ORIGINAL PAPER

A new model for bearing fault diagnosis based on optimized variational mode decomposition correlation coefficient weight threshold denoising and entropy feature fusion

Jing Yang · Yanping Bai · Yunyun Cheng · Rong Cheng · Wendong Zhang · Guojun Zhang

Abstract For the bearing fault diagnosis in small sample cases, a new model for signal denoising and entropy feature fusion (EFF) based on the wild horse optimizer (WHO) optimized variational mode decomposition (VMD) and correlation coefficient weight threshold (CCWT) is proposed (WHO–VMD–CCWT–EFF). For signal denoising, we first take the power spectrum entropy as the fitness function, and the WHO is used to optimize VMD parameters. Secondly, IMFs with correlation coefficient values less than 0.2 are removed and the correlation coefficient values as weights are applied to the corresponding IMF components, and then reconstruct them. Then, the refined composite multiscale dispersion entropy (RCMDE), refined composite multiscale fluctuation dispersion entropy (RCMFDE), refined composite multivariate generalized multiscale fuzzy entropy (RCmvMFE), refined composite multivariate generalized multiscale sample entropy (RCmvMSE), and multiscale permutation entropy (MPE) of the signal are calculated and fused. Finally, the Fisher discriminant classifier is used as the model for fault diagnosis. The proposed model achieves an accuracy of over 99% in 12 single working conditions and 30 multiple working conditions experiments using the case western reserve university (CWRU) dataset and the Paderborn dataset. Compared with existing feature fusion methods, the WHO–VMD–CCWT–EFF model only integrates five selected features, and can achieve accurate diagnosis of bearing faults in small sample experiments with 42 different artificial and real damages. This indicates that the model has good generalization ability between different datasets and working conditions.

Keywords Fault diagnosis · Denoising · Entropy feature fusion · Single working condition · Multiple working conditions · Small sample

J. Yang · Y. Cheng
School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Y. Bai (✉) · R. Cheng
School of Mathematics, North University of China, Taiyuan 030051, China
e-mail: baiyp666@163.com

W. Zhang · G. Zhang
State Key Laboratory of Dynamic Testing Technology, North University of China, Taiyuan 030051, China

1 Introduction

In recent years, the issues related to health monitoring in the mechanical industry urgently need to be addressed. Some scholars have begun to spend time monitoring the health status of structures such as beams and trusses [1, 2]. Of course, due to the development of automated machinery and equipment, people are increasingly interested in fault diagnosis of their components, such as bearings [3–5]. At present, bearing fault diagnosis mainly includes three

categories: signal processing and analysis, traditional fault diagnosis methods based on feature extraction, and self-extraction feature diagnosis methods based on deep learning [6]. In the actual research process, the three types of methods play their respective advantages and complement each other. From the perspective of signal processing, Li y et al. proposed a new time–frequency analysis (TFA) post-processing algorithm called local maximum high order time iterative synchrosqueezing (LHTIS) [7], and proved the effectiveness of this method by analyzing and processing fault signals. Haiyang Pan et al. [8] proposed multi-class fuzzy support matrix machine and successfully applied it to roller bearing fault diagnosis. In one study, VMD parameters and kernel fuzzy c-means (KFCM) were optimized respectively, and then the bearing fault types of small samples were identified [9]. Another study proposed a new method for bearing fault diagnosis based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm [10]. Another study proposed ensemble self-taught learning convolutional auto-encoders (STL-CAEs) [11], which can effectively solve the problem of few labeled data. The method of combining dempster-shafer (DS) evidence theory with support vector machines (SVM) has also appeared in other bearing fault diagnosis research [12]. A new fault diagnosis method called RSG is proposed in the literature [13]. Yang S et al. [14] combined two-dimensional convolutional neural network (2DCNN) feature extractor and random forest (RF) classifier to establish a fault diagnosis model for the problem of high-speed bearings in offshore wind turbines. And the experimental results were 99.5% when 700 training samples and 300 test samples were input to the model. Literature [15] intercepted 30 training samples and 30 test samples respectively, and then calculated and mixed the time-domain and frequency-domain features of the samples. Finally, the deep neural network was used to identify the fault type with an accuracy of 99.1%. In [16], the weighted signal difference average (WSDA) as a new fitness function was proposed to optimize VMD, and a one-dimensional neural network was used for rolling bearing fault diagnosis. In the experiment with 5000 samples as the training set and 1000 samples as the test set, the accuracy of bearing fault diagnosis is 99.6%. In [17], the few-shot learning method was successfully applied to the fault diagnosis of rolling bearings, and the experimental verification was carried out under the mixed working conditions. The results show that when the number of training samples is 60, 200, 900, and 19,800, and the number of test samples is 75, the accuracy rates are 82.8%, 94.32%, 98.55%, and 99.77%, respectively.

It can be seen from previous research that deep learning is widely used in the field of bearing fault diagnosis. Of course, this is also due to its advantages in feature extraction and classification. However, the accuracy of deep learning is often at the cost of increasing the number of training samples. In addition, some studies are only carried out for a single working condition, and the effect of applying the proposed model to other working conditions remains to be verified. Based on the above analysis, the WHO-VMD-CCWT-EFF is proposed in this paper. In order to verify the usability and universality of the model, the CWRU dataset [18] and Paderborn dataset [19] are used for various single and multiple working condition experiments. The experimental results indicate that WHO-VMD-CCWT-EFF can achieve good results with only 10 training samples and 90 test samples. The main contributions of this paper are summarized as follows:

1. A correlation coefficient weight threshold denoising method is proposed to denoise the fault signal decomposed by VMD.
2. To extract better classification features, an entropy feature fusion method is proposed and the new bearing fault diagnosis method named WHO-VMD-CCWT-EFF is verified in the experiment.
3. A new deviation metric is used to measure the stability of the model and validated in various experiments.
4. On the basis of completing the experiment of single working condition, the experiment of mixing multiple working conditions is carried out, and good results are obtained. And the WHO-VMD-CCWT-EFF is still applicable and stable in the case of a small amount of data.

The rest of this paper is organized as follows: Section 2 introduces the theoretical basis related to the model. Section 3 introduces the framework of the model. The CWRU dataset and the Paderborn dataset are used for the experiments and analyses in Sect. 4. Finally, Sect. 5 gives the conclusion.

2 Theoretical backgrounds

2.1 WHO-VMD

VMD [20] is a non-recursive and adaptive signal processing algorithm proposed based on algorithms such as empirical mode decomposition (EMD). It aims to dissect the original signal in the frequency domain and decompose it into intrinsic mode function (IMF) with limited bandwidth and center frequency. WHO is a meta-heuristic optimization algorithm proposed by Iraj Naruei [21]. Similar optimization algorithms, such as the improved grey wolf optimization (IGWO), ant lion optimizer (ALO), and marine predator algorithm (MPA), have also been well applied to various structural detection. This article selects the WHO to optimize VMD parameters [22–24].

The algorithm is mainly inspired by the special behavior of wild horses that is different from other animals, that is, foals leave their parent groups before puberty to join other parent groups to avoid mating between relatives. In addition to mating behaviors, wild horses renew their position through social behaviors such as grazing behavior, group leadership, and exchange and selection of leaders.

In the process of VMD decomposition of bearing fault signal, it is found that the number of IMFs in VMD, that is, the number of decomposed layers k and the value of multiplication factor α will directly affect the decomposition effect. To select the relatively optimal parameters, the parameter (k, α) of VMD is optimized by WHO, as shown in Fig. 1.

Step1: Set the parameters of WHO. The total number of wild horses N = 30, the maximum number of iterations Max_iter = 30, the crossover ratio PC = 0.13, the percentage of stallions in the group population PS = 0.2, the number of stallions Nstallion = N*PS, the number of foals in each group Nfoal = (N-Nstallion)/Nstallion. The parameters to be optimized $\alpha \in [100, 2000]$, $k \in [4, 8]$, $\alpha, k \in Z$.

Step2: Create populations, select leaders, and calculate fitness function values.

Step3: Search and update according to grazing behavior if Rand > PC, otherwise update by mating behavior. It should be noted that Rand is a random number with uniform distribution in the range [0, 1].

Step4: Group leaders as well as stallions are updated, respectively.

Step5: Determine whether the number of iterations is reached, if so, output (k, α), otherwise return to step 3.

2.2 Power spectrum entropy

The power spectral entropy [25] represents the uncertainty of signal energy under power spectral partitioning, which is a quantitative description of the complexity of signal energy distribution in the frequency domain. In actual industrial environments, bearing fault signals are collected in environments with different noise sources, resulting in complex frequency components of bearing fault signals. In order to effectively reflect the fault characteristics contained in each frequency component, the power spectrum entropy is used as the fitness function of the optimization algorithm. When the entropy value of the power spectrum is small, the frequency components in the signal are simple, and the power spectrum is concentrated on some frequency components, which can reflect the characteristics of the fault signal. In addition, the power spectral entropy values of each IMF component in the same fault state after VMD decomposition are relatively stable, and the power spectral entropy values vary in different fault states. This further proves that the power spectral entropy is suitable as the fitness function of the optimization algorithm.

Step1: Define the original fault signal sequence as $x(t) = \{x(1), x(2), x(3), \ldots x(L)\}$

$$P(i) = \frac{|x(w)|^2}{2\pi L} \tag{1}$$
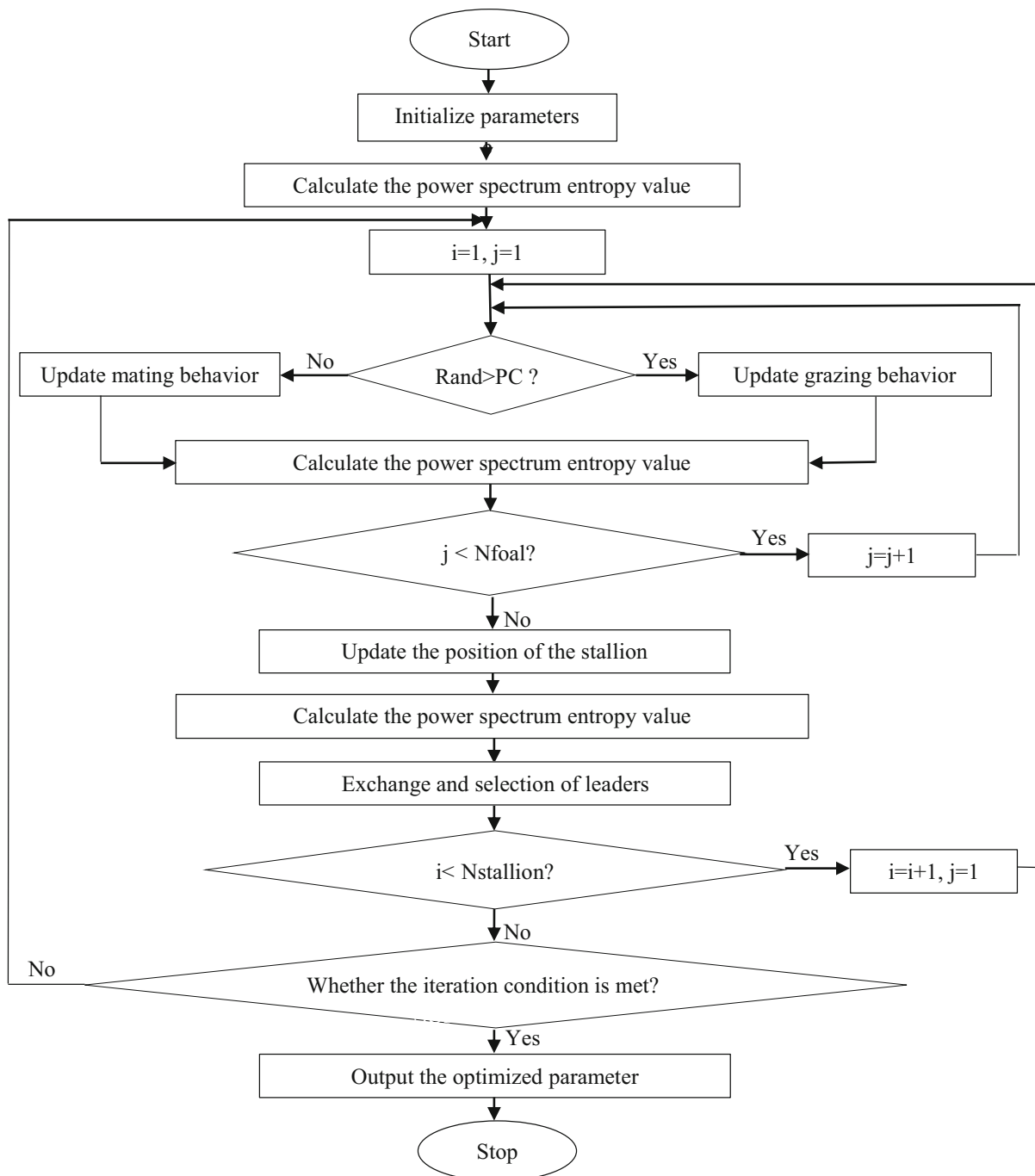
Fig. 1 Flow chart of WHO-VMD algorithm

where L is the length of the signal, $P(i)$ is the power spectrum of the signal. $x(w)$ is the Fourier transform of the signal.

Step 2: Obtain the power spectral density distribution function by normalization:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N} P(i)} \quad i = 0, 1, 2....N \qquad (2)$$

where N is the number of frequency components in the Fourier transform.

Step 3: Define the power spectrum entropy through the power spectral density distribution function as:

$$H = -\sum_{i=1}^{N} p(i)\log p(i) \qquad (3)$$

2.3 Correlation coefficient

The correlation coefficient is a description of the similarity between two random signals or deterministic signals. After the VMD decomposition of the bearing fault signal, the correlation between each IMF and the original bearing fault signal can be judged by calculating the correlation coefficient value. Then, it can be inferred from the correlation coefficient whether the IMF contains the main features of the original signal. Generally speaking, the closer the absolute value of the correlation coefficient is to 1, the higher the degree of correlation between the two, and the more obvious the features of the original signal contained in the IMF. The correlation coefficient R_k between the k-th IMF and the original signal is defined as:

$$R_k = \frac{E(u_k(t)f(t)) - E(u_k(t))E(f(t))}{\sqrt{D(u_k(t))}\sqrt{D(f(t))}} \qquad (4)$$

where $f(t)$ is the original signal and $u_k(t)$ is the k-th IMF. E and D represent expected values and variance.

2.4 Entropy features

2.4.1 RCMDE

RCMDE [26, 27] was first proposed and applied to biomedical signals in 2017. RCMDE is improved by multi-scale and coarse-graining based on dispersion entropy. The specific calculation steps are as follows:

Step 1: For the sequence X of length N, divide it into segments of length τ. The average value of each segment is calculated and arranged to obtain a coarse-grained sequence.

$$x_{k,j}^{\tau} = \frac{1}{\tau} \sum_{b=k+\tau(j-1)}^{k+\tau j-1} X_b, 1 \leq j \leq \left\lfloor \frac{N}{\tau} \right\rfloor, 1 \leq k \leq \tau \qquad (5)$$

where $x_k^{\tau} = \left\{ x_{k,1}^{\tau}, x_{k,2}^{\tau}, ... \right\}$ is the k-th coarse-grained sequence at the τ scale.

Step 2: Map the time series $x_{k,j}^{\tau}$ to $y_{k,j}^{\tau}$ by Eq. (6)

$$y_k^{\tau} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_k^{\tau}} e^{\frac{-(t-u)^2}{2\sigma^2}} dt \qquad (6)$$

where u represents the mathematical expectation of sequence y_k^{τ}, and σ represents the variance of sequence y_k^{τ}.

Step 3: Map the time series y_k^{τ} to Z_j^c by Eq. (7)

$$Z_j^c = \text{Round}\left(c \cdot y_k^{\tau} + 0.5\right) \qquad (7)$$

where Round() represents the rounding function, and c represents the number of categories.

Step 4: Calculate the embedding vector by Eq. (8).

$$z_i^{m,c} = \left\{ z_i^c, z_{i+d}^c, \cdots, z_{i+(m-1)d}^c \right\}$$

$$i = 1, 2, \cdots, N - (m-1)d \qquad (8)$$

where m is the embedding dimension and d is the time delay.

Step 5: Calculate the dispersion patterns and its corresponding probability. Assuming that $z_i^c = v_0$, $z_{i+d}^c = v_1$, and $z_{i+(m-1)d}^c = v_{m-1}$, the dispersion pattern corresponding to $z_i^{m,c}$ is $\pi_{v_0 v_1 \cdots v_{m-1}}$. Calculate the probability corresponding to the dispersion pattern according to Formula (9).

$$p(\pi_{v_0 v_1 \cdots v_{m-1}}) = \frac{\text{Number}(\pi_{v_0 v_1 \cdots v_{m-1}})}{N - (m-1)d} \qquad (9)$$

Step 6: Calculate the average value $\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}})$ of the probability of the dispersion pattern, and obtain the RCMDE value through $\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}})$.

$$\text{RCMDE}\,(x_k^{\tau}, m, c, d, \tau) = -\sum_{\pi=1}^{c^m} \overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}}) \ln(\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}}))$$

$$(10)$$

2.4.2 RCMFDE

RCMFDE is based on the study of Azami H et al. [28, 29] for dispersion entropy. The fluctuation dispersion entropy is superior to the dispersion entropy in that it takes into account the volatility of the time series while maintaining stable performance and less computation. Like RCMDE, RCMFDE obtains the dispersion pattern by formula (5–8). Calculate the probability corresponding to the dispersion pattern according to Formula (11).

$$p(\pi_{v_0 v_1 \cdots v_{m-1}}) = \frac{\text{count}\left\{\begin{array}{l} i | i \le N - (m-1)d, \\ z_i^{m,c} \text{pattern } \pi_{v_0 v_1 \cdots v_{m-1}} \end{array}\right\}}{N - (m-1)d} \quad (11)$$

Among them, count () is the number of maps from $z_i^{m,c}$ to $\pi_{v_0 v_1 \cdots v_{m-1}}$.

Calculate the average value of the dispersion pattern probabilities at scale τ, and the RCMFDE is obtained through $\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}})$.

$$\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}}) = \frac{1}{\tau} \sum_{k=1}^{\tau} p_k \quad (12)$$

$$E_{\text{RCMFD}}\left(x_k^\tau, m, c, d, \tau\right) = -\sum_{\pi=1}^{(2c-1)m-1} \overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}}) \cdot \ln[\overline{p}(\pi_{v_0 v_1 \cdots v_{m-1}})] \quad (13)$$

2.4.3 RCmvMFE

RCmvMFE [30] is a tool proposed in 2017 to analyze the complexity of multi-channel signals. The detailed description of RCmvMFE is as follows:

Step 1: For a multivariate signal $Y = \{y_{k,b}\}_{,b=1}^{C}$ containing p signals with a length of C, the coarse-grained operations are performed to obtain a time series, represented as $z_\alpha^{(\beta)} = \left\{x_{\alpha,k,i}^{(\beta)}\right\}$, where β is the time series scale.

$$x_{\alpha,k,i}^{(\beta)} = \frac{1}{\beta} \sum_{b=(i-1)\beta}^{i\beta+\alpha-1} y_{k,b}$$
$$1 \le i \le \left\lfloor \frac{C}{\beta} \right\rfloor = N, 1 \le k \le p, 1 \le \alpha \le \beta \quad (14)$$

Step 2: The multivariate embedded reconstruction is used.

$$X_m(i) = \big[x_{1,i}, x_{1,i+\tau_1}, \ldots, x_{1,i+(m_1-1)\tau_1}, x_{2,i},$$
$$x_{2,i+\tau_2}, \ldots, x_{2,i+(m_2-1)\tau_2}, \ldots, \quad (15)$$
$$x_{P,i}, x_{P,i+\tau_P}, \ldots, x_{P,i+(m_P-1)\tau_P}\big]$$

where $M = [m_1, m_2, \ldots m_p]$, $\tau = [\tau_1, \tau_2, \ldots \tau_P]$ are the embedding dimension and delay time, respectively, $n = \max\{M\} \times \max\{\tau\}$, $i = 1, 2, \ldots N - n$.

Step 3: Calculate the distance between $X_m(i)$ and $X_m(j)$, where $i \ne j$.

$$d[X_m(i), X_m(j)] = \max_{l=1,2,\ldots,m} \{|x(i+l-1) - x(j+l-1)|\} \quad (16)$$

Step 4: According to the given threshold r and fuzzy membership function $\theta(d,r)$, $\phi^m(r)$ with the embedding dimension m can be obtained:

$$\theta(d,r) = \exp\left(\frac{-(d)^{fp}}{r}\right)$$

$$\phi^m(r) = \frac{1}{(N-n)} \sum_{i=1}^{N-n} \frac{\sum_{j=1,i\ne j}^{N-n} \exp\left(\frac{-(d[X_m(i),X_m(j)])^{fp}}{r}\right)}{N-n-1} \quad (17)$$

Step 5: Let $m = m + 1$ and repeat steps 2–4. Calculate the average values $\overline{\phi}_{\beta,\alpha}^m$ and $\overline{\phi}_{\beta,\alpha}^{m+1}$ of Eq. (17). Then RCmvMFE can be calculated by Eq. (18)

$$\text{RCmvMFE}(Y, \beta, M, n, r) = -\ln\left(\frac{\overline{\phi}_{\beta,\alpha}^{m+1}}{\overline{\phi}_{\beta,\alpha}^m}\right) \quad (18)$$

2.4.4 RCmvMSE

The probability calculation method for RCmvMSE [31] varies when the embedding dimension is m.

$$B_i^m(r) = (N - n - 1)^{-1} P_i \quad (19)$$

$$B^m(r) = (N - n)^{-1} \sum_{i=1}^{N-n} B_i^m(r) \quad (20)$$

Let $m = m + 1$, and repeat the above steps to get $B^{m+1}(r)$. Calculate the mean values $\overline{B}_{\beta,\alpha}^m$ and $\overline{B}_{\beta,\alpha}^{m+1}$ in m

and $m + 1$ dimensions. RCmvMSE can be calculated by Eq. (21)

$$RCmvMSE(Y, \beta, M, n, r) = -\ln\left(\frac{\overline{B}_{\beta,\alpha}^{m+1}}{\overline{B}_{\beta,\alpha}^{m}}\right) \qquad (21)$$

2.4.5 MPE

To better study and analyze the dynamic characteristics of EEG, Ouyang G et al. [32] proposed a multiscale permutation entropy based on permutation entropy.

Step 1: A new time series is obtained by coarse-graining an original sequence Y of length N, where τ is the scale factor.

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, 1 \le j \le \left\lfloor \frac{N}{\tau} \right\rfloor \qquad (22)$$

Step 2: The phase space reconstruction is applied with $y^{(\tau)}$ to obtain the time series X_i.

$$X_i = \left(y_i, y_{i+\lambda}, \ldots, y_{i+(m-1)\lambda}\right) \qquad (23)$$

where m is the embedding dimension and λ is the delay time.

Step 3: The X_i is sorted in ascending order to generate a sequence of position indexes. For any kind of X_i, there are $m!$ permutations. The probability of each permutation is calculated according to Eq. (24).

$$P(\omega) = \frac{T(\omega)}{N - (m-1)\lambda} \qquad (24)$$

where $T(\omega)$ is the number of occurrences of permutation ω, $1 \le \omega \le m!$

Step 4: Define the multiscale permutation entropy by Eq. (25).

$$H_{PE} = -\sum P(\omega) \ln P(\omega)$$

$$H_{MPE} = [H_{P1}, H_{P2}...H_{P\tau}] \qquad (25)$$

2.5 Deviation

In order to measure the stability of the model, a new deviation indicator is defined. Suppose that for experiment A, the result of the i-th repeated experiment is A_i, $i = 1,2,3......N$.

$$Deviation_A = \max[A_i] - \min[A_i] \qquad (26)$$

where $\max[A_i]$ is to find the maximum value of A_i, $\min[A_i]$ is to find the minimum value of A_i.

3 The WHO–VMD–CCWT–EFF

The framework of the WHO–VMD–CCWT–EFF model is shown in Fig. 2. Three parts of fault signal denoising, feature extraction and fusion, and feature classification are included in this method.

Denoising: First, the VMD optimized by the WHO algorithm decomposes various bearing fault signals into IMFs; Secondly, the correlation coefficients between each IMF and the original bearing fault signal are calculated; Then, selecting IMFs with correlation function values greater than 0.2 with the original bearing signal; Finally, the correlation coefficients are used as the weight to multiply the corresponding IMFs to reconstruct the fault signal.

Feature extraction: RCMDE, RCMFDE, RCMvMFE, RCMvMSE, and MPE are extracted from the fault signal after denoising, and then the five entropy features extracted are fused.

Classification: The feature samples of the fault signals are divided into training and test sets according to the experimental requirements, and then the fault signals are classified by Fisher classifier.

4 Experimental analyses

In order to verify the effectiveness of the method proposed in this paper, two classical public datasets are used in the experiments.
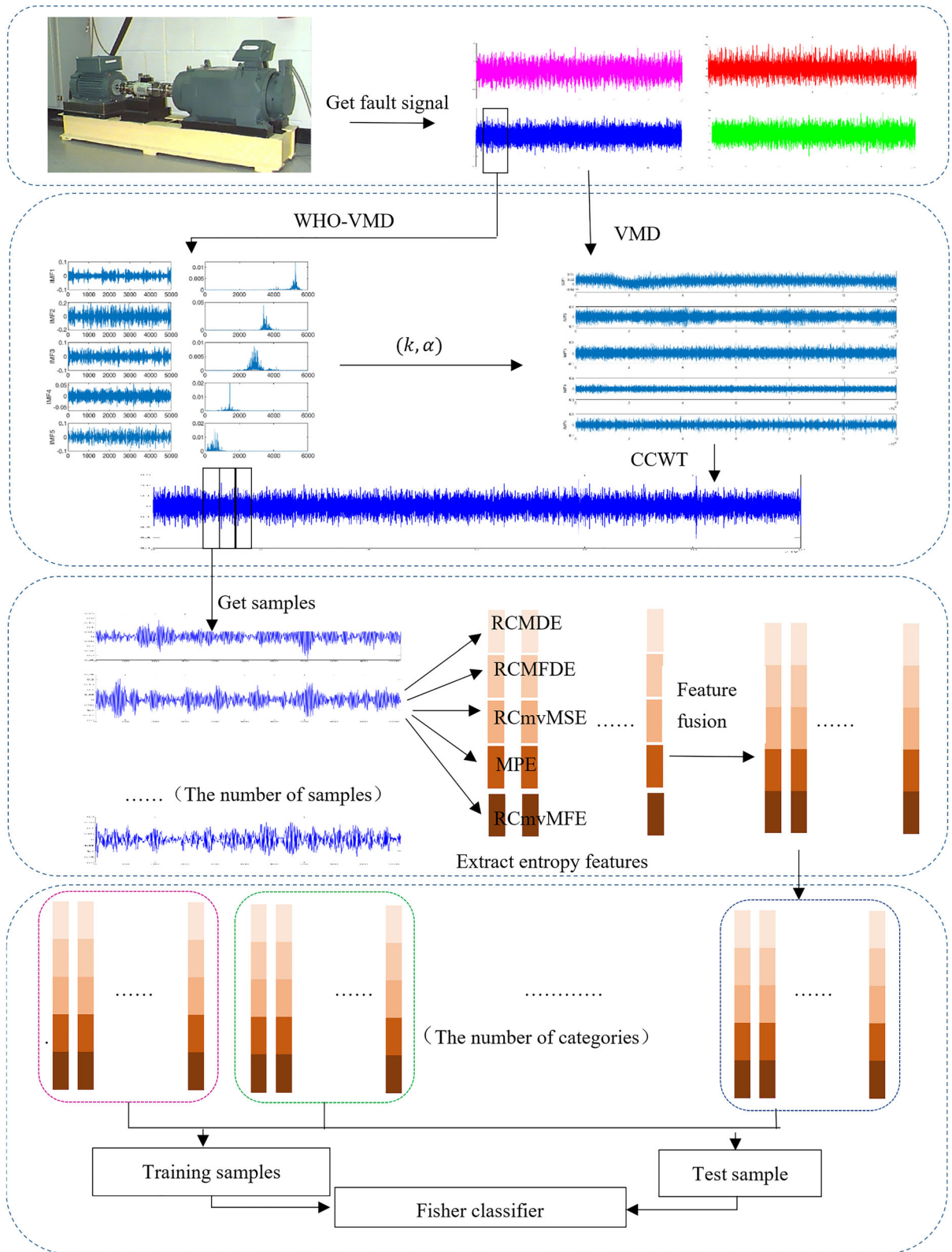
4.1 Analysis of bearing fault signal

4.1.1 WHO–VMD

To address the issue of VMD decomposition being greatly affected by parameters (k, α), the WHO is used to optimize the parameters. Figure 3 shows the convergence curve of partial artificial damage and real damage under the Paderborn dataset. It can be seen that the number of iterations required to achieve

convergence for each fault signal is inconsistent. Therefore, in order to consider as many fault signals as possible, 30 is chosen as the number of iterations. Similarly, the number of search agents is set to 30. From Fig. 4, it can be seen that when the number of search agents is 30, the convergence of the fault signal is better.

Taking the 0.3556 mm outer race bearing fault at 0HP as an example, the fitness curve of the WHO-VMD is shown in Fig. 5. As can be seen from the Fig. 5, the value of the fitness function for the first iteration is 7.2145×10^{-4}. The value of the fitness function is 7.0273×10^{-4} after a small decrease, and

(a) artificial damage

(b) real damage
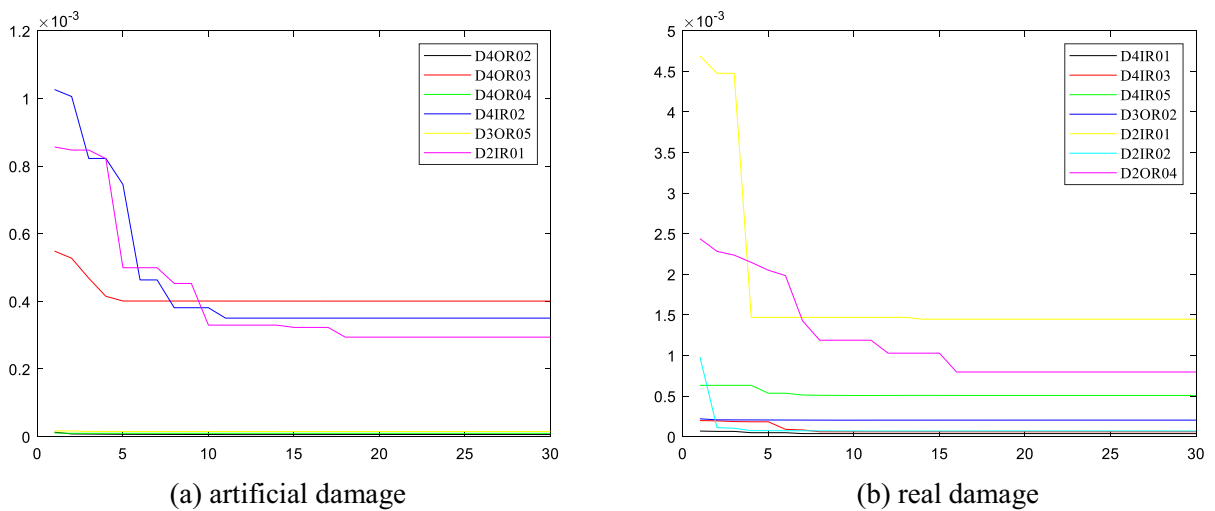
Fig. 3 Convergence curve of partial fault signal under the Paderborn dataset

(a) artificial damage

(b) real damage
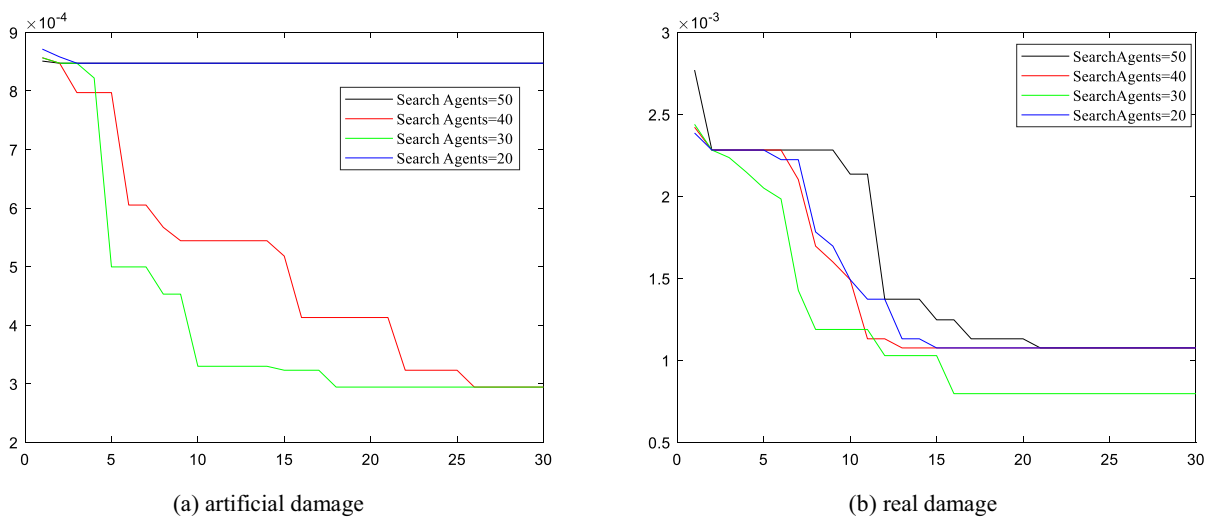
Fig. 4 Convergence curves of IR with Label = 2 in the Paderborn University dataset under different search agents
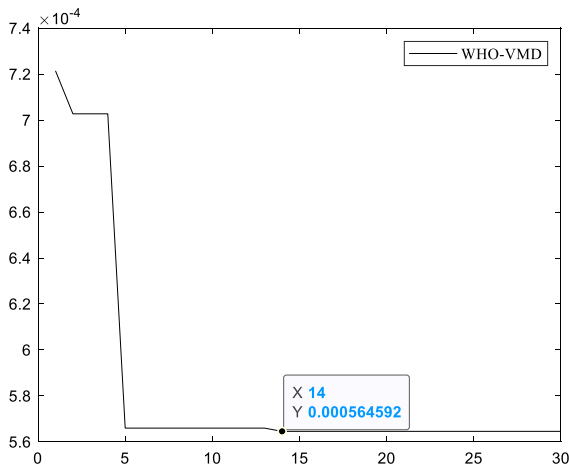
Fig. 5 The convergence curve of the WHO
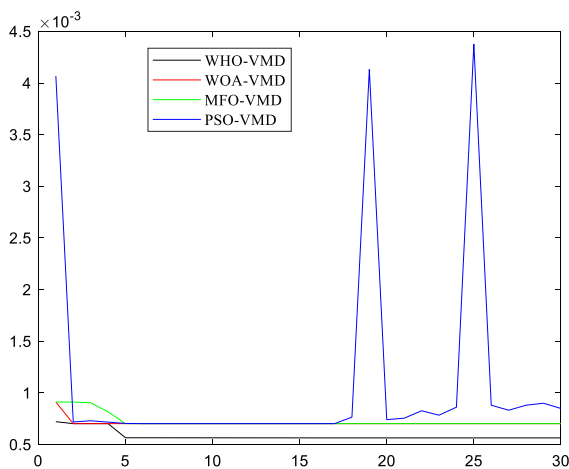
Fig. 6 VMD optimized by four different optimization algorithms

tends to be smooth after two iterations. The fitness function value is reduced to 5.66×10^{-4} after the fifth iteration, and to a minimum value of after the 14th iteration. The results show that WHO has a fast convergence rate in the process of VMD parameter optimization, which proves that WHO is suitable for optimizing the parameters of VMD. The parameters obtained by WHO-VMD in this experiment is $k = 5$, $\alpha = 1106$.

To verify the superiority of the WHO algorithm in optimizing the VMD parameters, the particle swarm optimization algorithm (PSO) [33], the whale optimization algorithm (WOA) [34], and the moth-flame optimization algorithm (MFO) [35] are used to optimize the VMD parameters. The number of optimization algorithm populations is set to 30, and the maximum number of iterations is 30, and then the fitness function convergence curve shown in Fig. 6 is obtained. As can be seen from Fig. 6, the convergence curve of PSO is unstable and the phenomenon of sudden high and low appears. The WOA and MFO converge after the 2nd and 6th iterations, respectively, corresponding to a power spectrum entropy function value of 7.0273×10^{-4}. It can be seen that WOA and MFO are faster in finding the optimum, and convergence can be achieved in fewer iterations. In contrast, WHO can reach the fitness termination value of WOA and MFO after the 2nd iteration, and still continue to iterate to 5.646×10^{-4} thereafter. This proves the superiority of WHO in optimizing VMD parameters.

4.1.2 CCWT

After determining the VMD parameters, the fault signal is decomposed to obtain IMF and denoised through CCWT. The correlation coefficient about IR and OR in the Paderborn real damage D2 dataset as shown in Fig. 7. When 0.3 is chosen as the denoising threshold for CCWT, half of the IMFs are filtered, which may lead to excessive denoising and loss of otherwise useful information. When 0.1 is chosen as the denoising threshold of CCWT, no IMFs are filtered, and the expected denoising effect cannot be achieved. In the paper, 0.2 is chosen as the denoising threshold of CCWT. The IMFs with correlation coefficients less than 0.2 are filtered out first, and then the correlation coefficients of the remaining are used as the weighting coefficients of the IMFs to reconstruct the original signal.

Taking the OR (Label = 3) of the Paderborn real damage D4 dataset as an example. For easy observation, 500 sample points are selected to compare the differences before and after denoising, as shown in

(a) OR (Label=3)   (b) IR (Label=6)
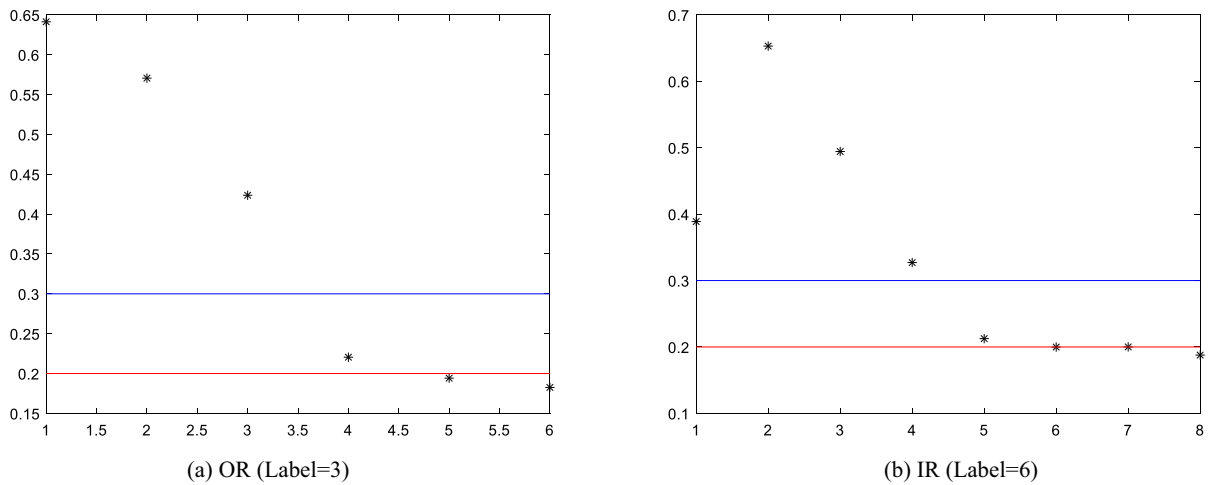
Fig. 7 The correlation coefficient about IR and OR in the Paderborn real damage D2 dataset. **a** OR (Label = 3), **b** IR (Label = 6)
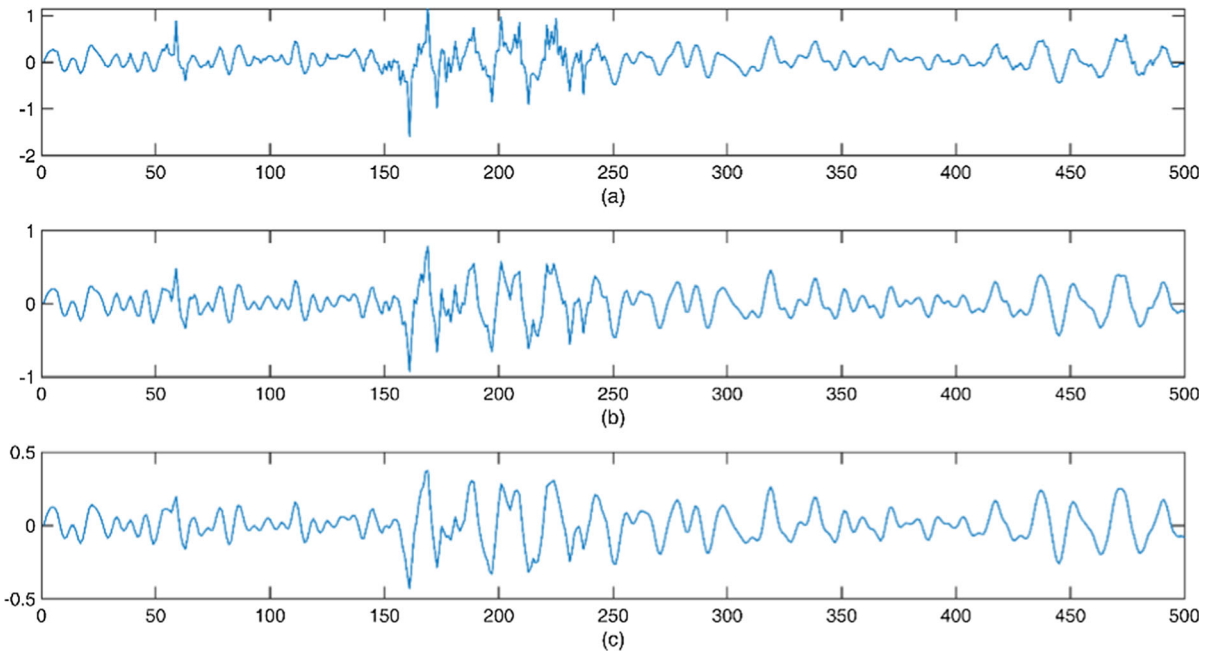
Fig. 8 Comparison of CCT and CCWT **a** original signal, **b** CCT denoising, **c** CCWT denoising

Fig. 8. When correlation coefficient threshold (CCT) is used, IMFs whose correlation coefficient value is less than 0.2 are removed and the remaining components are reconstructed. Based on the former, the CCWT applies the correlation coefficient values greater than 0.2 to the corresponding IMF components as weights. As can be seen from Fig. 8, compared with the signal of CCT denoising, the signal curve of
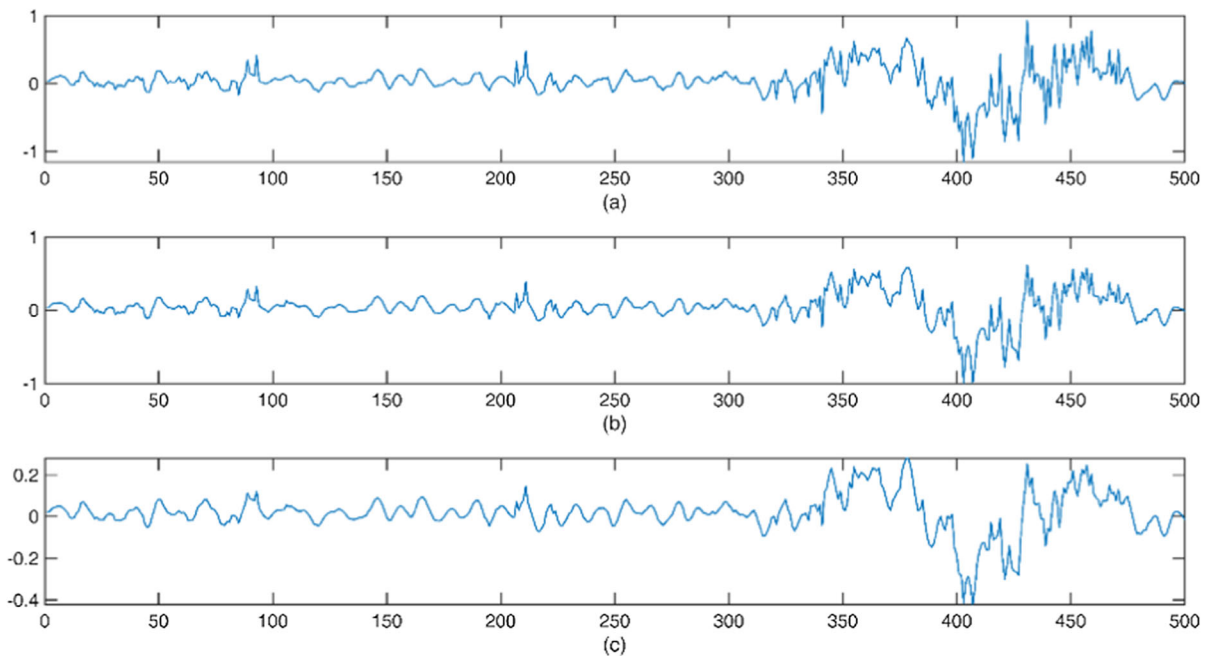
Fig. 9 Comparison of CCT and CCWT, **a** original signal, **b** CCT denoising, **c** CCWT denoising
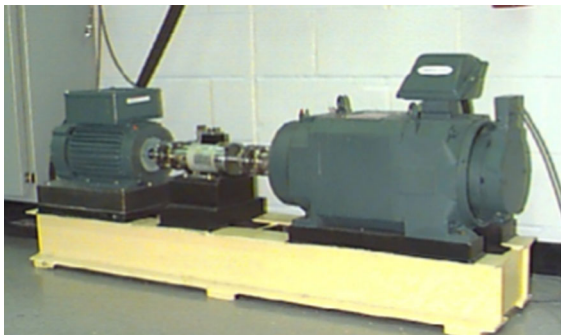
Fig. 10 The CWRU bearing data center bearing test stand

CCWT denoising is smoother and has less burrs. Therefore, we think the denoising effect of CCWT is better.

In addition, although the threshold is set to 0.2 considering the characteristics of most fault signals. Due to the large number of fault signals in both datasets, there are still some signals with all correlation coefficient values greater than 0.2. Taking the IR (Label = 8) of the Paderborn real damage D1 dataset as an example, the denoising effects of CCT and CCWT on such situations are explored. As can be seen from the Fig. 9, the signal after CCT denoising is almost indistinguishable from the original signal because all the correlation coefficient values are greater than 0.2. CCWT not only effectively avoids this defect, but also achieves good denoising effect. This is because the weighting operation on IMF not only enhances the useful signal, but also weakens the noisy signal. The larger the correlation coefficient between IMF and the original signal during CCWT denoising, the more useful information the signal contains. On the contrary, the smaller the value of correlation coefficient, the more the IMF is considered to contain noise. The operation of using the correlation coefficient values as weights is equivalent to amplifying the IMF that is considered to contain useful information and shrinking the IMF that contains noise. Therefore, the CCWT method is considered not only enhance the useful signal but also weaken the noisy signal.

Table 1 The CWRU dataset description

Motor load (HP)	Approx. motor speed (rpm)	Bearing fault type	Bearing fault data	Label
0	1797	Normal	0 mm	1
		Inner race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	2/3/4
		Outer race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	5/6/7
		Ball fault	0.1778 mm, 0.3556 mm, 0.5334 mm	8/9/10
1	1772	Normal	0 mm	1
		Inner race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	2/3/4
		Outer race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	5/6/7
		Ball fault	0.1778 mm, 0.3556 mm, 0.5334 mm	8/9/10
2	1750	Normal	0 mm	1
		Inner race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	2/3/4
		Outer race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	5/6/7
		Ball fault	0.1778 mm, 0.3556 mm, 0.5334 mm	8/9/10
3	1730	Normal	0 mm	1
		Inner race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	2/3/4
		Outer race fault	0.1778 mm, 0.3556 mm, 0.5334 mm	5/6/7
		Ball fault	0.1778 mm, 0.3556 mm, 0.5334 mm	8/9/10

Fig. 11 Diagnostic accuracy of three classifiers under four working conditions
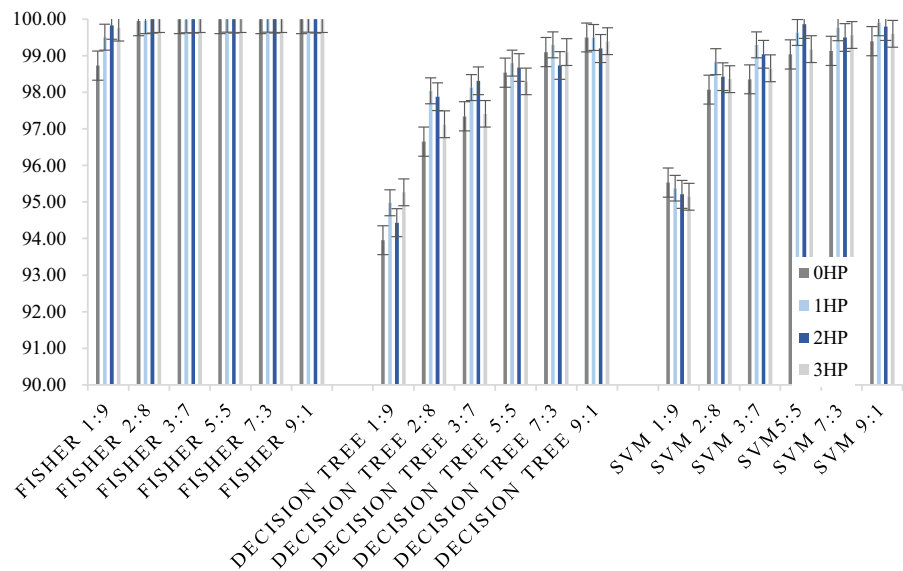
4.2 The CWRU dataset

4.2.1 The CWRU dataset description

The experimental data in this paper comes from the rolling bearing test stand shown in Fig. 10. The 6205-2RS JEM SKF deep groove ball bearing is used as the test bearing, and the data is collected under four loads of 0HP, 1HP, 2HP, and 3HP, and the sampling frequency is 12 kHz. Three damage faults made by electro-discharge machining (EDM), namely inner race fault, outer race fault, and ball fault, are included in the experiment. Each fault includes three different degrees of damage with diameters of 0.1778 mm, 0.3556 mm, and 0.5334 mm, as shown in Table 1. In
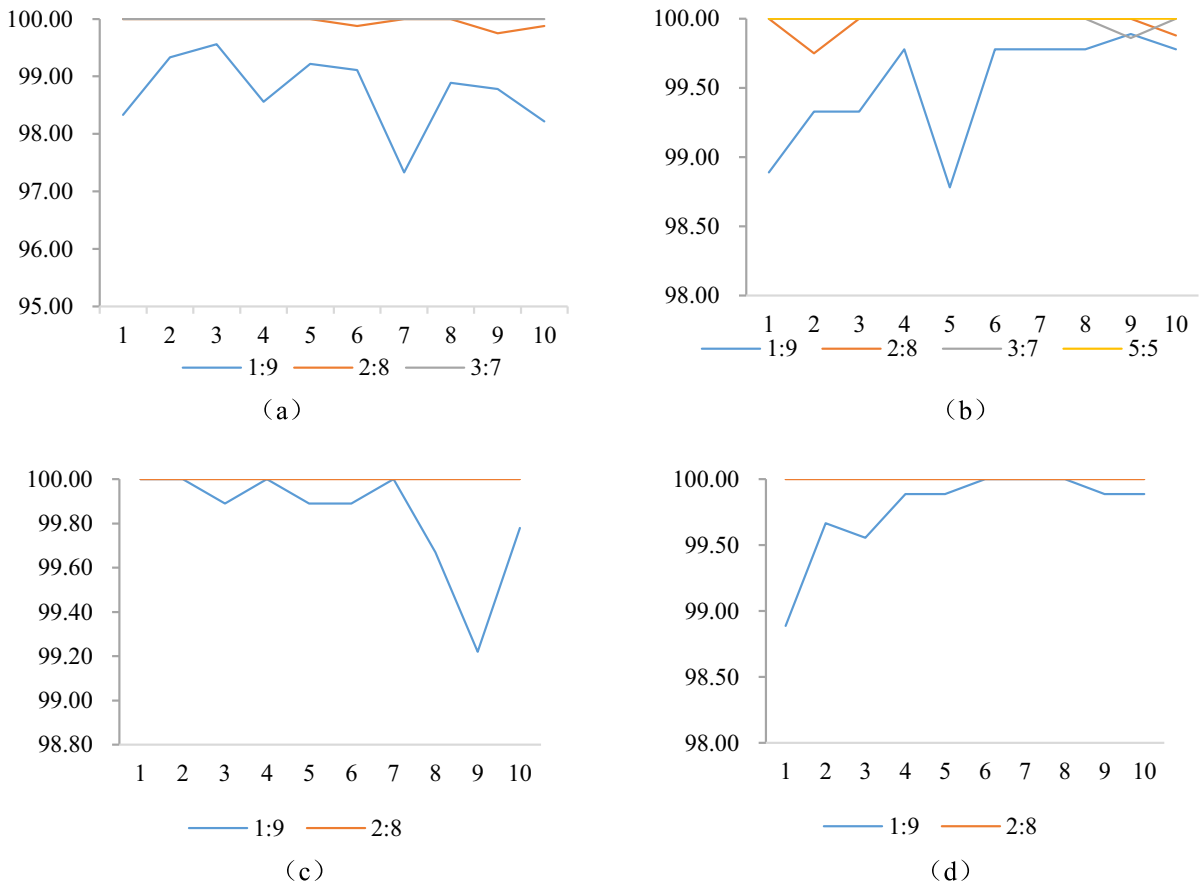
Fig. 12 The 10-time fault diagnosis accuracy of Fisher classifier under different training and testing ratios. **a** 0HP, **b** 1HP, **c** 2HP, **d** 3HP

the experiment, 100 samples are intercepted for each fault signal without overlap.

4.2.2 Experimental analysis of single working condition bearing fault diagnosis

After feature extraction of bearing fault signal, Fisher classifier is used to classify the fused features. The number of samples for each class of fault signals in the experiment is set to 100, and then the training set and test set are divided according to different proportions. Each result is the average of ten replicates. In order to verify the performance of the selected classifiers, three classifiers are selected for experiments with training samples and test samples at different ratios. As shown

in Fig. 11, the accuracy of all three classifiers under the four working conditions shows an increasing trend when the ratio of training samples to test samples is larger. The best performance for the decision tree classification is achieved with 2HP data, while the SVM has higher accuracy under 1HP. In Fisher classifier, except for the slightly lower performance of 0HP when the ratio of training samples to test samples is 1:9, the accuracy of other working conditions exceeds 99%. When the ratio of training samples and test samples is 1:9, the Fisher classifier is improved by 4.5–5.4% compared with the decision tree classifier, and 3.2–4.63% compared with the SVM classifier. It is clear that the Fisher classifier still shows

Table 2 Experimental results of four working conditions under different ratios of training samples and test samples (%)

	Fisher		Decision tree		SVM	
	Mean (%)	Deviations (%)	Mean (%)	Deviations (%)	Mean (%)	Deviations (%)
The ratio of training samples to test samples	0HP					
1:9	**98.73**	**2.23**	93.96	5	95.53	**4.56**
2:8	**99.95**	**0.25**	96.65	**5.75**	98.08	2.38
3:7	**100**	**0**	97.34	3.14	98.36	1.14
5:5	**100**	**0**	98.54	2	99.04	1
7:3	**100**	**0**	99.1	1.67	99.13	1.67
9:1	**100**	**0**	99.5	2	99.4	2
The ratio of training samples to test samples	1HP					
1:9	**99.51**	**1.11**	94.98	**8.45**	95.38	**5.66**
2:8	**99.96**	**0.25**	98.04	2.25	98.84	2.62
3:7	**99.99**	**0.14**	98.13	3.29	99.3	1.43
5:5	**100**	**0**	98.8	1.8	99.64	1
7:3	**100**	**0**	99.3	1.33	99.77	0.67
9:1	**100**	**0**	99.5	2	99.9	1
The ratio of training samples to test samples	2HP					
1:9	**99.83**	**0.78**	94.43	**5.11**	95.21	**5.56**
2:8	**100**	**0**	97.88	3	98.43	3.38
3:7	**100**	**0**	98.32	2	99.04	1.42
5:5	**100**	**0**	98.68	1.4	99.86	0.4
7:3	**100**	**0**	98.73	1	99.5	1.33
9:1	**100**	**0**	99.2	3	99.8	2
The ratio of training samples to test samples	3HP					
1:9	**99.77**	**1.11**	95.27	**6.22**	95.14	**3.33**
2:8	**100**	**0**	97.13	1.75	98.36	2.25
3:7	**100**	**0**	97.41	1.57	98.66	2.14
5:5	**100**	**0**	98.3	2	99.18	1.4
7:3	**100**	**0**	99.1	1.33	99.57	1
9:1	**100**	**0**	99.4	2	99.6	2

Bold values are used to quickly grasp the key points in the table when reviewing papers. In addition, this approach can further highlight the experimental effectiveness of the model proposed in this chapter

the superior performance when the number of samples is small, which is what we would like to see.

For the bearing fault diagnosis method, superior performance is the primary requirement, but the stability of the model is also crucial. To further illustrate the stability of the WHO-VMD-CCWT-EFF, the results of 10 experiments for four working conditions are recorded as shown in Fig. 12. The difference between the maximum and minimum values in 10 experiments are used as deviations to measure the stability of the WHO-VMD-CCWT-EFF. It can be seen from the Fig. 12 that when the ratio of training samples to test samples is 1:9, the deviation under 0HP is 2.23%, which is the largest deviation under the four working conditions. At the training sample to test sample ratio of 2:8, the deviation of the

Table 3 Dataset descriptions with different ratios of training data to test data under two working conditions

Fault location		Normal	Inner race			Out race			Ball		
	Label	1	2	3	4	5	6	7	8	9	10
	Load	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
1:9	Train	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
	Test	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90
2:8	Train	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20	20/20
	Test	80/80	80/80	80/80	80/80	80/80	80/80	80/80	80/80	80/80	80/80
3:7	Train	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30
	Test	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70
5:5	Train	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50
	Test	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50	50/50
7:3	Train	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70	70/70
	Test	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30	30/30
9:1	Train	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90
	Test	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10

0HP and 1HP are 0.25%, and the deviation of the 2HP and 3HP are 0. This indicates that the data under 0HP with few training samples is slightly less stable compared to the data under other working conditions. But when the proportion of training samples is slightly larger under the premise of small samples, the data under 0HP still shows very good performance.

Table 2 presents the experimental results of the four working conditions under different ratios of training samples and test samples. It can be seen that under the four working conditions, the Fisher classifier not only has higher accuracy but also maintains the smallest deviation, which further verifies the effectiveness and stability of the WHO-VMD-CCWT-EFF. At the same time, we can see that the accuracy of the other two classifiers reaches over 99% when the ratio of training samples to test samples is 9:1 and remains around 95% when the ratio of training samples to test samples is 1:9. This indicates that denoising and feature extraction are successful, for which feature extraction will be analyzed in detail later.

4.2.3 Experimental analysis of bearing fault diagnosis under multiple working conditions

Since the actual industrial environment is complex and changeable, it is impossible to ensure that the collected data are always under the same working conditions. So, a variety of bearing fault diagnosis experiments and analysis under multiple working conditions are carried out. The CWRU dataset includes four working conditions, which means that six mixed experiments of two working conditions and four mixed experiments of three working conditions are included in the experiments. Like the single working condition experiment, the training set and the test set are divided according to different proportions. In addition, each experiment is repeated ten times, and the average value is taken as the final experimental result. Details of the experimental data for the two working conditions and the three working conditions are described in detail in Tables 3 and 4 respectively (0HP + 1HP and 0HP + 1HP + 2HP as an example).

Figure 13 shows the experimental results under multiple working conditions. Compared with Fig. 11, it can be seen that the WHO-VMD-CCWT-EFF performs better in the multiple working conditions experiment. To explain this phenomenon, the experiments shown in Fig. 14 are performed. It is not difficult to see that the accuracy rates of the three classifiers under multiple working conditions are almost always higher than those under a single working condition. Therefore, we can conclude that

Table 4 Dataset descriptions with different ratios of training data to test data under three working conditions

Fault location		Normal	Inner race			Out race			Ball		
	Label	1	2	3	4	5	6	7	8	9	10
	Load	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2	0/1/2
1:9	Train	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10
	Test	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
2:8	Train	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20	20/20/20
	Test	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80	80/80/80
3:7	Train	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30
	Test	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70
5:5	Train	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50
	Test	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50	50/50/50
7:3	Train	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70	70/70/70
	Test	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30	30/30/30
9:1	Train	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
	Test	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10	10/10/10

Fig. 13 Experimental results of different ratios of training samples to test samples under multiple working conditions

the phenomenon is not related to the classifier and may be since the extracted features of the same fault in different working conditions are relatively similar or the experimental samples of multiple working conditions are increased.

To verify whether this phenomenon is related to the increase in the number of samples, experiments are carried out with the combination 0HP + 1HP as an example. Of course, the number of training samples and test samples for the selected combinations is the same as for the experiments under a single working condition. Table 5 presents the experimental data in detail.

Figure 15 shows the experimental results, the difference between the accuracy of fisher classifier in both cases is 0–0.41%, and the difference between decision tree and SVM is 0–2.9%. It is not difficult to draw a conclusion that for some classifiers, an increase in sample size has a certain impact on accuracy. But
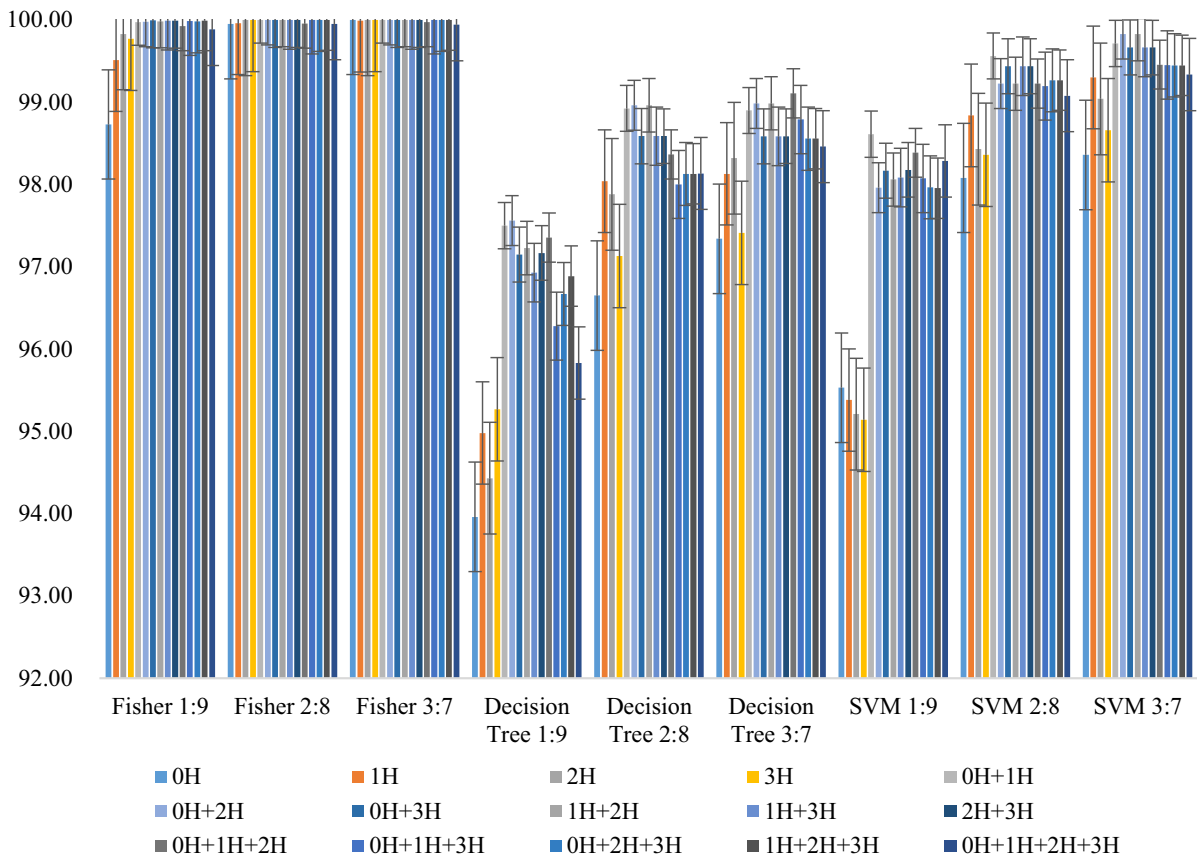
Fig. 14 Experiment comparison between a single working condition and multiple working conditions under three classifiers

Table 5 Description of the number of samples under two working conditions that are consistent with the number of samples under a single operating condition (0HP + 1HP as an example)

| Fault location		Normal	Inner race			Out race			Ball		
	Label	1	2	3	4	5	6	7	8	9	10
	Load	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1
1:9	Train	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5
	Test	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45
2:8	Train	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
	Test	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90	90/90
3:7	Train	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15
	Test	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35
5:5	Train	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25
	Test	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25	25/25
7:3	Train	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35	35/35
	Test	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15	15/15
9:1	Train	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45	45/45
	Test	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5	5/5
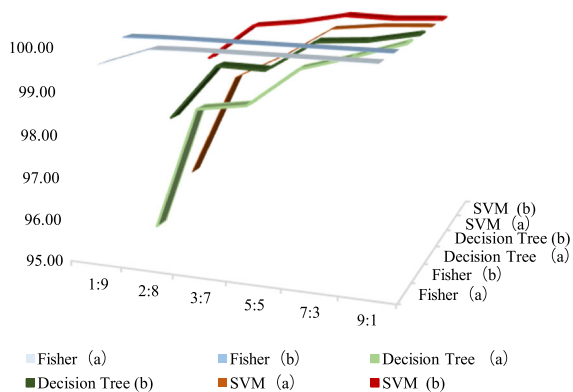
Fig. 15 Effect of using the same number of data samples and the same proportional of data samples on bearing fault diagnosis **a** the same sample size as the single working condition experiment **b** the same sample proportion as single working condition experiment

the proposed model almost overcomes this shortcoming. For further validation, experiments as shown in Fig. 16 are performed. The number of samples for the two working conditions and three working conditions in the figure are shown in the Tables 3 and 4. In Fig. 16, the ratio of training samples to test samples for each experiment is 1:9. From the figure, the accuracy rate of some three working conditions experiments are higher than that of two working conditions, while the rest are lower than that of two working conditions. And the accuracy rate of four working conditions experiments is the lowest. This indicates that for the proposed model, the increase of samples does not necessarily lead to the improvement of fault diagnosis rate.

4.3 Paderborn university dataset

4.3.1 Dataset description

The Paderborn dataset is proposed by Christian Lessmeier et al. of the Kat-Data Center, and the 6205 deep groove ball bearing is used as the test bearing for the collection of the artificial damage dataset and the real damage dataset. The rolling bearing test stand is shown in Fig. 17. As with the CWRU data, Paderborn dataset is collected under each of the four working conditions, as shown in Table 6. As shown in Tables 7 and 8, the artificial damage dataset contains a total of 8 faults, while the real damage dataset contains 9 faults. The bearing damage

for both data sets are inner ring (IR) and outer ring (OR). For each type of fault signal, 100 samples are intercepted without overlapping for the experiment.

#### 4.3.2 Experimental analysis of bearing fault diagnosis in single working condition

Whether the fault types of bearings can be diagnosed efficiently and accurately depends largely on the feature extraction. And the classification effect will be better if the extracted features of fault are more obvious. To confirm the advantages of the feature extraction method in this paper, the features of the artificial damage and the real damage under four working conditions are visualized.

In Fig. 18a–d and e–h are the feature visualization plots for four single working conditions under artificial damage and real damage data, respectively. As shown in Fig. 18, D3 performs the best in the feature visualization of the artificial damage, which means that D3 is more successful in feature extraction. In contrast, D1 performs relatively poorly, which explains the lower accuracy of D1 compared to other data in Table 9. Similarly, the same situation occurred with the real damage experiment. This shows that the relatively low accuracy of D1 compared to other data may be due to the acquisition conditions.

The experimental results of artificial damage and real damage are given in Tables 9 and 10, respectively. The accuracy rate is relatively low and the deviation is large when the ratio of the training sample to the test sample is low. With the increase of the ratio, all the data reaches 100% except for the real damage D1 which reaches 99.77%. This shows that the increase of the ratio of the training sample to the test sample can enhance the experimental effect. In addition, the real damage D2 and D3 reach 100% when the ratio of training samples to test samples is 2:8, and D4 reaches 100% at 3:7. This indicates that even with only a small number of samples, the WHO-VMD-CCWT-EFF can identify the type of bearing fault.

4.3.3 Experimental analysis of bearing fault diagnosis under multiple working conditions

As with the CWRU data, the experiments on artificial damage and real damage under multiple working conditions are also carried out after completing the experiments of single working conditions. Here the
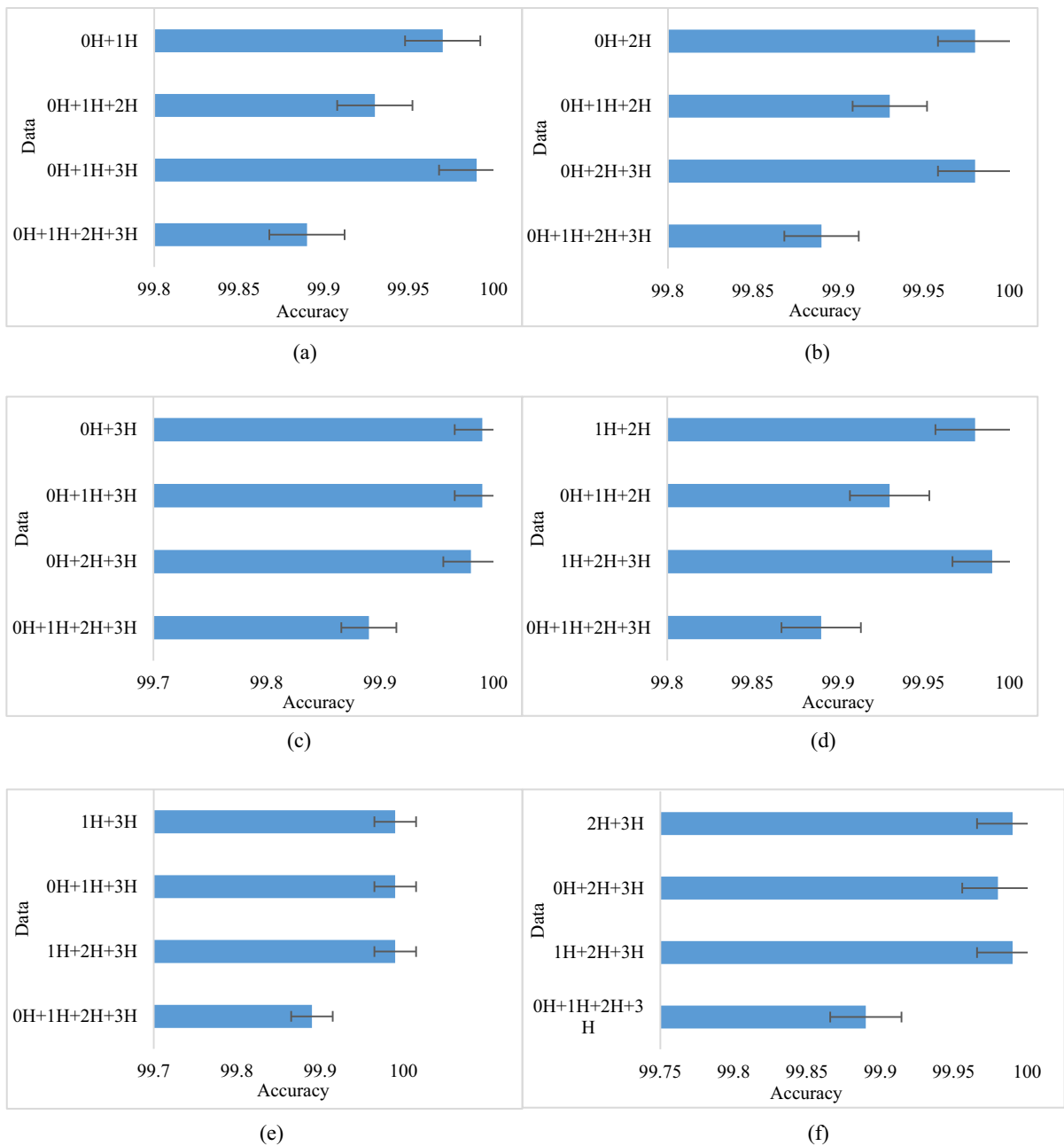
Fig.16 Effect of the phenomenon of increasing sample size due to increased working conditions on bearing fault diagnosis **a** 0H + 1H and its extended experiments **b** 0H + 2H and its extended experiments **c** 0H + 3H and its extended experiments **d** 1H + 2H and its extended experiments **e** 1H + 3H and its extended experiments **f** 2H + 3H and its extended experiments
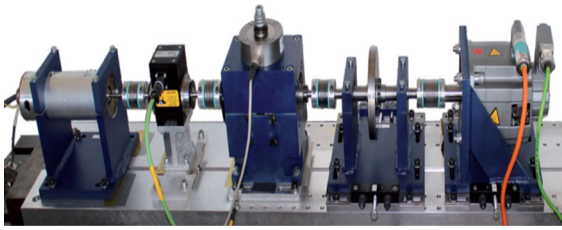
Fig. 17 Rolling bearing test stand

experimental data for the artificial damage dataset are analyzed specifically, and the same is true for the real damage dataset. The experimental data for the two working conditions and the three working conditions

are described in detail in Tables 11 and 12, respectively.

As shown in Tables 13 and 14, which are the results of the experiments on artificial damage and real damage under multiple working conditions, respectively. For both artificial damage and real damage, the accuracy rate reaches more than 99% when the ratio of training sample to test sample is 1:9. Compared to a single working condition, the accuracy of multiple working conditions is higher. This is consistent with the conclusions obtained from the CWRU data. In addition, when the ratio of training samples to test samples is 1:9, the deviation is relatively large, but it improves at 2:8 and then fluctuates as the ratio

Table 6 Four working conditions of bearing experimental data

Data name	Rotational speed (rpm)	Load torque (Nm)	Radial force (N)
D1	900	0.7	1000
D2	1500	0.1	1000
D3	1500	0.7	400
D4	1500	0.7	1000

Table 7 The artificial damage dataset description

Bearing code	Component	Extent of damage (level)	Damage method	Label
K006	No	No	No	1
KA01	OR	1	EDM	2 (OR01)
KA03	OR	2	Electric engraver	3(OR02)
KA05	OR	1	Electric engraver	4(OR03)
KA07	OR	1	Drilling	5(OR04)
KA08	OR	2	Drilling	6(OR05)
KI01	IR	1	EDM	7(IR01)
KI03	IR	1	Electric engraver	8(IR02)
KI07	IR	2	Electric engraver	9(IR03)

Table 8 The real damage dataset description

Bearing code	Component	Extent of damage (level)	Damage (main mode and symptom)	Characteristic of damage	Label
K006	No	No	No	No	1
KA04	OR	1	Fatigue, pitting	single damage	2(OR01)
KA15	OR	1	Plastic deform, Indentations	single damage	3(OR02)
KA16	OR	2	Fatigue, pitting	Repetitive damage	4(OR03)
KA30	OR	1	Plastic deform, Indentations	repetitive damage	5(OR04)
KI04	IR	1	Fatigue, pitting	Multiple damage	6(IR01)
KI16	IR	3	Fatigue, pitting	single damage	7(IR02)
KI17	IR	1	Fatigue, pitting	Repetitive damage	8(IR03)
KI18	IR	2	Fatigue, pitting	Single damage	9(IR04)
KI21	IR	1	Fatigue, pitting	Single damage	10(IR05)
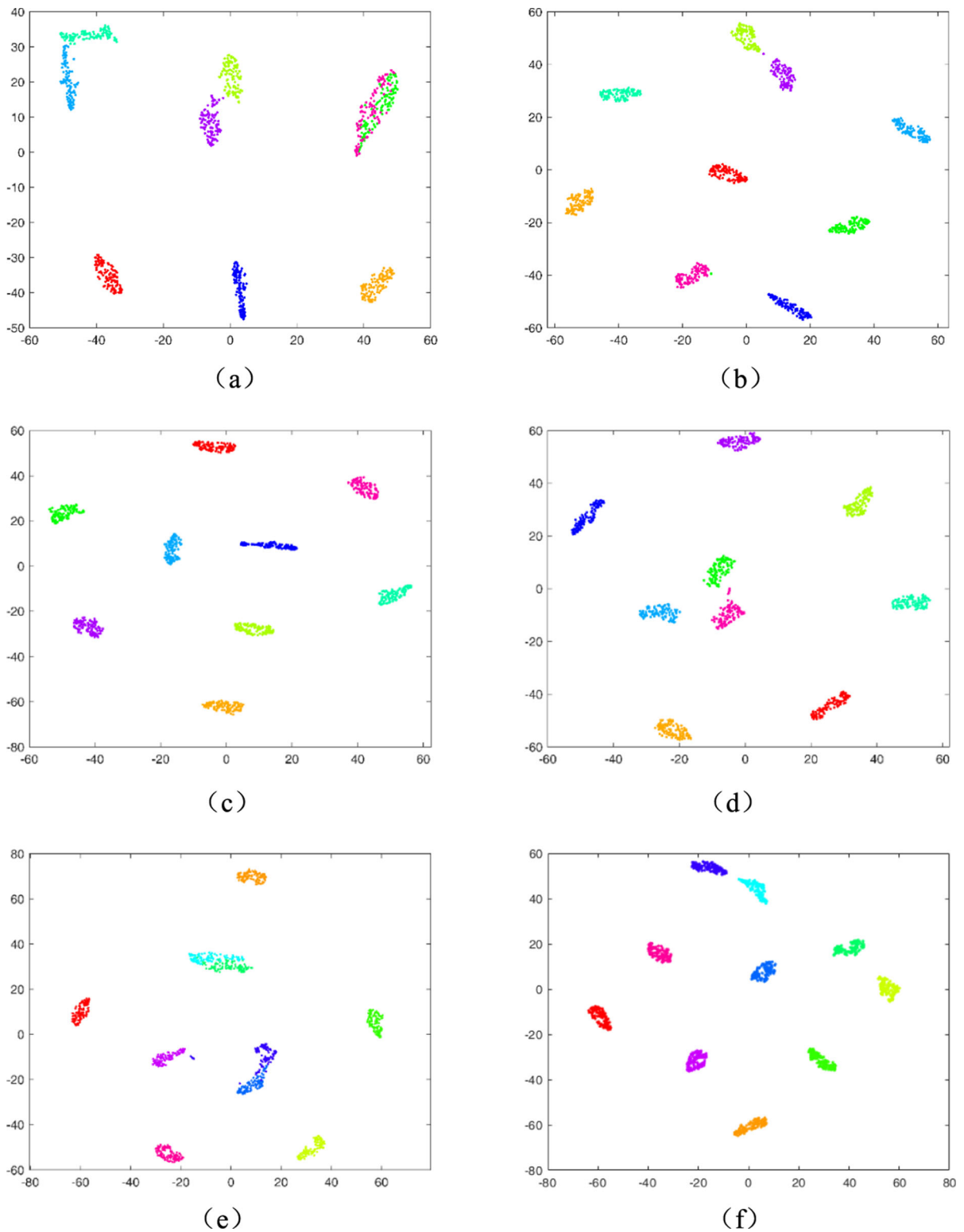
Fig. 18 Feature visualization **a** artificial damage D1, **b** artificial damage D2, **c** artificial damage D3, **d** artificial damage D4, **e** real damage D1, **f** real damage D2 **g** for real damage D3, **h** real damage D4

(g)

(h)
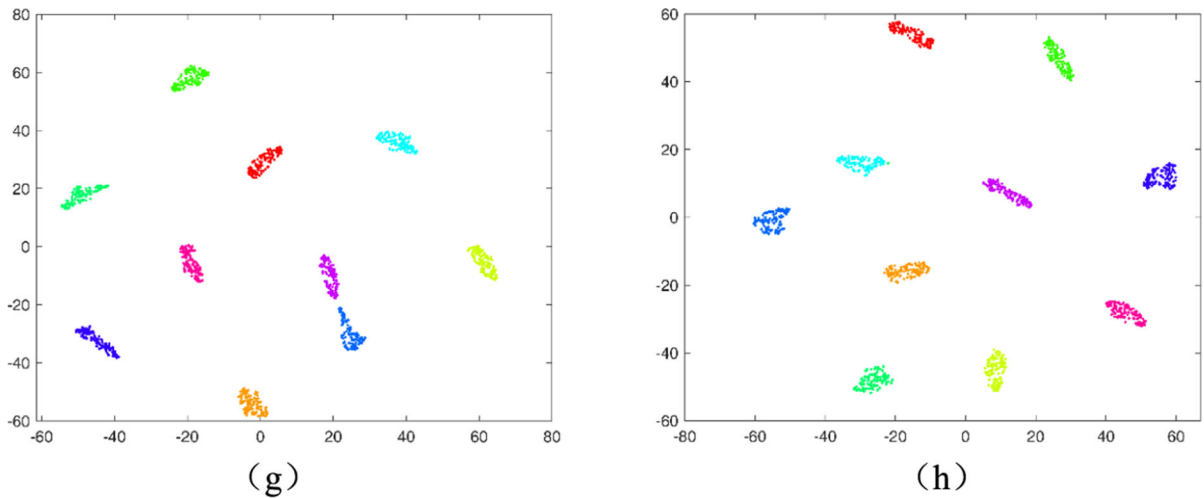
Fig. 18 continued

Table 9 Experimental results of artificial damage

The ratio of training samples to test samples	D1		D2		D3		D4	
	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)
1:9	97.41	4.07	98.07	3.58	99.56	1.73	98.96	2.59
2:8	99.83	0.42	99.78	0.56	100	0	99.99	0.14
3:7	99.94	0.32	99.97	0.32	100	0	100	0
5:5	100	0	99.98	0.22	100	0	100	0
7:3	100	0	100	0	100	0	100	0

Table 10 Real damage experimental results

The ratio of training samples to test samples	D1		D2		D3		D4	
	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)	Mean (%)	Deviation (%)
1:9	96.12	5	98.09	4	99.64	0.42	97.52	7.67
2:8	99.06	1.38	100	0	100	0	99.99	0.13
3:7	99.71	0.57	100	0	100	0	100	0
5:5	99.78	0.4	100	0	100	0	100	0
7:3	99.77	0.67	100	0	100	0	100	0
9:1	99.77	0.67	100	0	100	0	100	0

Table 11 Dataset descriptions with different ratios of training data to test data under two working conditions (D1 + D2 under artificial damage dataset as an example)

Fault location		Normal	IR				OR				
	Label Data name	1 D1/D2	2 D1/D2	3 D1/D2	4 D1/D2		5 D1/D2	6 D1/D2	7 D1/D2	8 D1/D2	9 D1/D2
1:9	Train	10/10	10/10	10/10	10/10		10/10	10/10	10/10	10/10	10/10
	Test	90/90	90/90	90/90	90/90		90/90	90/90	90/90	90/90	90/90
2:8	Train	20/20	20/20	20/20	20/20		20/20	20/20	20/20	20/20	20/20
	Test	80/80	80/80	80/80	80/80		80/80	80/80	80/80	80/80	80/80
3:7	Train	30/30	30/30	30/30	30/30		30/30	30/30	30/30	30/30	30/30
	Test	70/70	70/70	70/70	70/70		70/70	70/70	70/70	70/70	70/70
5:5	Train	50/50	50/50	50/50	50/50		50/50	50/50	50/50	50/50	50/50
	Test	50/50	50/50	50/50	50/50		50/50	50/50	50/50	50/50	50/50
7:3	Train	70/70	70/70	70/70	70/70		70/70	70/70	70/70	70/70	70/70
	Test	30/30	30/30	30/30	30/30		30/30	30/30	30/30	30/30	30/30
9:1	Train	90/90	90/90	90/90	90/90		90/90	90/90	90/90	90/90	90/90
	Test	10/10	10/10	10/10	10/10		10/10	10/10	10/10	10/10	10/10

Table 12 Dataset descriptions with different ratios of training data to test data under three working conditions (D1 + D2 + D3 under artificial damage dataset as an example)

Fault location		Normal	IR				OR				
	Label Data name	1 D1/D2/D3	2 D1/D2/D3	3 D1/D2/D3	4 D1/D2/D3		5 D1/D2/D3	6 D1/D2/D3	7 D1/D2/D3	8 D1/D2/D3	9 D1/D2/D3
1:9	Train	10/10/10	10/10/10	10/10/10	10/10/10		10/10/10	10/10/10	10/10/10	10/10/10	10/10/10
	Test	90/90/90	90/90/90	90/90/90	90/90/90		90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
2:8	Train	20/20/20	20/20/20	20/20/20	20/20/20		20/20/20	20/20/20	20/20/20	20/20/20	20/20/20
	Test	80/80/80	80/80/80	80/80/80	80/80/80		80/80/80	80/80/80	80/80/80	80/80/80	80/80/80
3:7	Train	30/30/30	30/30/30	30/30/30	30/30/30		30/30/30	30/30/30	30/30/30	30/30/30	30/30/30
	Test	70/70/70	70/70/70	70/70/70	70/70/70		70/70/70	70/70/70	70/70/70	70/70/70	70/70/70
5:5	Train	50/50/50	50/50/50	50/50/50	50/50/50		50/50/50	50/50/50	50/50/50	50/50/50	50/50/50
	Test	50/50/50	50/50/50	50/50/50	50/50/50		50/50/50	50/50/50	50/50/50	50/50/50	50/50/50
7:3	Train	70/70/70	70/70/70	70/70/70	70/70/70		70/70/70	70/70/70	70/70/70	70/70/70	70/70/70
	Test	30/30/30	30/30/30	30/30/30	30/30/30		30/30/30	30/30/30	30/30/30	30/30/30	30/30/30
9:1	Train	90/90/90	90/90/90	90/90/90	90/90/90		90/90/90	90/90/90	90/90/90	90/90/90	90/90/90
	Test	10/10/10	10/10/10	10/10/10	10/10/10		10/10/10	10/10/10	10/10/10	10/10/10	10/10/10

increases. This reminds us that the selection of training to test sample ratio is a very important when experimenting with multiple working conditions.

To verify the effectiveness of the entropy fusion method proposed in this paper, the experiments of single entropy and fusion entropy are carried out respectively. 9 entropies are selected for the experiments in Tables 15 and 16, which are RCMDE, RCMFDE, RCMvMFE, RCMvMSE, MPE, multiscale dispersion entropy (MDE), multiscale weight

Table 13 Experimental results of artificial damage under multiple working conditions

Data	Evaluation indicators	The ratio of training samples to test samples					
		1:9	2:8	3:7	5:5	7:3	9:1
D1 + D2	Mean (%)	99.54	99.96	99.98	100.00	100.00	100.00
	Deviations (%)	1.91	0.28	0.16	0.00	0.00	0.00
D1 + D3	Mean (%)	99.86	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	1.36	0.00	0.00	0.00	0.00	0.00
D1 + D4	Mean (%)	99.88	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.99	0.00	0.00	0.00	0.00	0.00
D2 + D3	Mean (%)	99.88	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	1.17	0.00	0.00	0.00	0.00	0.00
D2 + D4	Mean (%)	99.93	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.43	0.00	0.00	0.00	0.00	0.00
D3 + D4	Mean (%)	99.94	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.37	0.00	0.00	0.00	0.00	0.00
D1 + D2 + D3	Mean (%)	99.59	99.87	99.94	99.92	99.96	99.96
	Deviations (%)	1.56	0.19	0.11	0.15	0.12	0.37
D1 + D2 + D4	Mean (%)	99.68	99.96	99.96	99.97	99.95	99.96
	Deviations (%)	1.93	0.09	0.05	0.07	0.12	0.37
D1 + D3 + D4	Mean (%)	99.81	99.99	99.98	100.00	100.00	100.00
	Deviations (%)	0.74	0.05	0.05	0.00	0.00	0.00
D2 + D3 + D4	Mean (%)	99.82	99.96	99.96	99.97	99.95	99.96
	Deviations (%)	0.49	0.09	0.05	0.07	0.12	0.37
D1 + D2 + D3 + D4	Mean (%)	99.53	99.83	99.81	99.84	99.85	99.81
	Deviations (%)	0.90	0.17	0.28	0.39	0.28	0.56

permutation entropy (MWPE), multivariate fuzzy entropy (MVFE), and multivariate sample entropy (MVSE). Among them, MVFE and MVSE are single-scale entropy, and the remaining seven are multiscale entropy. Taking the multiple working conditions experiment as an example, the experimental effect of multiscale entropy both in artificial damage and real damage far exceeds that of single-scale entropy. In order to improve the accuracy and reduce deviations of the model, an entropy fusion method is proposed.

For the fusion of entropy features, the main goal is to improve the accuracy and minimize the deviation. According to Tables 15 and 16, the better-performing entropy features are fused in turn to obtain the experimental results of the artificial damage fusion entropy features in Table 17 and the experimental results of the real damage fusion entropy features in Table 18. Comparing Table 17 with Table 18, we can see that the WHO-VMD-CCWT-EFF is more applicable to the real damage dataset, which is the direction of our efforts. In the entropy fusion, the experimental performance of D1, D2, and D4 is not the best, but the difference with the indicator of the best effect is slight.

To make the model more convincing, the experimental results of the three working conditions experiments of D1, D2, and D3 with different ratios of

Table 14 Real damage experimental results under multiple working conditions

Data	Evaluation indicators	The ratio of training samples to test samples					
		1:9	2:8	3:7	1:9	7:3	9:1
D1 + D2	Mean (%)	99.12	99.68	99.85	99.91	99.93	99.90
	Deviations (%)	1.00	0.50	0.29	0.40	0.17	0.50
D1 + D3	Mean (%)	99.07	99.68	99.85	99.91	99.93	99.90
	Deviations (%)	1.00	0.50	0.29	0.40	0.17	0.50
D1 + D4	Mean (%)	99.09	99.68	99.85	99.91	99.93	99.90
	Deviations (%)	1.00	0.50	0.29	0.40	0.17	0.50
D2 + D3	Mean (%)	99.99	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.06	0.00	0.00	0.00	0.00	0.00
D2 + D4	Mean (%)	99.99	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.11	0.00	0.00	0.00	0.00	0.00
D3 + D4	Mean (%)	99.99	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.11	0.00	0.00	0.00	0.00	0.00
D1 + D2 + D3	Mean (%)	99.19	99.48	99.64	99.67	99.70	99.60
	Deviations (%)	0.89	0.42	0.43	0.40	0.67	1.00
D1 + D2 + D4	Mean (%)	99.51	99.71	99.76	99.75	99.74	99.87
	Deviations (%)	0.74	0.58	0.38	0.33	0.56	0.33
D1 + D3 + D4	Mean (%)	99.46	99.71	99.76	99.75	99.74	99.87
	Deviations (%)	0.70	0.58	0.38	0.33	0.56	0.33
D2 + D3 + D4	Mean (%)	99.98	100.00	100.00	100.00	100.00	100.00
	Deviations (%)	0.11	0.00	0.00	0.00	0.00	0.00
D1 + D2 + D3 + D4	Mean (%)	99.58	99.74	99.79	99.77	99.73	99.80
	Deviations (%)	1.11	0.16	0.25	0.30	0.25	0.50

Table 15 Accuracy of single entropy of artificial damage

Data	Evaluation indicators	RCMFDE	RCMDE	RcMVMFE	RcMVMSE	MPE	MDE	MWPE	MVFE	MVSE
D1 + D2 + D3	Mean (%)	90.74	90.01	98.22	96.91	94.34	83.67	94.43	49.06	33.75
	Deviations (%)	12.47	16.71	4.90	7.86	15.19	20.33	13.25	20.49	7.28
D1 + D2 + D4	Mean (%)	94.20	89.43	99.35	98.60	96.99	83.89	96.32	54.08	46.88
	Deviations (%)	6.79	10.70	3.95	4.53	10.33	15.10	11.85	35.80	30.86
D1 + D3 + D4	Mean (%)	91.87	93.37	99.33	98.02	94.16	82.13	92.97	42.49	45.53
	Deviations (%)	8.85	10.78	2.80	4.24	12.80	7.98	11.65	21.60	13.91
D2 + D3 + D4	Mean (%)	93.94	90.01	99.56	98.76	97.59	84.31	97.07	56.48	48.38
	Deviations (%)	9.34	4.90	1.85	2.92	4.36	10.86	4.32	11.81	15.88
D1 + D2 + D3 + D4	Mean (%)	86.81	86.16	97.51	96.27	92.43	76.23	92.61	47.48	39.76
	Deviations (%)	8.61	11.20	4.04	6.54	13.52	10.22	11.05	18.86	15.68

Table 16 Accuracy of single entropy of real damage

Data	Evaluation indicators	RCMFDE	RCMDE	RcMVMFE	RcMVMSE	MPE	MDE	MWPE	MVFE	MVSE
D1 + D2 + D3	Mean (%)	90.03	89.78	97.52	95.47	92.00	83.30	91.03	22.23	16.11
	Deviations (%)	10.30	8.26	3.26	5.74	9.59	11.11	9.19	6.59	4.37
D1 + D2 + D4	Mean (%)	93.42	95.20	99.11	97.25	95.16	89.29	93.00	20.80	18.25
	Deviations (%)	1.78	2.74	0.33	1.44	1.30	4.00	2.11	5.44	2.67
D1 + D3 + D4	Mean (%)	92.23	94.34	98.79	96.73	94.36	88.12	92.21	20.89	17.90
	Deviations (%)	13.56	11.26	3.56	6.63	9.33	15.70	10.04	4.48	3.00
D2 + D3 + D4	Mean (%)	96.99	97.54	99.83	98.64	97.99	93.19	97.77	51.26	46.99
	Deviations (%)	10.93	8.37	1.63	4.85	8.96	14.44	8.67	12.56	14.41
D1 + D2 + D3 + D4	Mean (%)	91.84	93.63	98.59	96.25	94.05	87.68	91.86	30.32	25.01
	Deviations (%)	11.64	9.36	4.97	6.61	10.31	13.78	10.50	3.28	4.14

Table 17 Accuracy of artificial damage fusion entropy

Data	Evaluation indicators	RcMVMFE + RcMVMSE	RcMVMFE + RcMVMSE + MPE	RcMVMFE + RcMVMSE + MPE + + RCMDE	The WHO-VMD-CCWT-EFF
D1 + D2 + D3	Mean (%)	98.92	99.49	99.59	**99.59**
	Deviations (%)	3.09	2.67	2.26	**1.56**
D1 + D2 + D4	Mean (%)	99.60	**99.73**	99.72	99.68
	Deviations (%)	2.43	1.81	**1.69**	1.93
D1 + D3 + D4	Mean (%)	99.57	99.81	99.77	**99.81**
	Deviations (%)	2.10	1.40	0.99	**0.74**
D2 + D3 + D4	Mean (%)	99.78	**99.89**	99.84	99.82
	Deviations (%)	0.66	**0.25**	0.58	0.49
D1 + D2 + D3 + D4	Mean (%)	98.92	99.52	99.43	**99.53**
	Deviations (%)	3.09	1.73	1.70	**0.90**

Bold values are used to quickly grasp the key points in the table when reviewing papers. In addition, this approach can further highlight the experimental effectiveness of the model proposed in this chapter

training samples to test samples are analyzed. The trend plots of entropy fusion experimental results for artificial and real damage at D1, D2 and D3 are given in Figs. 19 and 20, respectively. As the number of fused feature entropies increases the artificial damage accuracy increases and the deviation decreases. In the real damage, the performance of the 5 entropy fusions is not the best when the ratio of training samples to test samples is high. This shows that for the fusion of entropy features, it is not the more the better, but the category and number of entropy should be reasonably selected according to the experiment. In this paper, five entropy features are fused in the model.

Table 18 Accuracy of real damage fusion entropy

Data	Evaluation indicators	RcMVMFE + RcMVMSE	RcMVMFE + RcMVMSE + MPE	RcMVMFE + RcMVMSE + MPE + + RCMDE	The WHO-VMD-CCWT-EFF
D1 + D2 + D3	Mean (%)	98.16	98.86	99.13	**99.19**
	Deviations (%)	1.67	2.00	1.15	**0.89**
D1 + D2 + D4	Mean (%)	99.23	99.40	**99.52**	99.51
	Deviations (%)	0.56	0.56	**0.56**	0.74
D1 + D3 + D4	Mean (%)	99.09	99.31	99.40	**99.46**
	Deviations (%)	1.56	0.85	1.04	**0.70**
D2 + D3 + D4	Mean (%)	99.91	99.93	99.93	**99.98**
	Deviations (%)	0.85	0.59	0.59	**0.11**
D1 + D2 + D3 + D4	Mean (%)	99.18	99.45	99.42	**99.58**
	Deviations (%)	2.44	1.83	1.78	**1.11**

Bold values are used to quickly grasp the key points in the table when reviewing papers. In addition, this approach can further highlight the experimental effectiveness of the model proposed in this chapter
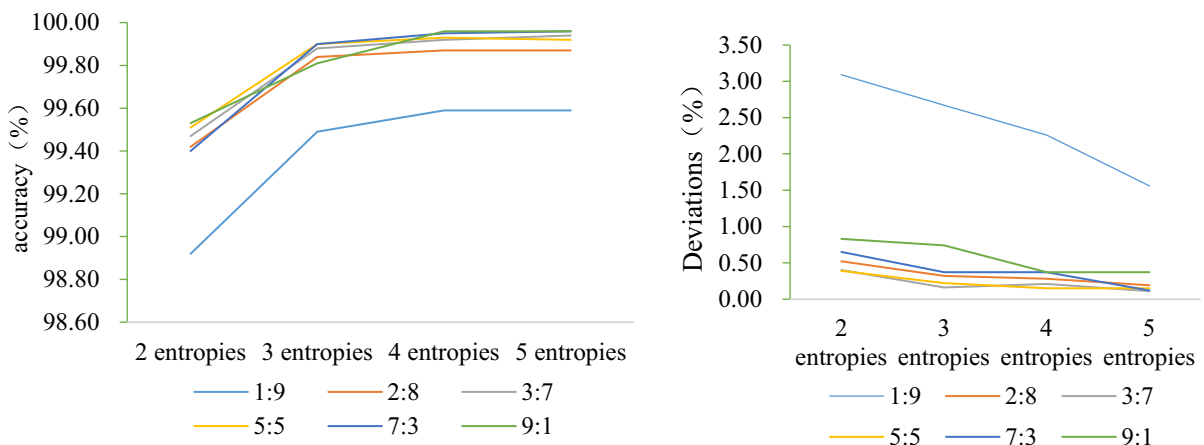
Fig. 19 Experimental results of entropy fusion of artificial damage D1 + D2 + D3

4.4 Model comparison

As shown in Table 19, in order to verify the performance of the proposed method, we selected existing models from recent years for comparison. In bearing fault diagnosis, the most classic CWRU dataset is used as the dataset for model comparison. The comparison results indicate that the proposed model can identify bearing faults with higher accuracy when the training and testing data are the same as other

literature. In addition, while maintaining the same level of accuracy, less data is used.

5 Conclusions

A bearing fault diagnosis method WHO-VMD-CCWT-EFF based on signal denoising and feature fusion is proposed to address the issues of low accuracy in traditional methods and the need for a large number of data samples in deep learning
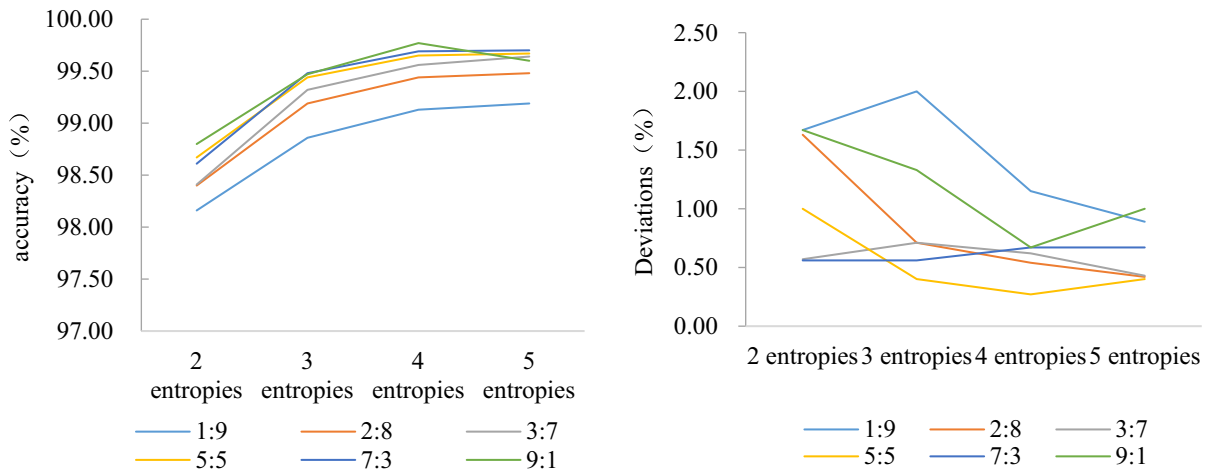
Fig. 20 Experimental results of entropy fusion of real damage D1 + D2 + D3

Table 19 Comparison of the proposed model with other models in the literature

Literature	Year	Data	Number of training samples/number of test samples (per category)	Accuracy (%)
[36]	2019	0HP	200/37	98.10
		1HP	200/37	97.62
		2HP	200/37	97.92
		3HP	200/37	95.76
[37]	2023	3HP	80/120	99.2
[38]	2023	0HP	10/90	97.98 ± 1.58
		1HP	10/90	96.29 ± 0.60
		2HP	10/90	95.84 ± 1.83
		3HP	10/90	97.49 ± 1.43
		0 HP + 1 HP + 2 HP + 3 HP	400/3600	98.72 ± 0.31
[15]	2017	2 HP	30/30	99.10
[39]	2021	0 HP (Three types)	20/20	99.12
[40]	2021	0 HP	240/60	99.83
		1 HP	240/60	99.83
		2 HP	240/60	100
		3 HP	240/60	100
[16]	2021	2 HP	500/100	99.69
The proposed method	–	0HP	10/90,20/80,30/70	98.73,99.95,100
		1HP	10/90,20/80,30/70,50/50	99.51,99.96,99.99,100
		2HP	10/90,20/80	99.83,100
		3HP	10/90,20/80	99.77,100
		0 HP + 1 HP + 2 HP + 3 HP	40/360,80/320,120/280,200/200	99.89,99.95,99.94,99.97

methods. This paper focuses on two aspects of bearing fault signal denoising and feature extraction. In order to verify the effectiveness and stability of the model, the CWRU dataset and Paderborn dataset are used for various single and multiple working experiments. The experimental results show that WHO-VMD-CCWT-EFF exhibits superior performance in both single and multiple working conditions when the training data of both datasets are small. The following experimental results confirm this conclusion.

(1) The WHO-VMD-CCWT-EFF model can accurately identify the fault status of bearings. It can be proved by the fact that experiments on the CWRU dataset and the Paderborn dataset (12 single operating conditions and 30 multiple operating conditions) achieve over 99% accuracy.

(2) In the Paderborn dataset, when the ratio of training samples to test samples is 1:9, the difference between real and artificial damage is 0.02% -1.44%. This indicates that even under small sample experimental conditions, the model has good stability and generalization ability.

(3) The fusion entropy feature vector is an effective method for extracting bearing fault features. In the experiments on the CWRU dataset, in addition to the Fisher discriminator achieving an accuracy of over 98% in small samples, the accuracy of decision tree and SVM classification in small samples also reach over 93.5%, which proves this point.

(4) Compared to the Paderborn dataset, the CWRU dataset performs better in the experiment. This indicates that differences in data from different devices can affect the performance of the model. Therefore, in the future, we will focus on researching cross equipment bearing fault diagnosis.

Data availability The datasets analyzed during the current study are available in the Case Western Reserve University Bearing Data Center and Paderborn University Kat-Data Center Website repository.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Saadatmorad, M., Talookolaei, R.A.J., Pashaei, M.H., et al.: Pearson correlation and discrete wavelet transform for crack identification in steel beams. Mathematics **10**(15), 2689 (2022)
2. Tiachacht, S., Khatir, S., Thanh, C.L., et al.: Inverse problem for dynamic structural health monitoring based on slime mould algorithm. Eng. Comput. 1–24 (2021).
3. Cheng, X.: The application of automation technology in mechanical design and manufacturing (2019)
4. Gültekin, Ö., Çinar, E., Özkan, K., et al.: A novel deep learning approach for intelligent fault diagnosis applications based on time-frequency images. Neural Comput. Appl. **34**(6), 4803–4812 (2022)
5. Wang, X., Si, S., Li, Y.: Multiscale diversity entropy: a novel dynamical measure for fault diagnosis of rotating machinery. IEEE Trans. Ind. Inform. **99**, 1–1 (2020)
6. Song, X., Liao, Z., Wang, H., et al.: Incrementally accumulated holographic SDP characteristic fusion method in ship propulsion shaft bearing fault diagnosis. Meas. Sci. Technol. **4**, 33 (2022)
7. Li, Y., Shi, Z., Lin, T.R., et al.: An iterative reassignment based energy-concentrated TFA post-processing tool and application to bearing fault diagnosis. Measurement **193**, 110953 (2022)
8. Pan, H., Xu, H., Zheng, J., et al.: Multi-class fuzzy support matrix machine for classification in roller bearing fault diagnosis. Adv. Eng. Inform. **51**, 101445 (2022)
9. Chang, Y., Bao, G., Cheng, S., et al.: Improved VMDkgFCM algorithm for the fault diagnosis of rolling bearing vibration signals. IET Signal Process. **15**, 238–250 (2021)
10. He, F., Ye, Q.: A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm. Sensors **22**(4), 1410 (2022)
11. Zhang, Y., Wang, J., Zhang, F., Lv, S., Zhang, L., Jiang, M., Sui, Q.: Intelligent fault diagnosis of rolling bearing using the ensemble self-taught learning convolutional auto-encoders. IET Sci. Meas. Technol. **16**(2), 130–147 (2022)

12. Hui, K.H., Lim, M.H., Leong, M.S., et al.: Dempster-Shafer evidence theory for multi-bearing faults diagnosis. Eng. Appl. Artif. Intell. **57**, 160–170 (2017)

13. Lyu, P., Zhang, K., Yu, W., et al.: A novel RSG-based intelligent bearing fault diagnosis method for motors in high-noise industrial environment. Adv. Eng. Inform. **52**, 101564 (2022)

14. Yang, S., Yang, P., Yu, H., Bai, J., Feng, W., Su, Y., Si, Y.: A 2DCNN-RF model for offshore wind turbine high-speed bearing-fault diagnosis under noisy environment. Energies **15**(9), 3340 (2022)

15. Muhammad, S., Cheol-Hong, K., Jong-Myon, K.: A hybrid feature model and deep-learning-based bearing fault diagnosis. Sensors **17**(12), 2876 (2017)

16. Wang, Q., Yang, C., Wan, H., et al.: Bearing fault diagnosis based on optimized variational mode decomposition and 1D convolutional neural networks. Meas. Sci. Technol. **32**(10), 104007 (2021)

17. Zhang, A., Li, S., Cui, Y., et al.: Limited data rolling bearing fault diagnosis with few-shot learning. IEEE Access **99**, 1–1 (2019)

18. Case Western Reserve University Bearing Data Center Website http://csegroups.case.edu/bearingdatacenter/home.

19. Christian Lessmeier et al., Kat-Data Center: mb.uni-paderborn.de/kat/datacenter, Chair of Design and Drive Technology, Paderborn University.

20. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. IEEE Trans. Signal Process. **62**(3), 531–544 (2014)

21. Naruei, I., Keynia, F.: Wild horse optimizer: a new metaheuristic algorithm for solving engineering optimization problems. Eng. Comput. **38**(Suppl 4), 3025–3056 (2022)

22. Ho, L.V., Nguyen, D.H., Mousavi, M., et al.: A hybrid computational intelligence approach for structural damage detection using marine predator algorithm and feedforward neural networks. Comput. Struct. **252**, 106568 (2021)

23. Ho, L.V., Trinh, T.T., De Roeck, G., et al.: An efficient stochastic-based coupled model for damage identification in plate structures. Eng. Fail. Anal. **131**, 105866 (2022)

24. Al Thobiani, F., Khatir, S., Benaissa, B., et al.: A hybrid PSO and Grey Wolf Optimization algorithm for static and dynamic crack identification. Theoret. Appl. Fract. Mech. **118**, 103213 (2022)

25. Ji, Y., Wang, X., Liu, Z., et al.: EEMD-based online milling chatter detection by fractal dimension and power spectral entropy. Int. J. Adv. Manuf. Technol. **92**(1781), 1–16 (2017)

26. Azami, H., Rostaghi, M., Abásolo, D., et al.: Refined composite multiscale dispersion entropy and its application to biomedical signals. IEEE Trans. Bio-Med. Eng. **99**, 1–1 (2017)

27. Rostaghi, M., Azami, H.: Dispersion entropy: a measure for time-series analysis. IEEE Signal Process. Lett. **23**(5), 610–614 (2016)

28. Azami, H., Escudero, J.: Amplitude-and fluctuation-based dispersion entropy. Entropy **20**(3), 210 (2018)

29. Azami, H., Arnold, S.E., Sanei, S., et al.: Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. IEEE Access **7**, 68718–68733 (2019)

30. Azami, H., Escudero, J.: Refined composite multivariate generalized multiscale fuzzy entropy: A tool for complexity analysis of multichannel signals. Physica A **465**, 261–276 (2017)

31. Ahmed, M.U., Mandic, D.P.: Multivariate multiscale entropy analysis. IEEE Signal Process. Lett. **19**(2), 91–94 (2012)

32. Ouyang, G., Li, J., Liu, X., et al.: Dynamic characteristics of absence EEG recordings with multiscale permutation entropy analysis. Epilepsy Res. **104**(3), 246–252 (2013)

33. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN'95-International Conference on Neural Networks. IEEE, vol. 4: pp. 1942–1948 (1995).

34. Mirjalili, S., Lewis, A.: The whale optimization algorithm. Adv. Eng. Softw. **95**, 51–67 (2016)

35. Mirjalili, S.: Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowl. Based Syst. **89**(Nov), 228–249 (2015)

36. Li, H., Huang, J., Ji, S.: Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network. Sensors **19**(9), 2034 (2019)

37. Yang, J., Xie, G., Yang, Y., et al.: A multilevel recovery diagnosis model for rolling bearing faults from imbalanced and partially missing monitoring data. Math. Biosci. Eng. **20**(3), 5223–5242 (2023)

38. Nie, G., Zhang, Z., Shao, M., et al.: A novel study on a generalized model based on self-supervised learning and sparse filtering for intelligent bearing fault diagnosis. Sensors **23**(4), 1858 (2023)

39. Chang, Y., Bao, G., Cheng, S., et al.: Improved VMD-KFCM algorithm for the fault diagnosis of rolling bearing vibration signals. IET Signal Proc. **4**, 15 (2021)

40. Zhao, J., Yang, S., Li, Q., et al.: A new bearing fault diagnosis method based on signal-to-image mapping and convolutional neural network. Measurement **176**(1), 109088 (2021)