Hindawi

Research Article

Rotating Machinery Fault Diagnosis Using HHO-RSSD and RCGmvMAAPE

Ning Zhang (ID)

School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan 430070, China

Correspondence should be addressed to Ning Zhang; qq081100@yandex.com

Rotating machinery has played an enormous role in industrial production, and its stable operation is related to whether production can proceed smoothly. At present, multichannel entropy-based methods are usually be adopted to analyze multichannel vibration signals. However, the collected signal may only have one channel in the actual situation. At this time, analyzing only a single channel signal cannot effectively utilize the advantages of multivariate analysis. For this reason, this paper presents a novel multivariate analysis approach and applies it to the fault diagnosis of machinery. Firstly, the parameter-optimized resonance sparse decomposition (RSSD) algorithm is adopted to decompose the single-channel vibration signal into high and low resonance components. Then, the two components are regarded as dual-channel vibration signals and input into the refined composite generalized multivariate multiscale amplitude aware permutation entropy (RCGmvMAAPE) method to gain fault features. Eventually, the features are input to the deep belief network (DBN) classifier to perform fault judgment. The experiments of rotating machinery are carried to verify the effectiveness of the developed approach. The results display that the proposed fault diagnosis method can achieve the classification accuracy of 100% and 98% when only a single-channel vibration signal is used, which is better than the fault diagnosis method based on a multichannel vibration signal and enjoys strong stability.

1. Introduction

Rotating machinery is the most widely adopted mechanical equipment in the industrial sector. However, since the working environment of rotating machinery is mostly harsh, various types of faults are prone to occur, resulting in serious personal and property losses [1, 2]. Therefore, it is necessary to study the general diagnostic techniques for rotating machinery. The internal structure of the rotating machinery will change if there are the faults, which exacerbate internal vibration. Consequently, the vibration signal of the rotating machinery contains information that can characterize the current state, which indicates that the signals can be used for analysis to determine the present status [3, 4].

The operating conditions of rotating machinery can be characterized by vibration during operation, but the vibration is nonlinear data [5]. In order to extract the state feature from the vibration data, effective methods must be adopted to amplify the characteristic information and eliminate the interference [6]. Signal decomposition algorithms are typical methods for processing this kind of signals. By decomposing raw signal into several components, these methods analyze the complexity of the signal on multiple time scales and reduce the interference of disturbance components such as noise on feature extraction [7]. In the current signal decomposition algorithms, the more typical ones include wavelet transform (WT) and empirical mode decomposition (EMD). Both of the above two algorithms have obvious defects, which affect the reliability of analysis. For example, WT cannot analyze the high frequency components. In addition, it lacks ability to adaptively process signals since its reliability is influenced by the wavelet basis function. EMD decomposes the signal based on the local characteristics of the signal itself, so it can realize the analysis without manually setting parameters, thereby has the advantage of adaptive analysis. However, EMD has serious modal aliasing and end effect defects, and the

physical meaning of some components is not obvious, which affects its reliability [8].

Resonance-based sparse signal decomposition (RSSD) is a signal decomposition method based on resonance properties of signals, which can realize accurate analysis and complexity measurement for nonlinear signals with complex components [9, 10]. Based on the tunable Q-factor wavelet transform (TQWT), RSSD uses the difference of quality factor Q between the continuous oscillation signal and the transient impact component to represent the complex signal sparsely with high quality factor and low quality factor [11]. Different from signal decomposition approaches based on frequency or time scale such as EMD, RSSD combines the frequency and bandwidth of the signal simultaneously, so it can perfectly separate the periodic pulse component and the transient nonoscillating component in the vibration signal [12]. Based on different quality factors, the approach divides the composition of the signal into periodic harmonics, fault impact, and noise and divides them into high resonance components and low resonance components [13]. Therefore, the RSSD algorithm has a significant advantage for analyzing the impact fault signals. Nevertheless, the excellent performance of the RSSD algorithm is influenced by the quality factor, the weight coefficient and the Lagrangian operator. The improper parameter settings will interfere with the performance of RSSD [14]. The Harris Hawk algorithm (HHO) is a novel heuristic optimization algorithm proposed by Mirjalili, which mainly simulate predation behavior of Harris Hawk in nature. Compared with several other typical optimization algorithms, the HHO algorithm performs better and has higher search efficiency [15]. Considering the excellent performance of HHO in the optimization problem, combined with the RSSD algorithm, this paper proposes an optimized RSSD based on HHO. This method not only can adaptively find the best combination of RSSD parameters, but also has high optimization efficiency and excellent generalization.

After processing the vibration signals, how to extract the highly distinguishable features is the key to the fault diagnosis of rolling bearings. With the development of nonlinear science, the feature extraction technology based on entropy theory, such as permutation entropy (PE), amplitude-aware permutation entropy (AAPE), and multiscale amplitude-aware permutation entropy (MAAPE) had been favored by a large number of researchers due to the good nonlinear data processing performance [16, 17]. Because of the good ability to extract the nonlinear fault information hidden in the vibration signal, Wu used multiscale permutation entropy (MPE) for the health detection of rolling bearings and obtained ideal results [18]. However, the permutation entropy does not consider the contribution of the amplitude of the time series to the entropy value in the calculation, which leads to inaccurate and sufficient analysis [19]. In this regard, Chen proposed multiscale AAPE (MAAPE) by replacing PE with AAPE and used it to excavate the fault characteristics of rolling bearings [20]. Although MAAPE has better feature extraction performance, there are still two shortcomings in the application process as follows: (1) the coarse-grained method adopted by MAAPE is achieved by calculating the mean value of each coarse-grained time series, which slows down the dynamic mutation trend of the original time series to some extent; (2) the stability of MAAPE will decrease significantly when the time series is short [21]. In view of the abovementioned shortcomings, this paper proposes a refined composite generalized coarse grained technology and thus proposes refined composite generalized multiscale amplitude aware permutation entropy (RCGMAAPE) method.

Although RCGMAAPE enjoys excellent performance, it is only applied to a single channel vibration data, thereby has insufficient characterization capacity for multichannel data, which reduces the quality of the fault information obtained to a certain extent [22]. Reasonable use of multiple channel fault information can achieve a more comprehensive diagnosis of rotating machinery faults. Based on the theory of multidimensional embedding reconstruction, RCGMAAPE is extended to multivariate, that is, the refined composite generalized multivariate multiscale amplitude aware permutation entropy (RCGmvMAAPE), which is adopted to realize the complexity measurement of multichannel data. At present, most of the feature extraction approaches based on vibration signals are univariate analysis techniques. These methods extract the single-entropy or multiscale entropy of multiple components to mine the fault characteristics of vibration signals, which can only effectively use the signal of a single channel. It has been proved by experiments that these methods also have good results and can achieve accurate classification of rotating machinery fault types. However, it can be noted that the features composed of multiscale entropy of multiple components usually are high-dimensional and contain more redundant information, so it is necessary to reduce the dimensionality to improve the classification efficiency and accuracy. To this end, this paper develops a new feature extraction model that can realize multivariate analysis using only a single-channel signal. The principle is to disintegrate the fault signal into a pair of high and low resonance components through the parameter optimized resonance sparse decomposition algorithm. After that, these two components are employed as multichannel data to form a multivariate signal. Finally, the proposed RCGmvMAAPE method is applied to extract the fault feature of the signal.

After obtaining the fault characteristics of rotating machinery, selecting a suitable classifier for fault identification is the very critical part. At present, the commonly used classifiers include support vector machine (SVM) and extreme learning machine (ELM), which are widely used in pattern recognition because of their good generalization and reliability. Nevertheless, the performance of SVM is easily affected by the parameters, which need to be optimized [23]. ELM has high classification efficiency and excellent performance. However, it is prone to large errors when dealing with high-dimensional nonlinear classification problems since the kernel function is not used [24]. With the continuous development of deep learning, the application of deep learning to deal with classification problems has gradually become a feasible solution. However, deep learning is mainly aimed at the classification and

identification of large batches of data, so the performance of classification problems for small samples is not as good as machine learning. Deep belief network (DBN) is an atypical structure of deep learning in the processing of small samples, which is composed of multilayer restricted Boltzmann machines [25, 26]. DBN can effectively avoid the problem of parameter selection by using pretraining and repeated fine-tuning. In addition, it can be effectively used for the pattern recognition problem of small samples, so this paper employs it for the fault recognition of rotating machinery.

In conclusion, the main contribution of this paper is to propose a new multivariate feature extraction method, RCGmvMAAPE, and apply it to the fault diagnosis of rotating machinery. In addition, considering that the vibration signal may have only one channel, it cannot effectively take advantage of the multivariate analysis method. Therefore, an optimized RSSD is proposed to convert single-channel vibration signals into dual channel signals, so as to make full use of the advantage that the multivariate analysis method can extract fault information from multichannel vibration signals synchronously. The structure of this paper is as follows: Sections 2.1 and 2.2 mainly introduce the principle of RSSD and the specific implementation process of optimized RSSD. Sections 2.3 mainly introduces the theory of RCGmvMAAPE and compares it with RCmvMPE, RCmvMSE and mvMAAPE. Sections 3 introduces the specific steps of the proposed fault diagnosis method. Sections 4 validates the effectiveness of the proposed method by using two typical rotating machinery data. Section 5 draws the conclusion of this paper.

2. Modified RSSD Method

2.1. Principle of RSSD.

Resonance is a property of the signal. The larger the resonance property, the better the frequency aggregation of the signal; the smaller the resonance property, the better the time aggregation of the signal. The signal resonance sparse decomposition method sparsely decomposes a complex signal into high and low resonance components according to the different signal resonance properties. The resonance property is represented by Q, and the formula is as follows:

$$Q = \frac{f_c}{B_W},\tag{1}$$

where f_c is the center frequency of the signal and B_W is the bandwidth.

The resonant sparse signal decomposition method combines the influence of frequency and bandwidth on the signal. It can effectively separate the signals with overlapping frequency bands and similar center frequencies using different quality factors. This method first uses the two-channel filter bank shown in Figure 1 to perform TQWT on the signal to obtain a base function library with high and low-quality factors. Here, $H_0(w)$ and $H_1(w)$ are low-pass and high-pass filters, respectively; $v_0(n)$ and $v_1(n)$ are the filtered subband signals, respectively. The low-pass scale factor α and the
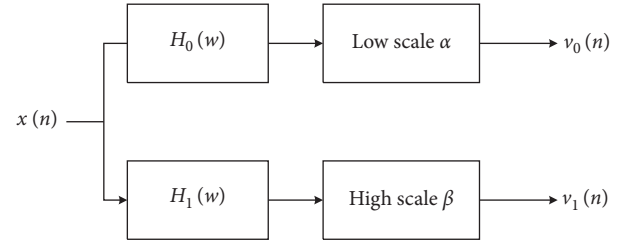

FIGURE 1: The two-channel filter banks.

high-pass scale factor β can be obtained through the quality factor Q and the redundancy r, as shown in (2).

$$\beta = \frac{2}{(Q+1)\alpha} = 1 - \frac{\beta}{r}.\tag{2}$$

Then, the corresponding coefficients were obtained by iteration, and the sparse decomposition objective function was established by morphological analysis method, as follows:

$$J(W_1, W_2) = \|x - S_1 W_1 - S_2 W_2\|_2^2 + \lambda_1 \|W_1\|_1 + \lambda_2 \|W_2\|_1,\tag{3}$$

where J is the objective function; W_1 and W_2 are the transformation coefficients of the subband signals x_1 and x_2 under the frameworks S_1 and S_2, respectively; and λ_1 and λ_2 are the regularization parameters.

The different values of λ_1 and λ_2 affect the energy distribution of sparse components. If only λ_1 increases, the energy of the corresponding component of λ_1 decreases, and the same is true for λ_2. If λ_1 and λ_2 are increased at the same time, the residual component energy will increase. In equation (3), the first norm is not differentiable, making it challenging to solve. For this reason, this paper adopts the split augmented Lagrangian search algorithm. The objective function is minimized by iterative updating, and finally, the high and low resonance components are separated.

$$\begin{cases} \widehat{x}_1 = S_1 W_1^*, \\ \widehat{x}_2 = S_2 W_2^*, \end{cases}\tag{4}$$

where W_1^* and W_2^* are, respectively, the transformation matrix of the high and low resonance components when the objective function J is the minimum; $\widehat{x}_1$ and $\widehat{x}_2$ are the estimated values of high and low resonance components, respectively.

2.2. HHO-RSSD Decomposition

2.2.1. Fitness Function.

Correlation kurtosis is an index used to evaluate the impact component content [27]. Compared with a single kurtosis index, correlation kurtosis introduces a correlation function based on kurtosis to perform auxiliary operations, which can characterize the transformation of shock components in a shock signal. The correlation kurtosis can be expressed as follows.

$$CK_M(T) = \sum_{n=1}^{N} \left(\frac{\prod_{m=0}^{M} x_{n-MT}}{\left(\sum_{n=1}^{N} x_n^2 \right)^{M+1}} \right), \tag{5}$$

where x_n represents the initial time series; N is the number of data points contained in the signal; T indicates the period of the required pulse signal; M is the number of offset periods.

The correlation kurtosis index is sensitive to the impact component in the signal, and its sensitivity is related to the settings of the parameters T and M. For a vibration signal, when the given parameter T matches the period of the initial signal, at this time, as the impact component in the signal increases, the magnitude of the correlation kurtosis also increases. In addition, in the decomposition process of RSSD, more identical components may be decomposed between high and low resonance components. To avoid this situation, the constraint condition of cross-correlation is introduced. Assuming there are two signals X and Y, the correlation coefficient between the two signals is expressed as follows:

$$C = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})} * \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}, \tag{6}$$

where C represents the correlation coefficient of the two signals, and the value range is $[-1,1]$. When C is -1, the two signals are negatively correlated; when C is 0, the two signals are not correlated; when C is 1, the two signals are positively correlated, and the two signals can be considered the same signal. Therefore, combining the advantages of the correlation kurtosis and the correlation coefficient, the maximum ratio of the correlation kurtosis value of the low resonance component to the correlation coefficient of the high resonance component is used as the fitness function K, as follows:

$$K = \frac{CK_M}{C}. \tag{7}$$

2.2.2. HHO Algorithm.
The Harris Hawk optimization algorithm is an intelligent optimization algorithm that simulates the predation behavior of Harris Hawk. It mainly consists of three parts: the search phase; and the conversion and development phase.

(1) Search Phase. Harris Hawks randomly roost somewhere and find their prey through two strategies:

$$X(t+1) = \begin{cases} X_{\text{rand}}(t) - r_1 |X_{\text{rand}}(t) - 2r_2 X(t)|, q \geq 0.5, \\ [X_{\text{rabbit}}(t) - X_m(t)] - r_3 [lb + r_4 (ub - lb)], q < 0.5, \end{cases} \tag{8}$$

where $X(t)$ and $X(t+1)$ are the positions of the individuals in the current and next iteration respectively; t is number of iterations; $X_{\text{rand}}(t)$ is the position of the randomly selected individual, $X_{\text{rabbit}}(t)$ is the position of the prey, that is, the position of the individual with the best fitness; r_1, r_2, r_3, r_4, and q are all random numbers between $[0,1]$. q is used to

select the strategy to be adopted, $X_m(t)$ is the average position of the individual; (ub, lb) refers to the range of the initial random position of the eagle and the expression is as follows:

$$X_m(t) = \sum_{k=1}^{M} \frac{X_k(t)}{M}, \tag{9}$$

where $X_k(t)$ is the position of the k-th individual in the population and M is the population size.

(2) Search and Development Conversion Phase. The HHO algorithm can switch between different development behaviors according to the escape energy of the prey. During the flight, the energy of the prey will be greatly reduced. In order to simulate this situation, the energy of the prey can be expressed as

$$E = 2E_0 \left(1 - \frac{t}{T} \right), \tag{10}$$

where E_0 is the initial energy of the prey, which is a random number between $[-1,1]$, which is automatically updated at each iteration, t is the number of iterations, and T is the maximum number of iterations.

(3) Development Phase. Define r as a random number between $[0,1]$, used to select different development strategies. When $0.5 \leq |E| < 1$ and $r \geq 0.5$, the soft siege strategy is adopted for position update:

$$X(t+1) = \Delta X(t) - E|JX_{\text{rabbit}}(t) - X(t)|, \tag{11}$$

where $\Delta X(t) = X_{\text{rabbit}}(t) - X(t)$ represents the difference between the position of the prey and the current position of the individual, and J is a random number between $[0, 2]$.

When $|E| < 0.5$ and $r \geq 0.5$, a hard siege strategy is adopted to update the position:

$$X(t+1) = X_{\text{rabbit}}(t) - E|\Delta X(t)|. \tag{12}$$

When $0.5 \leq |E| < 1$ and $r < 0.5$, the asymptotic fast swooping soft siege strategy is adopted for position update:

$$X(t+1) = \begin{cases} Y, f(Y) < f(X(t)), \\ Z, f(Z) < f(X(t)), \end{cases} \tag{13}$$

$$Y = X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}}(t) - X(t)|, \tag{14}$$

$$Z = Y + S * \text{LF}(2), \tag{15}$$

where f is the fitness function, S is a two-dimensional random vector, the elements are random numbers between $[0,1]$, and LF is the mathematical expression of Levi flight.

When $|E| < 0.5$ and $r < 0.5$, the position is updated by the hard encircling strategy of asymptotic fast swooping:

$$X(t+1) = \begin{cases} Y, f(Y) < f(X(t)), \\ Z, f(Z) < f(X(t)), \end{cases} \tag{16}$$

$$Y = X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}}(t) - X_m(t)|, \tag{17}$$

$$Z = Y + S * LF(2). \tag{18}$$

Algorithm steps:

Step 1: population initialization. According to the upper and lower bounds of each dimension of the search space, initialize each individual.

Step 2: calculate the initial fitness. Set the position of the individual with the best fitness as the current prey position.

Step 3: location update. First, update the escape energy of the prey, and then execute the corresponding location update strategy in the search or development behavior according to the escape energy and the generated random number.

Step 4: calculate fitness. Calculate the fitness of the individual after the location update, and compare it with the fitness value of the prey. If the fitness value of the individual after the location update is better than the prey, the individual location with the better fitness value is used as the new prey location.

Repeat Steps 3 and 4, when the number of iterations of the algorithm reaches the maximum number of iterations. Output the current position of the prey as the estimated position of the target.

2.2.3. HHO-Optimized RSSD Algorithm Steps. The technical flowchart of the proposed HHO optimization RSSD is shown in Figure 2 , and the specific optimization steps are as follows:

(1) The random position of the eagle was initialized, the number of iterations $T = 50$, the population size $N = 20$, and the three parameters were set between 0.001 and 200, and determine the fitness evaluation function K.

(2) Preset the value range of the parameters to be optimized, such as the quality factor Q, the weight coefficient A, and the Lagrangian multiplier u, and give an initial value randomly.

(3) The RSSD is used to decompose the vibration signal, and the three parameters of the RSSD are optimized through the HHO algorithm. After iteration, the local optimal parameters are retained.

(4) HHO updates the position of the eagle through different strategies, introduces it into the RSSD, obtains the fitness function value, compares it with the optimal fitness function value obtained in the previous iteration, and obtains the optimal parameters corresponding to this optimal fitness function.

(5) When the number of algorithm iterations reaches the maximum number of iterations, the global optimal fitness function value and the optimal parameter value are output.
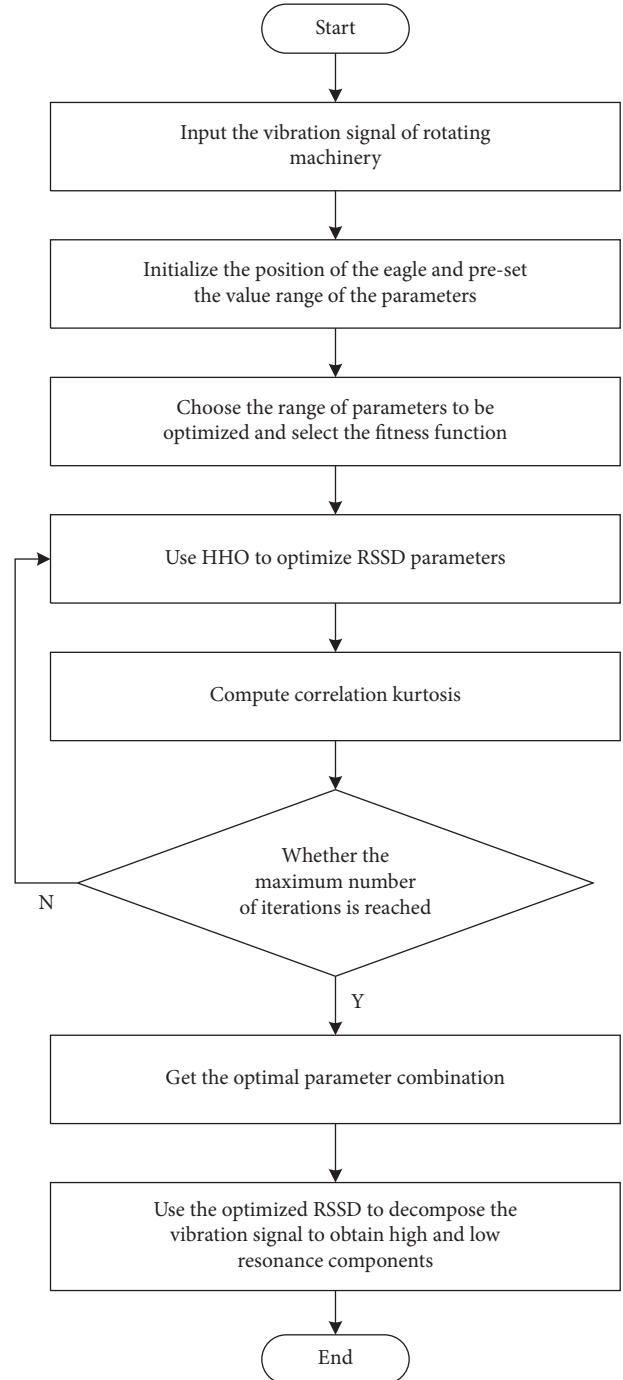

FIGURE 2: The technical flowchart of the proposed HHO optimization RSSD.

(6) Substitute the optimal parameter combination into the RSSD to realize the decomposition of the vibration signal.

2.3. RCGmvMAAPE

2.3.1. AAPE. AAPE is based on PE [28]. Therefore, its theory is very similar to PE similarity, and it is necessary to explain the specific improvement method of AAPE after describing

the theory of PE. The foundation principle of PE is as follows:

(1) For a one-dimensional time series $x = \{x_1, x_2, \ldots x_i, \ldots x_N\}$ of length N, at any time point t, the reconstruction vector of m dimension can be generated by the reconstruction of x

$$X_t^{m,d} = \left\{x_t, x_{t+d}, \ldots, x_{t+(m-2)d}, x_{t+(m-1)d}\right\}$$
$$t = 1, 2, \ldots, N - (m-1)d, \tag{19}$$

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

(2) In each reconstruction vector, according to the size of each element, in ascending order, the permutation of $\pi_{r_0, r_1, \cdots, r_{m-1}}$ can be obtained, which fulfills:

$$\left\{x_{t+(j_1-1)d}, x_{t+(j_2-1)d}, \ldots, x_{t+(j_{m-1}-2)d}, x_{t+(j_m-1)d}\right\}, \tag{20}$$

where j_* denote the index of the column of each element in the reconstructed component. Thus, when the embedding dimension is m, there are $m!$ possible ordinal patterns, of which the ith permutation is marked as π_i.

(3) The relative frequency of occurrence of π_i in each permutation pattern is described as

$$p(\pi_i) = \frac{g(\pi_i)}{N - (m-1)d}, \tag{21}$$

where $g(\pi_i)$ indicate a function that counts the number of occurrences of π_i in $X_t^{m,d}$. Whenever the permutation order of the internal elements of $X_t^{m,d}$ is π_i, the value of $g(\pi_i)$ linearly increases by 1.

(4) Therefore, according to the definition of entropy, PE can be described as

$$PE(x, m, d) = -\sum_{\pi_i=1}^{\pi_i=m!} p(\pi_i)\ln p(\pi_i). \tag{22}$$

Although PE has excellent performance, it is still found to have more serious defects, which makes it less reliable in quantifying the complexity of time series. First of all, based on the above description, PE only considers the contribution of the ordering structure of the time series to the complexity when calculating the probability, while the influence of the amplitude information of each data point in the time series on the entropy value is not calculated. Secondly, when there are components with equal amplitude in the time series, the influence of this group of elements on the entropy value is not clearly stated. For this reason, by enhancing the sensitivity to the amplitude and frequency of the time series, AAPE is more comprehensive and accurate in measuring the complexity of the time series. The principle of AAPE is reviewed as follows:

Supposing that the starting value of $p(\pi_i^{m,d})$ is 0, for the reconstruction vector $X_t^{m,d}$, when the time t adds from 1 to N-m+1 increasingly, the value of $p(\pi_i^{m,d})$ is updated when the permutation $\pi_i^{m,d}$ changed.

$$p^{\text{update}}\left(\pi_i^{m,d}\right) = p\left(\pi_i^{m,d}\right) + \left(\frac{\alpha}{m}\sum_{k=1}^{m}\left|x_{t+(k-1)d}\right| + \frac{1-\alpha}{m-1}\sum_{k=2}^{d}\left|x_{t+(k-1)d} - x_{t+(k-2)d}\right|\right), \tag{23}$$

where $\alpha \in [0, 1]$ denotes the adjustment coefficient, which is used to adjust the weight of the time series amplitude mean and the deviation between the amplitudes. Thus, the

probability of $p(\pi_i^{m,d})$ occurring in the whole time series is $\pi_i^{m,d}$.

$$p\left(\pi_i^{m,d}\right) = \frac{p^{\text{update}}\left(\pi_i^{m,d}\right)}{\sum_{t=1}^{N-m+1}\left(\alpha/m\sum_{k=1}^{m}\left|x_{t+(k-1)d}\right| + 1 - \alpha/m - 1\sum_{k=2}^{m}\left|x_{t+(k-1)d} - x_{t+(k-2)d}\right|\right)}. \tag{24}$$

The AAPE of time series can be computed as follows:

$$\text{AAPE}(x, m, d, \alpha) = -\sum_{\pi_k=1}^{\pi_k=m!} p(\pi_k)\ln p(\pi_k). \tag{25}$$

2.3.2. mvAAPE. In this part, the multivariate amplitude perception permutation entropy is developed to quantify the

complexity of multichannel time series. The principle of mvAAPE can be expressed as follows [29]:

(1) Given the multivariate data $X = \{X_{c,1}, X_{c,2}, \ldots, X_{c,i}, \ldots, X_{c,L}\}c = 1, 2, \ldots, q$ of q channels of length L to be analyzed. Perform phase space reconstruction on each sample, and the resulting matrix is as follows:

$$Z = \{X_{c,i}, X_{c,i+d}, \ldots, X_{c,i+(m-2)d}, X_{c,i+(m-1)d}\}. \qquad (26)$$

(2) Rearrange the reconstruction vector Z into $\{x_{c,i+(j_1-1)d} \leq x_{c,i+(j_2-1)d} \leq \cdots \leq x_{c,i+(j_{m-1}-1)d} \leq x_{c,i+(j_m-1)d}\}$ in ascending order. At this time, the possible sorting mode π_i exists $m!$

(3) Assume that the starting value of $p(\pi_{c,i}^{m,d})$ is zero. For the reconstruction vector Z, when the time i gradually increases from 1 to $L-m+1$, the value of $p(\pi_{c,i}^{m,d})$ is updated every time $\pi_{c,i}^{m,d}$ appears.

$$p^{update}(\pi_{c,i}^{m,d}) = p(\pi_{c,i}^{m,d}) + \left(\frac{\alpha}{m} \sum_{k=1}^{m} |x_{c,t+(k-1)d}| + \frac{1-\alpha}{m-1} \sum_{k=2}^{m} |x_{c,t+(k-1)d} - x_{c,t+(k-2)d}| \right). \qquad (27)$$

(4) Compute the probability of i-th sorting pattern $\pi_i (1 \leq i \leq m!)$ in c-th channel as

$$p(\pi_{c,i}^{m,d}) = \frac{p^{update}(\pi_{c,i}^{m,d})}{\sum_{t=1}^{N-m+1} \left(\alpha/m \sum_{k=1}^{m} |x_{c,t+(k-1)d}| + 1 - \alpha/m - 1 \sum_{k=2}^{m} |x_{c,t+(k-1)d} - x_{c,t+(k-2)d}| \right)}. \qquad (28)$$

For q-channel time series, $p(\pi_{c,i})$ satisfies $\sum_{c=1}^{q} \sum_{i=1}^{m!} p(\pi_{c,i}) = 1$.

(5) The probability of the i-th pattern π_i in q-channel time series can be computed as follows:

$$p(\pi_i) = \sum_{c=1}^{q} p(\pi_{c,i}). \qquad (29)$$

(6) According to the definition, mvAAPE can be expressed as

$$\text{mvAAPE}(X, m, d, \alpha) = -\sum_{i=1}^{m!} p(\pi_i) \ln p(\pi_i). \qquad (30)$$

mvAAPE mainly integrates data from multiple channels so that AAPE can extract more features, making the analysis more comprehensive and accurate. However, mvAAPE can only extract the features of the signal on a single scale. But, the effective information contained in the actual vibration signal is often presented on multiple scales, and it is difficult to fully extract the fault characteristics in the vibration signal by only carrying out a single-scale analysis. Therefore, in order to mine the fault information of the vibration signal from multiple scales and enhance the robustness of the analysis, the multivariate multiscale amplitude aware permutation entropy was developed.

2.3.3. mvMAAPE. The realization principle of mvMAAPE is to obtain multiple coarse-grained time series by performing coarse-grained processing on the multichannel time series. These coarse-grained time series respectively represent the vibration information of the original multichannel signal at various scales. Subsequently, based on mvAAPE to mine the

fault information in these coarse-grained time series to realize mvMAAPE analysis. The basic implementation principle of mvMAAPE is described as follows.

(1) For q channel time series $U = \{u_{k,1}, u_{k,2}, \ldots, u_{k,i}, \ldots, u_{k,L}\}$ with data points L. The multivariate coarse-grained time series at scale factor τ is computed as

$$y_{k,j}^{\tau} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} u_{k,i} \quad 1 \leq j \leq \frac{L}{\tau}, 1 \leq k \leq p, \qquad (31)$$

Where τ is the scale factor, the coarse grained time series is the raw time series when $\tau = 1$. When $\tau > 1$, the original time series is divided into coarse-grained time series of length L/τ.

(2) Compute the mvAAPE of each multivariate coarse-grained time series, and get the mvMAAPE of U as follows:

$$\text{mvMAAPE} = \text{mvAAPE}(y_{k,j}^{\tau}, m, d, \alpha). \qquad (32)$$

By extending mvAAPE from single-scale analysis to multiscale, more information can be obtained from multivariate coarse-grained time series of different scales, which is called multivariate multiscale amplitude aware permutation entropy analysis. However, in the abovementioned multivariate coarse-grained time series with a scale factor of τ, only the information of the multivariate coarse-grained time series starting from $u_{k,1}$ is considered, and the information of the remaining $\tau - 1$ multivariate time series is not used. mvMAAPE does not consider the relationship between adjacent coarse-grained time series, resulting in a lack of statistical information.

2.3.4. RCGmvMAAPE. To overcome the shortcomings of mvMAAPE, a new entropy method is proposed, which is called RCGmvMAAPE. Compared with mvMAAPE, RCGmvMAAPE mainly made two improvements. Firstly, to reduce the large variance of mvFE when the scale factor is large, this paper adopts refined composite analysis to achieve coarse graining of time series, which can reduce the dependence of entropy value on the length of time series data and achieve stable results even when the length of time series to be analyzed is short. Secondly, in order to accurately describe the dynamic changes of the time series, the second-order moment (root mean square) is used to replace the first-order moment (mean) used in the traditional coarse-graining method, so that it has a stronger fault feature extraction ability. The principle of RCGmvMAAPE is as follows:

(1) For the n-channel multivariate time series $X = \left\{x_{k,b}\right\}_{b=1}^{L}$, $k = 1, 2, \ldots, n$ with data point L, the multivariate coarse grained time series is calculated by using root mean square instead of mean value at scale factor τ. The elements in the a-th coarse-grained time series $Y_a^{\tau} = \left\{y_{k,i,1}^{\tau}, y_{k,i,2}^{\tau}, \ldots\right\}$ are expressed as follows:

$$y_{k,i,a}^{\tau} = \sqrt{\frac{1}{\tau} \sum_{b=a+\tau(i-1)}^{a+i\tau-1} x_{k,b}^2} \quad 1 \le i \le \frac{L}{\tau}, \, 1 \le k \le n, \, 1 \le a \le \tau. \tag{33}$$

For a scale factor τ, there will be τ diverse coarse-grained multivariate time series, as present in Figure 3.

(2) For each coarse-grained multivariate time series, calculate the marginal relative frequencies $p\left(\pi_j\right)$. Then the average relative frequency $\overline{p\left(\pi_j\right)}$ can be calculated as follows

$$\overline{p\left(\pi_j\right)} = \frac{1}{\tau} \sum_{a=1}^{\tau} p_a\left(\pi_j\right). \tag{34}$$

(3) Therefore, the RCGmvMAAPE of the multichannel time series can be described as follows:

$$\text{RCGmvMAAPE} = -\sum_{\pi_j=1}^{\pi_j=m!} \overline{p\left(\pi_j\right)} \ln \overline{p\left(\pi_j\right)}. \tag{35}$$

2.3.5. Parameter Selection and Performance Analysis. In the RCGmvMAAPE algorithm, there are five parameters that need to be set in advance, namely the embedding dimension m, the time delay d, the adjustment coefficient a, the length of the time series N and the scale factor S. For the embedding dimension m, too small value will result in too few states contained in the reconstruction vector, and the algorithm losses its effectiveness, making it impossible to detect dynamic mutations in the time series. Conversely, if m is too large, the reconstruction of the phase space will homogenize

the time series, which not only increases the amount of calculation but also fails to highlight subtle changes in the time series. Therefore, consider setting the embedding dimension m as 5. The time delay has little effect on the performance of the algorithm, so set it as $d = 1$. The value of the adjustment coefficient is usually $a = 0.5$. In addition, the scale factor S cannot be set too large; otherwise, it will produce more redundant information and affect the efficiency of the analysis. On the contrary, too small value will make the information extraction insufficient and affect the effectiveness of the analysis, so this article is set as $S = 20$. The length of the time series also has a certain degree of influence on the performance of the algorithm. Without loss of generality, three-channel Gaussian white noise signals are used for analysis, the lengths are respectively $N = 256, 512, 1024, 2048, 4096, 8192$, and their RCGmvMAAPE is calculated under the condition that other parameters are the same. Figure 4 shows the entropy values under different lengths. From Figure 1, when the length $N \ge 2048$, the entropy curve is smoother and the fluctuation is small. At this time, the RCGmvMAAPE of the white noise signals of different lengths has a small difference and the performance is relatively stable, so $N = 2048$ is selected.

This part is mainly based on simulation signals to verify the excellent performance of RCGmvMAAPE in measuring the complexity of multichannel vibration signals. The RCGmvMAAPE method is compared with other typical multivariate analysis methods through four different multichannel signals. White Gaussian noise (WGN) and $1/f$ noise are two time series used to construct multichannel simulation signals. The irregularity of WGN is higher than $1/f$. Compared with WGN, the power spectrum of $1/f$ noise is more complex, so more mode information is integrated. The generation of WGN is random, so the probability of its state transition matrix is approximately equal. On the contrary, $1/f$ is a long-range correlation signal, and the irregularity of $1/f$ noise is low than that of WGN. Therefore, $1/f$ noise is more complicated than WGN.

Without loss of generality, multichannel signals with three different channels are generated based on WGN and $1/f$ noise, which are (a) three channel WGN; (b) two channel WGN and one channel $1/f$; (c) one channel WGN and two channel $1/f$; (d) three channel $1/f$. RCGmvMAAPE, RCmvMPE, RCmvMSE and mvMAAPE were studied, respectively. The data length of each channel is $N = 2\,048$. The mean and standard deviation curves of RCGmvMAAPE, RCmvMPE, RCmvMSE and mvMAAPE of the four synthetic signals are shown in Figure 3. It can be seen from Figure 5 that compared with the other three methods, the standard deviation of RCGmvMAAPE is significantly smaller, which shows that RCGmvMAAPE is more stable when measuring the complexity of multichannel time series. In addition, RCGmvMAAPE and RCmvMPE methods can clearly distinguish four kinds of multichannel synthetic signals, while mvMAAPE method cannot effectively separate (a), (b), and (c), which indicates that the coarse grainization method based on refined composite generalized processing can obtain more accurate results, thus effectively measuring the complexity
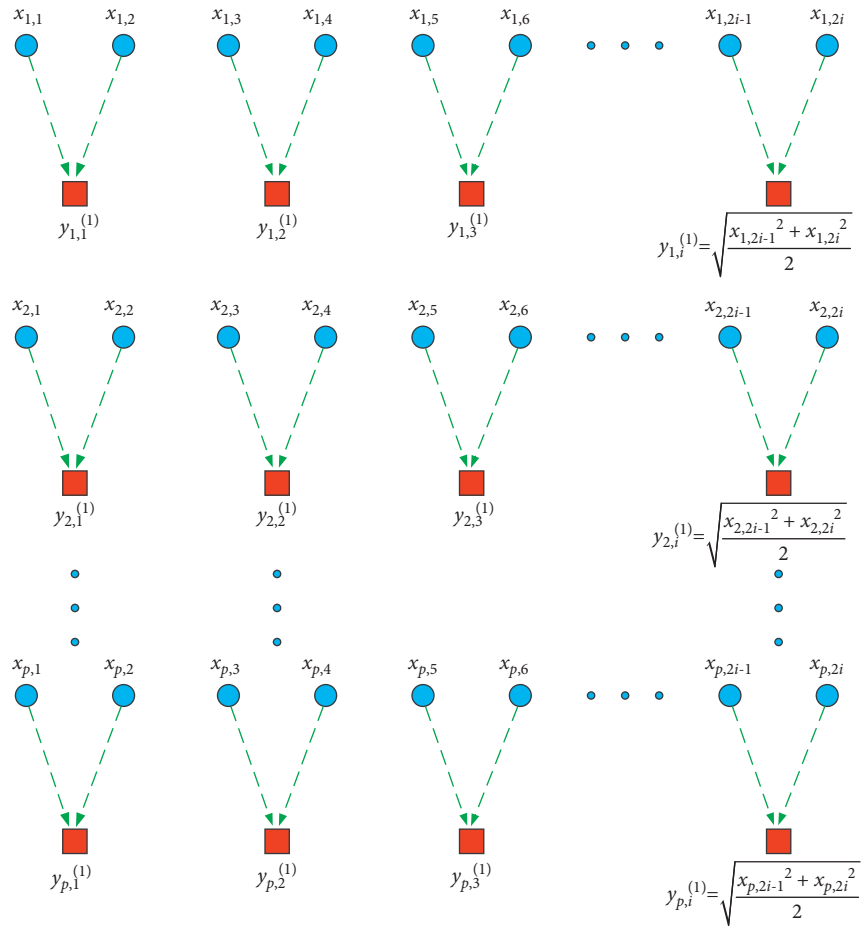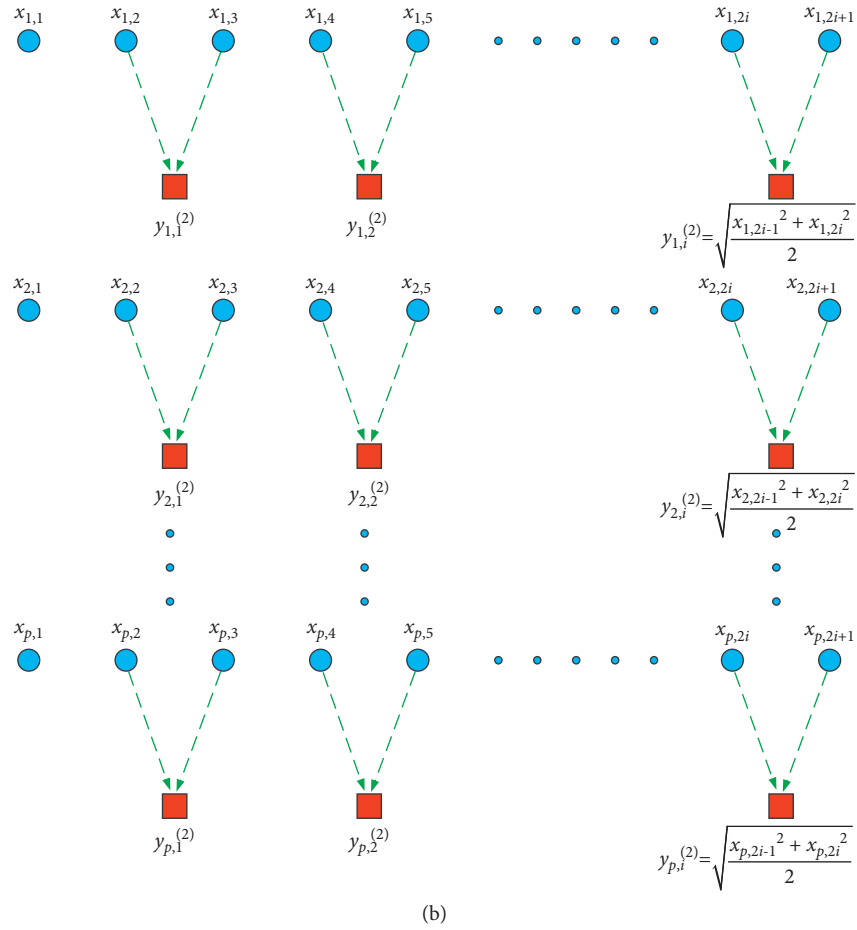
(a)

FIGURE 3: Continued.

(b)

FIGURE 3: Illustration of refined composite coarse-grained approach for multivariate time series with scale factor 2. (a) First coarse-grained time series; (b) second coarse-grained time series. Figure is reproduced from Fuming Zhou 2020.
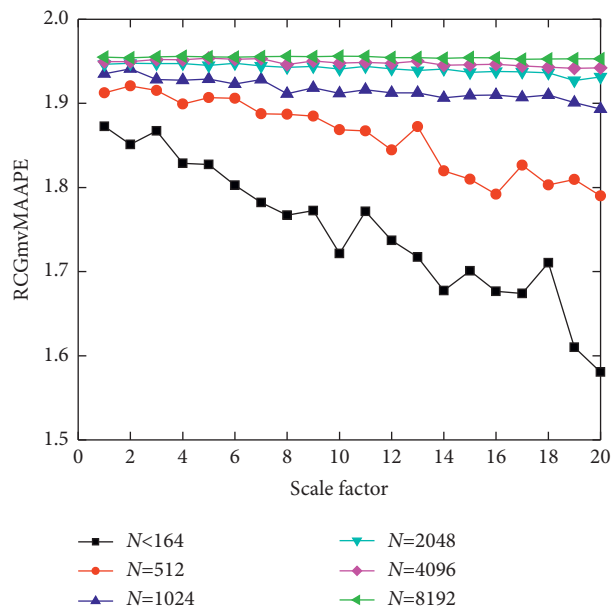

FIGURE 4: RCGmvMAAPE of three channel white noises under diverse lengths.
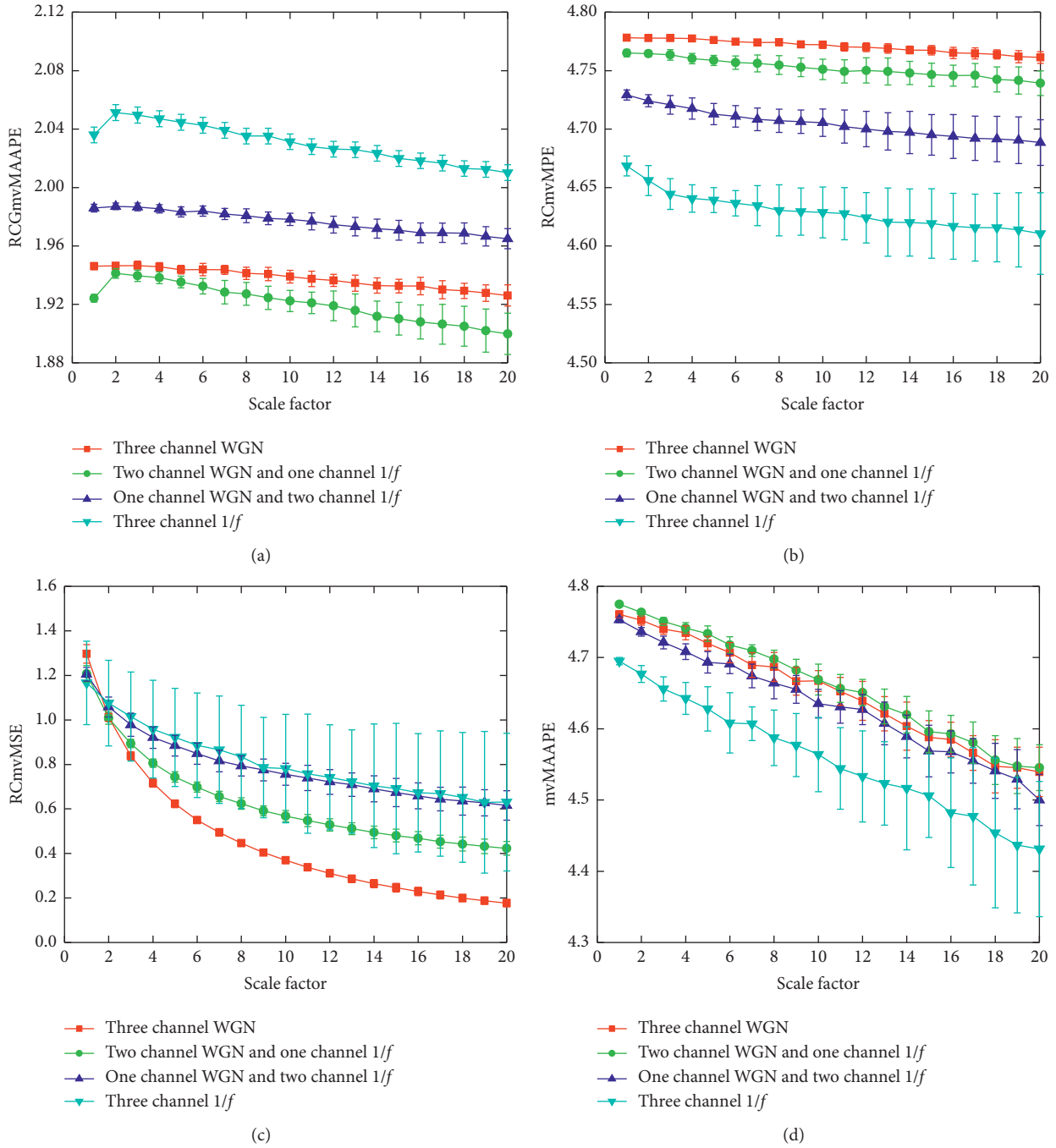
(a)

(b)

(c)

(d)

FIGURE 5: RCGmvMAAPE, RCmvMPE, RCmvMSE, and mvMAAPE of four multichannel synthetic signal.

of time series. Besides, RCmvMSE has a poor distinction between (c) and (d), which is mainly because the method is mainly based on multivariate sample entropy to realize its function, and sample entropy has many defects when processing time series, so RCmvMSE has a poor performance. In conclusion, compared with the other three multivariate analysis methods, RCGmvMAPE improves its ability to extract feature information from multi-channel vibration signals by adopting refined generalized composite coarsing processing, so it can better measure the complexity of multichannel signals.

3. The Proposed Fault Diagnosis Model

According to the previous analysis, RCGmvMAAPE can effectively measure the complexity of multichannel time series. The HHO-RSSD can adaptively decompose the vibration signal into high and low resonance components, and has excellent time-frequency analysis performance. Therefore, a new fault diagnosis technology for rotating machinery was developed. First, HHO-RSSD and RCGmvMAAPE are adopted to extract high-quality features that characterize the fault state from the signals of rotating machinery.

Subsequently, a deep belief network classifier with excellent generalization performance is used to identify the types of faults. The technical implementation process is shown in Figure 6. The detailed steps are as follows:

(1) Under a given sampling frequency, the vibration data of the rotating machinery in different fault states are collected through the accelerometer and divided into training samples and test samples.

(2) HHO algorithm is used to optimize the key parameters of RSSD and find the best combination of parameters. Subsequently, the optimized RSSD is used to decompose the vibration signal to obtain high and low resonance components containing rich vibration information to highlight the fault components.

(3) The high and low resonance components are used as multichannel data to construct a multivariate time series, and then the RCGmvMAAPE of the multivariate time series is calculated to generate fault features.

(4) The deep belief network classifier is trained through the training data set to obtain the best classifier model.

(5) The remaining test data set is input to the trained DBN classifier model for fault identification. According to the output result of the DBN classifier, the fault type of the rotating machinery is judged.

4. Experimental Verification

To validate the validity and reliability of the approach raised in this paper for the health recognition of general rotating machinery, experiments were carried out using two typical rotating machinery vibration data, rolling bearings and gears. The rolling bearing data is offered by the public data set, and the gear vibration data is collected on the QPZZ-II vibration test platform.

4.1. Case 1

4.1.1. Data Collection. To test the effectiveness of the raised approach of fault diagnosis for rotating machinery, firstly, experiments are carried out adopting rolling bearing data. The experimental data used the typical vibration data set of rolling bearings offered by the Electrical Laboratory of Case Western Reserve University [30]. The structure of the platform is shown in Figure 7. Seen from Figure 7, the vibration acquisition platform is composed of components such as a motor, a drive end bearing, a fan end bearing, and an accelerometer. The bearing model adopted in the experiment is 6205-2RS-JEM SKF. The running power of the motor is 0 horsepower and the rotating speed is 1797 rpm. The vibration data is collected by sensors installed at the drive end and the fan end. The sampling frequency is 12 kHz, and the sampling time for each working condition is 10 s. Different types of single-point faults are set up on the rolling bearings by EDM. The fault diameters are 0.1778 mm,
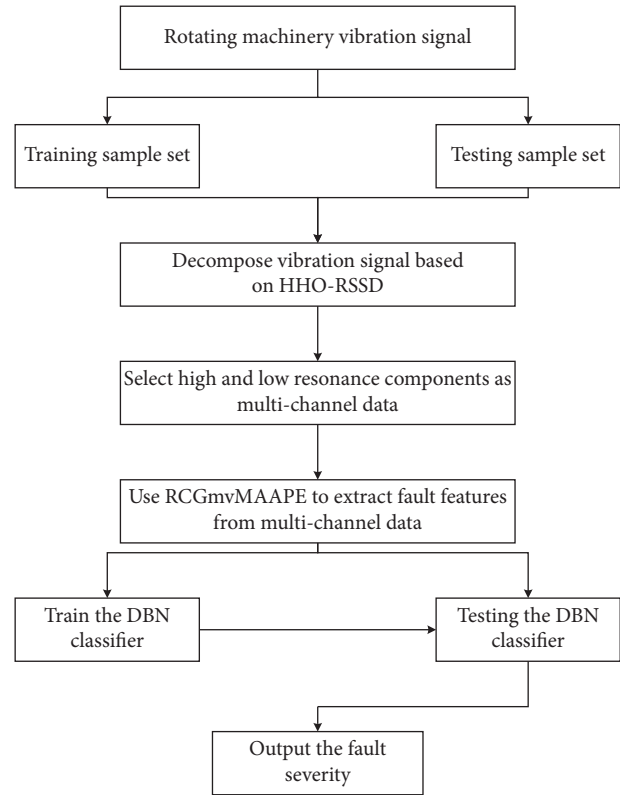
FIGURE 6: The technical implementation process.

0.3556 mm and 0.5334 mm, and the fault depth is 0.2794 mm. The fault diameter represents the severity of the fault of the rolling bearing. Experiments were performed for both fan and drive end bearings with outer race faults located at 6 o'clock. The fault data used in this experiment includes four types of normal, inner race fault, outer race fault, and ball fault. Each fault type contains three different severity, so a total of ten types of vibration data are included. The vibration data of each working condition is divided into 58 groups of nonoverlapping samples, and the number of data points contained in each sample is 2048. Randomly select 28 groups of samples as the training set, and the remaining 30 groups as the test set. The brief information of the data used is displayed in Table 1.

4.1.2. Analysis and Feature Extraction. Figure 8 presents the waveform of the vibration data adopted in the experiment. The waveform is a nonlinear modulation signal with complex components and contains a large number of irregular impact components. Therefore, it is hard to easily judge the status of fault of the rolling bearing by observing the waveform of the vibration signal, and further processing of the vibration signal is required to obtain more failure information.

This part mainly studies how to obtain the best RSSD algorithm. First input the rolling bearing vibration signal to HHO-RSSD to perform signal decomposition. Taking the normal state as an example, the iterative process ends when the value of the correlation kurtosis is the smallest.
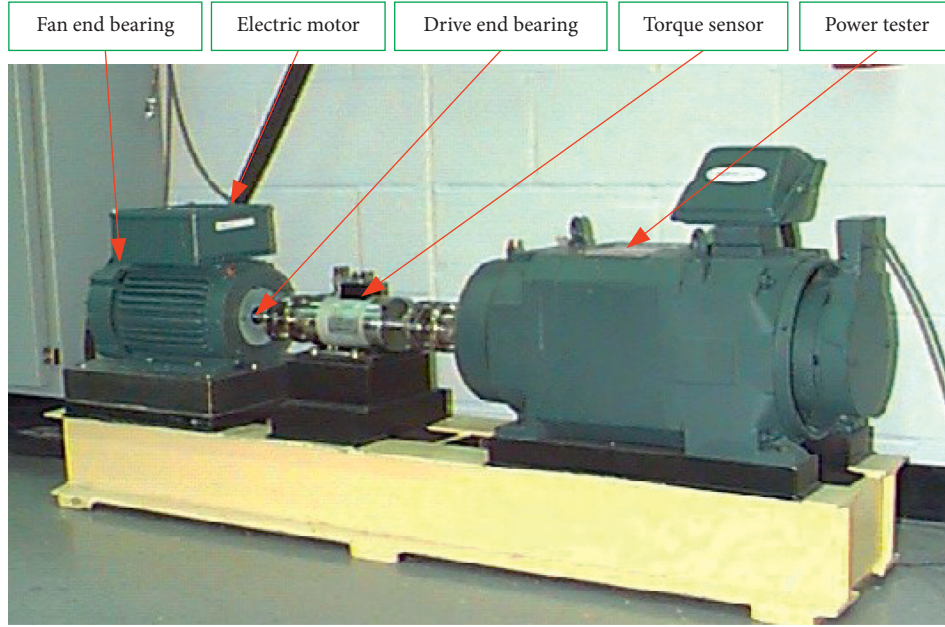
| Fan end bearing | Electric motor | Drive end bearing | Torque sensor | Power tester |

FIGURE 7: The fault simulation test platform of rolling bearing.

TABLE 1: The brief information of the data.

Working condition	Fault Diameter (mm)	Abbreviation	Training Sample	Testing Sample	Label
Normal		Nor	28	30	1
Inner race fault	0.1778	IRF1	28	30	2
	0.3556	IRF2	28	30	3
	0.5334	IRF3	28	30	4
Outer race fault	0.1778	ORF1	28	30	5
	0.3556	ORF2	28	30	6
	0.5334	ORF3	28	30	7
Ball fault	0.1778	BF1	28	30	8
	0.3556	BF2	28	30	9
	0.5334	BF3	28	30	10

Subsequently, after HHO optimized RSSD, a set of best parameter combinations were obtained, which are $Q_1 = 11.206$, $Q_2 = 0.958$, $A_1 = 0.92$, $A_2 = 0.16$, $u = 0.456$. Figure 9 is the evolution curve of fitness value in the optimization process of HHO. Seen from Figure 9, HHO can reach the local optimum relatively quickly, and then jump out quickly and reach the global optimum, and the final convergence value is also small, which shows that HHO has higher optimization performance. Then, the optimal parameter combination obtained by optimization is input into the RSSD, and the vibration signal is decomposed to acquire the high and low resonance components. The RSSD decomposition result of Nor is shown in Figure 10.

The high and low resonance components are taken as a multichannel time series, and then RCGmvMAAPE is adopted to excavate the fault features of the constructed multivariable data to construct the fault samples. In addition, to validate the effectiveness of the raised RCGmvMAAPE approach, it is compared with RCmvMPE, RCmvMSE, mvMAAPE and RCGMAAPE. The entropy results of seven methods are displayed in Figures 11(a)~ Figures 11(g). Here, Figures 11(a)–11(d) are the analysis results of four multivariate analysis methods on multivariate data composed of high resonance components and low resonance components; Figure 11(e) and Figure 11(f) are the results of using the univariate analysis method RCGMAAPE to analyze the high and low resonance components, respectively; Figure 11(g) is the analysis result of RCGmvMAAPE on the multivariate data composed of the vibration signals of the drive end bearing and the fan end bearing. By comparing Figures 11(a)–11(d), the advantages of RCGmvMAAPE in measuring the complexity of multichannel data over the other three methods can be validated. By comparing Figure 11(a) and Figures 11(e) and 11(f), it can be proved that using RCGmvMAAPE to analyze multichannel data is better than RCGMAAPE to analyze single-channel data. In addition, the comparison between Figure 11(a) and Figure 11(g) can prove that after proper processing, only a single-channel vibration signal can also achieve good results. Seen from Figure 11, compared with several other feature extraction models, the standard deviation of the entropy value of Figure 11(a) is smaller and the performance is more stable. On most scales, the ability of Figure 11(b) to distinguish between bearing faults is not
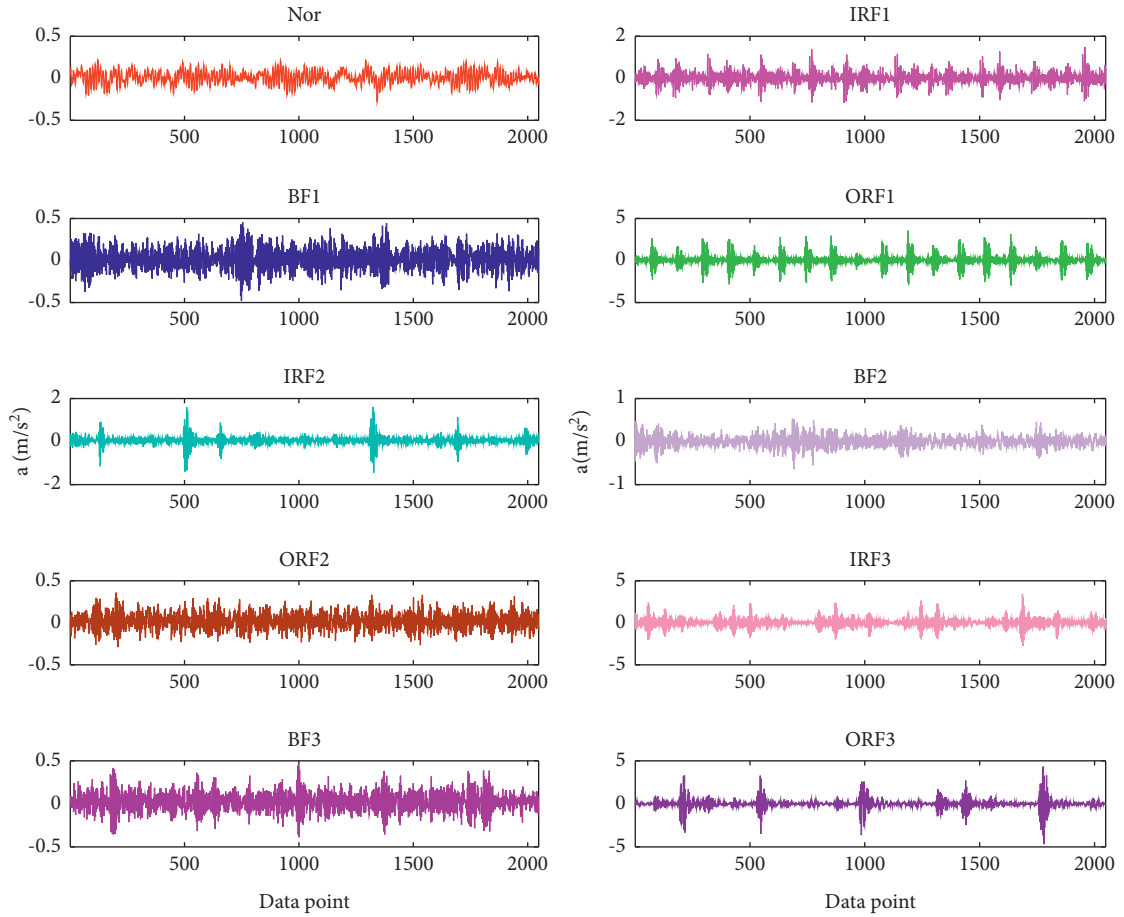
FIGURE 8: The vibration waveform of rolling bearing under ten fault conditions.
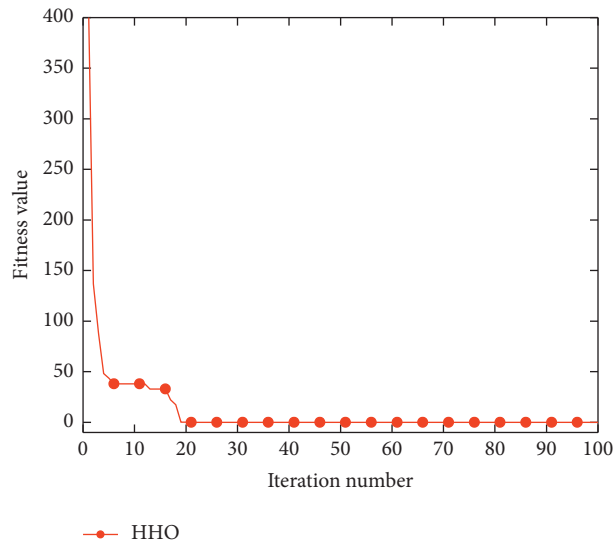
FIGURE 9: The evolution curve of fitness value.

satisfactory. Especially for IRF1 and BF1 samples, the curves of these two samples have obvious aliasing, so it is difficult to distinguish these two fault states. Figure 11(c) has a relatively obvious degree of discrimination, but its entropy deviation is

obviously larger, that is, the error is larger. Compared with Figure 11(a), Figure 11(d) has significantly worse ability to distinguish samples from each state, and the entropy deviation is also larger, which indicates that its performance is
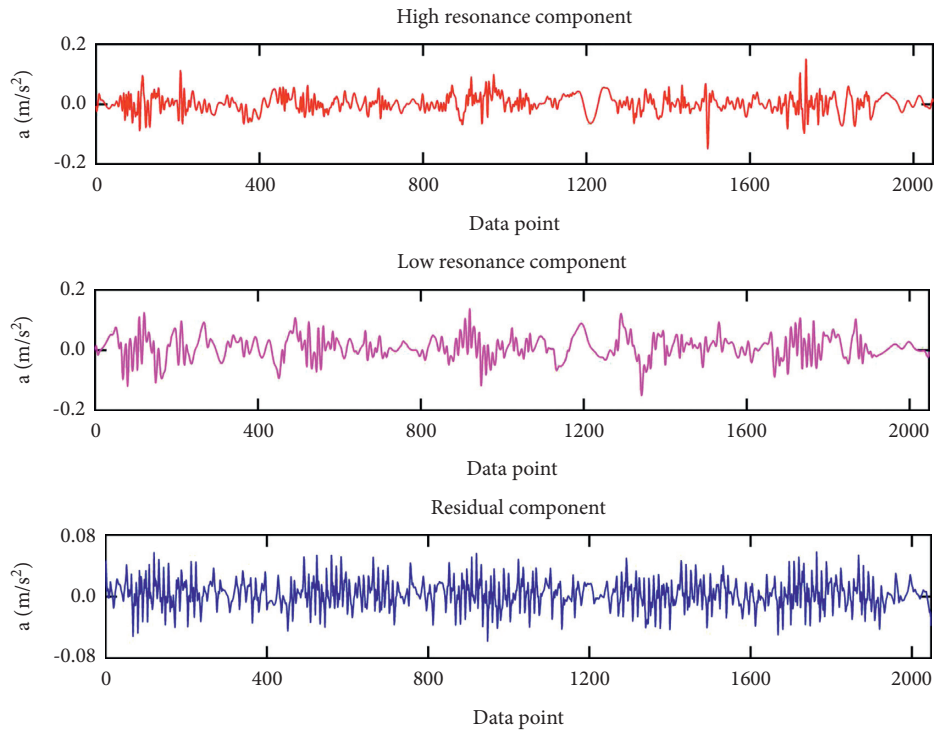
FIGURE 10: The RSSD decomposition result of vibration signal nor.

(a)

(b)
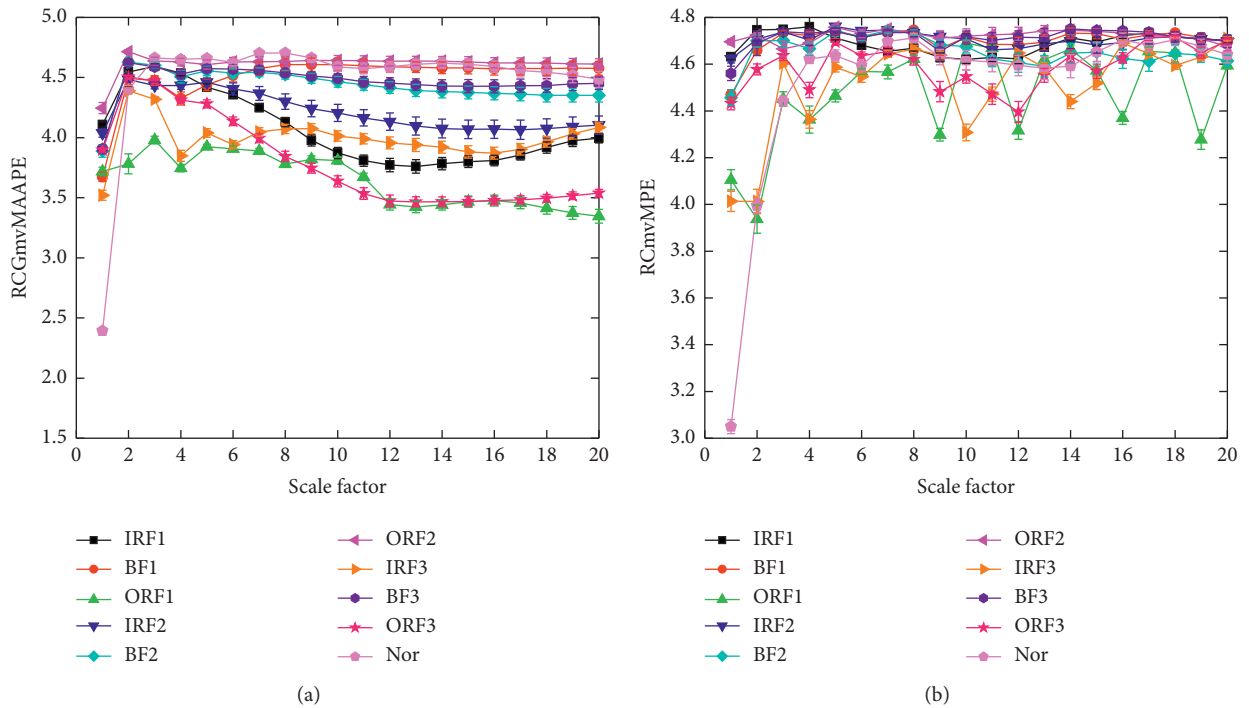
FIGURE 11: Continued.
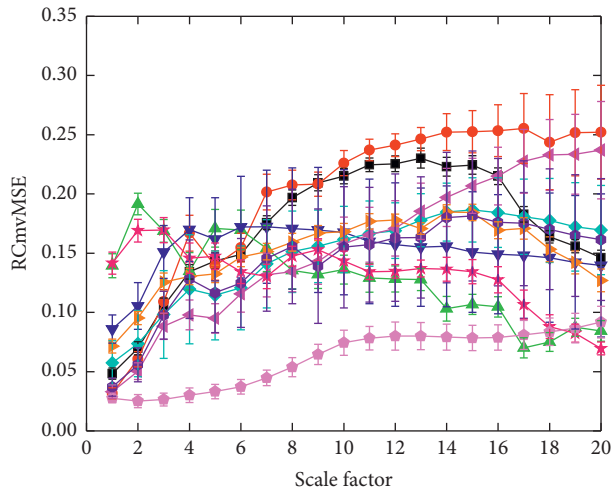
(c)
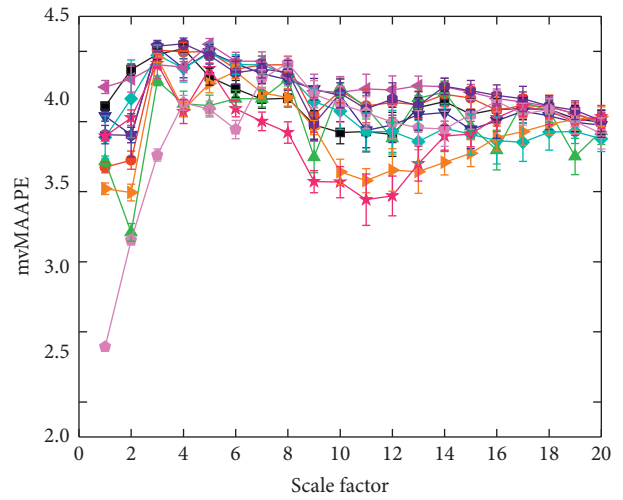
(d)
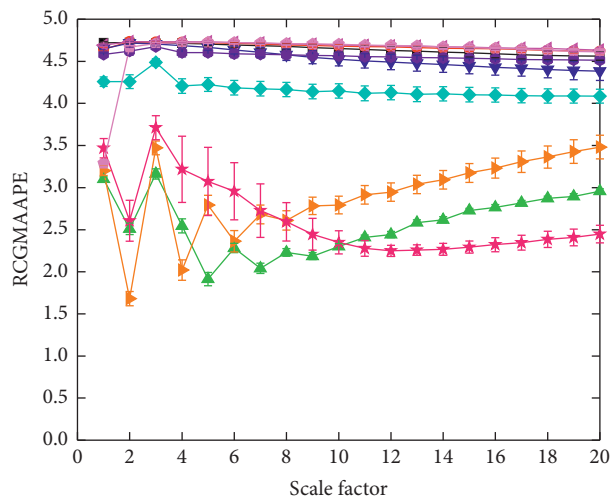
(e)
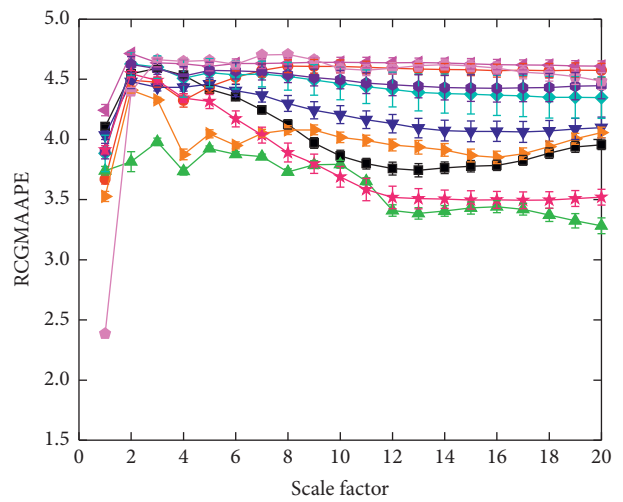
(f)

Figure 11: Continued.

(g)
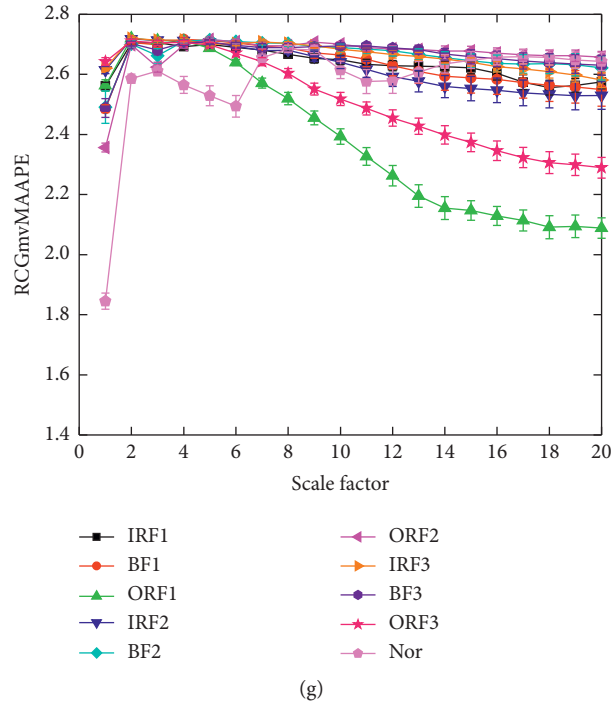
FIGURE 11: The entropy results of rolling bearing data analyzed by using seven methods.

unstable and its reliability is low. The entropy curves of the six samples in Figure 11(e) are obviously aliased, and the ability to distinguish these samples is very poor. Figure 11(f) has excellent performance, but it can be found that its entropy deviation is slightly larger than that of Figure 11(a), and the stability is insufficient, so its performance is weaker than Figure 11(a). Figure 11(g) can effectively distinguish between ORF1 and ORF3 samples, while the ability to distinguish other samples is weaker than Figure 11(a). But it can be found that the entropy deviation of Figure 11(g) is smaller, that is, the stability and reliability are better. This is mainly because the vibration signal of the fan end also contains the vibration information of the bearing during operation. As the Figure 11(g) integrates the information of two bearings, it has a relatively excellent effect. In summary, using RCGmvMAAPE to analyze multichannel data composed of high and low resonance components can achieve very excellent results, and the effect is even better than that of data composed of vibration signals from the drive end and the fan end.

To compare the performance of the abovementioned seven feature extraction models from a more intuitive perspective, the t-stochastic neighbor embedding (t-SNE) method is used for auxiliary analysis. The t-SNE approach is adopted to project the original features into a two-dimensional space. The visualization of the features extracted by the seven methods is 1 in Figure 12. Observed from Figure 12, the features of the same category in Figure 12(a) are accurately clustered, and samples of different categories are separated from each other, that is, the features are highly distinguishable. However, the distinguishability of features extracted based on other six models is weak, and some samples are aliased with each other, which makes it difficult to distinguish their categories. Comparing Figures 12(e) and 12(f), It can be observed that the visualization effect of Figure 12(f) is better than Figure 12(e), which shows that the low resonance component contains more fault information, so the extracted quality is higher. In addition, by comparing Figures 12(a) and 12(g), it can be found that the visualization effect of Figure 12(a) is better, while the BF2 and BF3 samples in Figure 12(g) show obvious aliasing, and the distribution of samples in the same category is relatively scattered, without obvious clustering center. Therefore, by visualizing the features, it can be proved that using RCGmvMAAPE to extract features from multivariate data composed of high resonance components and low resonance components has a better effect, which proves the reliability and effectiveness of the raised approach.

4.1.3. Fault Recognition. To quantify the performance of the above seven feature extraction models on rolling bearing fault diagnosis, the state features excavated by the seven approaches are input into the DBN recognizer for fault classification. The confusion matrix is a tool for describing the performance of a classification model. It contains information about the actual and predicted classifications completed by the classification model, which can be used to evaluate the performance of the classification model. By observing the confusion matrix, the detailed classification results of each category can be clear. The confusion matrix of the seven feature extraction models is displayed in Figure 13. Observed from Figure 13, the proposed fault diagnosis method achieves the best fault recognition rate, and samples
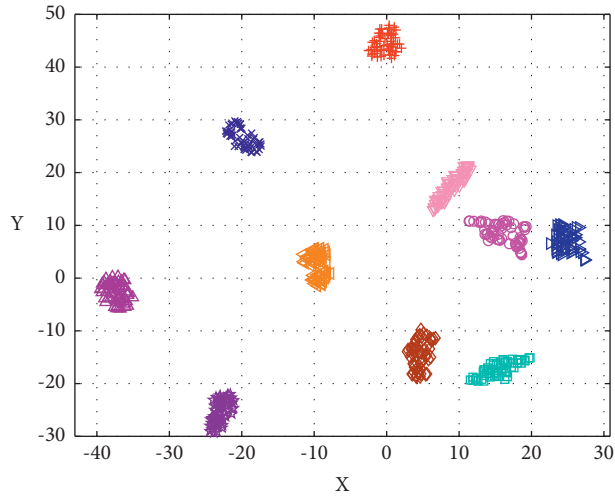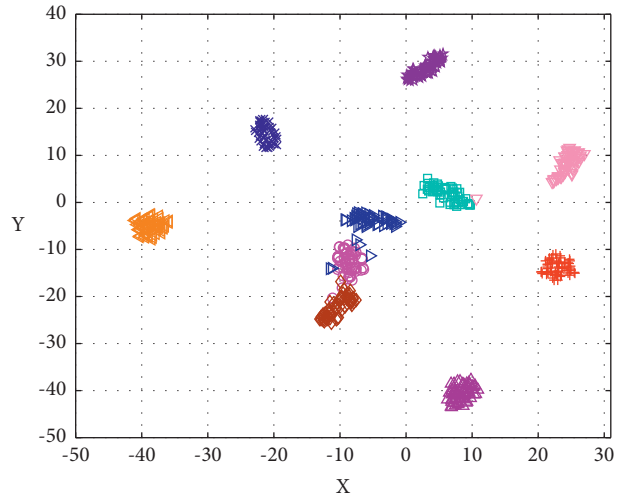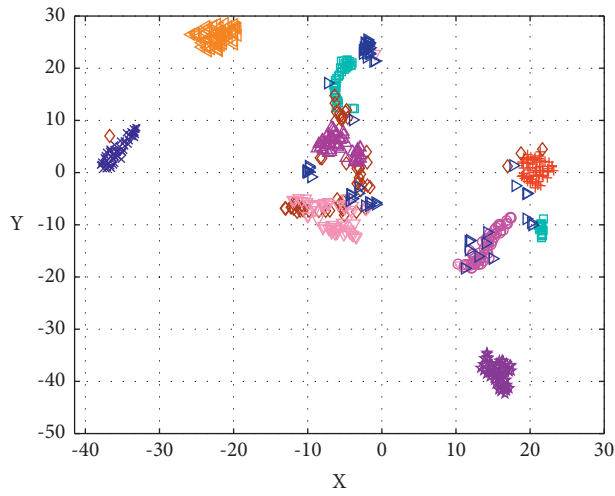
(a)

(b)

(c)
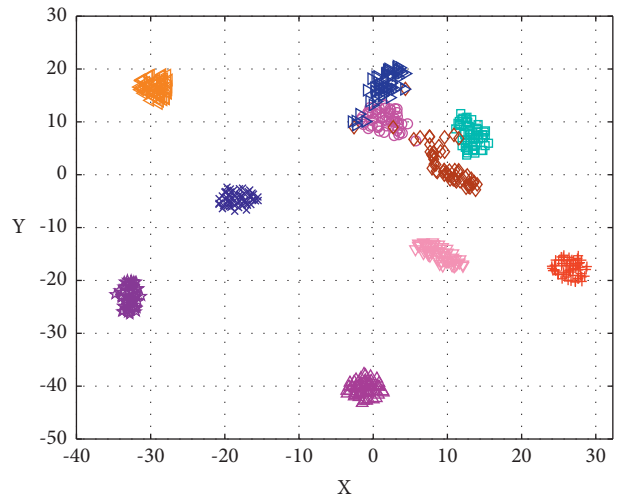
(d)
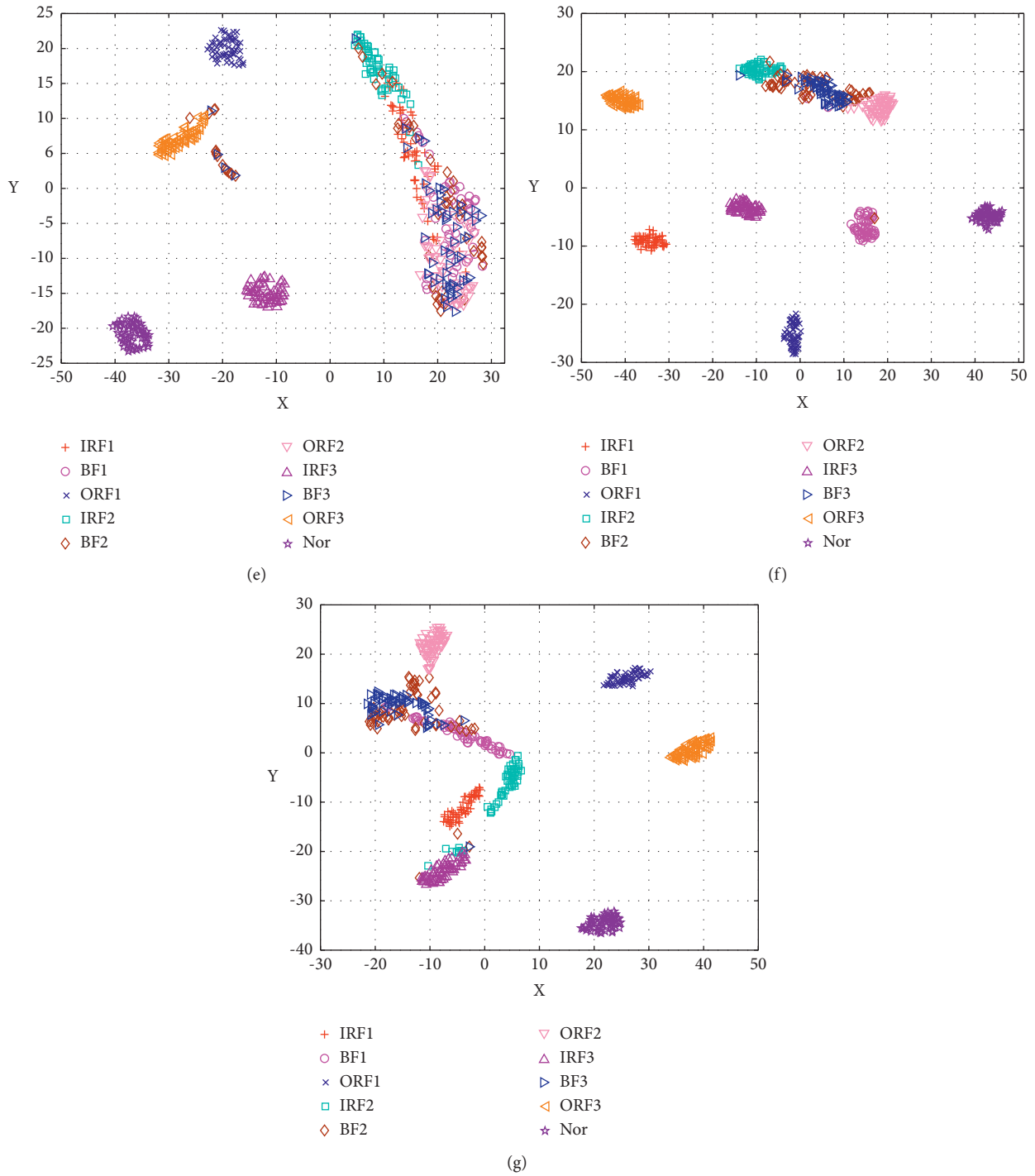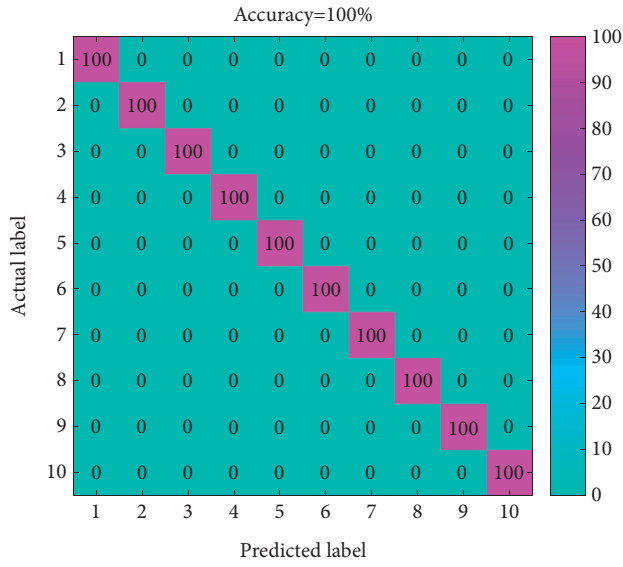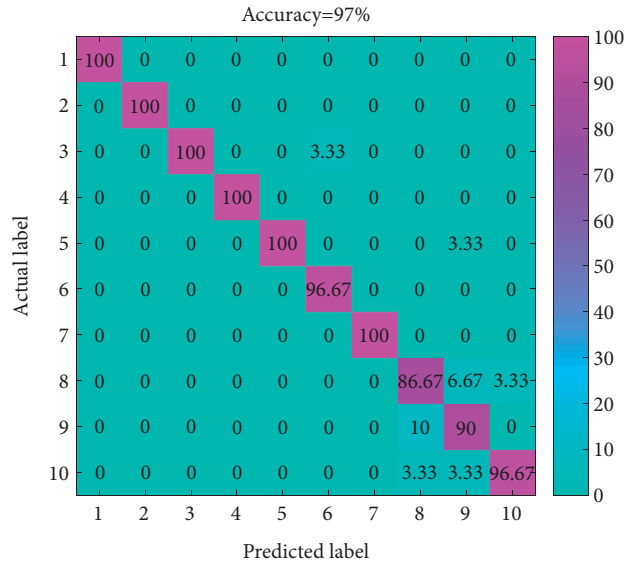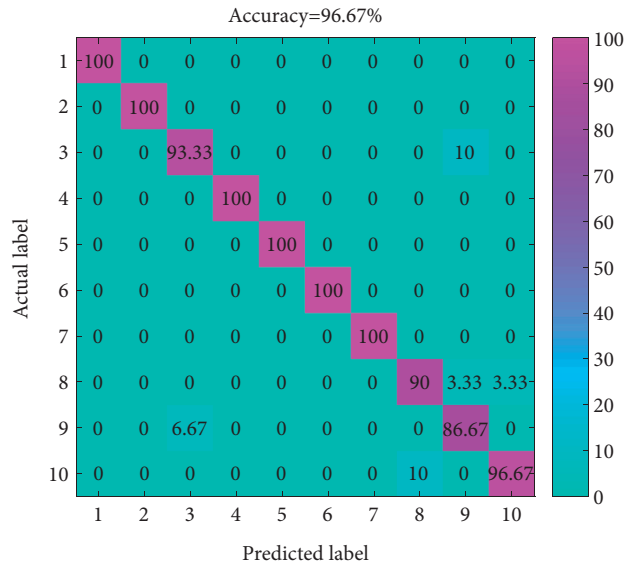
FIGURE 12: Continued.

(e)

(f)

(g)

FIGURE 12: The two-dimensional feature visualization map obtained by t-SNE.

of all categories are accurately classified. However, the classification accuracy of several other feature extraction models is lower than the proposed method. Corresponding to the previous analysis, the classification accuracy of Figure 13(e) is poor, and only a fault recognition rate of 90.33% has been achieved. Except for the Nor, IRF3 and ORF1 samples, the fault recognition rates of the other categories of samples are all lower than 100%. This is because

after RSSD decomposes the vibration signal, most of the vibration information is concentrated in the low resonance component, and the high resonance component contains less fault information, so the features extracted from the high resonance component have lower quality. In addition, by comparing Figures 13(a) and 13(g), it can be noticed that the multichannel data composed of vibration signals from the driver end and fan end does not achieve the best recognition
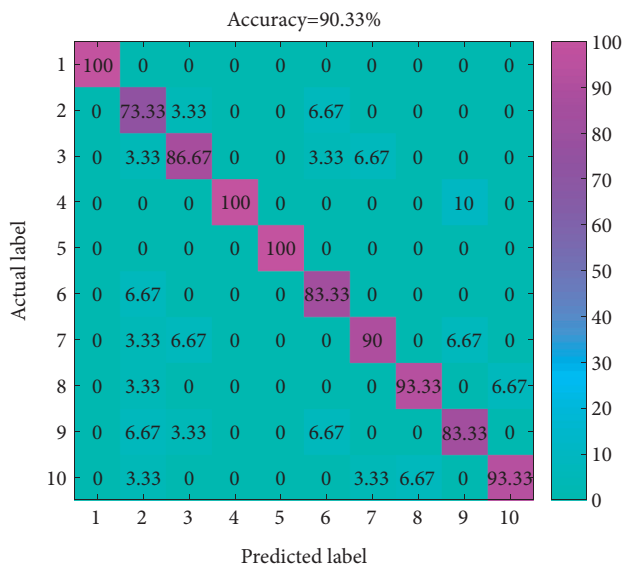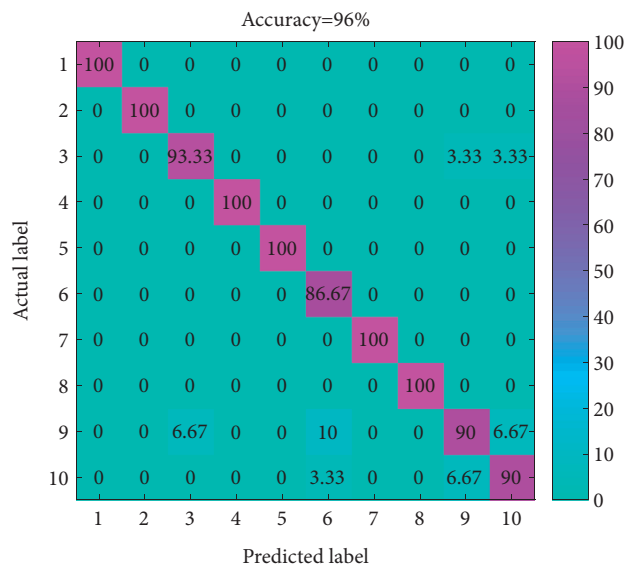
(a)

(b)
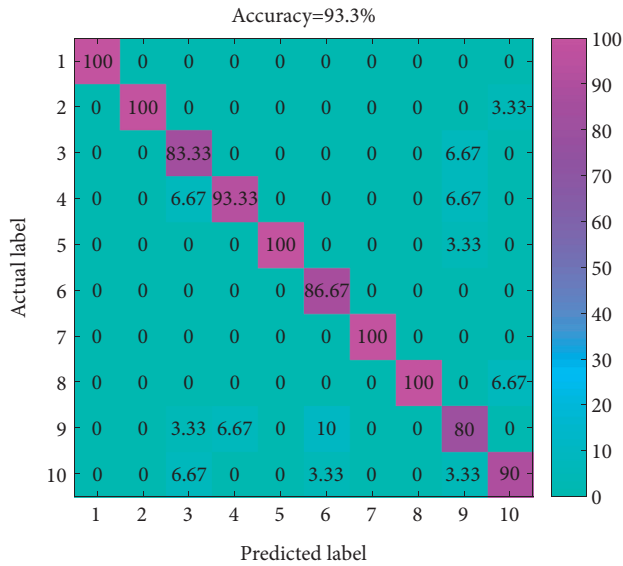
(c)

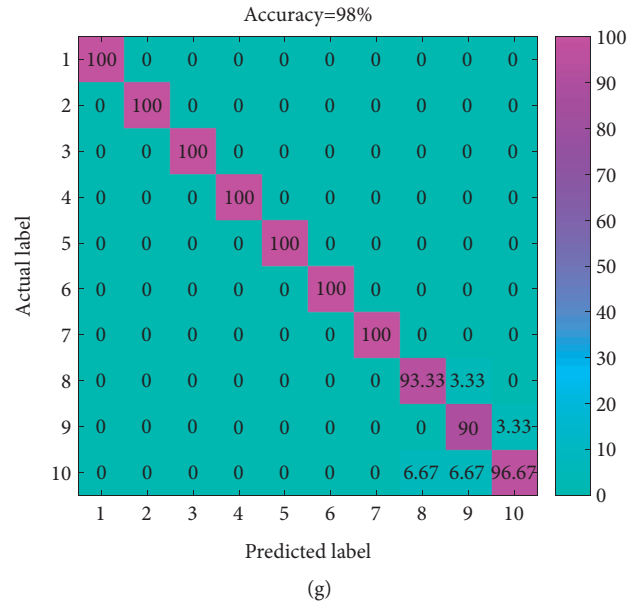(d)

(e)

(f)

FIGURE 13: Continued.

FIGURE 13: The confusion matrix of the seven feature extraction models.

effect, and its performance is weaker than the multichannel data composed of high resonance components and low resonance components. This proves that the RSSD decomposition can eliminate the interference in the signal. In summary, the proposed feature extraction model has excellent performance and can accurately identify various types of faults.

There may be errors in performing only a single classification experiment, and the performance of the proposed method cannot be reliably evaluated. Therefore, 20 trials was repeated to reduce the deviation caused by randomness and other factors. The results of seven feature extraction models in 20 trials are shown in Figure 14 and Table 2. Seen from Figure 14 and Table 2, the raised approach has the highest accuracy rate, and the average accuracy rate is 100%, that is, there are no misclassified samples in each classification. However, the classification accuracy of the other methods fluctuates, and the effect of each classification cannot be accurately estimated, that is, the stability is poor. Besides, comparing the feature extraction methods based on RCmvMSE and mvMAAPE, it can be found that the performance of the latter is better than the former. This shows that although the former adopts a fine composite coarse-graining process with excellent performance, mvAAPE has a stronger feature extraction performance than mvSE, so it can make up for the shortcomings of the traditional coarse-graining process. In addition, the diagnostic performance of each model is consistent with the previous visual analysis, that is, the performance of the model can be roughly judged by observing the distribution of each feature. In general, the proposed feature extraction model still has the best performance after many experiments, which proves its reliability.

This part mainly verifies the advantages of using RSSD to preprocess the signal. Typical signal decomposition techniques such as EEMD, LMD, and LCD are used to process vibration signals. Here, the first two components decomposed by each method are regarded as the components containing the main fault information to construct a multivariate signal. The subsequent processing steps are same with the presented approach. The diagnostic results of the four signal decomposition approaches under 20 trials are shown in Figure 15. Seen from Figure 15, the signal decomposition method based on HHO-RSSD achieves the best results, proving that the parameter-optimized RSSD has great application potential. The accuracy of the other three methods fluctuates, and the possibility of misclassification appears in each trial. The reason for this phenomenon is that the components decomposed by these three methods are of low quality, which affects the quality of extracted features. In short, as long as the parameters of the RSSD are reasonably selected, it can achieve very excellent results.

To explore the superiority of the DBN over other typical classifiers, the state features excavated by the proposed method are input into the typical recognizer. The selected classifiers are SVM, ELM, and Back Propagation Neural Network (BP). For convenience, the previous seven feature extraction models are marked as (a)–(g). The number of training samples and test samples remains the same. The results of these seven feature extraction models using different classifiers are shown in Table 3. Seen from Table 3, the DBN recognizer used is the best. The average recognition accuracy of DBN for the seven feature extraction models is 96.38%, which is higher than the other three classifiers, which proves its effectiveness and advantages. In addition, no matter what classifier is used, the recognition rate of the feature extraction model (a) is also the highest, with an average accuracy rate of 99.34%, which proves the advantages of this model over other models once again.
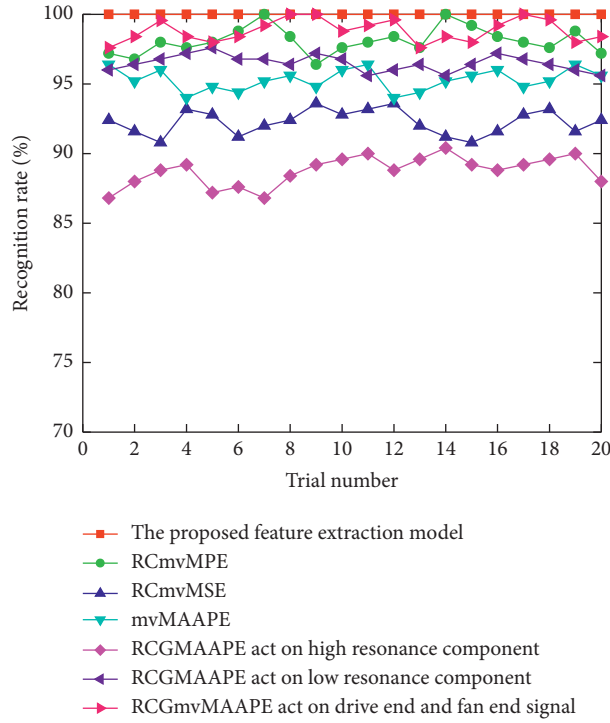
FIGURE 14: The diagnostic result of the seven feature extraction models in 20 trials.

TABLE 2: Recognition result of seven methods in 20 trials.

Methods	Accuracy obtained using diverse approaches (%)			SD
	Max	Min	Mean	
The presented approach	100	100	100	0
RCmvMPE	100	97	98.42	0.786
RCmvMSE	94.67	92.33	93.55	0.744
mvMAAPE	97	95	96.08	0.639
RCGMAAPE_High	92	89	90.63	0.885
RCGMAAPE_Low	98	96.33	97.08	0.483
RCGmvMAAPE_Drive_Fan	100	98	99.02	0.669

4.2. Case 2

4.2.1. Data Acquisition. The gearbox vibration data was collected on the QPZZ experimental platform [31]. The appearance and structure of the gearbox platform are displayed in Figure 16. The platform is made up of gearboxes, motors, bases and sensors. The sensor is arranged directly above the gear box. The rotating speed of the motor is set to 880 rpm. Five operating states were set up in the experiment: normal, gear pitting fault, gear broken tooth fault, gear wearing fault, gear pitting fault coupling with wearing fault. The brief information of the experimental data is displayed in Table 4. The sampling frequency of the sensor is set to 5.12 kHz, and the sampling time is 6 s. Due to the small amount of data, to ensure the accuracy of analysis, a sliding sampling method is adopted to select samples. The signal of the bearing Y on the motor side of the input shaft is used for analysis. In the subsequent multichannel analysis, the
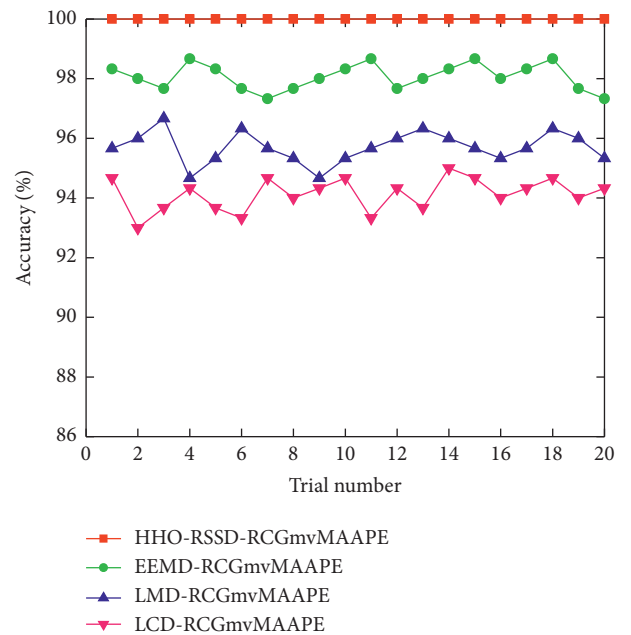

FIGURE 15: The diagnostic result of four signal decomposition methods.

vibration signals of the input shaft motor side bearing Y and the output shaft load side bearing Y are selected for analysis. The collected vibration signals are divided into 52 groups of samples after sliding sampling. Each group contains 2048 data points, of which 22 groups are adopted as the training data, and the remaining samples are adopted as the test data.

TABLE 3: The diagnostic results obtained by combining seven models with four classifiers.

Classifier model	The accuracy of test sample with different feature extraction model (%)							Average accuracy
	(a)	(b)	(c)	(b)	(e)	(f)	(g)	
DBN	100	98.67	93.33	96	91.33	96.67	98.67	96.38
SVM	100	97.67	94.67	93.33	90	95.33	97.33	95.48
ELM	99.67	95.33	91	95.33	88.67	94.67	97	94.52
BP	97.67	92	89.33	92.67	90.67	93.33	96	93.10
Average accuracy	99.34	95.92	92.08	94.33	90.18	95	97.25	—
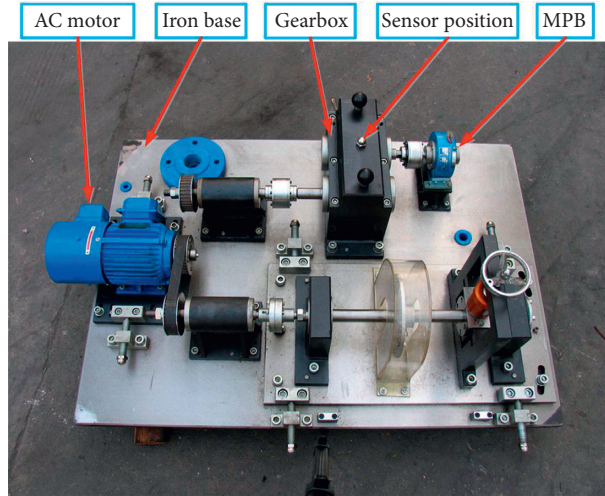
FIGURE 16: The gearbox test platform.

TABLE 4: The brief information of gear vibration data.

Fault state	Abbreviation	Training sample	Testing sample	Label
Normal	Nor	22	30	1
Gear wearing fault	WF	22	30	2
Gear pitting fault	PF	22	30	3
Gear tooth breaking fault	TBF	22	30	4
Gear pitting & wearing fault	PWF	22	30	5

4.2.2. Feature Excavation. Figure 17 shows the waveforms of vibration data in five states of the gearbox. Similarly, due to the lack of obvious rules and characteristics of the waveform, it is hard to easily judge the fault status. Therefore, it is necessary to carry out subsequent processing on the data to acquire more and more distinguishable features.

Similarly, this part first studies how to obtain the best parameters of RSSD. First, input the gear vibration signal into HHO-RSSD for decomposition, and execute the parameter optimization process. Taking the Nor signal as an example, the optimization process ends when the correlation kurtosis value is the smallest. Then, after parameter optimization, a set of optimal parameters was obtained, $Q_1 = 9.336$, $Q_2 = 1.319$, $A_1 = 1.23$, $A_2 = 0.11$, and $u = 0.377$, respectively. Figure 18 is the evolution curve of fitness value in the optimization procedure of HHO. Seen from Figure 18, HHO could quickly reach a local optimal value, and this value is finally determined to be the global optimal value. Therefore, this shows that HHO can optimize the target from the global scope, so as to find an optimal

solution. Then, the optimal parameter combination obtained by optimization is input into the RSSD, and the vibration signal is decomposed to acquire the high and low resonance components. The RSSD decomposition result of Nor is shown in Figure 19.

The high and low resonance components are taken as a multichannel time series. Then, RCGmvMAAPE is adopted to excavate the fault features of the constructed multivariable data to construct the fault samples. Moreover, to validate the superiority of the raised RCGmvMAAPE approach, it is compared with RCmvMPE, RCmvMSE, mvMAAPE, and RCGMAAPE. The entropy results of seven methods are displayed in Figures 20(a)~20(g). Here, the method used in each figure is consistent with the previous experiment. Seen from Figure 20, compared with several other feature extraction models, the standard deviation of the entropy value of Figure 20(a) is smaller and the performance is more stable. The distinguishability of several other features is also very strong, but the entropy deviation is generally large, and the error bar has obvious aliasing. This phenomenon proves that
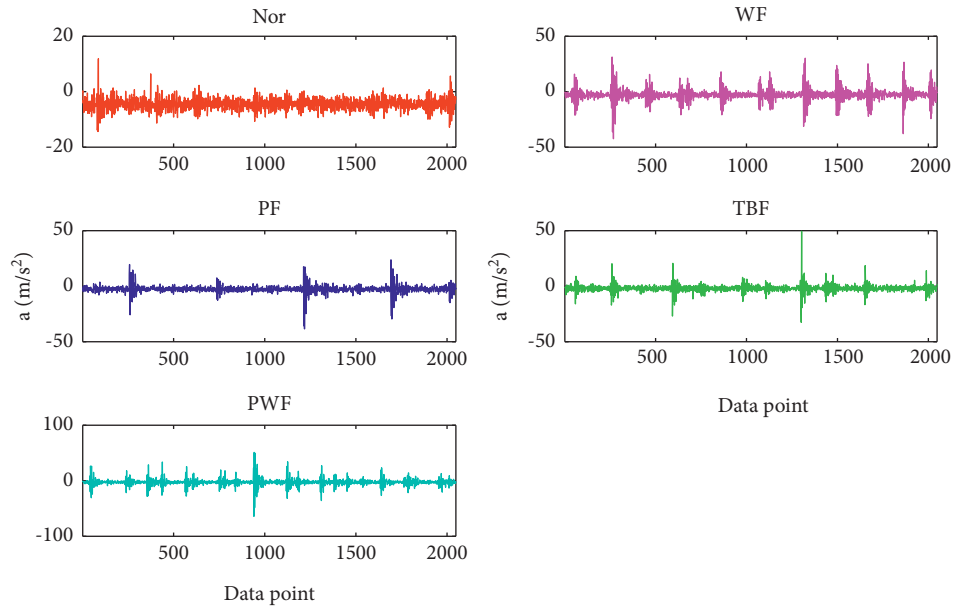
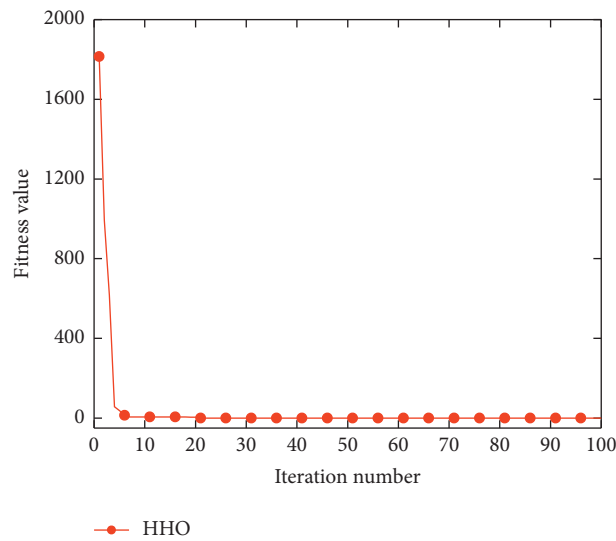Figure 17: The vibration waveform of gearbox under five fault conditions.

Figure 18: The evolution curve of fitness value.

although the features have obvious discrimination, the performance fluctuates greatly, which is not conducive to subsequent classification. The proposed model can distinguish each fault state better, and has a small entropy deviation on most scales, so it has not only strong separability, but also has stable performance.

Similarly, t-SNE is used for auxiliary analysis to intuitively compare the performance of the above seven feature extraction models. The visualization of the features extracted by the seven methods is displayed in Figure 21. Observed from Figure 21, the WF sample and the TBF sample in Figure 21(a) are partially aliased, and the distribution of these two categories is relatively scattered, that is, the samples of these two categories have poor separability. The distinguishability of features extracted based on other

feature extraction models is worse, and some samples do not even have cluster centers. Comparing Figures 21(a) and 21(g), Figure 12(g) has a better visualization effect. The clusters of the five categories are relatively scattered, but the distribution of samples of the same category is relatively scattered, and there is no obvious cluster center. By visualizing the features, the quality of the features extracted by each model can be roughly judged, and then the performance of the model can be judged. Therefore, it can be verified that the features of Figure 21(a) have better quality, which proves the superiority of the presented approach.

4.2.3. Fault Recognition. For the sake of quantifying the performance of the above seven fault feature extraction
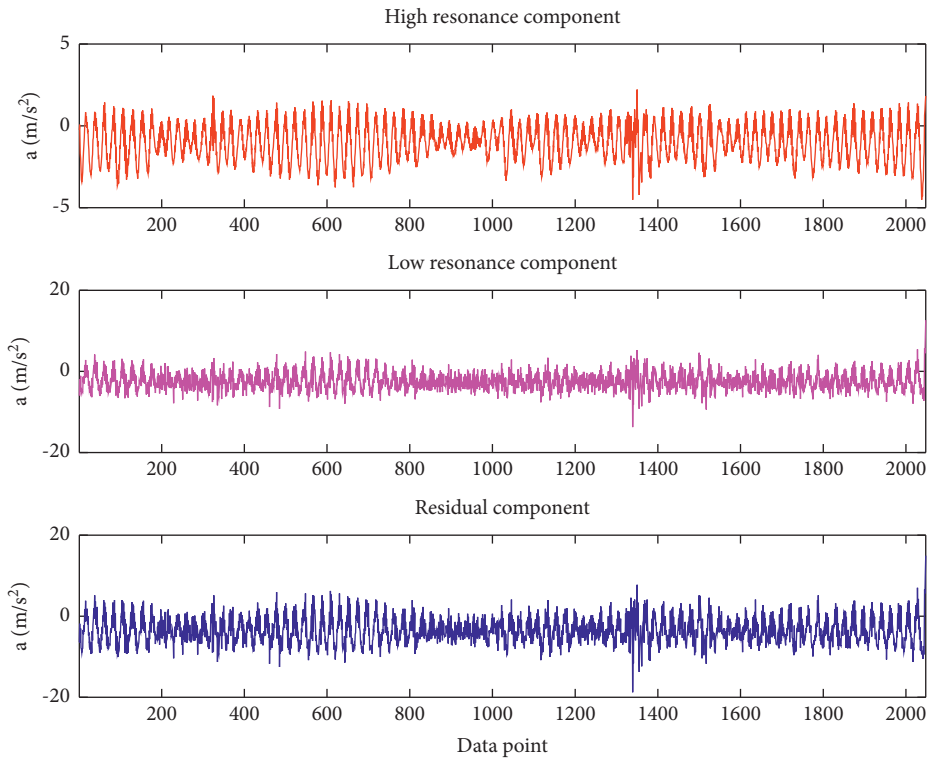
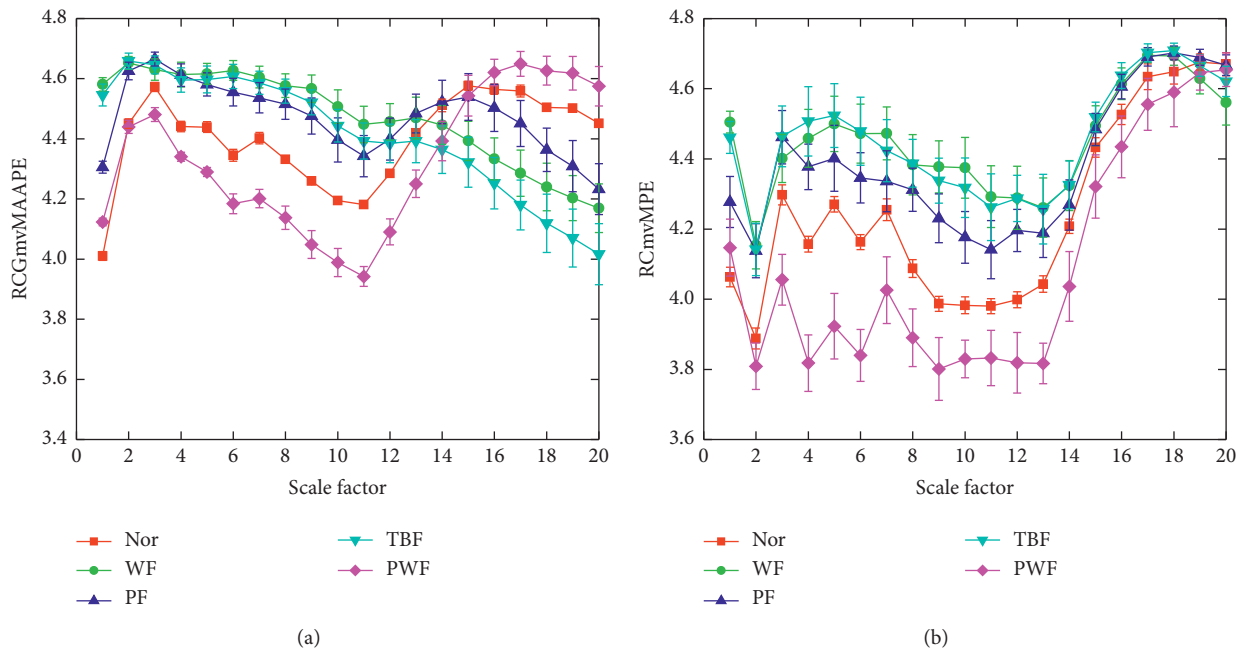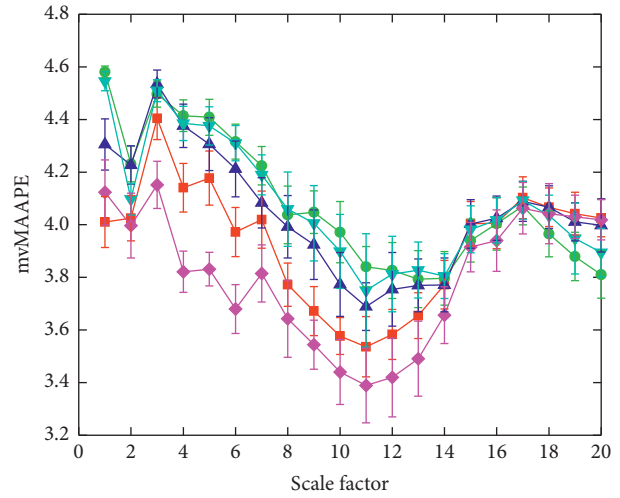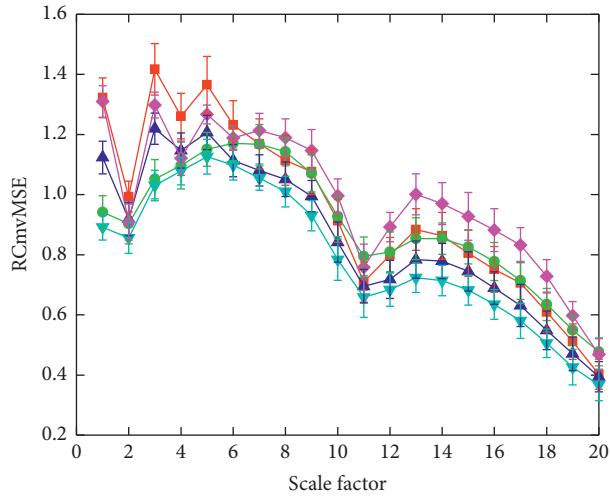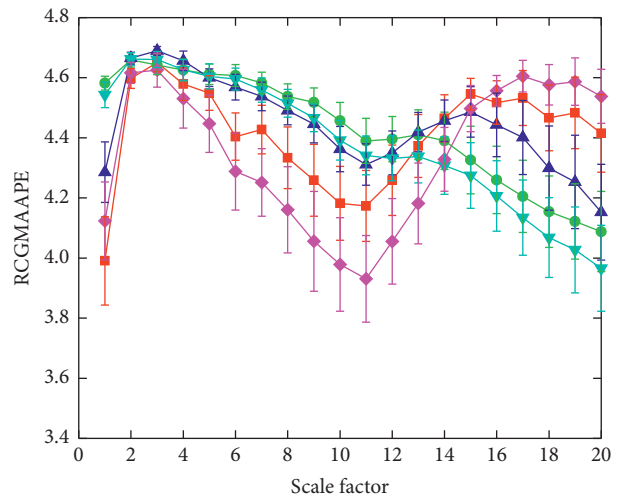FIGURE 19: The RSSD decomposition result of vibration signal nor.
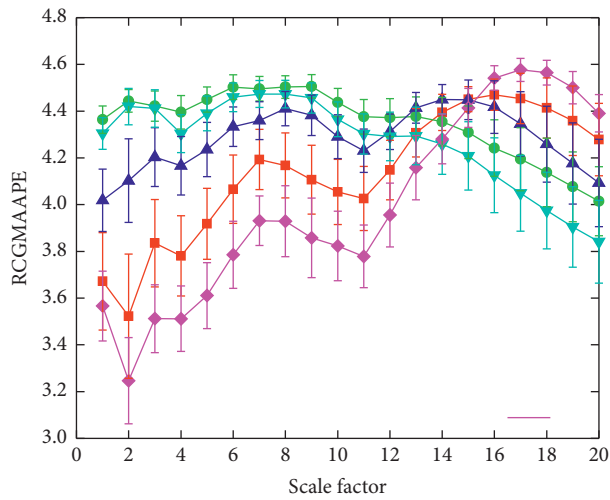

(a)

(b)

FIGURE 20: Continued.

(c)

(d)
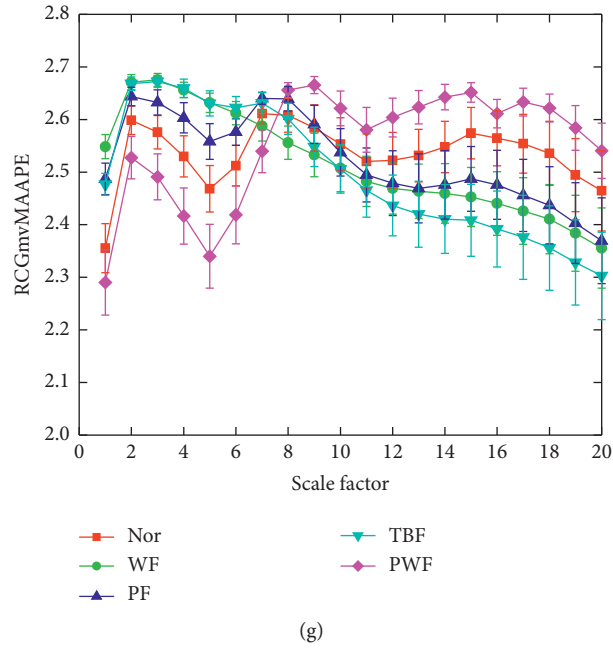

(e)

(f)

FIGURE 20: Continued.

(g)

FIGURE 20: The entropy results of gearbox data analyzed by adopting seven approaches.
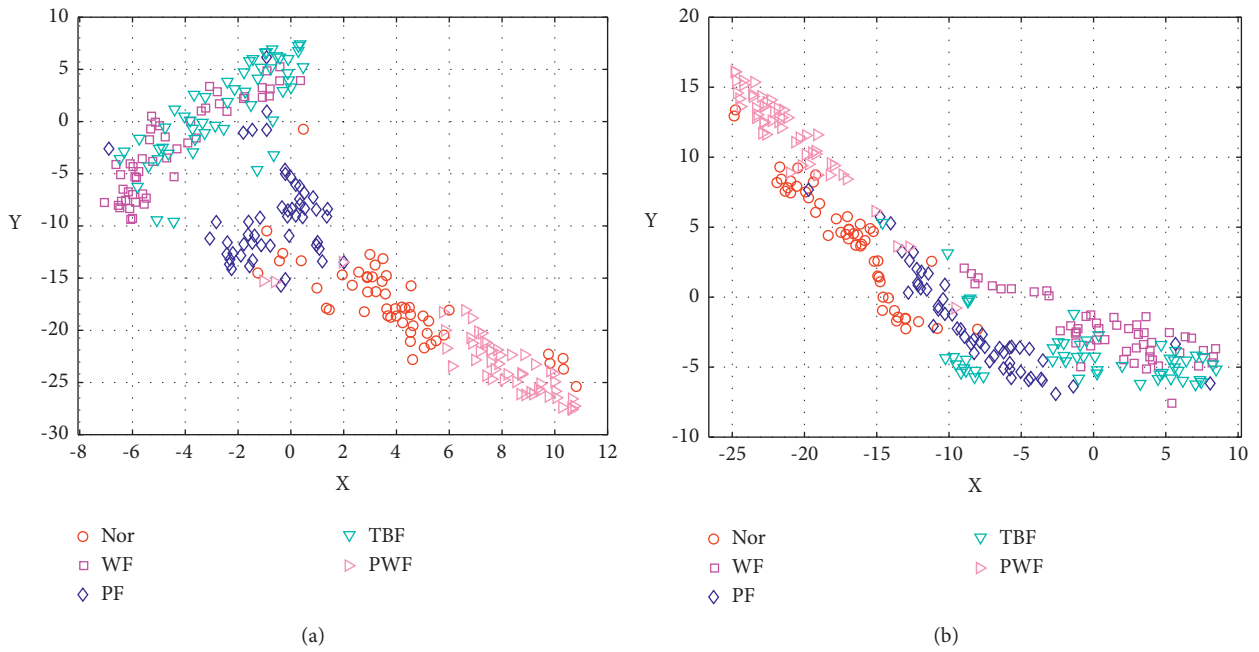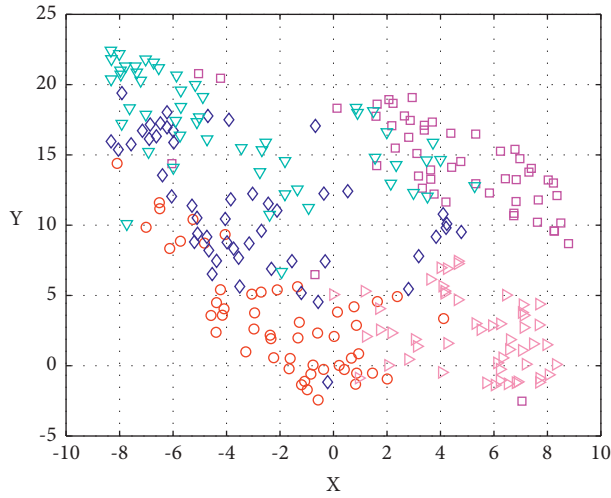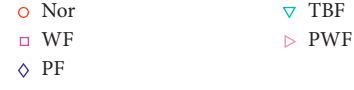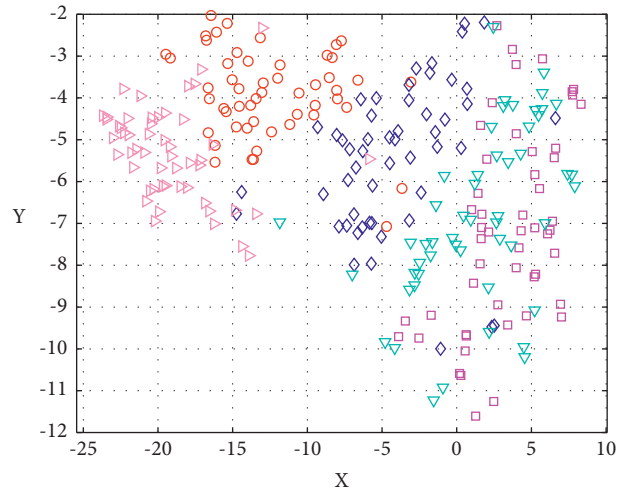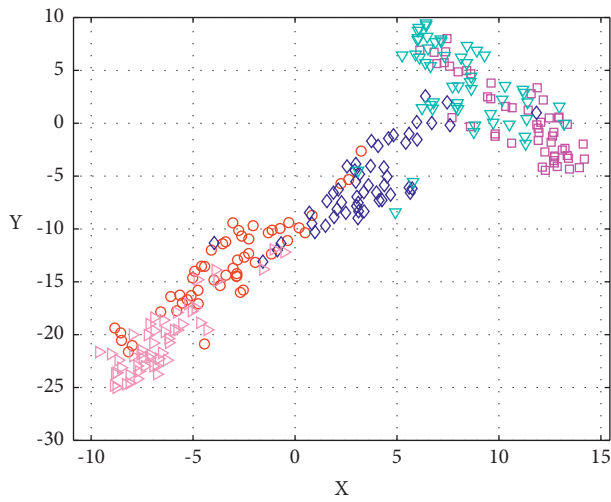

(a)

(b)
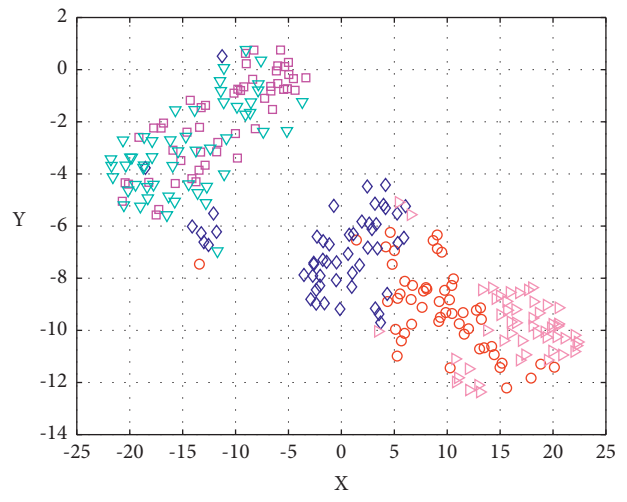
FIGURE 21: Continued.

(c)
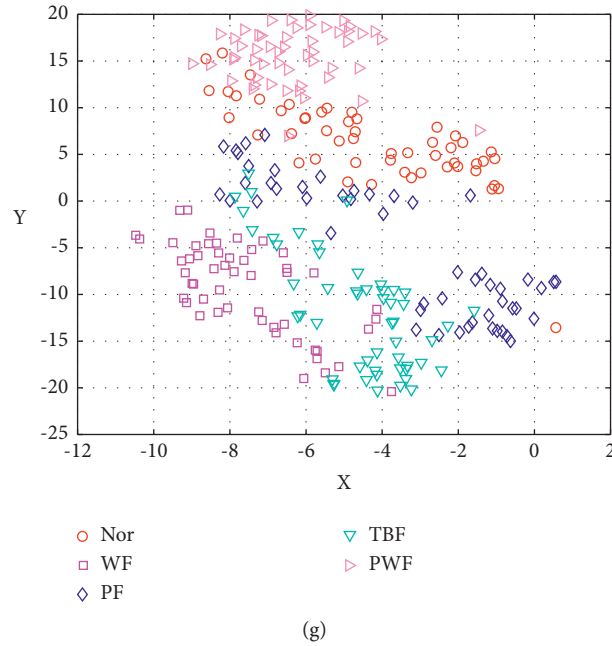
(d)

(e)

(f)

Figure 21: Continued.

(g)

FIGURE 21: The two-dimensional feature visualization map obtained by t-SNE.

models on rolling bearing fault diagnosis, the features excavated by the seven approaches are input into the DBN recognizer for fault classification. The confusion matrix of the seven feature extraction models is presented in Figure 22. Observed from Figure 22, some WF samples and TBF samples were misclassified. One WF sample was misclassified to TBF, and two TBF samples were misclassified to WF. The accuracy of a single classification is 98%, which is still reliable. The performance of several other feature extraction models is weaker than the proposed model, which is also consistent with the previous t-SNE analysis. In addition, the fault recognition rate of Figure 22(g) is better than Figure 21(a), reaching 98.67%. This is mainly because the gear vibration data usually includes multiple channels, and its operating information is distributed in multiple directions. Key feature will inevitably be missed when the signal of a single channel is used for analysis. Although Figure 22(g) only analyzes the original multichannel vibration signal without corresponding processing, the rich vibration information contained in the multichannel signal can provide enough features for judging the fault state. The proposed feature extraction model only uses a single-channel vibration signal, but can achieve a fault recognition rate of 98%, which is satisfactory to a certain extent. Comparing Figure 22(e) and Figure 22(f), the analysis of low resonance components has achieved better results, which is also consistent with the results of the previous experiment. This shows that the main fault information after RSSD decomposition is concentrated on the low resonance component, while the high resonance component contains less fault information.

By comparing Figures 22(a) and 22(f), it can be found that only a single experiment may not be able to estimate the effectiveness of the approach reliably, that is, a single

experiment has strong randomness. Therefore, 20 trials was repeated to reduce the deviation caused by randomness and other factors. The results of seven feature extraction models in 20 trials are shown in Figure 23 and Table 5. Seen from Figure 23 and Table 5, the proposed model achieves the best classification results, with an average accuracy rate of 98.10%, of which the highest is 100% and the lowest is 96%. Although the recognition rate has obvious fluctuations, it is generally reliable. The other six methods fluctuate sharply, especially the fifth feature extraction model, with a standard deviation as high as 2.191. Such a high deviation proves that the performance of the method is quite unstable, and the classification result is not very reliable. In addition, it can be found that the performance of the method proposed in this paper is better than that of the seventh model, which is consistent with the previous analysis, that is, a single experiment is not convincing. As the number of classifications increases, the proposed model has higher stability and performance. Therefore, it can be guaranteed that it is reliable in most classifications, while the performance stability of the seventh model is weaker than the proposed model, so the overall performance of the proposed method is excellent.

This part mainly verifies the advantages of using RSSD to preprocess the signal. Typical signal decomposition techniques such as EEMD, LMD, and LCD are used to process vibration signals. Here, the first two components decomposed by each method are regarded as the components containing the main fault information to construct a multivariate signal. The subsequent processing steps are same with the presented approach. The diagnostic results of the four signal decomposition approaches under 20 trials are shown in Figure 24. Seen from Figure 24, the signal decomposition method based on HHO-RSSD achieves the best results, proving that the parameter-optimized RSSD has
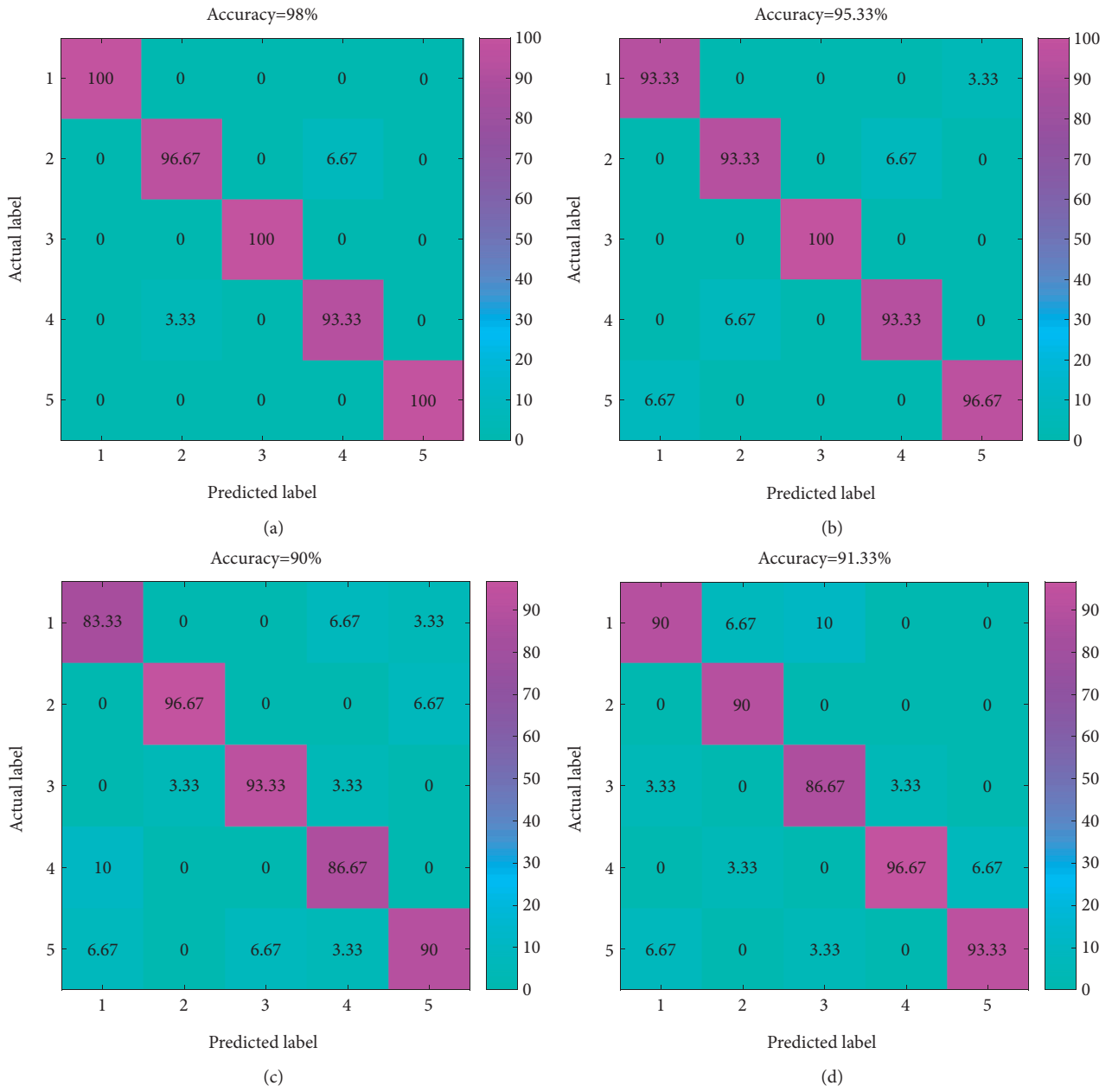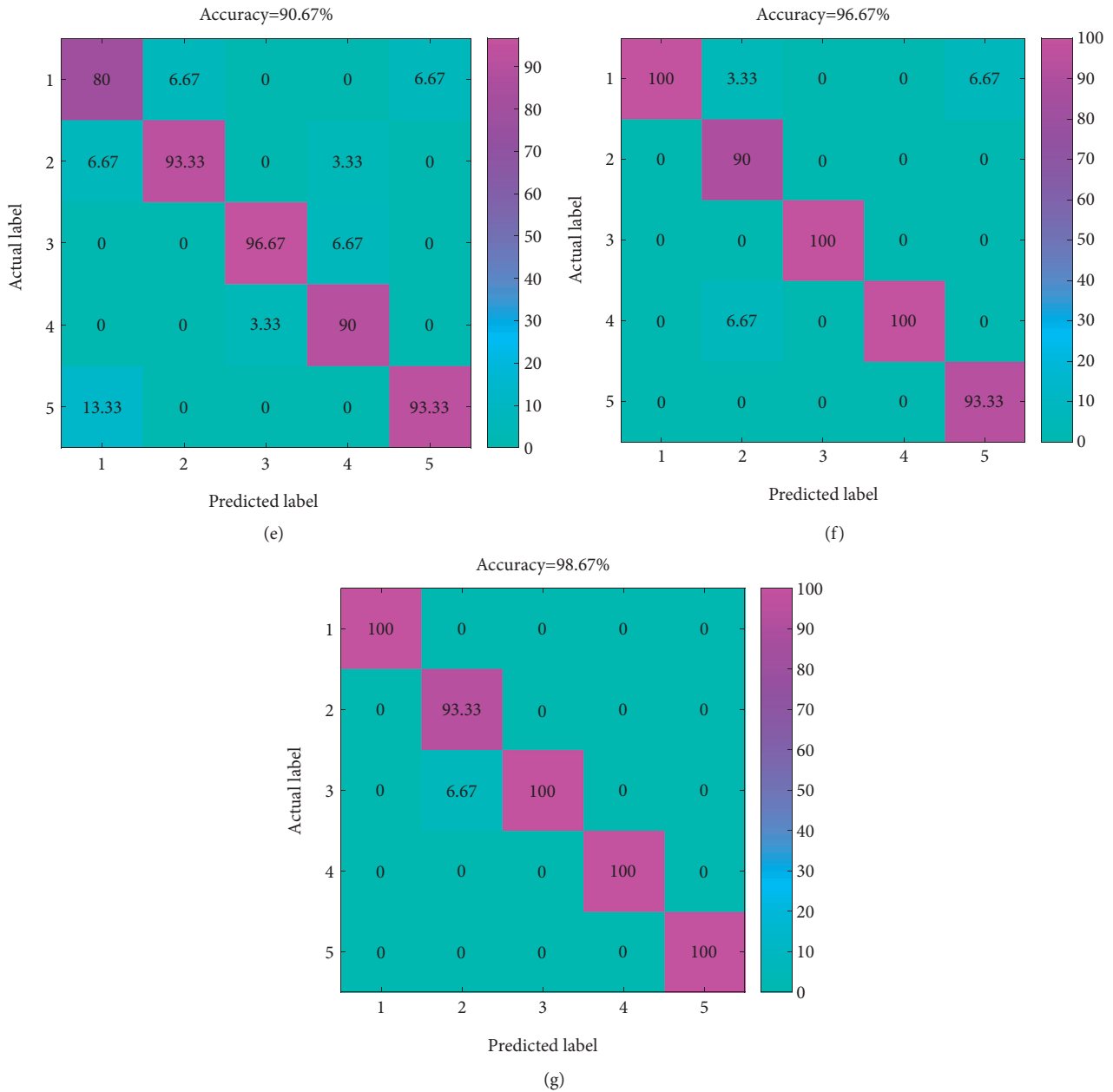
Figure 22: Continued.

(e)

(f)

(g)

FIGURE 22: The confusion matrix of the seven feature extraction models.

great application potential. The accuracy of the other three methods fluctuates, and the possibility of misclassification appears in each trial. This proves that HHO-RSSD has excellent signal analysis performance. By decomposing the signal, it can reduce the influence of interference components in the signal on feature extraction. Therefore, it is necessary and effective to use HH0-RSSD to process the signal.

Similarly, this part is used to study the advantages of choosing DBN as a classifier, so three typical classifiers are also selected for comparison. Here, the ratio of the test and training samples remains the same. Similarly, for convenience, the previous seven feature extraction models are marked as (a)–(g). The results of these seven feature extraction models using different classifiers are shown in Table 6. Seen from Table 6, the DBN recognizers obtain the highest recognition rate. The average recognition accuracy of DBN for the seven feature extraction models is 93.71%, which is higher than the other three classifiers, which proves its effectiveness and advantages. In addition, no matter what classifier is used, the recognition rate of the feature extraction model (a) is also the highest, with an average accuracy rate of 99%, which proves the advantages of this model over other models once again.
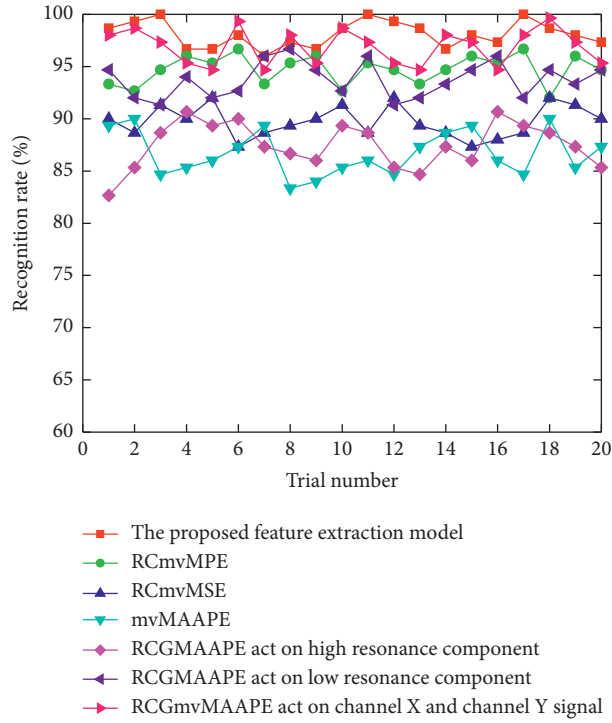
FIGURE 23: The diagnostic result of the seven feature extraction models in 20 trials.

TABLE 5: Recognition result of seven methods in 20 trials.

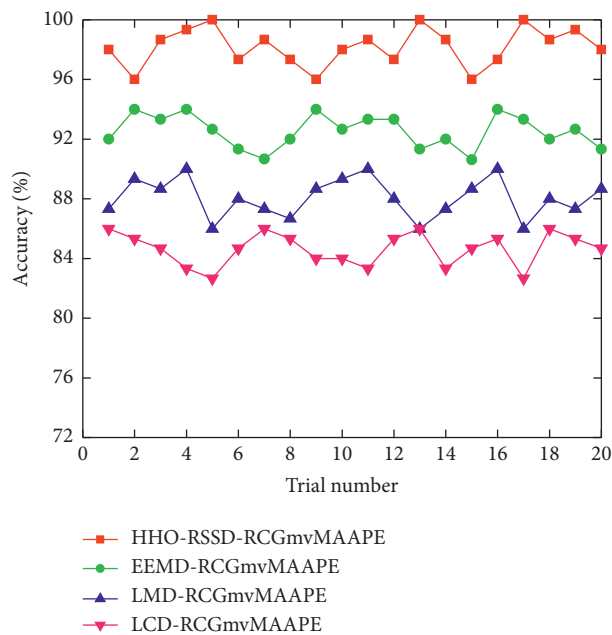| Methods	Accuracy obtained using diverse approaches (%)			SD
	Max	Min	Mean	
The presented approach	100	96	98.10	1.247
RCmvMPE	96.67	92	94.73	1.400
RCmvMSE	92	87.33	89.73	1.520
mvMAAPE	90	83.33	86.70	2.124
RCGMAAPE_High	90.67	82.67	87.47	2.191
RCGMAAPE_Low	96.67	91.33	93.73	1.696
RCGmvMAAPE_Drive_Fan	99.63	94.67	96.88	1.701

FIGURE 24: The diagnostic result of four signal decomposition methods.

TABLE 6: The recognition results of seven models with four recognizers (%).

Classifier model	The recognition rate of the test sample with different feature extraction model							Average accuracy
	(a)	(b)	(c)	(b)	(e)	(f)	(g)	
DBN	99.33	94.67	92	89.33	88.67	94	98	93.71
SVM	100	93.33	92.67	86.67	87.33	92	96	92.57
ELM	98	95.33	90.67	88	84.67	94	96.67	92.48
BP	98.67	92.67	88.67	86	88.67	91.33	94	91.43
Average accuracy	99	94	91	87.5	87.34	92.83	96.17	—

5. Conclusion

At present, typical health detection approach on the basis of signal processing and entropy are usually as follows: (1) Multiscale entropy of a single component; (2) Single entropy of multiple components; (3) Multiscale entropy of multiple components; These three types have some defects that can be improved. For example, the fault features extracted by the first two approaches are not comprehensive and sufficient, which may cause information omission. Although the third method can extract very comprehensive features, it may cause the dimensionality to be too large, and usually requires dimensionality reduction. Thus, a novel multiscale feature extraction method is proposed. First, the RSSD algorithm optimized by HHO is adopted to decompose the single-channel signal into high and low resonance components. Then use these two components as multichannel data and perform RCGmvMAAPE analysis to extract fault features. Eventually, the features are input to the DBN classifier for identification. Based on two rotating machinery vibration data sets, six different feature extraction models are utilized to compare with the presented approach. Experimental results show that the raised model can obtain a higher fault recognition rate and a higher utilization rate of information when only using a single channel vibration signal. Subsequently, to prove the superiority of the RSSD, three classic signal decomposition algorithms were used for comparative analysis, and the results proved that HHO-RSSD has satisfactory performance.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[2] K. Li, L. Su, J. Wu, H. Wang, and P. Chen, "A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine," *Applied Sciences*, vol. 7, no. 10, 2017.

[3] Z. Feng and M. J. Zuo, "Vibration signal models for fault diagnosis of planetary gearboxes," *Journal of Sound and Vibration*, vol. 331, no. 22, pp. 4919–4939, 2012.

[4] L. You, W. Fan, Z. Li, Y. Liang, M. Fang, and J. Wang, "A fault diagnosis model for rotating machinery using VWC and MSFLA-SVM based on vibration signal analysis," *Shock and Vibration*, vol. 2019, 2019.

[5] X. Yang, J. Yan, H. Zhu, Y. Cao, and L. Shao, "Research on fault diagnosis of hydraulic pump using convolutional neural network," *Journal of Vibroengineering*, vol. 18, no. 8, pp. 5141–5152, 2016.

[6] W. Zhang and J. Zhou, "A comprehensive fault diagnosis method for rolling bearings based on refined composite multiscale dispersion entropy and fast ensemble empirical mode decomposition," *Entropy*, vol. 21, no. 7, p. 680, 2019.

[7] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, pp. 108–126, 2013.

[8] W. Y. Liu, Q. W. Gao, G. Ye, R. Ma, X. N. Lu, and J. G. Han, "A novel wind turbine bearing fault diagnosis method based on Integral Extension LMD," *Measurement*, vol. 74, pp. 70–77, 2015.

[9] I. W. Selesnick, "Resonance-based signal decomposition: A new sparsity-enabled signal analysis method," *Signal Processing*, vol. 91, no. 12, pp. 2793–2809, 2011.

[10] B. Chen, B. Shen, F. Chen et al., "Fault diagnosis method based on integration of RSSD and wavelet transform to rolling bearing," *Measurement*, vol. 131, pp. 400–411, 2019.

[11] P. Ma, H. Zhang, W. Fan, and C. Wang, "Early fault diagnosis of bearing based on frequency band extraction and improved tunable Q-factor wavelet transform," *Measurement*, vol. 137, pp. 189–202, 2019.

[12] W. Huang, H. Sun, and W. Wang, "Resonance-based sparse signal decomposition and its application in mechanical fault diagnosis: A review," *Sensors*, vol. 17, no. 6, p. 1279, 2017.

[13] N. Chai, M. Yang, Q. Ni, and D. Xu, "Gear fault diagnosis based on dual parameter optimized resonance-based sparse signal decomposition of motor current," *IEEE Transactions on Industry Applications*, vol. 54, no. 4, pp. 3782–3792, 2018.

[14] Y. Lu, J. Du, and X. Tao, "Fault diagnosis of rolling bearing based on resonance-based sparse signal decomposition with optimal Q-factor," *Measurement and Control*, vol. 52, pp. 1111–1121, 2019.

[15] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[16] Y. Li, X. Wang, Z. Liu, X. Liang, and S. Si, "The entropy algorithm and its variants in the fault diagnosis of rotating machinery: review," *Ieee Access*, vol. 6, pp. 66723–66741, 2018.

[17] Z. Huo, M. Martinez-Garcia, Y. Zhang, R. Yan, and L. Shu, "Entropy measures in machine fault diagnosis: Insights and applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2607–2620, 2020.

[18] S.-D. Wu, P.-H. Wu, C.-W. Wu, J.-J. Ding, and C.-C. Wang, "Bearing fault diagnosis based on multiscale permutation

entropy and support vector machine," *Entropy*, vol. 14, pp. 1343–1356, 2012.

[19] B. Fadlallah, B Chen, A Keil, and J Príncipe, "Weighted-permutation entropy: complexity measure for time series incorporating amplitude information," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 87, no. 2, p. 022911, 2013.

[20] Y. Chen, Zhang, Zhao, Luo, and Sun, "Fault diagnosis of rolling bearing using multiscale Amplitude-aware permutation entropy and random forest," *Algorithms*, vol. 12, no. 9, p. 184, 2019.

[21] J. Zheng, H. Pan, J. Tong, and Q. Liu, "Generalized refined composite multiscale fuzzy entropy and multicluster feature selection based intelligent fault diagnosis of rolling bearing," *ISA Transactions*, 2021.

[22] C. Li, J. Zheng, H. Pan, J. Tong, and Y. Zhang, "Refined composite multivariate multiscale dispersion entropy and its application to fault diagnosis of rolling bearing," *IEEE Access*, vol. 7, pp. 47663–47673, 2019.

[23] A. Subasi, "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders," *Computers in Biology and Medicine*, vol. 43, no. 5, pp. 576–586, 2013.

[24] A. Iosifidis, A. Tefas, and I. Pitas, "On the kernel extreme learning machine classifier," *Pattern Recognition Letters*, vol. 54, pp. 11–17, 2015.

[25] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[26] G. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[27] M. Sedláček, P. Bojan, and J. Vižintin, "Correlation between standard roughness parameters skewness and kurtosis and tribological behaviour of contact surfaces," *Tribology International*, vol. 48, pp. 102–112, 2012.

[28] H. Azami and J. Escudero, "Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, 2016.

[29] F. Zhou, "An integrated health condition detection method for rotating machinery using refined composite multivariate multiscale Amplitude-aware permutation entropy," *Mathematical Problems in Engineering*, vol. 2020, 2020.

[30] "Case western Reserve university bearing data center website," 2018, http://csegroups.case.edu/bearingdatacenter/home.

[31] Y. Chen, Zhang, Zhao, Luo, and Lin, "Rotating machinery fault diagnosis based on improved multiscale amplitude-aware permutation entropy and multiclass relevance vector machine," *Sensors*, vol. 19, no. 20, p. 4542, 2019.