

## Research Article

# Fault Diagnosis Method for Rotating Machinery Based on Hierarchical Amplitude-Aware Permutation Entropy and Pairwise Feature Proximity

Ling Shu <sup>1</sup>, Jinxing Shen <sup>2</sup>, and Xiaoming Liu <sup>3</sup>

<sup>1</sup>School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China

<sup>2</sup>School of Field Engineering, Army Engineering University of PLA, Nanjing 210007, China

<sup>3</sup>School of Automation, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Jinxing Shen; 565423803@qq.com

Received 30 July 2021; Accepted 19 November 2021; Published 2 December 2021

Academic Editor: Xavier Chimentin

Copyright © 2021 Ling Shu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With a view to solving the defect that multiscale amplitude-aware permutation entropy (MAAPE) can only quantify the low-frequency features of time series and ignore the high-frequency features which are equally important, a novel nonlinear time series feature extraction method, hierarchical amplitude-aware permutation entropy (HAAPE), is proposed. By constructing high and low-frequency operators, this method can extract the features of different frequency bands of time series simultaneously, so as to avoid the issue of information loss. In view of its advantages, HAAPE is introduced into the field of fault diagnosis to extract fault features from vibration signals of rotating machinery. Combined with the pairwise feature proximity (PWFP) feature selection method and gray wolf algorithm optimization support vector machine (GWO-SVM), a new intelligent fault diagnosis method for rotating machinery is proposed. In our method, firstly, HAAPE is adopted to extract the original high and low-frequency fault features of rotating machinery. After that, PWFP is used to sort the original features, and the important features are filtered to obtain low-dimensional sensitive feature vectors. Finally, the sensitive feature vectors are input into GWO-SVM for training and testing, so as to realize the fault identification of rotating machinery. The performance of the proposed method is verified using two data sets of bearing and gearbox. The results show that the proposed method enjoys obvious advantages over the existing methods, and the identification accuracy reaches 100%.

## 1. Introduction

With the rapid development of modern manufacturing, rotating machinery is widely applied in various large-scale precision equipment such as wind turbines, aeroengines, and driverless cars and plays an important role in bearing loads and transmitting power [1]. The most representative rotating machinery is bearings and gears. Generally speaking, the working environment of such components is relatively harsh, and they often work in the state of high speed and high load, which are easy to produce faults. Once the fault occurs, it will cause equipment downtime at least and even huge economic losses and casualties [2]. Therefore, the research of efficient and accurate new fault diagnosis method of rotating machinery has important

practical significance for ensuring the safe and stable operation of equipment and reducing accidents.

Rotating machinery fault diagnosis based on vibration signal analysis is the current research hotspot, while feature extraction is the most critical step in the fault diagnosis process [3]. The fault vibration signals of rotating machinery are typical nonlinear and nonstationary signals, and the traditional linear processing methods cannot effectively extract the fault features representing the working state of rotating machinery. Therefore, many nonlinear time-frequency analysis methods have been proposed and applied to fault diagnosis domain, such as ensemble empirical mode decomposition (EEMD), local mean decomposition (LMD), and wavelet packet transform (WPT). Feng et al. [4] employed EEMD to process the bearing vibration signal to

get the corresponding intrinsic mode function (IMF) and then screened the sensitive IMF for reconstruction, so as to identify different fault types. Chen et al. [5] used wavelet packet decomposition to process the original signal and then calculated the energy entropy value of each subsignal as the fault feature for fault identification of bearings.

In addition to the above time-frequency analysis method, in recent years, nonlinear dynamic parameters based on entropy theory have developed rapidly and are widely used in feature extraction of nonlinear time series. Commonly used entropy methods include sample entropy (SE) [6], fuzzy entropy (FE) [7], and permutation entropy (PE) [8]. As effective feature extraction tools, these methods have been widely used in the domain of medical signal analysis, image processing, and vibration signal analysis. However, these methods also have certain shortcomings: SE has low computational efficiency and is highly dependent on the length of time series; the computational process of FE is similar to SE, so its computational efficiency is also low; Compared with FE and SE, PE enjoys simpler principles and higher computational efficiency. Nevertheless, PE only considers the order structure of the time series, but ignores the amplitude information of each element in the time series, which leads to the poor robustness of PE against noise. In addition, PE does not well solve the problem of inaccurate evaluation of the permutation pattern caused by the presence of elements with the same amplitude in the time series. In order to overcome the abovementioned shortcomings of PE, Azami et al. [9] proposed amplitude-aware permutation entropy (AAPE). This method fully considers the amplitude information of time series and the amplitude difference between different elements and is more sensitive to the amplitude and frequency of the signal. Moreover, by improving the measurement method of the same pattern in PE, AAPE can better deal with the problem of equal element amplitude. In the literature [9], the simulation signal and biomedical signal experiments proved that compared with SE, FE, and PE, AAPE enjoys better dynamic feature evaluation ability.

Unfortunately, AAPE can only evaluate the complexity of time series on a single scale, while the important information hidden in other scales is discarded; meanwhile, single-scale analysis cannot consider the long correlation of time series, resulting in inaccurate complexity assessment. With a viewing to solving this defect, based on the multiscale entropy theory [10], multiscale amplitude-aware permutation entropy (MAAPE) [11] was proposed, which extended the dynamic feature quantification based on AAPE to multiple time scales. However, multiscale expansion adopts the traditional averaging operation to obtain multiple coarse-grained time series, which enjoys poor stability. To improve the performance of MAAPE, Chen et al. proposed an improved MAAPE (IMAAPE) method [12], which uses the improved coarse-grained method instead of the traditional method so as to significantly improve the stability and feature extraction performance of MAAPE. However, both methods can only extract the dynamic feature information of the low-frequency components, but ignores the equally important information hidden in the high-frequency components. With a view to solving the defect that multiscale entropy cannot extract high-frequency features, Jiang et al. [13] proposed the concept of hierarchical

entropy and divided the original time series into different frequency bands by constructing high-frequency and low-frequency operators, thus obtaining the corresponding high-frequency and low-frequency components. In view of this, this paper combines hierarchical entropy with AAPE and proposes hierarchical amplitude-aware permutation entropy (HAAPE). The method adopts AAPE to quantify the complexity of each subseries obtained through hierarchical decomposition, which can extract the low-frequency and high-frequency features of the time series at the same time, thereby overcoming the drawback of high-frequency information loss in MAAPE.

In view of the theoretical advantages of HAAPE, this paper introduces it into the field of fault diagnosis to extract fault features from vibration signals of rotating machinery. However, the original fault feature vectors extracted by HAAPE are high-dimensional, which not only contains sensitive information closely related to the working state but also inevitably has redundant information that does not contribute to fault identification. Meanwhile, high-dimensional feature vectors mean high computational cost and possible overfitting phenomenon of the classifier. Therefore, it is necessary to choose an effective feature selection method to screen sensitive features from the original feature vectors and reduce the impact of redundant information [14].

Pairwise feature proximity (PWFP) [15] is a dimensionality reduction tool recently proposed by SL Happy et al. In this method, according to the principle of keeping minimum within-class distance and maximum between-class distance, all features are ranked by statistical method and the features with low scores are the required sensitive features. In literature [15], the dimensionality reduction performance of PWFP was verified through several real-world data sets, and the results showed that, compared with traditional methods such as Relief-F, Laplace Score, and Fisher Score, the performance of PWFP was superior when processing high dimension data with low sample size.

SVM [16] is a nonlinear supervised learning classifier and enjoys excellent generalization ability when dealing with small sample classification problems. However, the performance of SVM depends on the setting of two key parameters, namely, penalty factor and kernel function parameter. To give full play to the best performance of SVM, many optimization algorithms have been applied to the parameter selection of SVM. Among them, the gray wolf optimization (GWO) [17] enjoys great convergence speed and global search ability, so we employ it in this paper to determine the best parameters of SVM by iterative optimization. Based on the above, this paper proposes a new fault diagnosis method for rotating machinery based on HAAPE, PWFP and GWO-SVM. Firstly, HAAPE is used to extract the original fault features representing the working state of rotating machinery from the vibration signals. Then, PWFP is used to rank the original features according to the feature sensitivity, and the best features are selected to form the low-dimensional sensitive feature vectors to reduce the impact of redundant information. Finally, the sensitive feature vectors are input into the GWO-SVM-based classifier for training and recognition, and the rotating machinery fault diagnosis is completed. With a view to proving the effectiveness and universality of the proposed method, the performance of

the proposed method is verified using two fault data sets of bearing and gear. The results display that the proposed method can effectively complete the fault identification of rotating machinery, and the accuracy reaches 100%. At the same time, to prove the superiority of our method, we compare the proposed HAAPE with some comparison methods such as MAAPE, IMAAPE, multiscale permutation entropy (MPE), hierarchical permutation entropy (HPE), multiscale sample entropy (MSE), and hierarchical sample entropy (HSE). The results show that the feature extraction performance of HAAPE is superior.

All in all, the main contributions and innovations of this paper are as follows:

- (1) HAAPE, a new nonlinear time series feature extraction method, is proposed in this paper and applied to the field of rotating machinery fault diagnosis, which shows better feature extraction performance than comparative methods such as MAAPE, MPE, HPE, and so on.
- (2) Combining HAAPE-based feature extraction method, PWFP-based feature selection method and GWO-SVM-based classifier, a new rotating machinery fault diagnosis method is proposed.
- (3) Two experiments, namely bearing fault diagnosis experiment and gear fault diagnosis experiment are carried out to verify the performance of the proposed method, as well as verifying its superiority over other methods.

The rest of the paper is organized as follows: Section 2 explains the basic principle of AAPE, MAAPE and HAAPE, verifies the advantages of HAAPE over MAAPE through simulation signal experiment and discusses the choice of HAAPE's parameters. In Section 3, the diagnostic procedure of the proposed method is described in detail; meanwhile, the principles of PWFP and GWO-SVM are also summarized in this section. In Section 4, two data sets of bearing and gearbox are used to verify the performance of the proposed method. Finally, the whole paper is summarized in Section 5.

## 2. The Basic Principle of HAAPE

**2.1. AAPE.** In order to solve the problem that PE cannot consider the amplitude of each element in the time series and the amplitude difference between adjacent elements, which leads to the inaccurate evaluation of dynamic characteristics, AAPE is proposed. In other words, the principle of AAPE is based on PE. Therefore, to clearly explain the calculation process of AAPE, we firstly summarize the principle of PE as follows:

Given a random time series  $Y = \{y_i\}$ ,  $i = 1, 2, \dots, N$ . Based on the time delay  $\lambda$  and embedding dimension  $m$ ,  $Y$  can be transformed into multiple reconstruction series as follows:

$$Y_j^{m,\lambda} = [y_j, y_{j+\lambda}, \dots, y_{j+(m-2)\lambda}, y_{j+(m-1)\lambda}], \quad (1)$$

$$j = 1, 2, \dots, N - (m - 1)\lambda.$$

According to the size relationship of each element in the reconstructed series, the ascending sorting is carried out,

and different permutation patterns  $\pi_{v_0, v_1, \dots, v_{m-1}}$  are obtained as follows:

$$\left[ y_{j+(i_1-1)\lambda}, y_{j+(i_2-1)\lambda}, \dots, y_{j+(i_{m-1}-1)\lambda}, y_{j+(i_m-1)\lambda} \right], \quad (2)$$

where  $i_*$  represents the position index of each element in the reconstruction series. When the embedding dimension is  $m$ , the length of the reconstructed series is also  $m$ . Therefore, there are  $m!$  possible permutation patterns, and the  $k$ -th permutation pattern is denoted as  $\pi_k$ .

The occurrence probability of each permutation pattern can be calculated by

$$p(\pi_k) = \frac{\text{Num}(\pi_k)}{N - (m - 1)\lambda}, \quad (3)$$

where  $\text{Num}(\pi_k)$  represents the number of  $\pi_k$  pattern in all reconstruction subseries. Specifically, the number of  $\text{Num}(\pi_k)$  will increase by 1 whenever the permutation pattern of  $Y_j^{m,\lambda}$  is  $\pi_k$ .

Based on the Shannon entropy theory, the PE value of time series  $Y$  can be calculated as follows:

$$PE(Y, m, \lambda) = - \sum_{k=1}^{m!} p(\pi_k) \ln p(\pi_k). \quad (4)$$

It can be seen from the above that PE sorts different elements according to their amplitudes to obtain different permutation patterns, and the final entropy value is calculated through the statistics of the occurrence probability of different permutation patterns. There are two drawbacks in this process. First of all, PE only considers the order information of time series, but ignores the amplitude information of each element. Secondly, when elements with the same amplitude appear in the time series, PE numbers them according to their occurrence order, which leads to the imprecise statistics of the permutation patterns. In order to solve the above defects, AAPE is proposed. Compared with PE, the specific improvements of AAPE are as follows:

Set the initial value of  $p(\pi_k)$  as 0. When the permutation pattern of  $Y_j^{m,\lambda}$  is  $\pi_k$ , different from the counting rule of PE, AAPE adopts equation (5) to calculate the contribution degree of the different pattern. Among which,  $A \in [0, 1]$  denotes the adjustment coefficient which is utilized to adjust the weight of the time series amplitude average and the deviation between the amplitudes.  $l$  refers to the number of elements with the same amplitude in  $Y_j^{m,\lambda}$ . For example, when there are three elements with the same amplitude in  $Y_j^{m,\lambda}$ ,  $l = 6$ , and when there are no elements with the same amplitude in  $Y_j^{m,\lambda}$ ,  $l = 1$ .

$$p^{\text{update}}(\pi_k) = p(\pi_k) + \frac{1}{l!} \left( \frac{A}{m} \sum_{k=1}^m |y_{j+(k-1)\lambda}| \right. \\ \left. + \frac{1-A}{m-1} \sum_{k=2}^m |y_{j+(k-1)\lambda} - y_{j+(k-2)\lambda}| \right). \quad (5)$$

Correspondingly, the relative probability of the permutation pattern  $\pi_k$  can be calculated as follows:

$$p(\pi_k) = \frac{p^{\text{update}}(\pi_k)}{\sum_{j=1}^{N-(m-1)\lambda} A/m \sum_{k=1}^m |y_{j+(k-1)\lambda}| + 1 - A/m - 1 \sum_{k=2}^m |y_{j+(k-1)\lambda} - y_{j+(k-2)\lambda}|} \quad (6)$$

Therefore, the AAPE value of the time series  $Y$  can be calculated by

$$\text{AAPE}(Y, m, \lambda, A) = - \sum_{k=1}^m p(\pi_k) \ln p(\pi_k). \quad (7)$$

**2.2. MAAPE.** To improve the feature extraction performance of AAPE, MAAPE is proposed based on multiscale entropy theory.

For time series  $Y = \{y_i\} \quad i = 1, 2, \dots, N$ , coarse-graining process is conducted on it using equation (8) firstly.

$$y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} y_i, \quad 1 \leq j \leq \frac{N}{\tau}, \quad (8)$$

where  $\tau$  is the scale factor and  $y_j^\tau$  is the obtained coarse-grained time series.

MAAPE can be obtained by calculating AAPE values of  $\tau$  coarse-grained time series.

$$\text{MAAPE} = \text{AAPE}(y_j^\tau, m, \lambda, A). \quad (9)$$

**2.3. HAAPE.** As shown in the principle of MAAPE, the essence of multiscale extension is to calculate the mean value of  $\tau$  adjacent elements in the time series, so as to obtain different coarse-grained time series. Each subseries obtained by this processing method is the low-frequency component of the time series, so MAAPE can only extract the low-frequency information of the time series, but ignores the high-frequency information. To solve this issue, this paper proposes hierarchical amplitude-aware permutation entropy (HAAPE). This method adopts hierarchical processing method instead of multiscale extension, which can extract the high frequency and low-frequency features of time series simultaneously. The specific principle of HAAPE is summarized as follows:

- (1) For the time series  $Y = \{y_i\}$  of length  $N = 2^n$ , low-frequency operator  $Q_0$  and high-frequency operator  $Q_1$  are defined as follows:

$$\begin{cases} Q_0(y) = \frac{y_{2i+1} + y_{2i+2}}{2}, \\ Q_1(y) = \frac{y_{2i+1} - y_{2i+2}}{2}, \quad i = 0, 1, \dots, 2^{n-1} - 1, \end{cases} \quad (10)$$

where  $Q_0(y)$  and  $Q_1(y)$  are the low-frequency and high-frequency components of the time series, respectively.

The matrix representation of  $Q_0$  and  $Q_1$  is as follows, where  $j = 0$  or  $1$ .

$$Q_j = \begin{bmatrix} \frac{1}{2} & \frac{(-1)^j}{2} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{(-1)^j}{2} & \dots & 0 & 0 \\ \vdots & & & & & & \\ 0 & 0 & 0 & 0 & \dots & \frac{1}{2} & \frac{(-1)^j}{2} \end{bmatrix}. \quad (11)$$

- (2) When the number of hierarchical layers is  $k$ ,  $k$  dimension vector  $[s_1, s_2, \dots, s_k]$  is constructed, where  $s_k = 0$  or  $1$ . The hierarchical node  $l$  can be calculated using the following equation:

$$l = \sum_{m=1}^k 2^{k-m} s_m. \quad (12)$$

- (3) Based on  $[s_1, s_2, \dots, s_k]$ , repeat step (1) to obtain the hierarchical component  $Y_{k,l}$  corresponding to node  $l$  when the number of hierarchical layers is  $k$ :

$$Y_{k,l} = Q_{s_k}^k \cdot Q_{s_{k-1}}^{k-1} \dots Q_{s_1}^1 \cdot Y. \quad (13)$$

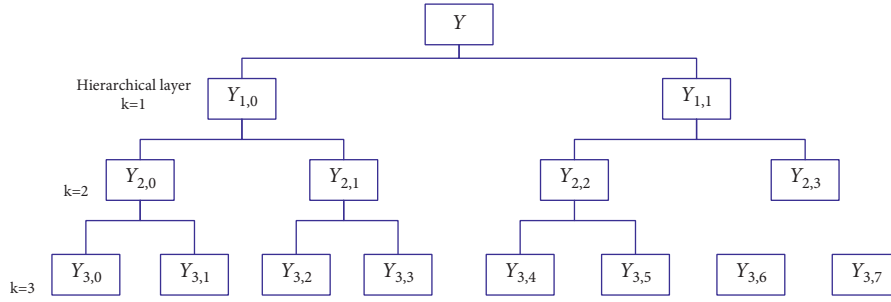
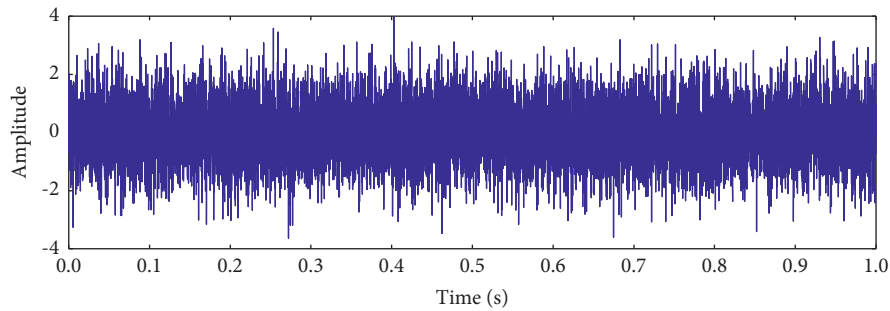
- (4) Based on the above, the HAAPE of the original series  $Y = \{y_i\}$  can be calculated as follows:

$$\text{HAAPE}(Y, k, l, m, \lambda, A) = \text{AAPE}(Y_{k,l}, m, \lambda, A). \quad (14)$$

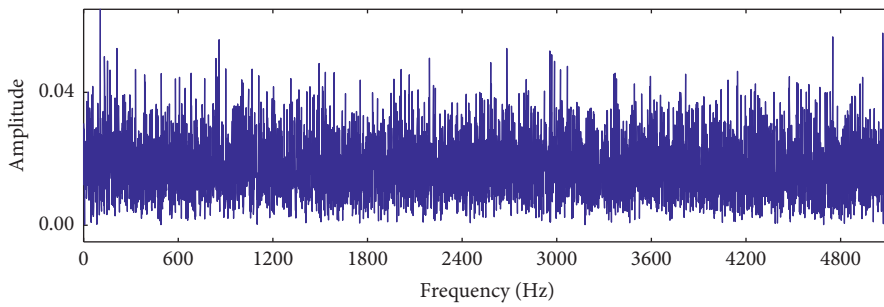
To intuitively explain the process of hierarchical decomposition, Figure 1 shows the schematic diagram when the number of hierarchical layers is 3, where  $Y_{1,0}$  and  $Y_{1,1}$ , respectively, represent the high-frequency component and low-frequency component of  $Y$  when the hierarchical layer is 1.

According to the above definition, the high-frequency operator  $Q_1$  and low-frequency operator  $Q_0$  of hierarchical entropy correspond to the low pass and high pass filter of the Harr wavelet, respectively. Thus, the essence of HAAPE is firstly, wavelet packet decomposition based on Harr wavelet is used to decompose the original signal to obtain subseries of different frequency bands, and then the AAPE value of each subseries is calculated to obtain corresponding feature vectors.

**2.4. Simulation Signal Experiment Analysis.** With a view to testing the performance of the proposed HAAPE method as well as comparing it with the existing MAAPE method, 50 independent random white noise signals with a length of 2048 are adopted for comparative experiments. Figure 2 displays the time domain and frequency domain waveforms of random white noise. Obviously, the complexity of white

FIGURE 1: Schematic diagram of hierarchical decomposition when  $k = 3$ .

(a)



(b)

FIGURE 2: (a) Time domain waveform of white noise. (b) Frequency domain waveform of white noise.

noise remains roughly constant throughout the frequency band, so theoretically, the entropy values of white noise will maintain stable [13]. HAAPE and MAAPE are used to extract the entropy values of all white noise samples, where the parameters of HAAPE and MAAPE are set to  $m = 5, \lambda = 1, A = 0.5, k = 3$ , and  $m = 5, \lambda = 1, A = 0.5, \tau = 8$ , respectively. The error bar curves corresponding to the two different methods are shown in Figure 3, where the mean entropy value and standard deviation are well shown. It can be seen that, for MAAPE, the entropy value of white noise shows an obvious downward trend with the increase of scale factor, which indicates that MAAPE can only extract the low-frequency information of time series, but cannot evaluate the complexity of high-frequency band well. On the contrary, for HAAPE, the entropy value of white noise on each node is basically unchanged, which is consistent with the reality, proving that HAAPE can effectively prevent the loss of high-frequency information. In addition, observation shows that the standard deviation of HAAPE method is smaller than that of MAAPE, indicating that HAAPE

method is more stable. As a whole, HAAPE enjoys better feature extraction performance than MAAPE.

**2.5. The Parameter Selection of HAAPE.** As described above, the main parameters of HAAPE include embedding dimension  $m$ , time delay  $\lambda$ , adjustment coefficient  $A$ , and hierarchical layer  $k$ . The average amplitude and amplitude difference of rotating machinery vibration signals are all very important for the extraction of fault features, so adjustment coefficient is set to  $A = 0.5$  in this paper, which is also in line with the recommendation of literature [9]. Generally speaking, time delay has little influence on the final entropy value, which is generally set to  $\lambda = 1$  [18]. For the embedded dimension  $m$ , if  $m$  is too small, HAAPE cannot effectively detect the dynamic features of time series, while if  $m$  is too large, the calculation cost will increase. In literature [6],  $m$  is recommended to be set between 3 and 7. For the hierarchical layer, if  $k$  is too small, it will lead to insufficient extraction of high frequency and low-frequency information of the time

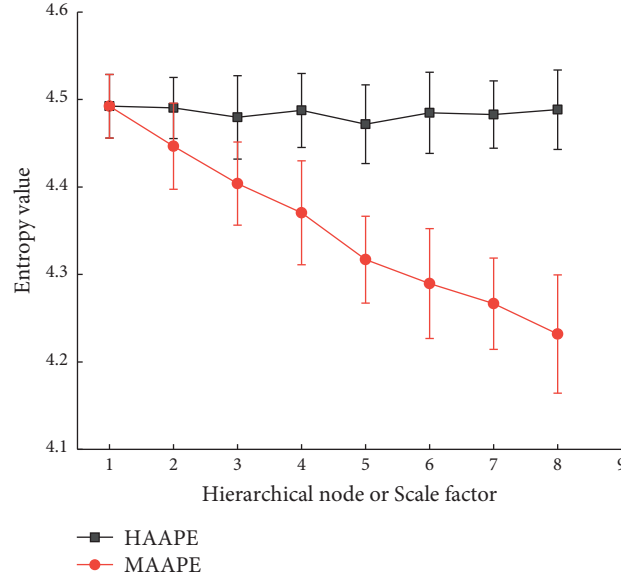


FIGURE 3: Entropy value curves of two different methods.

series; if  $k$  is too large, the calculation cost will be greatly increased and the practicability of the method will be reduced [19]. Considering feature extraction performance and calculation cost, as well as the common parameter settings for hierarchical entropy and AAPE [9, 11, 20], this paper finally sets the parameters of HAAPE as follows:  $m = 3, \lambda = 1, A = 0.5, k = 3$ .

### 3. The Proposed Fault Diagnosis Method for Rotating Machinery

**3.1. PWF Feature Selection Method.** By constructing low-frequency and high-frequency operators, HAAPE can simultaneously extract the high-frequency and low-frequency features which represent the state of time series, thereby effectively avoiding the drawback of high-frequency information loss in MAAPE. However, the original feature vectors extracted by HAAPE are high-dimensional, and there is inevitably redundant information that affects the final pattern recognition. In order to eliminate the influence of redundant information and improve the separability of feature vectors, it is necessary to use an effective feature selection method to screen sensitive features from original feature vectors. The PWF feature selection method is a dimensionality reduction tool proposed recently which is suitable for the classification of high dimension data with low sample size, whose criteria is based on the minimum within-class distance and maximum between-class distance. Different from the traditional methods, which take all feature samples of the same class as a whole for distance measurement, this method adopts two feature samples of the same or different classes as subsets and counts the number of optimal features corresponding to each subset. After that, each feature is scored and ranked, while the feature corresponding to the low score is the sensitive feature with better separability. In literature [15], it is proved through comparative experiments that PWF

enjoys better dimensionality reduction effect than Fisher Score, Laplace Score, and Relief-F. In view of the excellent performance of PWF, this paper employs it to further process the original high-dimensional features extracted using HAAPE. The specific principles of PWF can be summarized as follows:

- (1) Suppose a  $d$ -dimensional feature vector set  $\{x_i^d, y_i\}$  with a sample number of  $M$ , where  $x_i^d$  represents the feature vector and  $y_i$  is the corresponding sample label,  $y_i = 1, 2, \dots, c$ ,  $i = 1, 2, \dots, m$  and  $c$  represents the number of labels. For two different samples under the same label  $(x_j, y_j), (x_k, y_k), y_j = y_k$ , calculate the Manhattan distance between them, and get the distance vector  $p_{jk} = [s_1, s_2, \dots, s_d]$ , where  $s_d$  represents the distance between the  $d$ -th feature of the two samples. The sorted vector  $p_{jk}^* = [s_1^*, s_2^*, \dots, s_d^*]$  is obtained by ascending sorting according to the size of each element in  $p_{jk}$ , where  $s_1^*$  represents the position index of the nearest feature in the original feature vector.
- (2) Select the first  $d$  features from  $p_{jk}^*$  as a subset  $p^*$ . And, the same method is adopted to get the feature subset  $p^*$  corresponding to all paired samples with the same label. Then, the occurrence times of all features are counted to get the statistic vector  $\tilde{p} = [c_1, c_1, \dots, c_d]$ , where  $c_d$  represents the occurrence times of the  $d$ -th feature in the feature subset.
- (3) For two samples  $(x_j, y_j), (x_k, y_k), y_j \neq y_k$  under different labels, calculate the Manhattan distance between them and get the distance vector  $q_{jk} = [s_1, s_2, \dots, s_d]$ . The sorted vector  $q_{jk}^* = [s_1^*, s_2^*, \dots, s_d^*]$  is obtained by descending sorting according to the size of each element in  $p_{jk}$ , where  $s_1^*$  represents the position index of the farthest feature in the original feature vector.

- (4) Select the first  $d$  features from  $q_{jk}^*$  as a subset. And, the same method is used to obtain the feature subset  $q^*$  corresponding to all paired samples with different labels. Then, the occurrence times of all features are counted to obtain the statistical vector  $\tilde{q} = [c_1, c_1, \dots, c_d]$ .
- (5) Use equation (15) to calculate the occurrence probability of the required features in all paired samples:

$$\begin{aligned} P &= \frac{1}{N_p} \tilde{p}, \\ N_p &= \sum_{k=1}^c \binom{m_k}{1}, \\ Q &= \frac{1}{N_q} \tilde{q}, \\ N_q &= \sum_{j,k;j \neq k} m_j \cdot m_k, \end{aligned} \quad (15)$$

where  $m_k$  stands for the number of samples labeled  $k$ .

- (6) The criteria shown in equation (16) is adopted to screen the optimal features:

$$S(i) = \left| \frac{P - Q}{P + Q} \right|, \quad i = 1, 2, \dots, d, \quad (16)$$

where  $S(i)$  represents the score of each feature, and selecting the optimal  $m$  features means selecting the  $m$  features with the lowest score in the original feature vector.

**3.2. GWO-SVM.** SVM is a machine learning method which is widely used in the field of pattern recognition. It adopts kernel function to map the nonseparable samples in low-dimensional space to high-dimensional space and constructs a hyperplane to make the samples linearly separable in high-dimensional space, so as to improve the generalization ability. For small sample classification problem, SVM is generally the preferred method, so SVM has been widely used in the field of fault diagnosis. For reasons of space, the principle of SVM is not described in detail in this paper.

The performance of SVM depends on the setting of penalty factor  $c$  and kernel function parameter  $g$ . Reasonable parameters setting can bring out the best classification effect of SVM. On the contrary, unreasonable parameters setting will easily lead to overfitting or underfitting. In order to determine the optimal parameters of SVM adaptively and avoid the influence of artificial parameters setting, with average error obtained by trifold cross-validation of training samples as the fitness value, the gray wolf optimization (GWO) with strong global search ability is employed to determine the optimal parameters of SVM. The specific steps of GWO-SVM are as follows:

- (1) Set the parameters of GWO, and generate the initial wolf pack with  $[c, g]$  as the wolf's position.

- (2) The corresponding fitness values of all wolves in the initial wolf pack are calculated, and the wolves corresponding to the three minimum fitness values are set as  $\alpha, \beta, \sigma$ , respectively, while their positions are set as  $X_\alpha, X_\beta, X_\delta$ . The other wolves are all  $\omega$  wolves, and the positions are set as  $X$ .
- (3) According to  $X_\alpha, X_\beta, X_\delta, X$  is updated by the following equation to obtain new wolf pack:

$$\begin{cases} D_j = |C_j \cdot X_n^j - X_n^k|, \\ X_t = X_n^j - A_j \cdot D_j, \\ X_{n+1}^k = \frac{\sum_{t=1}^3 X_t}{3}, \end{cases} \quad (17)$$

where  $j = \alpha, \beta, \delta$ ,  $X_n^k$  represents the position of the  $k$ -th wolf in the  $n$ -th generation wolf pack,  $D_j$  represents the distance between  $\omega$  wolf and  $j$  wolf, and  $X_{n+1}^k$  represents the position of updated wolf pack.  $A$  and  $C$  stand for calculation coefficients, which are calculated as follows:

$$\begin{cases} A = 2a \cdot r_1 - a, r_1 \in [0, 1], \\ C = 2 \cdot r_2, r_2 \in [0, 1]. \end{cases} \quad (18)$$

- (3) Repeat steps 2 and 3 after getting new wolf pack.
- (4) Iterate until the maximum number of cycles  $T_{\max}$  is satisfied, and output the best  $[c, g]$  and corresponding classification accuracy.

The parameters of GWO-SVM in this paper are set as shown in Table 1:

**3.3. The Diagnostic Process of the Proposed Method.** Based on HAPPE, PWFP and GWO-SVM, this paper proposes a new intelligent fault diagnosis method for rotating machinery. The specific steps can be summarized as follows:

- (1) Data acquisition: the acceleration sensor is used to collect vibration signals of  $m$  different working states of rotating machinery, and  $N$  nonoverlapping samples with length  $L$  are taken for each state.
- (2) Feature extraction: HAAPE is employed to extract original fault features of all samples. According to Section 2.3, the parameters of HAAPE are set to  $m = 5, \lambda = 1, A = 0.5, k = 3$  in this paper. In order to enrich the number of features and facilitate further screening of the optimal features, the entropy values corresponding to the hierarchical components of each layer are taken into account, that is, the original feature vector obtained is  $[E_X, E_{1,0}, E_{1,1}, E_{2,0}, E_{2,1}, E_{2,2}, E_{2,3}, E_{3,0}, E_{3,1}, E_{3,2}, E_{3,3}, E_{3,4}, E_{3,5}, E_{3,6}, E_{3,7}]$ , where  $E_X$  represents the AAPE value of the original signal, and  $E_{1,0}$  represents the AAPE value corresponding to the hierarchical component at node 0 of the first layer. Therefore, the original feature vector obtained is a high-dimensional vector of length 15.

TABLE 1: Parameters setting for GWO-SVM

The size of wolf pack	$T_{\max}$	Initial fitness value	$c$	$g$
20	50	$+\infty$	[0.01 1000]	[0.01 1000]

- (3) Feature dimensionality reduction: PWF is used to sort the original features according to the separability of different features, and the features with low scores are selected as sensitive features to form the low-dimensional sensitive feature vector. In this paper, the number of sensitive features is set to 8, that is, the sensitive feature vector is a low-dimensional vector with length of 8.
- (4) Fault identification: for each state, sensitive feature vectors of  $n$  samples are randomly selected as training samples, and sensitive feature vectors of the rest  $N - n$  samples are selected as testing samples. The training samples are input into GWO-SVM for training, so as to determine the optimal parameters of SVM, and the fault classifier is constructed. Then the testing samples are input into the trained classifier for identification, and the fault diagnosis of rotating machinery is completed.

Figure 4 shows the flowchart of the proposed fault diagnosis method.

#### 4. Experimental Verification

For the purpose of verifying the effectiveness of the proposed fault diagnosis method, in this section, we firstly adopt the bearing fault data set of Case Western Reserve University, which is widely used in the fault diagnosis field, to test the performance of the presented method. In addition, to prove the universality of our method, the measured gearbox fault data set is also used to verify the performance of the proposed method. With a view to highlighting the advantages of the proposed method, we conduct a series of comparative experiments to verify the advantages of the proposed HAAPE feature extraction method and the necessity of using the PWF feature selection method.

##### Case 1. Fault diagnosis based on rolling bearing

In this section, the publicly available Case Western Reserve University bearing fault data set is used to verify the diagnostic performance of the proposed method. The experimental platform is displayed in Figure 5, which is mainly composed of motor, torque sensor, and dynamometer. The test bearing is located at the motor drive end, and the model is SKF6205. Using EDM technology to deal with normal bearings to simulate different fault types and degrees of bearings. The inner ring fault bearing, outer ring fault bearing, and rolling ball fault bearing with fault diameters of 0.1778, 0.3556, and 0.5334 mm are obtained, so there are ten different working states, including normal working state and nine fault states. During the experiment, the speed of the motor is 1797r/min, and the load is 0. The accelerometer is

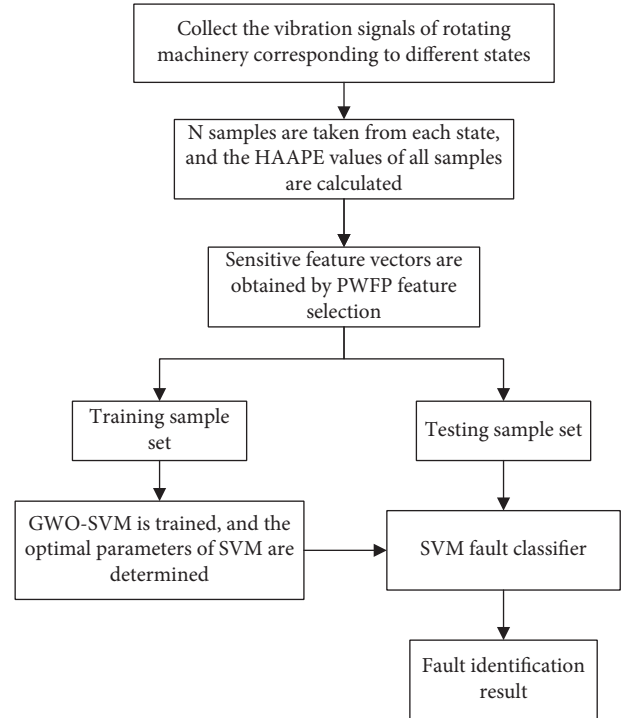


FIGURE 4: The flowchart of the proposed method.

installed at the drive end to collect vibration signals of rolling bearings under different working states with a sampling frequency of 12 kHz. Figure 6 shows the time-domain waveforms corresponding to different states. For each working state, 50 nonoverlapping samples with a length of 2048 are taken. Among them, 20 samples are randomly selected as training samples and the remaining 30 samples are testing samples. So there are 200 training samples and 300 testing samples. Table 2 lists the details of different working states of bearings.

According to the fault diagnosis process described in Section 3.3, HAAPE is firstly employed to extract the fault features of all samples to obtain the original high-dimensional fault feature vectors. The mean entropy curves corresponding to different states are shown in Figure 7(a). Then, PWF is used to rank the obtained original fault features, and the results are shown in Figure 7(b). The first eight features after ranking (i.e., features with positions of 12, 2, 3, 10, 13, 1, 14, and 9 in the original feature vectors) are selected as sensitive features in this paper to form sensitive feature vectors. The criteria of PWF feature selection method are based on the minimum within-class distance and maximum between-class distance. Therefore, the eight sensitive features represent the most separable features in the original feature vector. The sensitive feature vectors of the training samples are input into the GWO-SVM-based fault classifier for training to construct the optimal SVM classifier. Then, sensitive feature vectors of testing samples are input for identification. The running results are shown in Figure 8. It can be seen that the proposed method can effectively distinguish different fault types and fault degrees of rolling bearings, and the fault recognition accuracy is 100%, which



proves that the proposed method enjoys good fault identification effect.

In order to highlight the feature extraction advantages of the proposed HAAPE method, four multiscale entropy methods, namely, MAAPE, IMAAPE, MPE, and MSE, and two hierarchical entropy methods, namely, HPE and HSE, are, respectively, used to replace HAAPE in the proposed method for fault feature extraction, while other diagnostic processes remain unchanged. Parameter setting of different methods is displayed in Table 3. The identification results of different methods are shown in Table 4 and Figure 9, where “time” in Table 4 refers to the time required to extract the fault features of a single sample using the corresponding method. The experimental results show that both HAAPE method proposed in this paper and IMAAPE method proposed by Chen et al. enjoy the highest recognition accuracy, reaching 100%. However, the computational efficiency of HAAPE is significantly higher than that of IMAAPE, which can extract fault features of samples at a faster speed. In addition, it can be observed that the recognition accuracy of hierarchical entropy methods is higher than that of other multiscale entropy methods (HAAPE > MAAPE, HPE > MPE, HSE > MSE). This reason is that hierarchical entropy can extract the high-frequency and low-frequency features of vibration signal simultaneously and overcome the defect of high-frequency information loss in multiscale entropy, so the feature extraction performance is better. In terms of computational efficiency, the PE-based method is significantly more efficient than the SE-based method. The calculated cost of the proposed HAAPE method is slightly higher than that of MAAPE, HPE, and MPE, but it is sufficient to meet the needs of practical application. Overall, compared with other methods, the feature extraction performance of the proposed HAAPE is better.

For the purpose of studying the relationship between the number of features and the recognition accuracy as well as to prove the necessity of using the PWFP method to reduce the dimension of features, after using PWFP to rank the original features, the relationship curve between the number of features and the recognition accuracy is drawn, as depicted in Figure 10. It can be seen that, with the increase of the number of features, the recognition accuracy corresponding to different methods first increases, and then remains basically stable. It is worth noting that the higher the number of features, the higher the training cost. Therefore, it is not necessary to use all the features for training and recognition. The ideal situation is to obtain the highest recognition accuracy with the least number of features, which indicates that it is necessary to adopt the PWFP method for feature selection. It can be found that when HAAPE, MAAPE, and IMAAPE achieve the highest accuracy, the corresponding feature number is the least, both of which are 8. However, the highest recognition accuracy of HAAPE and IMAAPE is 100%, and that of MAAPE is 97.67%. This phenomenon proves that compared with other methods, HAAPE and IMAAPE enjoys stronger robustness and can perfectly realize bearing fault identification with only a small number of features. In addition, according to Table 4, the calculation cost of HAAPE is significantly lower than that of IMAAPE.

Therefore, overall, HAAPE enjoys the best feature extraction performance among these methods.

#### Case 2. Fault diagnosis based on gearbox

In the previous section, the rolling bearing fault data set is used to verify the performance of the proposed method. Experimental results show that our method can effectively identify different fault types and fault degrees of bearings and has advantages in recognition accuracy and robustness compared with other methods. In order to further verify the universality of the proposed fault diagnosis model, we use the gear fault data set to further verify the performance of the method in this section.

Gear fault data is collected from QPZZ-II rotating machinery fault experimental platform, and its structure is shown in Figure 11. The gear box used in the experiment is a single reduction gear unit, where the pinion is driven wheel with 55 teeth, and the gear wheel is driven wheel with 75 teeth. We use different fault types of gears to replace normal gears in the gearbox to simulate different fault states of the gearbox. There are four different faults: pinion wear, gear wheel pitting, gear wheel broken teeth, and gear wheel pitting + pinion wear. In the experiment, the motor is connected to the synchronous belt to drive the shaft to provide power for the gearbox. The motor speed is 880r/min, and the load is 0. The acceleration sensor installed on the bearing seat on the motor side of the gearbox output shaft is used to collect the vibration signals of the gearbox under different states. The sampling frequency is 5120 Hz, and the sampling time is 10.4s. Figure 12 shows the time-domain waveforms corresponding to different working states. Due to the limited data length, this paper adopts a sliding window with a length of 2048 and a step size of 1024 for sampling, and 50 samples with a length of 2048 are taken for each working state, among which 20 samples are randomly selected as training samples and the remaining 30 samples are testing samples. Table 5 lists the details of the different working states of the gearbox.

Same as Case 1, the HAAPE values of all samples are extracted to obtain the original fault feature vectors. The mean entropy curves corresponding to different states are shown in Figure 13(a). It can be seen that not all features have good separability, so feature selection is necessary. The features sorted by PWFP are shown in Figure 13(b). The first 8 features are selected (that is, the features whose positions in the original feature vectors are 7, 12, 8, 15, 1, 5, 11, and 2) to form sensitive feature vectors. The sensitive feature vectors of training samples are input into GWO-SVM for training, and the optimal parameters of SVM are determined as  $c = 361.09$ ,  $g = 140.34$ . Then, the sensitive feature vectors of testing samples are input into the trained SVM for recognition. At the same time, in order to compare the performance of different feature extraction methods, MAAPE, IMAAPE, HPE, MPE, HSE, and MSE are used to replace HAAPE for feature extraction, respectively. The parameters setting of different methods are shown in Table 3 and the diagnostic results are depicted in Table 6 and Figure 14. It can be seen that the identification accuracy of the proposed method reaches 100%, which

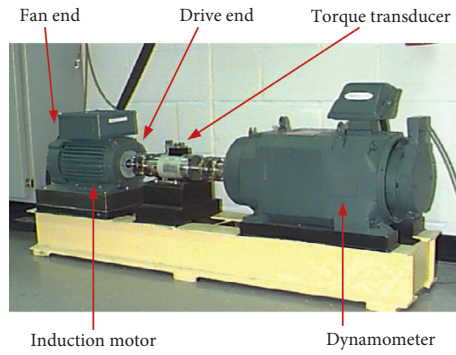


FIGURE 5: The experiment platform.

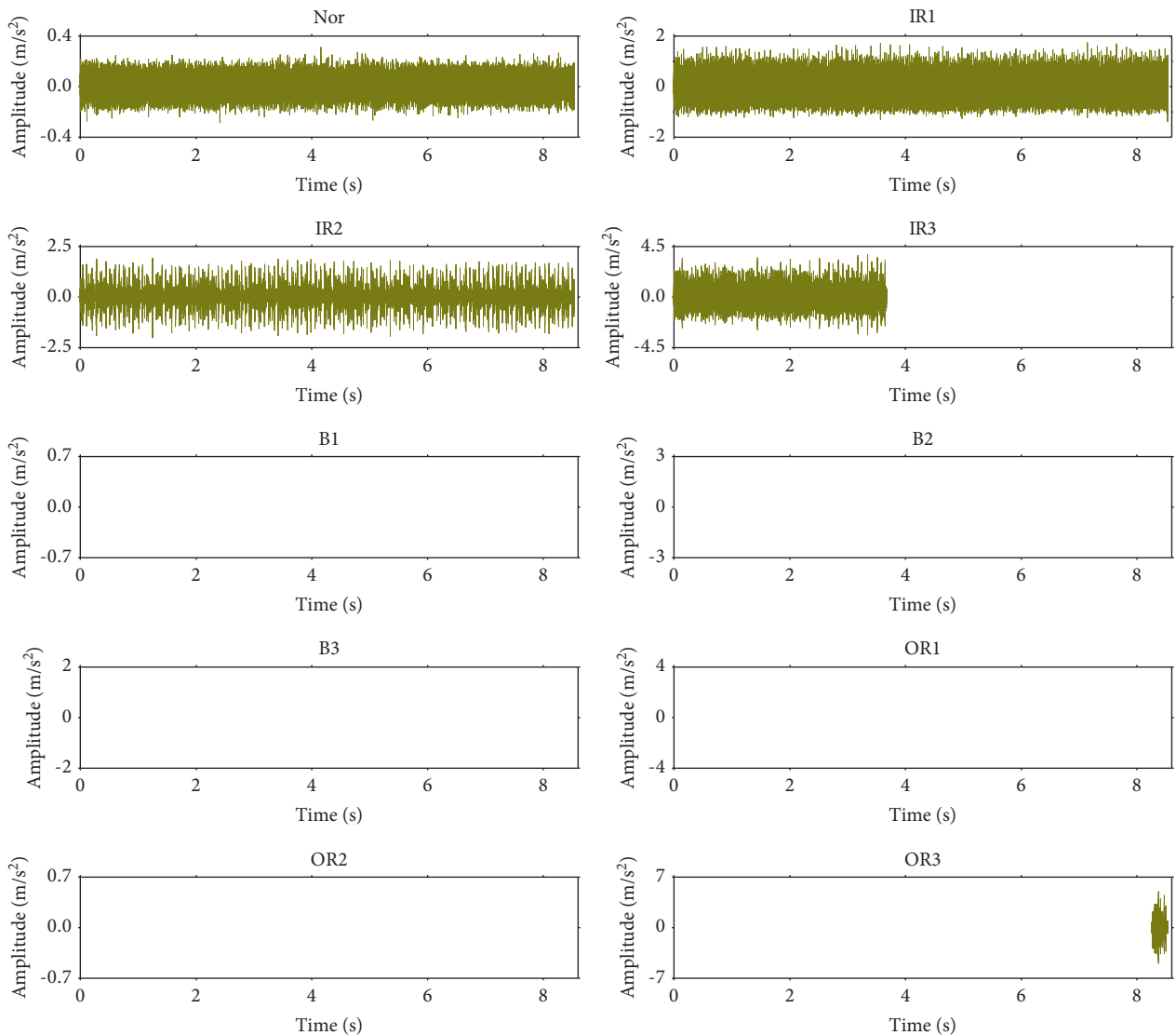


FIGURE 6: Time-domain waveforms corresponding to different states.

indicates that the method can effectively realize the identification of different gear fault types. Considering the two aspects of calculation efficiency and recognition accuracy, the proposed method obviously has superior performance. In addition, it can be seen from the recognition

accuracy of different methods that the feature extraction performance of hierarchical entropy is better than that of multiscale entropy, which is the same as the conclusion obtained by Case 1, thus fully proving the advantage of hierarchical entropy over multiscale entropy.

TABLE 2: Detailed information of different working states of bearings.

Label	Bearing state	Abbreviations	Fault diameter (mm)	The number of training samples	The number of testing samples
1	Normal state	Nor		20	30
2	Inner race fault	IRF1	0.1778	20	30
3		IRF2	0.3556	20	30
4		IRF3	0.5334	20	30
5	Ball fault	BF1	0.1778	20	30
6		BF2	0.3556	20	30
7		BF3	0.5334	20	30
8	Outer race fault	ORF1	0.1778	20	30
9		ORF2	0.3556	20	30
10		ORF3	0.5334	20	30

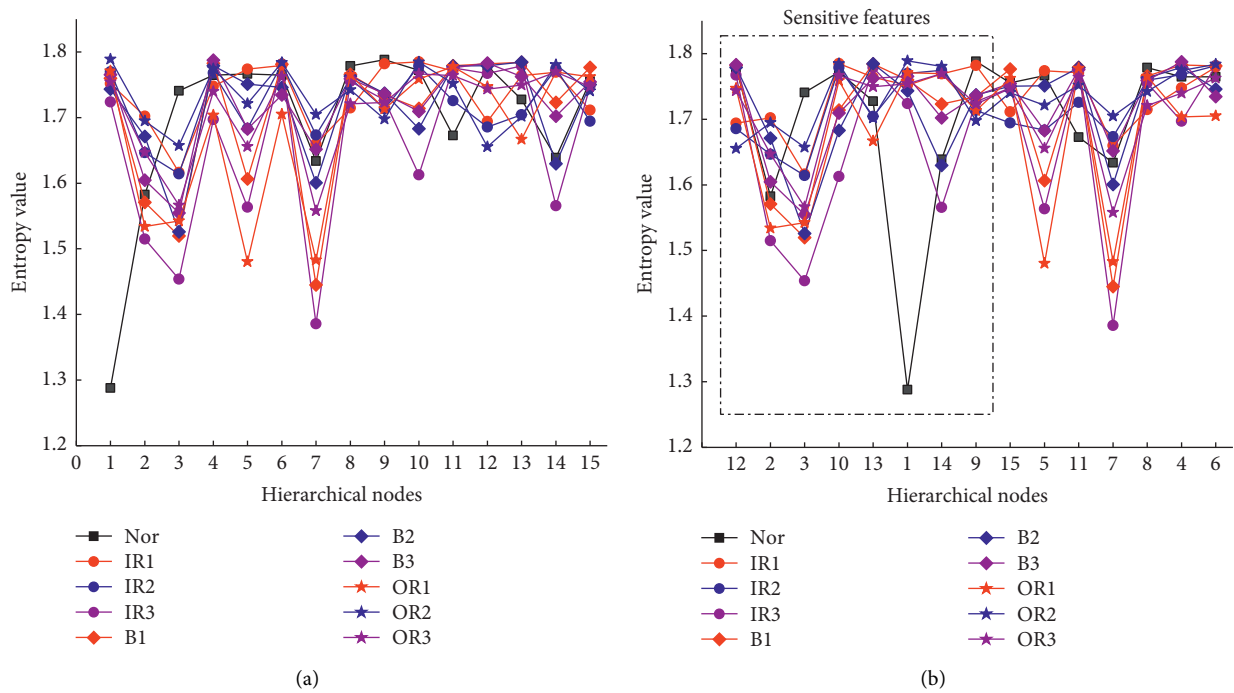


FIGURE 7: The original feature vectors and the sensitive feature vectors. (a) Original fault features; (b) fault features after ranking.

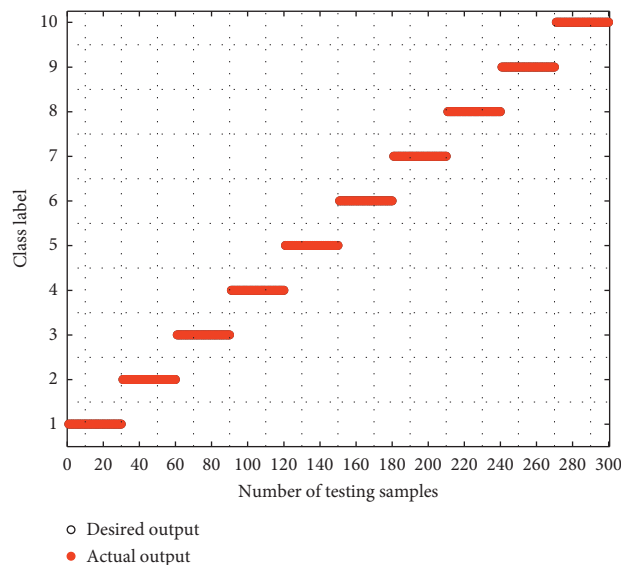


FIGURE 8: The fault identification results of the presented method.

TABLE 3: Parameters settings for different methods.

	HAAPE	MAAPE	IMAAPE	HPE	MPE	HSE	MSE
Embedding dimension	$m = 3$	$m = 3$	$m = 3$	$m = 3$	$m = 3$	$m = 2$	$m = 2$
Adjustment coefficient	$A = 0.5$	$A = 0.5$	$A = 0.5$				
Similar tolerance						$r = 0.15$	$r = 0.15$
Time delay	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$
Scale factor		$\tau_{\max} = 15$	$\tau_{\max} = 15$		$\tau_{\max} = 15$		$\tau_{\max} = 15$
Hierarchical layer	$k = 3$			$k = 3$		$k = 3$	
Number of sensitive feature	8	8	8	8	8	8	8

TABLE 4: The comparison of different feature extraction methods.

Different feature extraction methods	Time (s)	Recognition accuracy (%)	The number of misclassifications	Best parameters of SVM
HAAPE	0.058	100	0	$c = 264.38, g = 14.81$
MAAPE	0.043	97.67	7	$c = 762.61, g = 4.36$
IMAAPE	0.165	100	0	$c = 435.63, g = 0.09$
HPE	0.051	98	6	$c = 507.48, g = 313.94$
MPE	0.032	86.67	40	$c = 576.22, g = 226.14$
HSE	0.309	92.33	23	$c = 1.87, g = 0.79$
MSE	0.273	83	51	$c = 490.11, g = 0.06$

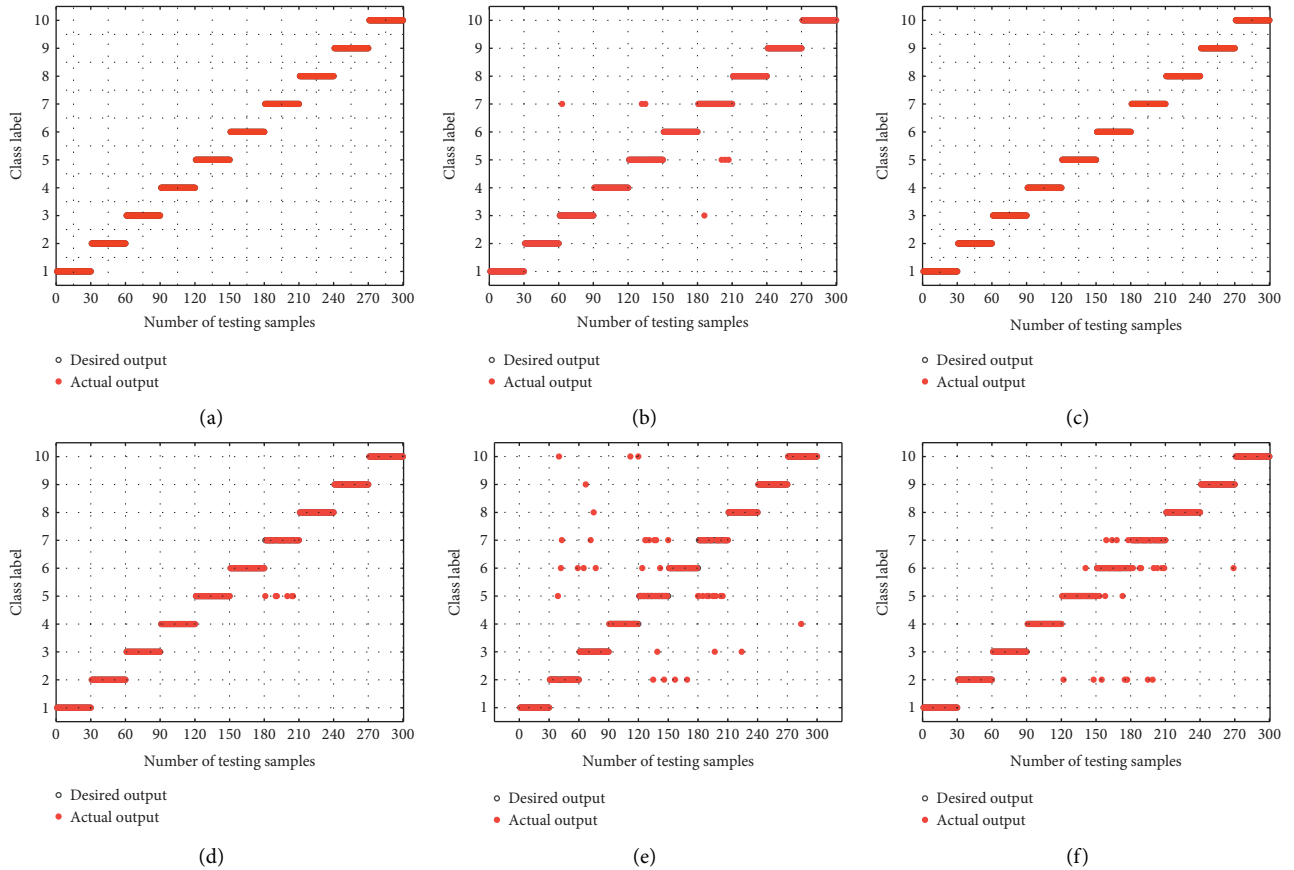


FIGURE 9: Continued.

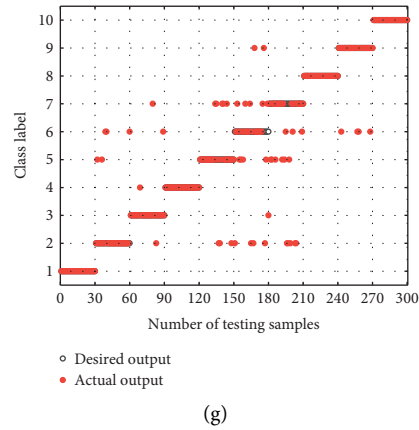


FIGURE 9: The recognition results corresponding to different feature extraction methods.

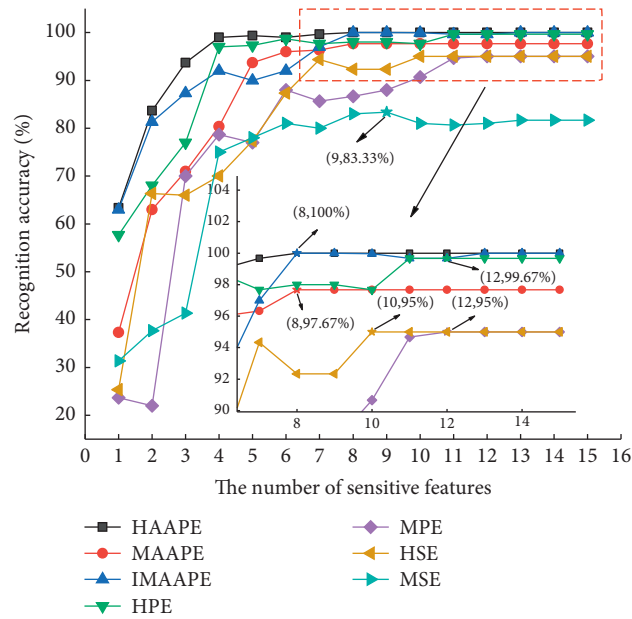


FIGURE 10: The relationship between the number of features and the recognition accuracy.

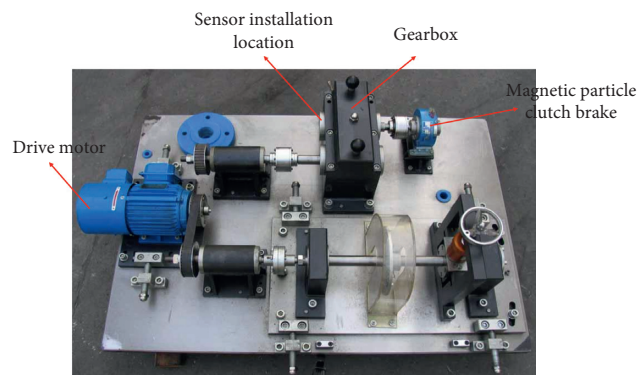


FIGURE 11: The experimental platform.

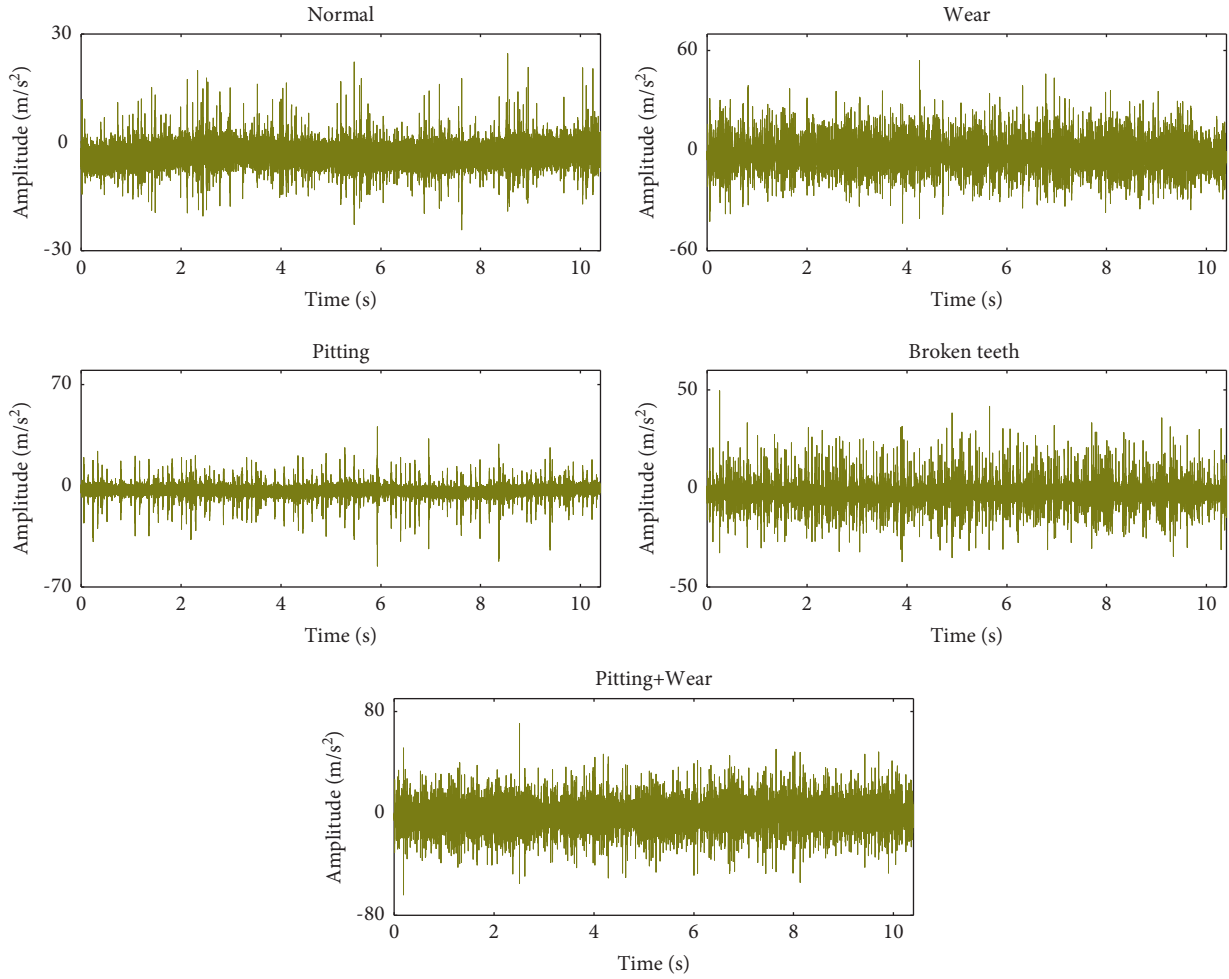


FIGURE 12: Time-domain waveforms corresponding to different states.

TABLE 5: Detailed information of different working states of gearbox.

Label	Gear state	Abbreviations	The number of training samples	The number of testing samples
1	Normal state	Normal	20	30
2	Pinion wear	Wear	20	30
3	Gear wheel pitting	Pitting	20	30
4	Gear wheel broken teeth	Broken teeth	20	30
5	Gear wheel pitting + pinion wear	Pitting + Wear	20	30

Similar to Case 1, to study the relationship between the number of features and recognition accuracy, after ranking the original features using PWF, the relationship curve between the number of features and recognition accuracy is drawn, as shown in Figure 15. It can be observed that the proposed method only needs seven features to achieve a recognition accuracy of 100%, which not only demonstrates

that this method is less dependent on the number of features and enjoys strong robustness but also proves the necessity of using PWF for feature selection. In general, Case 2 once again proves the effectiveness of the proposed method and its advantages over other methods, which indicates that our method has good universality and provides a new idea for the fault diagnosis of rotating machinery.

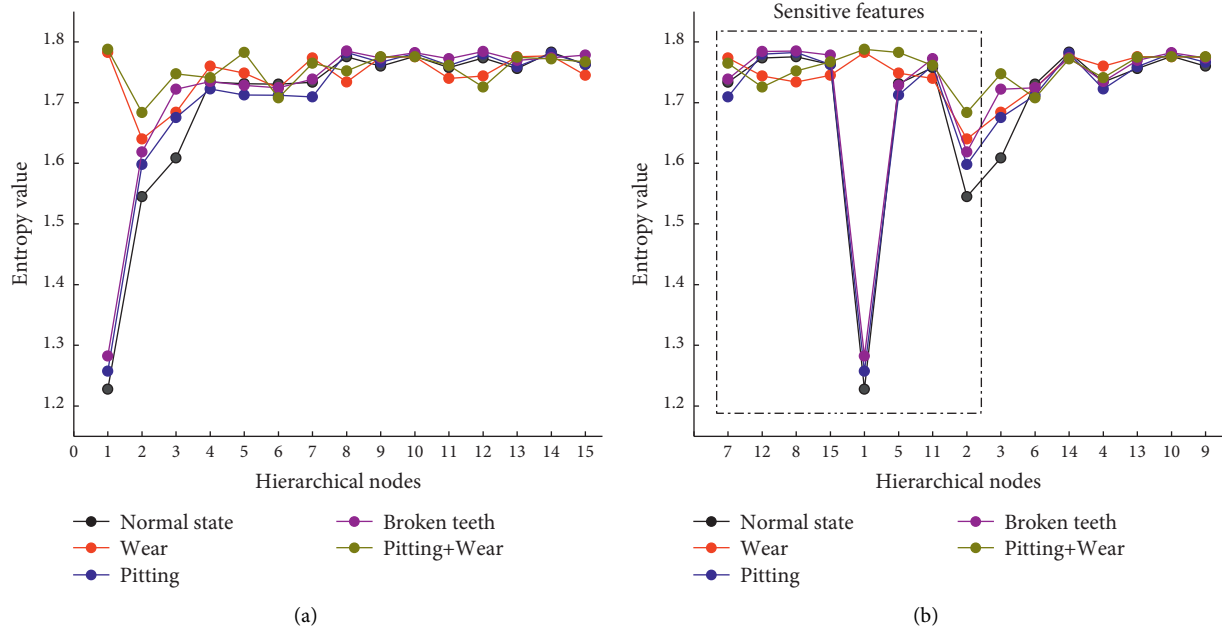


FIGURE 13: The original feature vectors and the sensitive feature vectors. (a) Original fault features; (b) fault features after ranking.

TABLE 6: The comparison of different feature extraction methods.

Different feature extraction methods	Time (s)	Recognition accuracy (%)	The number of misclassifications	Best parameters of SVM
HAAPE	0.056	100	0	$c = 361.09, g = 140.34$
MAAPE	0.044	96	6	$c = 20.84, g = 40.13$
IMAAPE	0.163	99.33	1	$c = 53.56, g = 36.08$
HPE	0.053	90.67	14	$c = 6.55, g = 154.45$
MPE	0.031	89.33	16	$c = 144.18, g = 17.01$
HSE	0.312	88	18	$c = 471.14, g = 5.18$
MSE	0.275	71.33	43	$c = 133.2, g = 0.026$

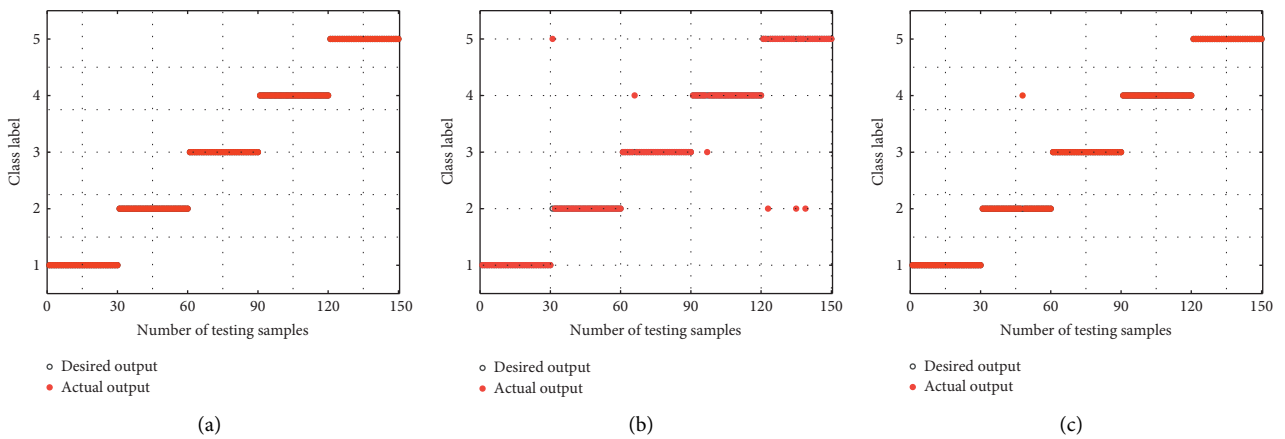


FIGURE 14: Continued.

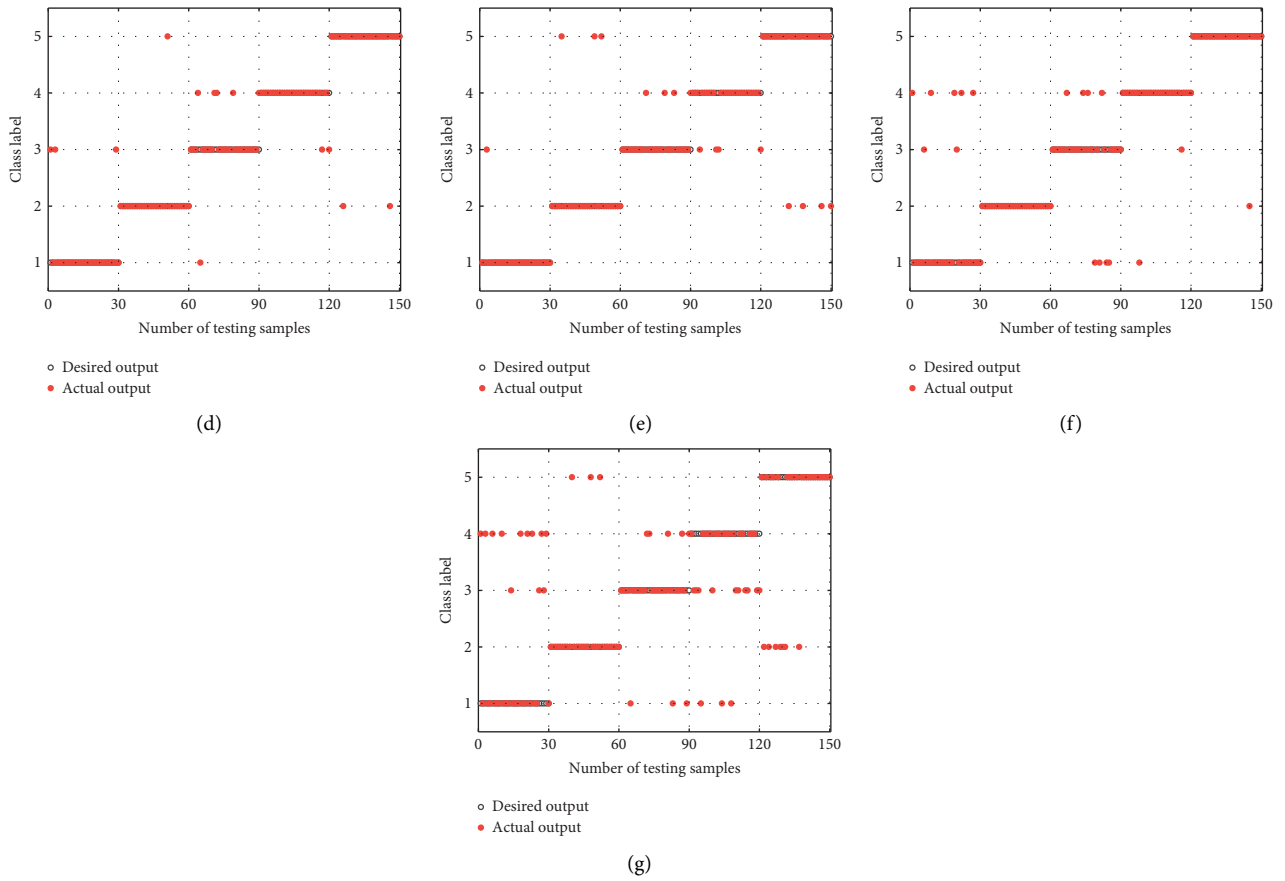


FIGURE 14: The recognition results corresponding to different feature extraction methods.

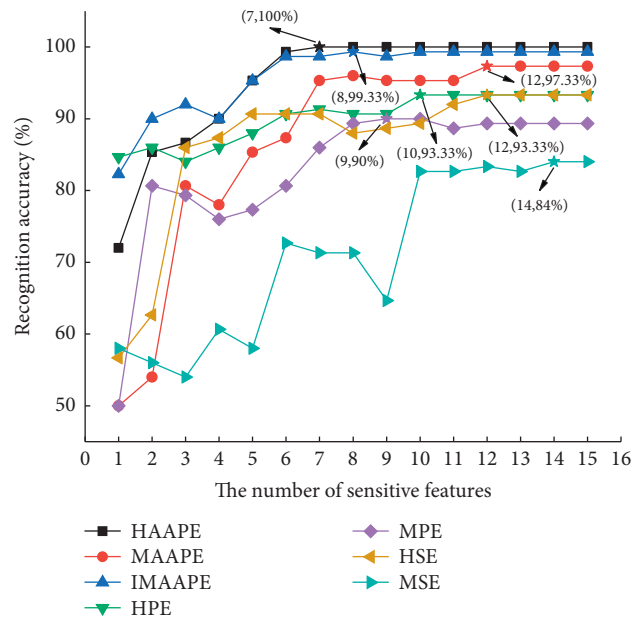


FIGURE 15: The relationship between the number of features and the recognition accuracy.



## 5. Conclusion

With a view to accurately identifying different fault states of rotating machinery, a new fault diagnosis method based on HAAPE, PWFP, and GWO-SVM is proposed in this paper, and the performance of the proposed method is verified using two fault data sets of bearing and gearbox. The main work of this paper can be summarized as follows:

- (1) Aiming at the shortcoming that multiscale amplitude-aware permutation entropy (MAAPE) can only extract low-frequency features of time series, but ignores the high-frequency features, this paper proposes hierarchical amplitude-aware permutation entropy (HAAPE). By constructing high-frequency and low-frequency operators, HAAPE can extract the high-frequency and low-frequency information of time series at the same time, which effectively overcomes the information loss problem in MAAPE.
- (2) Using HAAPE as feature extraction method, combining PWFP feature selection method and GWO-SVM classifier, a new fault diagnosis method for rotating machinery is proposed. First of all, the fault features of rotating machinery are extracted by HAAPE, then the original features are ranked by PWFP and sensitive features are screened to form sensitive feature vectors. Finally, the obtained sensitive feature vectors are input into GWO-SVM classifier for training and recognition. The performance of the proposed method is verified using two data sets of bearing and gearbox, and it is proved that the proposed method could efficiently and accurately identify different fault types of rotating machinery.
- (3) In order to compare the advantages of the proposed method over the existing methods, a series of comparative experiments are carried out. The results show that HAAPE is superior to MAAPE, IMAAPE, HPE, MPE, HSE, and MSE in feature extraction performance and robustness. In addition, we also study the relationship between the number of features and the recognition accuracy and prove the necessity of using the PWFP-based dimensionality reduction method.

## Data Availability

The rolling bearing data are provided by the Case Western Reserve University and can be downloaded in <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

Thanks are due to the Case Western Reserve University for supplying the bearing data.

## 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] D. Dou, J. Yang, J. Liu, and Y. Zhao, "A rule-based intelligent method for fault diagnosis of rotating machinery," *Knowledge-Based Systems*, vol. 36, pp. 1–8, 2012.
- [3] S. Riaz, H. Elahi, K. Javaid, and T. Shahzad, "Vibration feature extraction and analysis for fault diagnosis of rotating machinery-A literature survey," *Asia Pacific Journal of Multidisciplinary Research*, vol. 5, no. 1, pp. 103–110, 2017.
- [4] Z. Feng, M. Liang, Y. Zhang, and S. Hou, "Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation," *Renewable Energy*, vol. 47, pp. 112–126, 2012.
- [5] Y. Chen, T. Zhang, Z. Luo, and K. Sun, "A novel rolling bearing fault diagnosis and severity analysis method," *Applied Sciences*, vol. 9, no. 11, p. 2356, 2019.
- [6] J. S. Richman and M. J. Randall, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology - Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [7] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, 2007.
- [8] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [9] Hamed, J. Azami, and Escudero, "Amplitude-aware permutation entropy: illustration in spike detection and signal segmentation," *Computer Methods and Programs in Biomedicine*, vol. 128, pp. 40–51, 2016.
- [10] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, pp. 705–708, 2007.
- [11] Z. Chen, L. Zhao, and Sun, "fault diagnosis of rolling bearing using multiscale Amplitude-aware permutation entropy and random forest," *Algorithms*, vol. 12, no. 9, p. 184, 2019.
- [12] Y. Chen, T. Zhang, W. Zhao, Z. Luo, and H. Lin, "Rotating machinery fault diagnosis based on improved multiscale Amplitude-aware permutation entropy and multiclass relevance vector machine," *Sensors*, vol. 19, 20 pages, 2019.
- [13] Y. Jiang, C.-K. Peng, and Y. Xu, "Hierarchical entropy analysis for biological signals," *Journal of Computational and Applied Mathematics*, vol. 236, no. 5, pp. 728–742, 2011.
- [14] J. Li, K. Cheng, S. Wang et al., "Feature selection: a data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017.
- [15] S. Happy, R. Mohanty, and A. Routray, "An effective feature selection method based on pair-wise feature proximity for high dimensional low sample size data," in *Proceedings of the 2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 1574–1578, IEEE, Kos, Greece, September 2017.
- [16] W. S. Noble, "What is a support vector machine?" *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [17] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.
- [18] J. Zheng, J. Cheng, and Y. Yang, "Multiscale permutation entropy based rolling bearing fault diagnosis," *Shock and Vibration*, vol. 2014, no. 1, 8 pages, Article ID 154292, 2014.

- [19] E. Song, Y. Ke, C. Yao, Q. Dong, and L. Yang, "fault diagnosis method for high-pressure common rail injector based on IFOA-VMD and hierarchical dispersion entropy," *Entropy*, vol. 21, no. 10, p. 923, 2019.
- [20] K. Zhu, X. Song, and D. Xue, "A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm," *Measurement*, vol. 47, pp. 669–675, 2014.