, \dots, Z_{i+(m-1)d}^c = v_{m-1}$. c^m is the number of possible patterns.

5. For each dispersion pattern, the relative frequency of potential dispersion patterns can be calculated as

$$p(\pi_{v_0, v_1, \dots, v_{m-1}}) = \frac{\text{number}(\pi_{v_0, v_1, \dots, v_{m-1}})}{N - (m - 1)d} \tag{31}$$

where number $\pi_{v_0, v_1, \dots, v_{m-1}}$ represent the number of dispersion patterns. $p(\pi_{v_0, v_1, \dots, v_{m-1}})$ equal to the number of $Z_i^{m,c}$ mapped to $\pi_{v_0, v_1, \dots, v_{m-1}}$ divided by the number of elements in $Z_i^{m,c}$.

6. The dispersion entropy is calculated as:

$$DE(x, m, c, d) = - \sum_{\pi=1}^{c^m} P(\pi_{v_0, v_1, \dots, v_{m-1}}) \ln(P(\pi_{v_0, v_1, \dots, v_{m-1}})) \tag{32}$$

2.6.2 Permutation entropy

Bandt and Pompe [49] proposed an approach called PE; this approach allows to analyze the complexity of the signal and detect dynamic changes in time series using the comparison of neighboring values. Its advantages are simplicity, robustness, and stability in addition to good noise resistance. The

mathematical theory of PE is described briefly below. For a time series $x = \{x_1, x_2, \dots, x_N\}$, the m -dimensional embedding vector is constructed as:

$$X_i = \{x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)\}, i = 1, 2, \dots, N - (m - 1)\tau \tag{33}$$

where m represents the embedding dimension and τ is the time delay.

Each X_i can be rearranged in an increasing order as

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \leq \dots \leq x(i + (j_m - 1)\tau) \tag{34}$$

If there exist two elements in X_i that have the same value, like:

$$x(i + (j_1 - 1)\tau) = x(i + (j_2 - 1)\tau) \tag{35}$$

Then their order can be denoted as:

$$x(i + (j_1 - 1)\tau) \leq x(i + (j_2 - 1)\tau) \quad j_1 < j_2, \tag{36}$$

For any X_i , it can be mapped onto a group of symbols as

$$S(g) = (j_1, j_2, \dots, j_m) \tag{37}$$

where $g = 1, 2, \dots, k$ and $k \leq m!$. $m!$ is the largest number of distinct symbols, and $S(g)$ is the one of $m!$ symbol sequence. The probability distribution of each symbol sequence is calculated as P_1, P_2, \dots, P_k . The PE for the time series is defined as follows:

$$H_p(m) = - \sum_{j=1}^k P_j \ln P_j \tag{38}$$

where $0 \leq H_p(m) \leq \ln(m!)$, $H_p(m)$ reaches the maximum $\ln(m!)$ when $P_j = \frac{1}{m!}$. $H_p(m)$ can be further normalized as:

$$H_p = \frac{H_p(m)}{\ln(m!)} \tag{39}$$

2.6.3 Singular value decomposition

SVD is a very important matrix decomposition technique in linear algebra proposed by Beltrami in 1873. It has been regularly used in feature extraction and signal processing. SVD of an $m \times n$ matrix A is given by:

$$A = USV^T = \sum_{i=1}^P A_i = \sum_{i=1}^P u_i \sigma_i v_i^T \tag{40}$$

where $U = (m \times m)$ and $V = (n \times n)$ are the orthogonal matrix, and $S = (m \times n)$ being a matrix containing the singular values σ_i on the main diagonal and 0 elsewhere.

3 Fault feature classification using MPA-LSSVM

3.1 Marine predators algorithm

The MPA is a nature-inspired heuristic optimization algorithm proposed by A. Faramarzi [50]. The detailed steps of the algorithm are presented as follows:

3.1.1 Stage 1

This is the most important process in the initial iteration of improvement where exploration is important; at this point, when the movement of the predator is faster than the prey, the best strategy for the predator is to not move. So this stage is represented mathematically as follows:

while $Iter < \frac{1}{3} Max_Iter$

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_B \otimes (\overrightarrow{Elite}_i - \overrightarrow{R}_B \otimes \overrightarrow{Prey}_i) & i = 1, \dots, n \\ \overrightarrow{Prey}_i = \overrightarrow{Prey}_i + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \end{cases} \quad (41)$$

where \overrightarrow{R}_B represents a random vector indicating Brownian motion. \otimes is the entry-wise multiplications, and Prey and Elite are the prey locations and best predator, respectively. Moreover \overrightarrow{R} is a vector of uniform random numbers in [0,1]. And P is a constant equal to 0.5. Max_iter is the maximum iteration and Iter represents the current iteration. This scenario is mainly encountered in the first third of iterations when step size or movement velocity is high.

3.1.2 Stage 2

At this stage, predator and prey movements are at the same pace. This simulates that they are both searching for the prey. This situation occurs mainly in the middle of the improvement process as exploration is gradually replaced by exploitation. Therefore, the population is divided into two parts. The first part was chosen for exploration while the second part was chosen for exploitation, in another way. The prey is responsible for the exploitation process, while the predatory animal is responsible for the exploration process.

While $\frac{1}{3} Max_Iter < Iter < \frac{2}{3} Max_iter$

For the first half of the population

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{Elite}_i - \overrightarrow{R}_L \otimes \overrightarrow{Prey}_i) & i = 1, \dots, n/2 \\ \overrightarrow{Prey}_i = \overrightarrow{Prey}_i + P \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \end{cases} \quad (42)$$

where \overrightarrow{R}_L is a vector including random numbers based on Lévy distribution. The multiplication of \overrightarrow{R}_L by Prey represents the simulation of the movement of prey in the manner

of Luff. Since the step size of the Lev distribution is usually smaller, this will aid in exploration. The rule applied to the second half of the populations can be written as

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_B \otimes (\overrightarrow{R}_B \otimes \overrightarrow{Elite}_i - \overrightarrow{Prey}_i) & i = n/2, \dots, n \\ \overrightarrow{Prey}_i = \overrightarrow{Elite}_i + P \cdot CF \otimes \overrightarrow{stepsize}_i \end{cases} \quad (43)$$

$$CF = (1 - \frac{Iter}{Max_Iter})^{(2 \frac{Iter}{Max_Iter})}$$

where CF represents the adaptive parameter to control the step size for predator. The multiplication of \overrightarrow{R}_B by the Elite represents the simulation movement of predator in Brownian, where the position of the prey is updated based on the movement of predator.

3.1.3 Stage 3

At this stage (low velocity ratio), the predator’s movement is faster than the prey. This phenomenon occurs in the last stage of the optimization process when the utilization capacity is high. Additionally, levy movement is the predator’s best strategy. The mathematical model of the last stage is represented as follows:

while $Iter > \frac{2}{3} Max_Iter$

$$\begin{cases} \overrightarrow{stepsize}_i = \overrightarrow{R}_L \otimes (\overrightarrow{R}_L \otimes \overrightarrow{Elite}_i - \overrightarrow{Prey}_i) & i = 1, \dots, n \\ \overrightarrow{Prey}_i = \overrightarrow{Elite}_i + P \cdot CF \otimes \overrightarrow{stepsize}_i \end{cases} \quad (44)$$

3.1.4 Stage 4

Eddy formation or fish aggregation devices (FADs) are an example of environmental issues affecting the behavior of marine predators as FADs represent local optima in the field of exploration. The mathematical model that describes the effects of FAD is formulated as follows:

$$\overrightarrow{Prey}_i = \begin{cases} \overrightarrow{Prey}_i + CF [\overrightarrow{X}_{min} + \overrightarrow{R} \otimes (\overrightarrow{X}_{max} - \overrightarrow{X}_{min})] \otimes \overrightarrow{U} & \text{if } r < FAD_s \\ \overrightarrow{Prey}_i + [FAD_s(1 - r) + r] (\overrightarrow{Prey}_{r1} - \overrightarrow{Prey}_{r2}) & \text{if } r > FAD_s \end{cases} \quad (45)$$

where $FAD = 0.2$, and U is a binary solution with values 0 or 1; $r1$ and $r2$ are the indices of the prey, and X_{max} and X_{min} represent the upper and lower bounds, respectively.

3.1.5 Stage 5

Predators have a good and detailed memory of successful foraging sites. This ability is simulated using MPA, and

the matrix fitness can be evaluated to refresh Elite after updating the Prey and the FADs. The fit value of each solution is compared with the previous value, so the optimum solution is saved in memory. Computational complexity of MPA algorithm is $O(t(nd + \text{cof} \times n))$, where cof is the evaluation function, t is the iteration number, d is the optimization problem dimension, and n is the number of populations. The pseudo-codes of MPA is shown as follows:

Pseudo-code of the MPA algorithm

```

Initialize search agent (Prey) populations  $i=1,2,\dots,n$ 
While termination criteria are not met
    Calculate the fitness, construct the Elite matrix and accomplish memory saving
    If  $\text{Iter} < \text{Max\_Iter}/3$ 
        Update prey according to Eq.41
    Else if  $\text{Max\_Iter}/3 < \text{Iter} < 2 * \text{Max\_Iter}/3$ 
        For the first half of the populations ( $i=1,\dots,n/2$ )
            Update prey according to Eq.42
        For the other half of the populations ( $i=n/2,\dots,n$ )
            Update prey according to Eq.43
    Else if  $\text{Iter} > 2 * \text{Max\_Iter}/3$ 
        Update prey according to Eq.44
    End if
    Accomplish memory saving and Elite update
    Applying FADs effect and update based on Eq.45
End while
    
```

3.2 LSSVM algorithm

SVM is a machine learning method that can be used for data classification or regression. It is based on statistical learning and the minimization of structural risks [34]. Its main role is to create an optimal hyperplane to maximize the separation margin between the two classes. The LSSVM is an improved algorithm based on SVM, where it converts the inequality constraints in SVM to equality and utilizes the least square linear formula as the loss function rather than the quadratic programming method used in SVM. The optimization objective function of LSSVM is defined as follows:

$$f(x) = \text{sgn}\{\omega \cdot \phi(x) + b\} \tag{46}$$

where ω is the weight vector, $\phi(x)$ represents nonlinear mapping function, and b is deviation vector, the final optimization problem becomes

$$\begin{cases} \min J(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + \gamma \sum_{i=1}^N \xi_i^2 \\ \text{s.t. } y_i [\omega^T \phi(x_i) + b] = 1 - \xi_i \\ i = 1, 2, \dots, N. \end{cases} \tag{47}$$

J is the optimized objective function, ξ is the error variable, and γ represents the penalty factor. To solve the problem of optimization and obtain a better classification model, the Lagrange multiplier α_i was introduced and the Lagrange function was constructed as follows: (48).

$$L(\omega, b, \xi, \alpha_i) = J(\omega, \xi_i) - \sum_{i=1}^N \alpha_i \{y_i [\omega^T \phi(x_i) + b] - 1 + \xi_i\} \tag{48}$$

By using the constraints on the Karush–Kuhn–Tucker (KKT) condition, at the extreme points sought, the related parameters of the Lagrange function are independently subjected to a partial derivative operation, with the result being 0. The linear matrix expression that results is as follows:

$$\begin{bmatrix} 0 & E^T \\ E & \psi \psi^T + E \cdot \gamma^{-1} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{49}$$

where $E = [1, 1, \dots, 1]^T$, $E, y = [y_1, y_2, \dots, y_n]^T$, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]^T$, and $\psi = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)]^T$. In the LSSVM algorithm, the decision function for classification is given as follows

$$f(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i K(x, x_i) + b \right] \tag{50}$$

where $k(x, x_i)$ is the kernel function satisfying the mercer condition.

3.2.1 MPA-LSSVM

The LSSVM is an improved algorithm based on a SVM, which can be widely used in the mechanical fault diagnosis domain

to solve classification and regression problems. However, the classification performance of LSSVM is mostly affected by the selection of parameters c and g [51]. To improve the classification feasibility of LSSVM, an algorithm should be used to optimize the parameters (c and g). MPA is a new optimization algorithm that was illustrated through benchmark functions and practical engineering problems. The results have demonstrated that MPA can obtain the optimal solution with a lower numerical cost compared to available optimization algorithms [50]. These comparison results encouraged us to use the MPA algorithm to optimize LSSVM parameters. The fitness function of MPA-LSSVM algorithm is shown in Eq. (51). In this study, the range of parameters c and g is $[0, 1000]$ chosen according to Refs. [21] and [52]; the populations size = 20 and max iterations = 100 based in [55] and the initial parameter settings of MPA algorithm are shown in Table 1. The flowchart of the MPA-LSSVM proposed is shown in Fig. 6, and the specific steps of this proposed method are as follows:

Step 1: Input the training set and test set normalized to the interval $[0, 1]$.

Step 2: Initialize the MPA and LSSVM parameters. Set the number of populations to 20, maximum number of iterations to 100. Then setting the range of (c , g), where the lower and upper bounds of c and g are 0 and 1000, respectively.

The position of each predator is defined as (c , g).

Step 3: Use the following equation as fitness function of MPA-LSSVM.

$$\text{fitness} = 1 - \frac{N_t}{N_t + N_f} \quad (51)$$

where N_t and N_f represent the number of true and false classification, respectively.

Evidently, a lower fitness value recorded a higher classification accuracy. The aim of the LSSVM parameter optimization problem is to minimize the fitness function.

Step 4: Update the predators and preys position according to Eqs. (41)–(45) in Sect. 3.1.

Step 5: Fitness evaluations:

- Updating the prey position.

The effect of updating the prey position on the fitness value can be evaluated using the FADs effects reported in stage 5 (Sect. 3.1). If the fitness value of the new prey position is lower compared to the previous position, then the new prey position replaces the previous one.

Step 6: Export the optimal values of (c_{best} , g_{best}) once the maximum number of iterations is reached and trained LSSVM model.

Step 7: The trained model is used to identify and classify the test dataset.

4 Fault diagnosis method based on WOAGWO-VMD and MPA-LSSVM

In order to improve the bearing fault identification accuracy, a novel intelligent fault diagnosis method is proposed in this paper. The structural framework of the proposed method is illustrated in Fig. 7. The specific steps of this method are summarized as follows:

Step 1: The bearing vibration acceleration signals in various states (i.e., normal, ball fault, inner race fault and outer race fault) are collected using the acceleration sensors.

Step 2: The hybrid algorithm (WOAGWO) is used to search the optimal parameters combination $[K_0, \alpha_0]$ of VMD.

Step 3: Utilize VMD with optimal parameter combination $[K_0, \alpha_0]$ to decompose the vibration signal into several IMFs.

Step 4: Analyze the correlation between each IMF component and the original signal by calculating the sensitive indicator (SI) of each IMF component and the original signal.

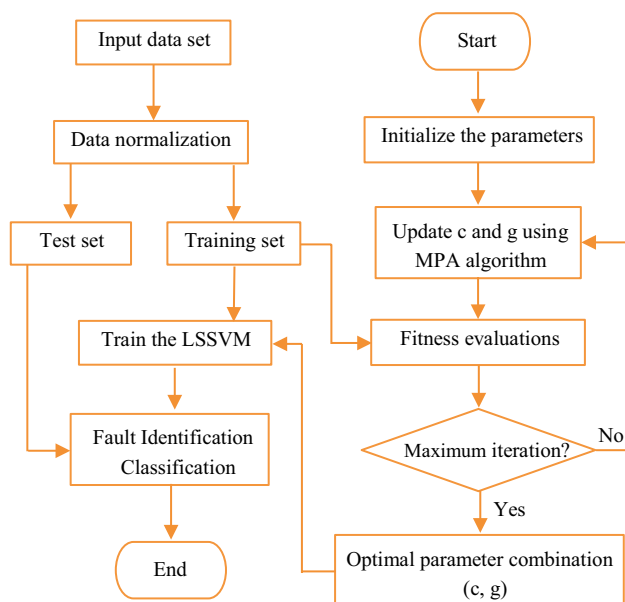


Fig. 6 Flowchart of MPA-LSSVM model

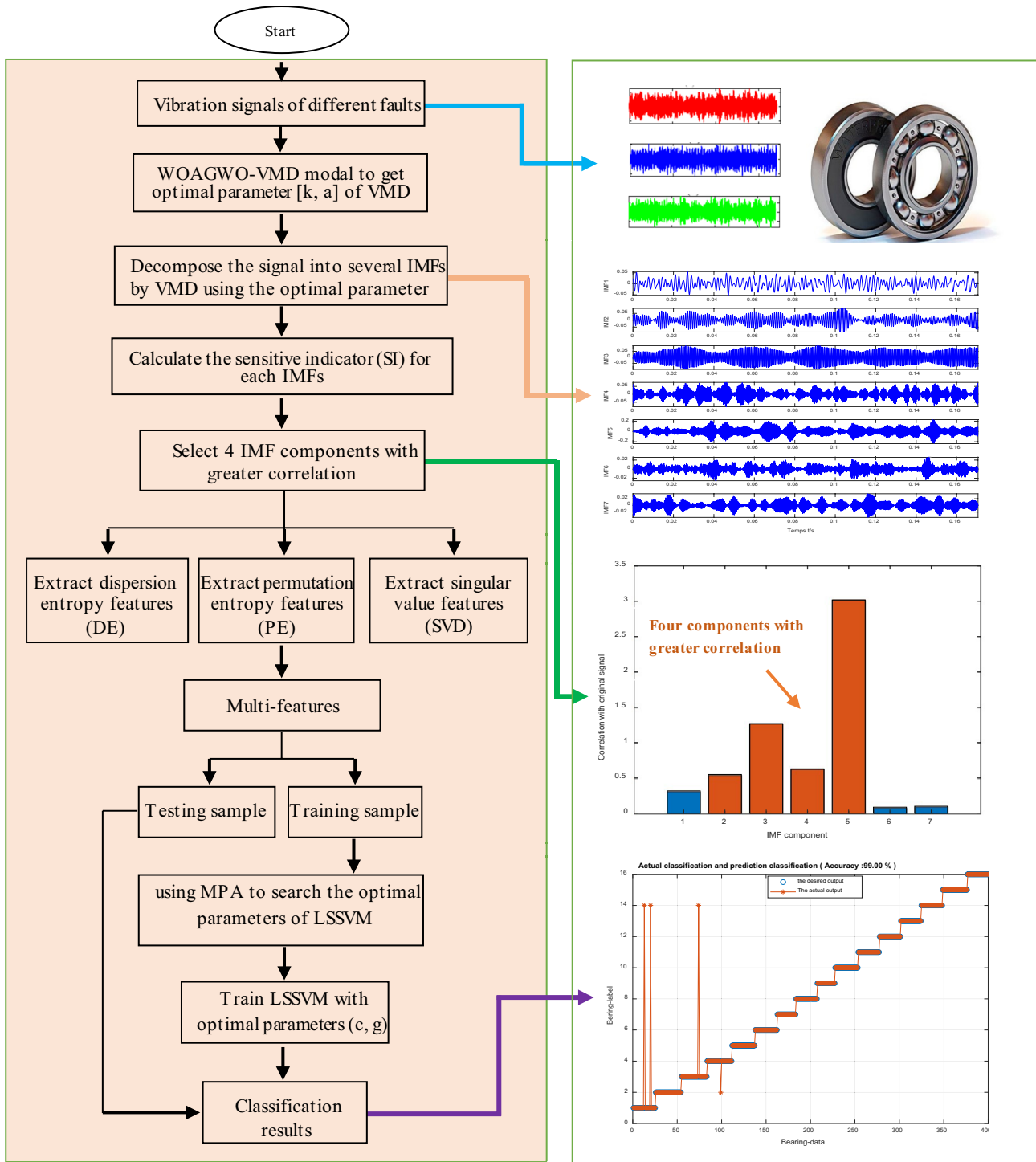


Fig. 7 Fault diagnosis method based on WOAGWO-VMD and MPA-LSSVM

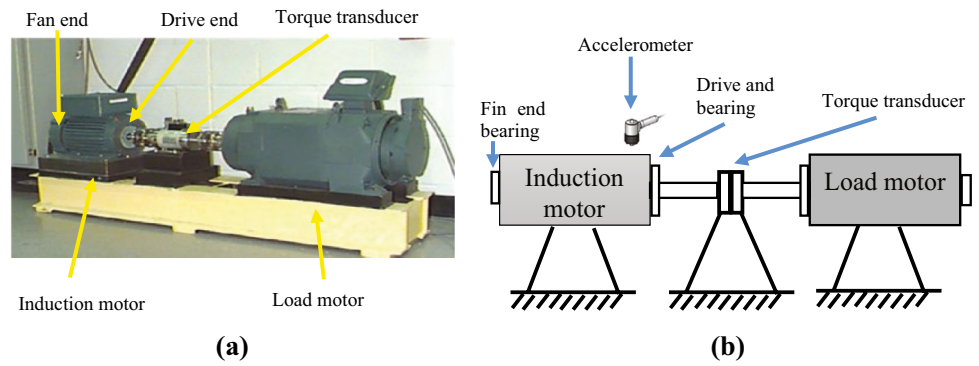
Step 5: Select four IMF components with greater correlation with the original signal, and calculate DE, PE, and SVD of each IMFs, then construct the multiple features vector.

Step 6: Normalize the sample feature value to [0, 1], using function mapminmax in matlab.

Step 7: The obtained feature vectors normalized are randomly divided into two groups, a training sample set and a testing sample set.

Step 8: The training samples is used as input to the MPA-LSSVM classifier, for obtaining the best LSSVM prediction model, while the testing samples is included in the

Fig. 8 The bearing test stand (a), and its Schematic diagram (b)



prediction model to recognize and classifier different fault types.

5 Experimental verification and result analysis

The experimental data of rolling bearings used in our paper is provided by Case Western Reserve University (CWRU) USA [53] in order to verify the validity of the suggested method in diagnosing rolling bearing

faults. The test stand of the rolling bearing experiment is shown in Fig. 8. It consists mainly of a 2 HP motor (left), a torque sensor (middle), a dynamometer (right), and control electronics. The rolling bearing used in the experiment is a SKF6205 deep groove ball bearing. The vibration acceleration signal of the bearing is obtained from the driving end under the condition of a rotation speed of 1730 r/min, a load of 3HP, and a sampling frequency of 12 kHz. The bearing vibration signals are first classified into four categories, namely ordinary rolling

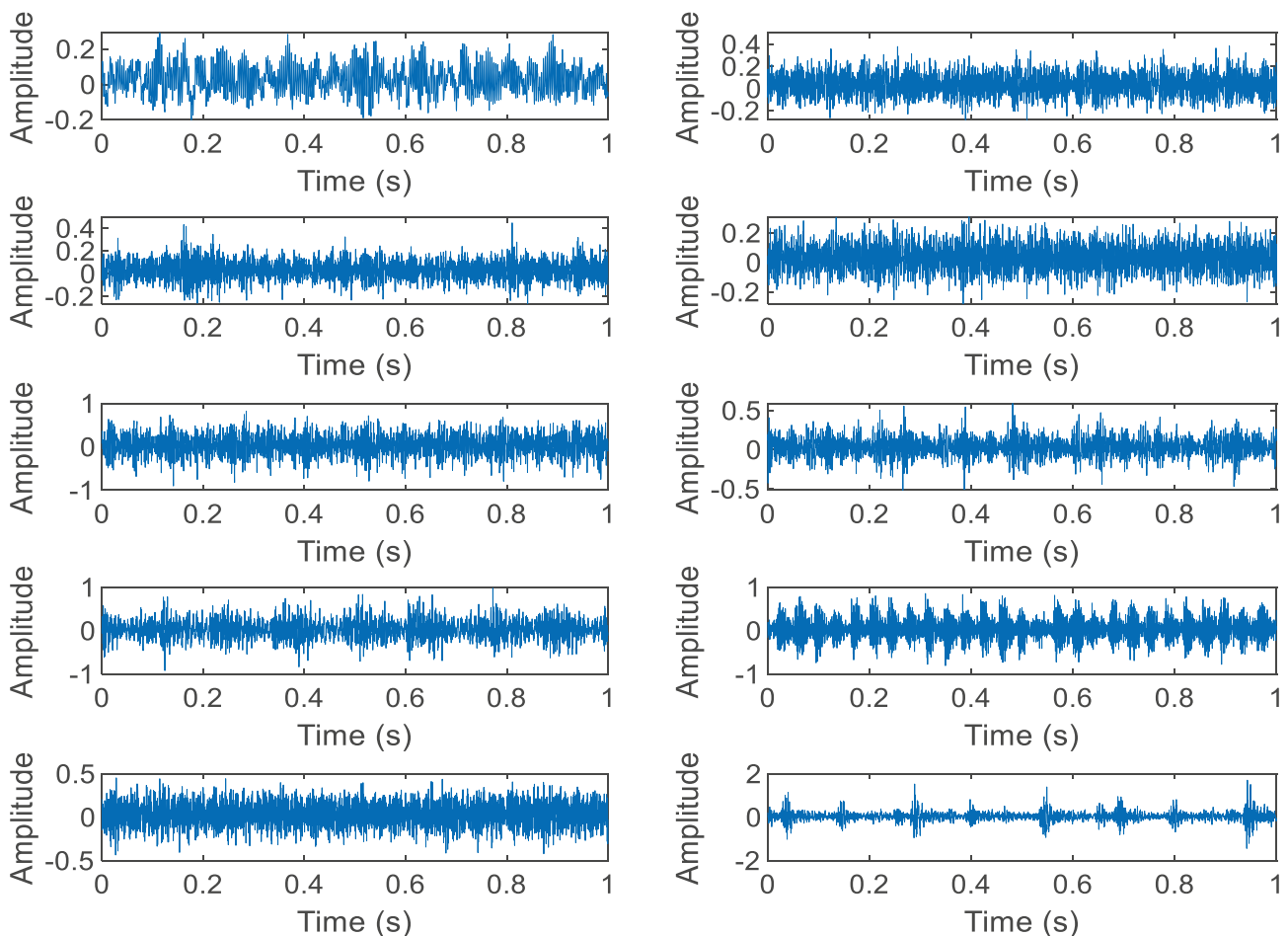


Fig. 9 Time domain waveform of the original bearing vibration

Table 2 Detailed description of the considered bearing working conditions

Fault diameter (inch)	health status	Conditions	damage point	Number of Training Samples	Number of Test Samples	Assigned labels	Category	
0	Normal			30	20	NO	1	
0.007	Inner race fault	ball fault	outer race fault	3 o'clock position	30	20	IR_7	2
				6 o'clock position	30	20	B_7	3
				12 o'clock position	30	20	OR_7_3	4
					30	20	OR_7_6	5
					30	20	OR_7_12	6
					30	20	IR_14	7
0.014	Inner race fault	ball fault	outer race fault		30	20	B_14	8
					30	20	OR_14	9
					30	20	IR_21	10
0.021	Inner race fault	ball fault	outer race fault	3 o'clock position	30	20	B_21	11
				6 o'clock position	30	20	OR_21_3	12
				12 o'clock position	30	20	OR_21_6	13
					30	20	OR_21_12	14
					30	20	IR_28	15
					30	20	B_28	16

bearings (normal) and rolling bearings with ball failure (B), outer ring failure (OR), and inner ring failure (IR). The faulty bearing is formed on the normal bearing by

using electro-discharge machining (EDM). The diameter of the fault is 0.007, 0,014, 0.021, and 0.028 in. respectively. The damage points of the bearing outer ring are

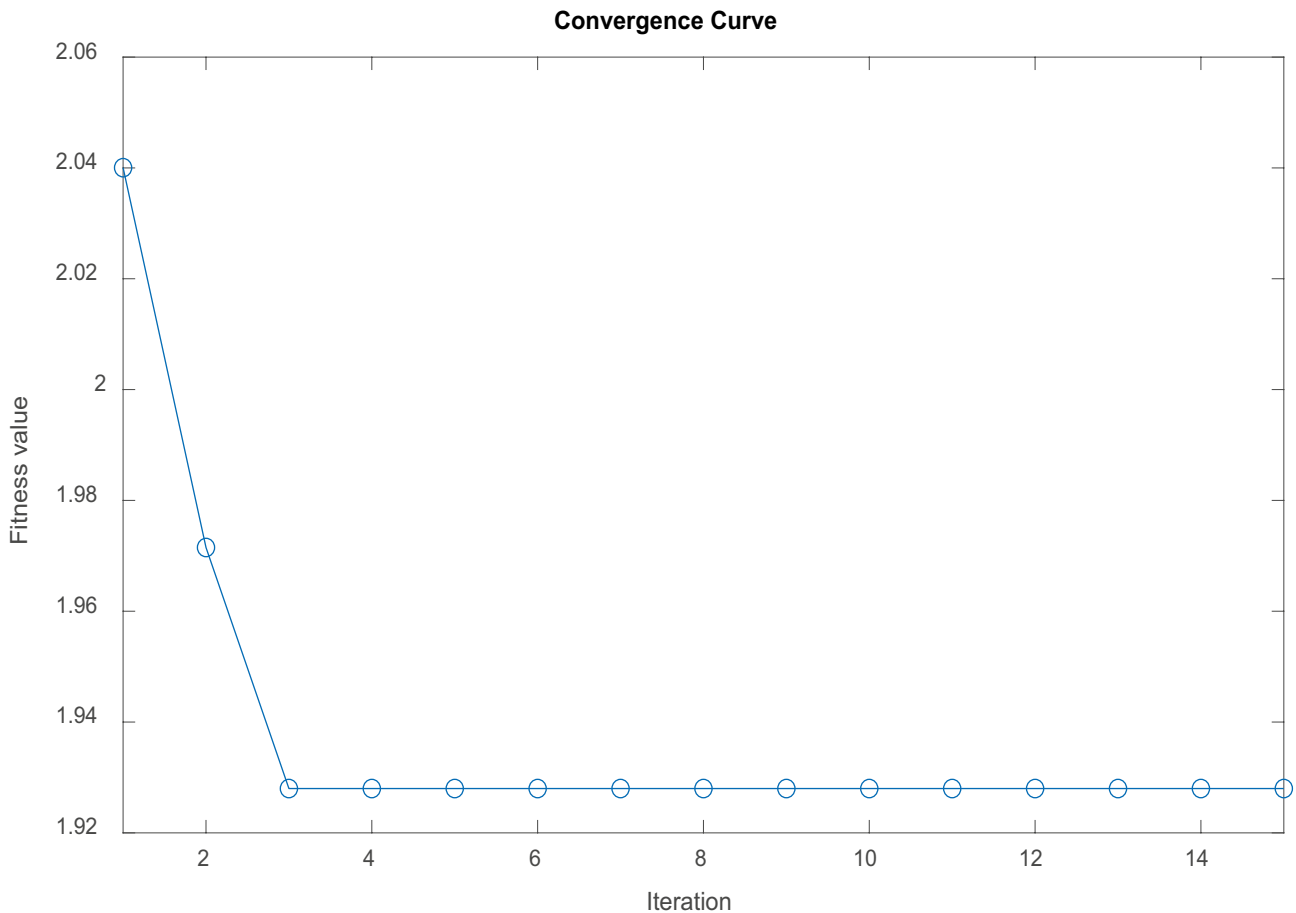


Fig. 10 Fitness value curve. Signal under different health conditions

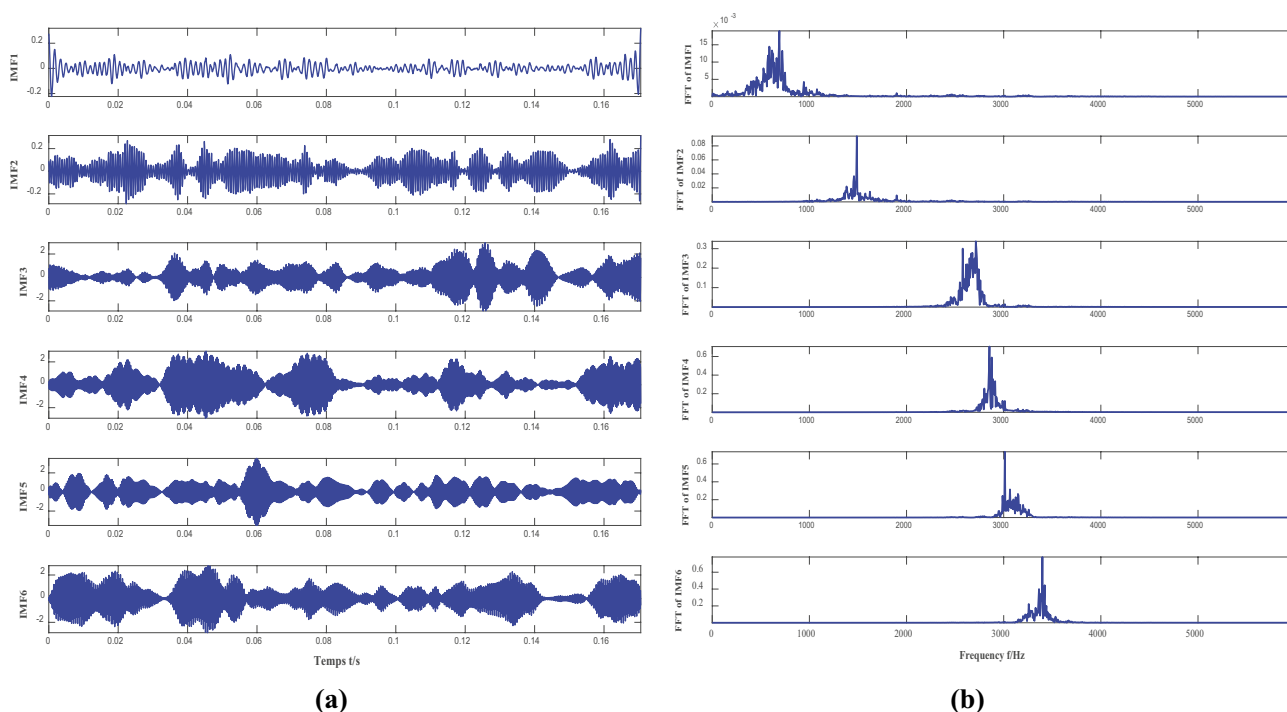


Fig. 11 Waveform and spectrum of 6 IMF components. (a) Time domain, (b) FFT spectrum

at 3 o'clock, 6 o'clock, and 12 o'clock respectively. The bearing vibration signal is classified into of 16 different types of failures. Each class has 50 samples with 4096 points for a total of 800 samples. Four hundred eighty groups are randomly selected for the training set and 320 groups are selected for the test sets. The time domain waveforms of vibration signal of rolling bearing are shown in Fig. 9. The detailed description of the class label is given in Table 2. This paper takes the ball fault signal with a fault severity of 0.028 in. as an example. The WOAGWO is utilized to optimize the parameters $[K, \alpha]$ of the VMD algorithm. The change curve of the minimum value of the average of the weighted permutation entropy

with the number of iterations is shown in Fig. 10. It can be seen from Fig. 10 that the minimum value of the average of the weighted permutation entropy of 1.9280 appeared in the third iteration, which indicates that the optimization algorithm is quickly converging and has global optimization capabilities, and is appropriate for searching for the best parameter combination of VMD. The corresponding optimal parameter combination $[K, \alpha]$ is $[6, 3835]$. This obtained parameters are entered into the parameter settings of the VMD. Figure 11 shows the time domain diagram and spectrum diagram of 6 IMF components after the decomposition of the inner ring fault signal by VMD.

For the vibration signals of 16 status, the best optimal combination $[K_0, \alpha_0]$ obtained by the hybrid optimization algorithm WOAGWO is shown in Table. 3

In order to demonstrate the superiority and efficacy of the multi-feature feature extraction method, the VMD method with the optimal parameter combination $[K, \alpha]$ obtained by WOAGWO-VMD algorithm is used to decompose different bearing vibration signals. Four IMF components with large correlation with the original signal are selected by the sensitive indicator (SI). For the IMF components selected, DE, PE, SVD, and multi-features are extracted from three perspectives of different fault types, different damage points, and different severity levels. Dimensionality reduction visualization t-SNE is used to visualize and compare the effects of the four feature

Table 3 Optimal parameter combination $[K, \alpha]$

Health status condition	$[k, \alpha]$	Health status condition	$[k, \alpha]$
Normal	[5 3084]	OR_14	[7 2543]
IR_7	[6 1503]	IR_21	[6 3249]
B_7	[5 2164]	B_21	[7 2455]
OR_7_3	[7 2668]	OR_21_3	[7 1972]
OR_7_6	[8 3990]	OR_21_6	[7 4000]
OR_7_12	[7 3875]	OR_21_12	[7 3645]
IR_14	[6 2604]	IR_28	[7 1154]
B_14	[7 1734]	B_28	[6 3835]

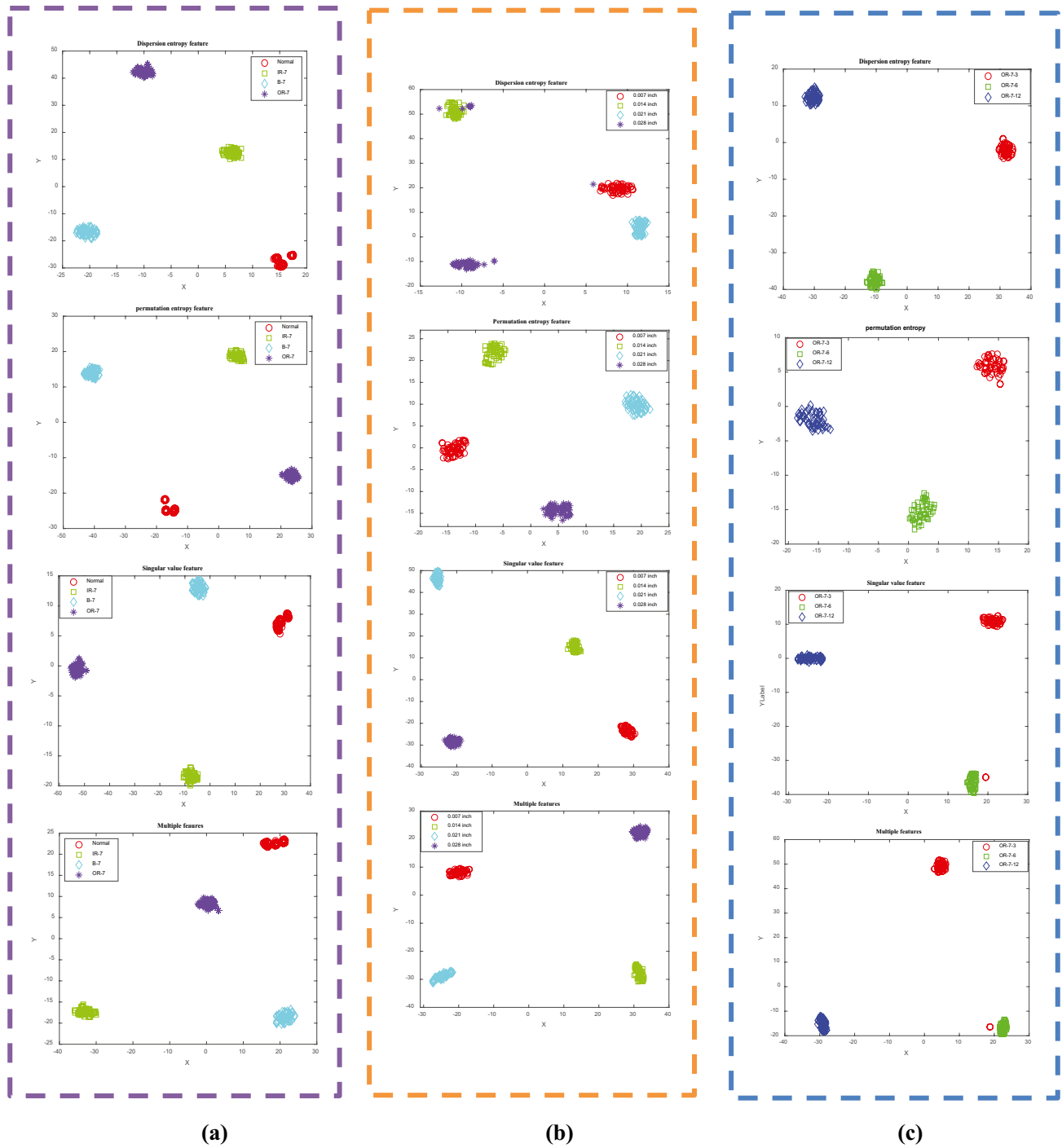


Fig. 12 Low-dimensional fault feature distribution. (a) Different fault types, (b) different severity, (c) different damage points

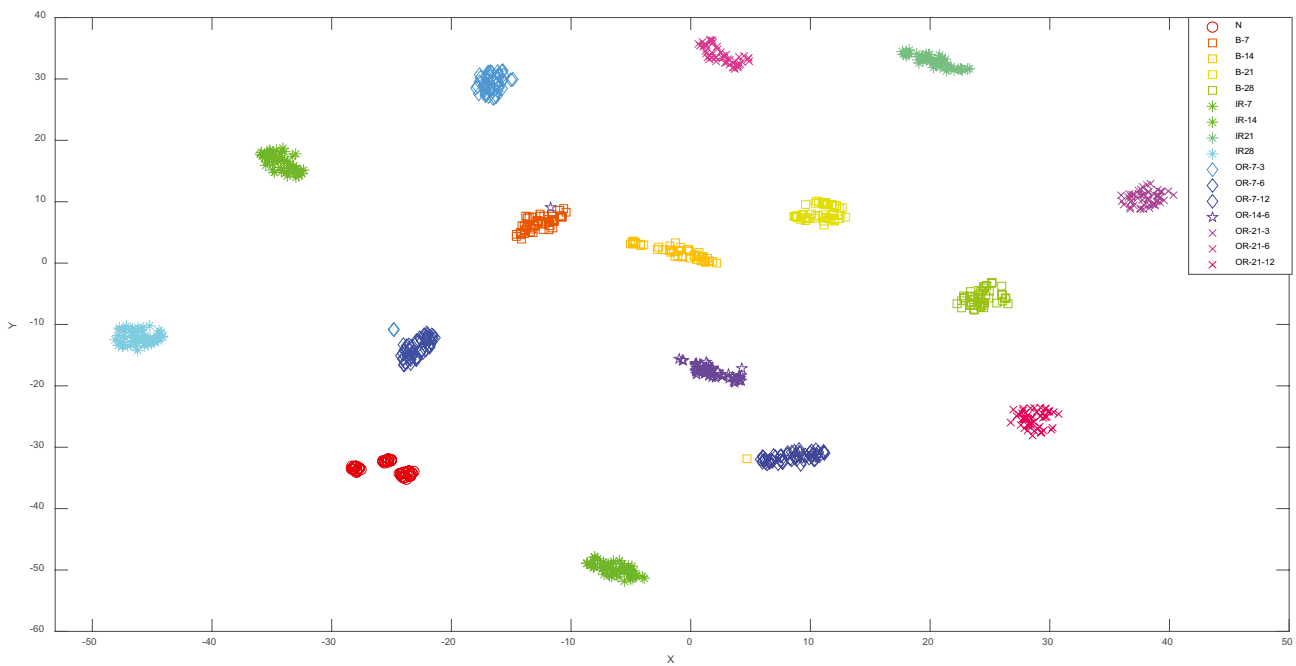


Fig. 13 Low-dimensional feature distribution of 16 kinds of bearing signals

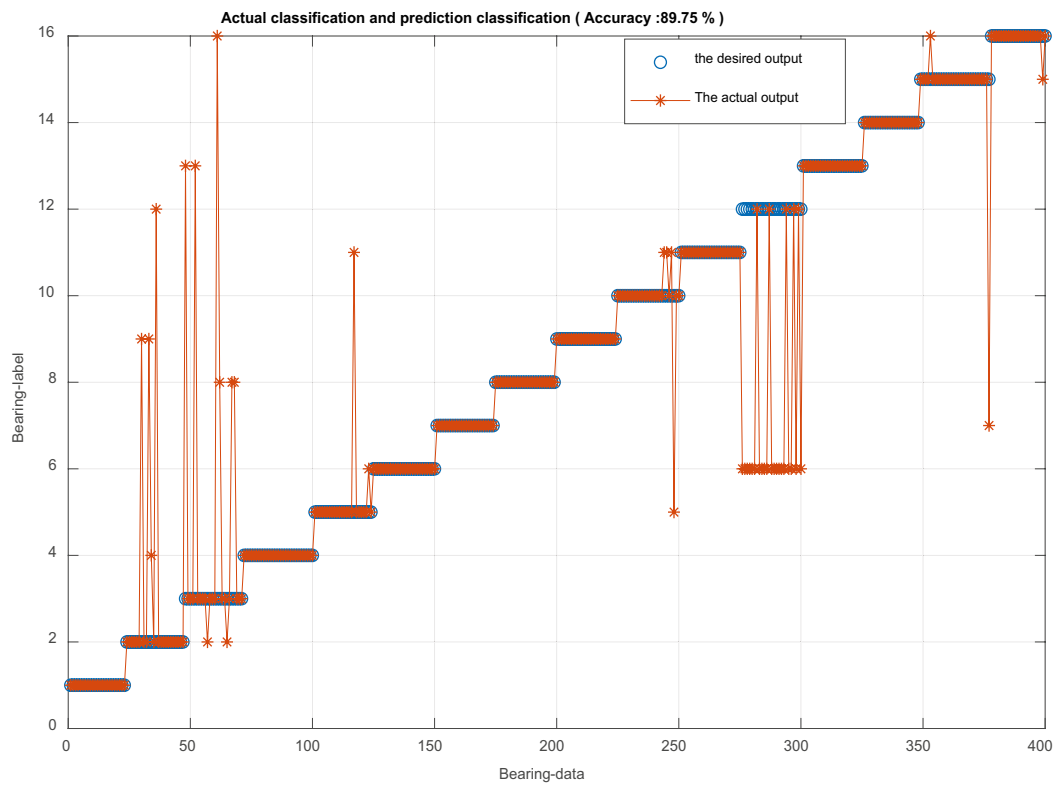
extraction methods, such as shown in Fig. 12. Through t-SNE visual comparison in Fig. 12, it is shown that the multiple features have good intra-class aggregation and inter-class separation from three perspectives of different fault types, different damage points, and different severity levels, and exhibit the dispersion entropy feature, permutation entropy feature, and singular value feature results solely. Using the t-SNE dimensionality reduction method, we can see the distribution of multi-features extracted from 16 different bearing vibration signals previously analyzed by the WOAGWO-VMD algorithm, as shown in Fig. 13.

It can be observed from Fig. 13 that the distinction of 16 types of data features is clear and the aggregation of data

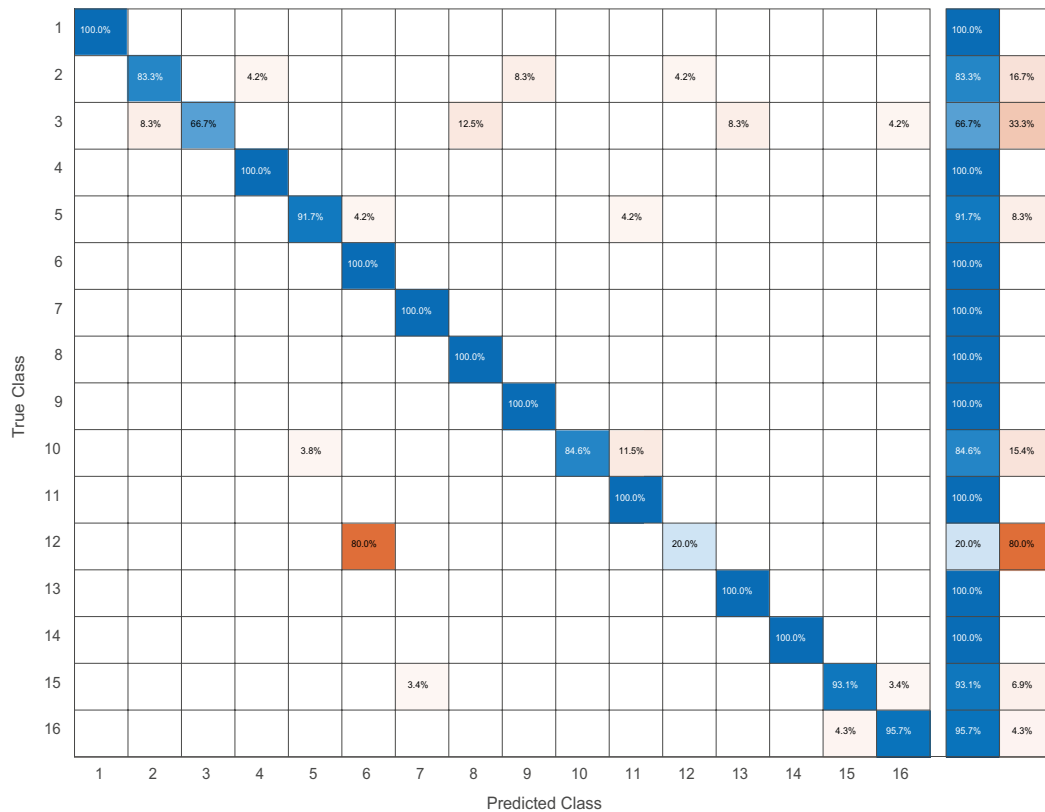
features is accurate. In summary, from Figs. 12 and 13, we conclude that multi-features can characterize the fault information of bearing signal accurately. To achieve the intelligent bearing fault diagnosis, the matrix of multi-features vector obtained were input to the MPA-LSSVM classifier for fault classification and recognition. The optimal LSSVM parameters (c_{best} , g_{best}) are obtained using MPA which are 67.12 and 72.64 respectively. To check the capacity of the proposed approach, it is compared with seven relevant methods. Fault diagnosis results of these methods are exhibited in Table 4. The recognition results and confusion matrix of the VMD-LSSVM method are illustrated in Fig. 14, while the recognition results and confusion matrix of the proposed method WOAGWO-VMD-MPA-LSSVM are illustrated in

Table 4 Comparison data of the comprehensive performance using different methods

Fault diagnosis method	Fitness value	Training accuracy (%)	Test accuracy (%)
VMD-LSSVM		95.21	89.75
WOAGWO-VMD-ELM		96.31	93.53
WOAGWO-VMD-SVM		97.25	94.75
WOAGWO-VMD-LSSVM		97.92	96.47
WOAGWO-VMD-PSO-LSSVM	0.0167	98.54	97.09
WOAGWO-VMD-GA-LSSVM	0.0104	98.96	97.92
WOAGWO-VMD-GWO-LSSVM	0.0042	99.58	98.75
WOAGWO-VMD-MPA-LSSVM	0.0020	100	99.00

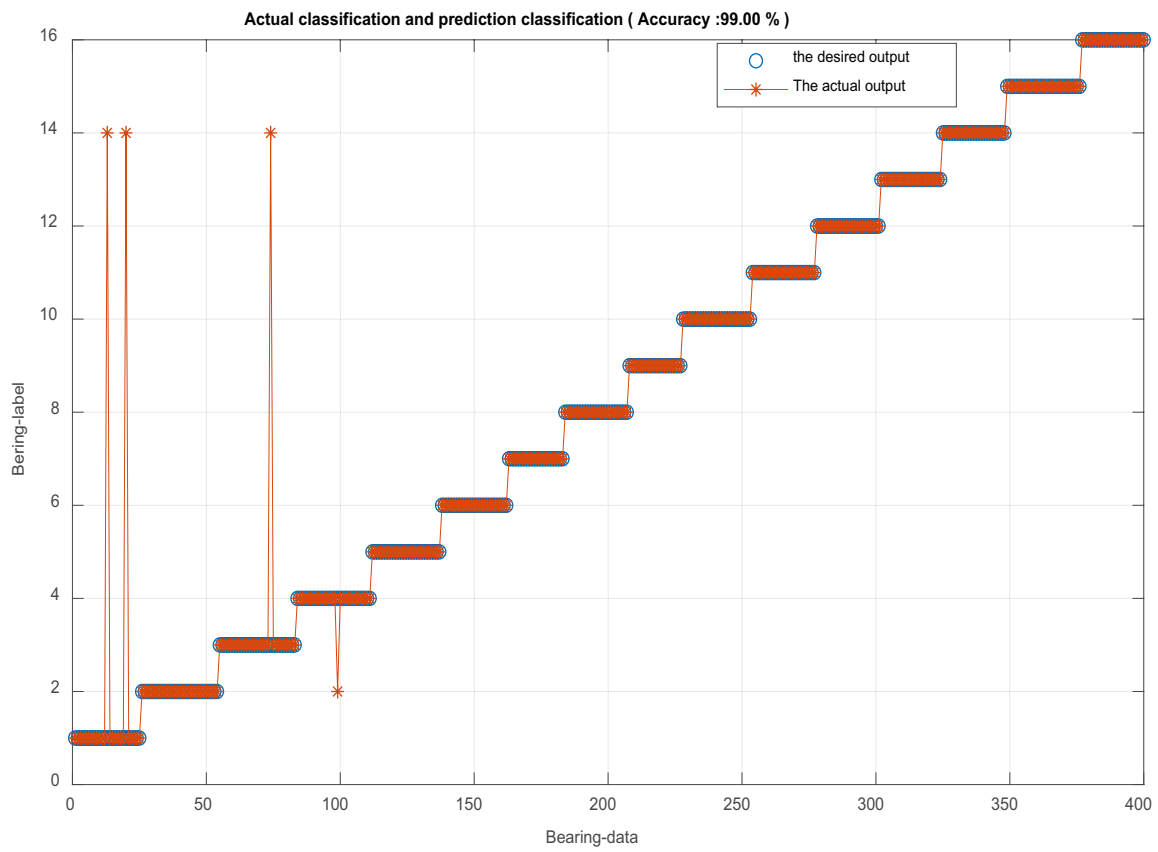


(a)

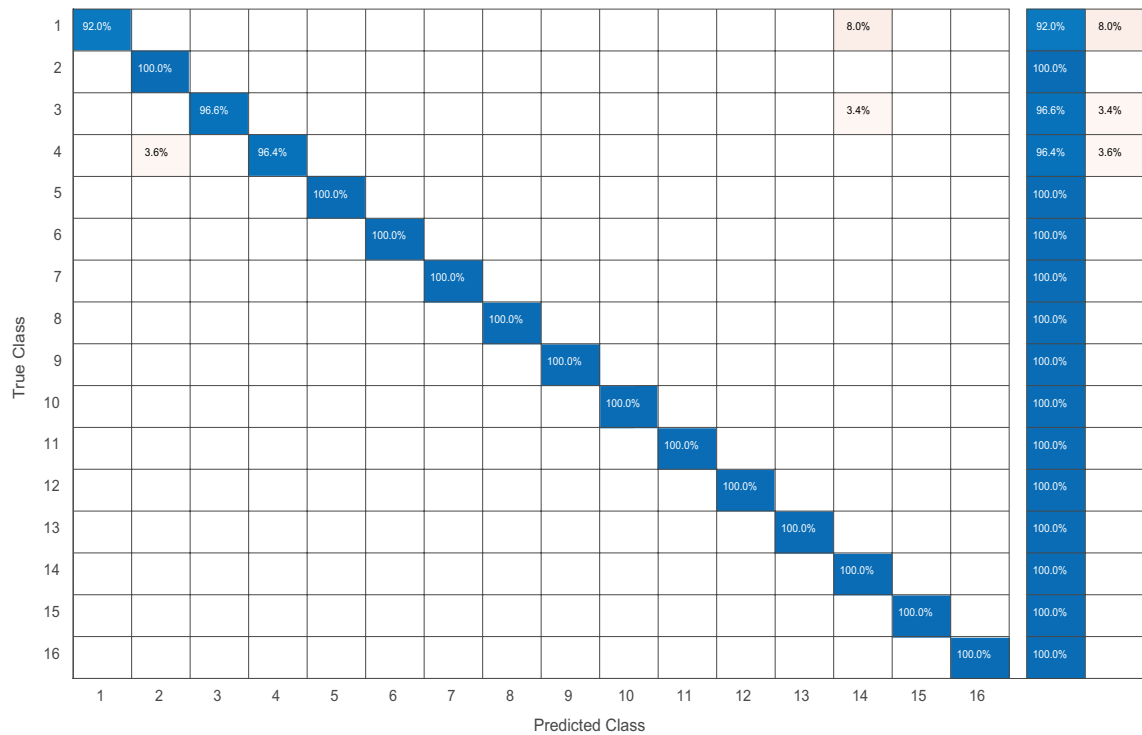


(b)

Fig. 14 (a) Recognition results, and (b) confusion matrix (%) of the VMD-LSSVM method



(a)



(b)

Fig. 15 (a) Recognition results, and (b) confusion matrix (%) of the WOAGWO-VMD-MPA-LSSVM method

Table 5 Comparison of our proposed method with some related published literature for diagnosis of bearing faults

Literature	Feature extraction	Classification	No. of classes	Accuracy
[54]	PSO-VMD+CMPE	ELM	4	87.67%
		RCFOA-ELM	4	98.34%
Our Work	WOAGWO-VMD+multi-features	ELM	16	93.53%
		MPA-LSSVM	16	99.00%
[6]	ESMD+CMWPE	LSSVM	10	88.60%
		MACGSA-LSSVM	10	95.80%
Our Work	WOAGWO-VMD+multi-features	LSSVM	16	96.47%
		MPA-LSSVM	16	99.00%
[21]	GA-VMD+Sample Entropy (SE)	PSO-LSSVM	10	95.04%
		VNWOA-LSSVM	10	97.00%
Our Work	WOAGWO-VMD+multi-features	PSO-LSSVM	16	97.09%
		MPA-LSSVM	16	99.00%
[25]	MIGA-VMD+multi-features	PSO-SVM	16	95.31%
		ANN	16	96.25%
Our Work	WOAGWO-VMD+multi-features	PSO-SVM	16	96.04%
		MPA-LSSVM	16	99.00%

Fig. 15. We can notice from Figs. 14 and 15, and Table 4 that:

1. The diagnosis effect of using the optimized VMD is better than that of the unoptimized VMD method, which indicates the optimized VMD can more accurately extract the fault feature information of the rolling bearing.
2. The proposed method achieved a higher classification accuracy in the diagnosis of various types of faults than the other methods. The classification accuracy of the proposed method reached 99%. In addition, compared with other optimization algorithms, MPA-LSSVM achieves higher classification accuracy, which proves the good performance of MPA algorithm in parameter optimization. Therefore, the validity and superiority of the proposed approach in bearing fault diagnosis are confirmed.

In order to further verify the effectiveness and superiority of the proposed method for diagnosis of bearing faults, we verify the superiority of our method used for feature extraction (WOAGWO-VMD decomposition and multi-features) as first approach by using the fault classifiers ELM, LSSVM, PSO-LSSVM, and PSO-SVM applied in the literature [54], [06], [21], and [25] respectively to diagnose and identify 16 bearing signals. The results are illustrated in Table 5. To ensure superiority of our fault classification method as well, it was compared with the classifiers also used in the previously mentioned literature as second approach. The results of the comparison are shown in Table 5. From the results table we can highlight the following:

1. When we use the same classifier mentioned in references [54], [6], [21], and [25], it turns out that our classification accuracy is always better than the classification accuracy of each reference. The present findings confirm the superiority of our feature extraction method.
2. When comparing the classifier we used MPA-LSSVM with other classifiers (RCFOA-ELM, MACGSA-LSSVM, VNWOA-LSSVM, ANN) employed in the same previous literature, it became clear that the classification accuracy of our classifier is better than the other classifiers. The result now provides evidence to effectiveness and superiority of the proposed fault classification method.

6 Conclusions

This paper presented a novel intelligent rolling bearing fault diagnosis method based on WOAGWO-VMD and MPA-LSSVM. In this method, the WOAGWO-VMD algorithm and multi-features were used for fault feature extraction, and the MPA-LSSVM algorithm for fault classification. In summary, this paper argued the following:

1. Through the simulated signal that was analyzed, the WOAGWO-VMD algorithm can extract the fault feature information of the signal more powerfully and significantly compared to EMD, which is critical for the diagnosis of rolling bearing faults.
2. In the WOAGWO-VMD algorithm, picking the minimum value of the average weighted permutation entropy as the fitness function helped to obtain the optimal parameter combination $[k, \alpha]$ of VMD quickly and efficiently.

3. Through the results obtained, the multi-features composed of dispersion entropy feature, permutation entropy feature, and singular value feature can more accurately characterize the fault information of the bearing signal based on t-SNE dimensionality reduction visualization.
4. The proposed MPA-LSSVM adaptively selects the optimal parameters $[c, g]$ for fault recognition of rolling bearings and demonstrates superior classification accuracy compared with the GWO-LSSVM, PSO-LSSVM, GA-LSSVM, SVM, and ELM.
5. The effectiveness and feasibility of the proposed method in this paper is verified by using experimental data of rolling bearing fault vibration signal. It can be also applied to the fault diagnosis of other mechanical parts for upcoming works.

Availability of data and material The data and materials that support the findings of this study are available on request.

Code availability All codes that support the findings of this study are available on request.

Declarations

Consent to participate Informed consent was obtained from all individual participants included in this study.

Consent for publication All individual participants declare that they accept to submit this paper for publication.

Conflict of interest The authors declare no competing interests.

References

1. Wang X, Sui G, Xiang J et al (2020) Multi-domain extreme learning machine for bearing failure detection based on variational modal decomposition and approximate cyclic correntropy. *IEEE Access* 8:197711–197729
2. Li Y, Wang J, Duan L et al (2019) Association rule-based feature mining for automated fault diagnosis of rolling bearing. *Shock Vib* 2019:1–12
3. Yan X, Liu Y, Jia M (2020) A fault diagnosis approach for rolling bearing integrated SGMD, IMSDE and multiclass relevance vector machine. *Sensors* 20:4352
4. Liu F, Gao J, Liu H (2020) A fault diagnosis solution of rolling bearing based on MEEMD and QPSO-LSSVM. *IEEE Access* 8:101476–101488
5. Mohd Ghazali MH, Rahiman W (2021) Vibration analysis for machine monitoring and diagnosis: a systematic review. *Shock Vib* 2021:1–25
6. He C, Wu T, Liu C et al (2020) A novel method of composite multiscale weighted permutation entropy and machine learning for fault complex system fault diagnosis. *Measurement* 158:107748
7. Ge J, Niu T, Xu D et al (2020) A rolling bearing fault diagnosis method based on EEMD-WSSST signal reconstruction and multiscale entropy. *Entropy* 22:290
8. Zhu H, He Z, Wei J et al (2021) Bearing fault feature extraction and fault diagnosis method based on feature fusion. *Sensors* 21:2524
9. Lee C-Y, Hung C-H (2021) Feature ranking and differential evolution for feature selection in brushless DC motor fault diagnosis. *Symmetry* 13:1291
10. Grover C, Turk N (2020) Rolling element bearing fault diagnosis using empirical mode decomposition and Hjorth parameters. *Procedia Comput Sci* 167:1484–1494
11. Nishat Toma R, Kim C-H, Kim J-M (2021) Bearing fault classification using ensemble empirical mode decomposition and convolutional neural network. *Electronics* 10:1248
12. Amarouayache IIE, Saadi MN, Guersi N et al (2020) Bearing fault diagnostics using EEMD processing and convolutional neural network methods. *Int J Adv Manuf Technol* 107:4077–4095
13. Babouri MK, Ouelaa N, Kebabsa T et al (2021) Diagnosis of mechanical defects using a hybrid method based on complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and optimized wavelet multi-resolution analysis (OWMRA): experimental study. *Int J Adv Manuf Technol* 112:2657–2681
14. Han M, Wu Y, Wang Y et al (2021) Roller bearing fault diagnosis based on LMD and multi-scale symbolic dynamic information entropy. *J Mech Sci Technol* 35:1993–2005
15. Li Z, Ma J, Wang X et al (2019) MVMD-MOMEDA-TEO model and its application in feature extraction for rolling bearings. *Entropy* 21:331
16. Ding J, Xiao D, Li X (2020) Gear fault diagnosis based on genetic mutation particle swarm optimization VMD and probabilistic neural network algorithm. *IEEE Access* 8:18456–18474
17. Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. *IEEE Trans Signal Process* 62:531–544
18. Civera M, Surace C (2021) A comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark. *Sensors* 21:1825
19. Ye M, Yan X, Jia M (2021) Rolling bearing fault diagnosis based on VMD-MPE and PSO-SVM. *Entropy* 23:762
20. Fu W, Shao K, Tan J et al (2020) Fault diagnosis for rolling bearings based on composite multiscale fine-sorted dispersion entropy and SVM with hybrid mutation SCA-HHO algorithm optimization. *IEEE Access* 8:13086–13104
21. Li J, Chen W, Han K et al (2020) Fault diagnosis of rolling bearing based on GA-VMD and improved WOA-LSSVM. *IEEE Access* 8:166753–166767
22. Yao G, Wang Y, Benbouzid M et al (2021) A hybrid gearbox fault diagnosis method based on GWO-VMD and DE-KELM. *Appl Sci* 11:4996
23. Zhang Q, Chen S, Fan ZP (2021) Bearing fault diagnosis based on improved particle swarm optimized VMD and SVM models. *Adv Mech Eng* 13:168781402110284
24. Feng G, Wei H, Qi T et al (2021) A transient electromagnetic signal denoising method based on an improved variational mode decomposition algorithm. *Measurement* 184:109815
25. Liang T, Lu H (2020) A novel method based on multi-island genetic algorithm improved variational mode decomposition and multi-features for fault diagnosis of rolling bearing. *Entropy* 22:995
26. Zhang W, Wang Y, Tan Y et al (2021) Application of adaptive local iterative filtering and permutation entropy in gear fault recognition. *Math Probl Eng* 2021:1–12
27. Jin Z, He D, Chen Y et al (2021) Research on fault diagnosis method of train rolling bearing based on variational modal decomposition and bat algorithm-support vector machine. *J Phys Conf Ser* 1820:012170

28. Cheng H, Zhang Y, Lu W et al (2019) A bearing fault diagnosis method based on VMD-SVD and Fuzzy clustering. *Int J Patt Recogn Artif Intell* 33:1950018
29. Tarek K, Abdelaziz lakehal, Zoubir C, et al (2021) Optimized multi layer perceptron artificial neural network based fault diagnosis of induction motor using vibration signals. *Diagnostyka* 22:65–74
30. Sikder N, Mohammad Arif AS, Islam MMM et al (2021) Induction motor bearing fault classification using extreme learning machine based on power features. *Arab J Sci Eng* 46:8475–8491
31. Wang Y, Xu C, Wang Y et al (2021) A comprehensive diagnosis method of rolling bearing fault based on CEEMDAN-DFA-improved wavelet threshold function and QPSO-MPE-SVM. *Entropy* 23:1142
32. Lu Y, Xie R, Liang SY (2020) CEEMD-assisted kernel support vector machines for bearing diagnosis. *Int J Adv Manuf Technol* 106:3063–3070
33. Ma J, Li C, Zhang G (2021) Rolling bearing fault diagnosis based on deep learning and autoencoder information fusion. *Symmetry* 14:13
34. Gao X, Wei H, Li T et al (2020) A rolling bearing fault diagnosis method based on LSSVM. *Adv Mech Eng* 12:168781401989956
35. Heidari M, Homaei H, Golestanian H (2015) Fault diagnosis of gearboxes using LSSVM and WPT 5:10
36. Gao S, Li T, Zhang Y (2020) Rolling bearing fault diagnosis of PSO-LSSVM based on CEEMD entropy fusion. *Trans Can Soc Mech Eng* 44:405–418
37. Nguyen V, Nguyen D et al (2021) Feature selection based on EEMD-LaS and optimized BSO-LSSVM classifier model in bearing fault diagnosis. *IJIES* 14:247–258
38. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Computat* 1:67–82
39. Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51–67
40. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
41. Mohammed H, Rashid T (2020) A novel hybrid GWO with WOA for global numerical optimization and solving pressure vessel design. *Neural Comput Appl* 32:14701–14718
42. Yan X, Liu Y, Zhang W et al (2020) Research on a novel improved adaptive variational mode decomposition method in rotor fault diagnosis. *Appl Sci* 10:1696
43. Li H, Fan B, Jia R et al (2020) Research on multi-domain fault diagnosis of gearbox of wind turbine based on adaptive variational mode decomposition and extreme learning machine algorithms. *Energies* 13:1375
44. Zhou F, Yang X, Shen J et al (2020) Fault diagnosis of hydraulic pumps using PSO-VMD and refined composite multiscale fluctuation dispersion entropy. *Shock Vib* 2020:1–13
45. Fadlallah B, Chen B, Keil A et al (2013) Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information. *Phys Rev E* 87:022911
46. Liang T, Lu H, Sun H (2021) Application of parameter optimized variational mode decomposition method in fault feature extraction of rolling bearing. *Entropy* 23:520
47. Grover C, Turk N (2020) Optimal statistical feature subset selection for bearing fault detection and severity estimation. *Shock Vib* 2020:1–18
48. Rostaghi M, Azami H (2016) Dispersion entropy: a measure for time-series analysis. *IEEE Signal Process Lett* 23:610–614
49. Bandt C, Pompe B (2002) Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett* 88:174102
50. Faramarzi A, Heidarinejad M, Mirjalili S et al (2020) Marine predators algorithm: a nature-inspired metaheuristic. *Expert Syst Appl* 152:113377
51. Fu W, Wang K, Zhang C et al (2019) A hybrid approach for measuring the vibrational trend of hydroelectric unit with enhanced multi-scale chaotic series analysis and optimized least squares support vector machine. *Trans Inst Meas Control* 41:4436–4449
52. Wu T, Liu CC, He C (2019) Fault diagnosis of bearings based on KJADE and VNWOA-LSSVM algorithm. *Math Probl Eng* 2019:1–19
53. CWRU (2008) Case Western Reserve University Bearing Date Center Website; CWRU: Cleveland, OH, USA. Available online: <https://engineering.case.edu/bearingdatacenter/download-data-file>
54. He C, Wu T, Gu R et al (2021) Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm – extreme learning machine. *Measurement* 173:108636
55. Zhu X, Huang Z, Chen J et al (2021) Rolling bearing fault diagnosis method based on VMD and LSSVM. *J Phys Conf Ser* 1792:012035

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.