Hindawi
Shock and Vibration
Volume 2020, Article ID 8851310, 18 pages
https://doi.org/10.1155/2020/8851310

Research Article

An Integrated Fault Identification Approach for Rolling Bearings Based on Dual-Tree Complex Wavelet Packet Transform and Generalized Composite Multiscale Amplitude-Aware Permutation Entropy

Wuqiang Liu, Xiaoqiang Yang, and Shen Jinxing

Field Engineering College, Army Engineering University of PLA, Nanjing 210007, Jiangsu, China

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

Received 13 July 2020; Revised 25 September 2020; Accepted 26 September 2020; Published 27 November 2020

Academic Editor: Chaoqun Duan

The health condition of rolling bearings, as a widely used part in rotating machineries, directly influences the working efficiency of the equipment. Consequently, timely detection and judgment of the current working status of the bearing is the key to improving productivity. This paper proposes an integrated fault identification technology for rolling bearings, which contains two parts: the fault predetection and the fault recognition. In the part of fault predetection, the threshold based on amplitude-aware permutation entropy (AAPE) is defined to judge whether the bearing currently has a fault. If there is a fault in the bearing, the fault feature is adequately extracted using the feature extraction method combined with dual-tree complex wavelet packet transform (DTCWPT) and generalized composite multiscale amplitude-aware permutation entropy (GCMAAPE). Firstly, the method decomposes the fault vibration signal into a set of subband components through the DTCWPT with good time-frequency decomposing capability. Secondly, the GCMAAPE values of each subband component are computed to generate the initial candidate feature. Next, a low-dimensional feature sample is established using the t-distributed stochastic neighbor embedding (t-SNE) with good nonlinear dimensionality reduction performance to choose sensitive features from the initial high-dimensional features. Afterwards, the featured specimen representing fault information is fed into the deep belief network (DBN) model to judge the fault type. In the end, the superiority of the proposed solution is verified by analyzing the collected experimental data. Detection and classification experiments indicate that the proposed solution can not only accurately detect whether there is a fault but also effectively determine the fault type of the bearing. Besides, this solution can judge the different faults more accurately compared with other ordinary methods.

1. Introduction

The working condition of bearings, as a vital part in rotating machinery, is closely related to the stable operation of equipment [1, 2]. Hence, real-time monitoring and prediction of the working status of the bearing is quite important to ensure safe production [3]. At present, there are a lot of mature and reliable methods to realize the fault diagnosis of bearing, such as vibration analysis, acoustic analysis [4, 5], oil analysis, and temperature analysis. Vibration signal is easy to collect and analyze, so it has been widely used and researched in the field of fault diagnosis. The fault diagnosis procedure of mechanical equipment based on vibration signal normally includes three steps: (1) collecting vibration data of equipment; (2) extracting the feature of a vibration signal to engender the initial feature; and (3) feeding the feature sample into the classifier for fault identification. Among them, the most important is the feature extraction, which is also the hotspot of current research. The quality of the extracted features directly affects the subsequent fault classification. The vibration signals of bearings are generally nonlinear [6]. Thus, it is the focus to explore the appropriate method to analyze the nonlinear vibration signal.

Due to the adverse factors such as friction, impact, and structural deformation in the working environment of bearing, the vibration signal is nonlinear and nonstationary. Therefore, how can reliable features be obtained from nonlinear data is the focus of research. With the further study of entropy-based theory [7], it becomes possible to process and analyze nonlinear data. For instance, Yan and Gao used approximate entropy (APE) for the first time in fault diagnosis to monitor the running status of bearings [8]. However, APE relies heavily on the length of data when processing signals with shorter data length, and the calculated entropy value may be less than the real. Richman and Moorman proposed sample entropy (SE) to settle the defect [9]. Unfortunately, SE may produce inaccurate estimations and undefined values. Afterwards, Bandt and Pompe presented permutation entropy (PE) [10], which measures the complexity by comparing the differences between adjacent data. Compared with other entropy-based methods, PE depends less on the model, and the calculation speed is fast and simple [11], whereas the amplitude information is neglected in the process of computing the PE. Consequently, the two time series would have with significantly different amplitudes but possibly with the same sort mode; meanwhile, the calculated PE value has an apparent error. To introduce amplitude information into the calculation process of permutation entropy, Azami and Escudero put forward amplitude-aware permutation entropy (AAPE) [12, 13] by introducing crucial information such as the amplitude and frequency of the signal into the calculation. Compared with PE, AAPE algorithm adds the deviation between amplitude and mean value of signal into the calculation process, contributing to further enhancing the stability and robustness of the algorithm.

However, both PE and AAPE are single-scale analysis methods. The actual vibration signal to be analyzed often contains information at multiple scales. The loss of a large amount of potentially useful information will be inevitably caused if only a single scale of analysis is conducted. Given the shortcomings of single-scale analysis, Costa et al. proposed multiscale entropy (MSE) [14], which can quantify the complexity of time series from multiple scales by dividing the original signal into multiple coarse-grained time series. Nevertheless, the multiscale approach adopted by MSE still has some defects: for example, the stability of the conventional multiscale computing method relies on the appropriate data length. Regarding short time series and larger-scale factors, a large entropy deviation will appear and cause the calculation result to be unreliable. Therefore, a composite multiscale method is employed in this paper to resolve the shortcomings of the traditional coarse-grained method. Meanwhile, the first-order moment (mean value) is expanded to the second-order moment (variance) during the process of coarse grained [15, 16]. Combined with AAPE, a generalized composite multiscale amplitude-aware permutation entropy (GCMAAPE) is proposed and utilized to subsequently extract the fault feature.

However, the direct usage of GCMAAPE to analyze the original signal cannot reveal the inherent characteristics such as the impact component contained in the vibration signal [17]. Thus, the entropy-based method is usually connected with the time-frequency processing method to reach a more comprehensive and detailed analysis for sake of highlighting the inherent characteristics of the vibration signal while extracting multiscale features [18]. Fourier transform (FT) is a commonly used signal analysis method while it cannot analyze the signal's time domain and frequency domain part simultaneously due to the uncertainty principle. The wavelet transform (WT) is a multiresolution analysis approach that can amplify the instantaneous changes in the signal through the window function. Nevertheless, WT is unable to analyze the high-frequency components of the signal. The wavelet packet transform (WPT) is an improved algorithm of the wavelet transform [19], which can process signals adequately and carefully, with a preferable time-frequency positioning capability. However, both WT and WPT have the same limitations, that is, the criteria for selecting the wavelet basis function cannot be well determined and it is difficult to set an appropriate decomposition level, resulting in restricting their further usage. Empirical mode decomposition (EMD) is an adaptive signal processing approach that can decompose complex signals into multiple intrinsic mode functions (IMFs). Each IMF includes features on different time scales of the raw signal [20] while EMD has serious disadvantage such as mode mixing and end effects. Besides, dual-tree complex wavelet transform (DTCWT) proposed by Kingsbury [21] possesses excellent performance such as nearly shift invariance and excellent directional selectivity while it cannot segment the high-frequency part of the signal. With the purpose of resolving the shortcoming of insufficient signal decomposition of DTCWT, dual-tree complex wavelet packet transform [22] (DTCWPT) is put forward. It can decompose the high-frequency part of the signal and solve the frequency aliasing effect of DTCWT.

Generally, the feature sample acquired through the time-frequency multiscale approach is high-dimensional and redundant, containing plenty of features that have no concern with the fault information. If directly used for classification, it not only reduces the classification efficiency but also seriously affects the identification accuracy. Consequently, it is necessary to select the features with a strong correlation to acquire sensitive low-dimensional fault features. T-distributed stochastic neighbor embedding (t-SNE) is a manifold dimensionality reduction algorithm with high nonlinear dimensionality reduction property [23]. Based on the probability distribution of random walk on the neighborhood graph, the structural relationship can be discovered in the raw data. Thus, this paper adopts t-SNE to reduce the dimensionality of the features to make up the final feature. After the final feature sample is obtained, it needs to be identified to determine the fault state. In terms of accurately estimating the current condition of the bearing, the accuracy of recognition is the first task needing to be considered. BP neural network is liable to get trapped in the local optimal value, and the convergence speed is slow. The approximation and generalization of the model are too dependent on the typicality of the selected sample. In the artificial neural network (ANN) [24], a large number of parameters need to

be set, such as weight values and initial thresholds; besides, the learning time is too long and even caught in a loop without learning purpose. Support vector machine (SVM) is widely utilized owing to its excellent generalization ability and the advantage of processing small samples. SVM has parameter optimization problems that the classification performance would be severely affected by the selection of penalty coefficient and kernel function parameters [25]. With the deepening of artificial intelligence research, deep learning has been extensively studied. Deep network is a neural network simulating human brain processing information and has multiple hidden layers and multiple perceptrons. The research achievement has been successfully applied to the fields such as image recognition, speech processing, and text processing. However, these applications are all aimed at big data, and the application in small sample recognition is deficient. Deep belief network (DBN) [26–30] is a typical structure of the deep network, which is composed of multilayer restricted Boltzmann machines. The difficulty of parameter selection can be effectively avoided by adopting pretraining and fine-tuning training procedures [31]. Meanwhile, it can be trained with only a few samples, exhibiting obvious advantages in small sample recognition.

In summary, the focus of this paper is to propose an integrated fault diagnosis method based on DTCWPT, GCMAAPE, t-SNE, and DBN. The four tools (DTCWPT, GCMAAPE, t-SNE, and DBN) are employed to implement its four main targets (fault predetection, signal preprocessing, fault feature extraction, and fault pattern recognition), respectively. The main contributions and innovations of this paper can be summarized as follows:

(1) A fault diagnosis method integrating fault predetection and fault identification is presented. Different from most single step fault diagnosis methods, the proposed stepwise fault diagnosis strategy realizes the non-disassembly health detection of rolling bearing and avoids the secondary damage caused by the uncertainty of the subsequent pattern recognition, making it more consistent with the practical engineering applications.

(2) After a fault is detected in the rolling bearing, DTCWPT is used to process the vibration signal of the fault bearing, eliminating the noise and highlighting the vibration characteristics.

(3) GCMAAPE, a new nonlinear dynamic method, was used to extract fault characteristics at multiple scales from bearing vibration signals, not only overcoming shortcomings of MAAPE in the description of signal complexity but also better extracting fault characteristics by adopting generalized coarse-grained methods.

(4) DBN is introduced to automatically identify different fault types and severity and exhibits excellent generalization performance and classification efficiency compared to RF and SVM.

(5) The experimental data are used to verify the proposed method by comparing it with other methods. The effectiveness and operability of the proposed method are demonstrated by the experimental result.

2. Theory of Experimental Methods

2.1. Dual-Tree Complex Wavelet Packet Transform (DTCWPT). The DTCWPT is a modified algorithm based on the theory of DTCWT by scholars Bayram and Selesnick, overcoming the shortcomings of the DTCWT algorithm that cannot decompose the high-frequency component of signal. DTCWPT is an extension of the traditional wavelet packet transform and adopts two parallel and independent discrete wavelet packets of the low-pass filter and high-pass filter to implement signal decomposition and reconstruction. The two discrete wavelet packets are called DTCWPT's real tree and virtual tree. During the signal decomposition and reconstruction, the delay interval between the real tree and the virtual tree filter is exactly a sample value, and the sampling point of the virtual tree is kept exactly in the middle of the real tree to form the complementarity of the information, contributing to obtaining a nearly shift invariance and reduced loss of information. Besides, DTCWPT utilizes two parallel and independent discrete wavelet packets to decompose the low-frequency and high-frequency parts, exhibiting extremely high resolution while also effectively suppressing the frequency aliasing phenomenon. The decomposition and reconstruction of DTCWPT [32] are presented in Figure 1.

$S(t)$ is the input original signal; $\widehat{S}(t)$ is the reconstructed signal; $f_{1\text{-}1}$ is the high-frequency filter of the first layer decomposition of the real tree; $f_{1\text{-}0}$ is the low-frequency filter of the first layer decomposition of the real tree;$f_{2\text{-}1}$ is the high-frequency filter of the first layer of the virtual tree; $f_{2\text{-}0}$ is the low-frequency filter of the first layer of the virtual tree; $a_R(1,2)$, $a_R(1,1)$ are the decomposed components of the first layer of the real tree; $a_{\mathrm{Im}}(1,2)$, $a_{\mathrm{Im}}(1,1)$ are the decomposed components of the first layer of the virtual tree; h_1,h_0 are the filters for real tree decomposition after the second layer, g_1, g_0 are filters for virtual tree decomposition after the second layer; $a_R(2,4),\ldots,a_R(2,1)$ are the components of the second-level decomposition of the real tree; $a_{\mathrm{Im}}(2,4),\ldots,a_{\mathrm{Im}}(2,1)$ are the components of the second-level decomposition of the virtual tree; h_1', h_0' are filters for real tree reconstruction outside the second layer; g_1', g_0' are filters for virtual tree reconstruction other than the second layer; $f'_{1\text{-}1}$ is the reconstructed high-frequency filter of the first layer of the real tree; $f'_{1\text{-}0}$ is the low-frequency filter reconstructed from the first layer of the real tree; $f'_{2\text{-}1}$ is the high-frequency filter reconstructed from the first layer of the virtual tree; and $f'_{2\text{-}0}$ is the low-frequency filter reconstructed from the first layer of the virtual tree; $2\downarrow$ besides, $2\downarrow$ denotes interval sampling, and $2\uparrow$ indicates incremental sampling.

2.2. AAPE and GCMAAPE

2.2.1. AAPE. PE was proposed by Bandt in 2002 to analyze the complexity of time series, revealing that the more complex the signal being analyzed, the larger the

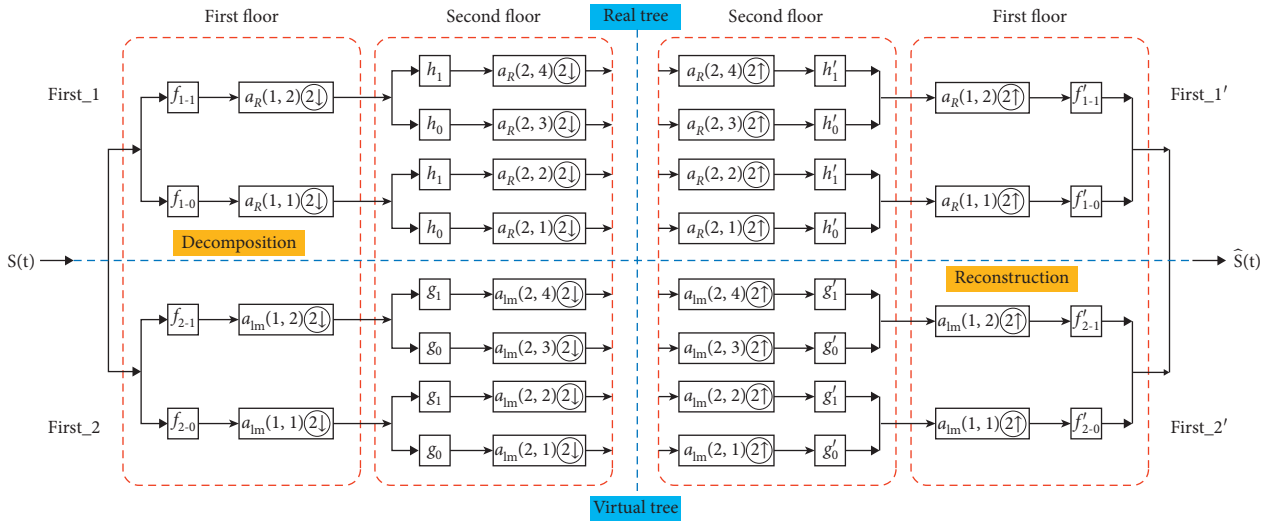

Figure 1: The decomposition and reconstruction process of DTCWPT.

permutation entropy value. For instance, the permutation entropy of white noise is greater than the permutation entropy of the cosine signal. The realization regulation of PE is described as follows [10].

Assuming a time series $x = \{x_1, x_2, \ldots, x_N\}$ of length N, a m-dimensional vector at any time t is constructed.

$$x_t^m = \{x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)\}, \quad t = 1, 2, \ldots, N-(m-1)\tau, \tag{1}$$

where m and τ denote embedding dimension and time delay, respectively. Define the permutation $\pi_j = (r_1, r_2, \ldots, r_{m-1})$ of $\{0, 1, \ldots, m-1\}$ when fulfils

$$x(t+r_o\tau) \le x(t+r_1\tau) \le \cdots \le x(t+r_{m-1}\tau). \tag{2}$$

When formula (2) holds, x_t^m has a permutation of π_j, where $0 \le r_n \le m-1$; $r_{n-1} < r_n$ holds when $x(t+r_{n-1}\tau) = x(t+r_n\tau)$.

For each permutation π_j, the relative frequency of $1 \le j \le m!$ can be calculated as

$$p(\pi_j) = \frac{\#\{t | t \le N-(m-1)\tau, x_t^m \text{ has type } \pi_j\}}{N-(m-1)\tau}, \tag{3}$$

where # represents the number of x_t^m belonging to type π_j. The PE of time series can be defined as

$$H_{\mathrm{PE}}(x, m, \tau) = -\sum_i^{m!} p(\pi_j) \ln p(\pi_j). \tag{4}$$

According to the principles introduced above, PE ignores the diversity of amplitude in the uniform sort mode and may lose the signal's amplitude information. Regarding the vibration signal acquired from the rolling bearings, the amplitude includes a great deal of information related to the working state, which is the most important feature that represents the current operating condition; thus it cannot be ignored. For instance, the sequences {1, 2, 3, 4, 5, 6} and {1, 2, 3, 4, 5, 96} are the same when mapping while the mapping differences between 5-6 and 5–96 are very large. A case where different types of time series are mapped to the same sort mode by a mapping function under the embedding dimension $m = 4$ is illustrated in Figure 2. It can be observed from the figure that the distances between the four points of diverse types of time series are not equal, suggesting that the amplitudes of the vibration signals are not consistent. Nevertheless, the sort mode (1 and 2) should be the same according to the principle of PE [33, 34].

Given the problems of PE, AAPE is proposed to increase the impact of key information such as amplitude and frequency on PE calculations to enhance the stability and robustness of PE. The calculation flowchart of AAPE algorithm [17] is presented in Figure 3.

Assuming that the initial value of $p(\pi_j^{m,\tau})$ is 0, the probability of its occurrence $p(\pi_j^{m,\tau})$ for the time series $X_t^{m,\tau}$ should be recalculated whenever $\pi_j^{m\tau}$ appears when t gradually increases from 1 to $N-m+1$.

$$p(\pi_j^{m,\tau}) = p(\pi_j^{m,\tau}) + \left(\frac{A}{m} \sum_{k=1}^{m} \left| x_{t+(k-1)\tau} \right| + \frac{1-A}{m-1} \sum_{k=2}^{m} \left| x_{t+(k-1)\tau} - x_{t+(k-2)\tau} \right| \right), \tag{5}$$

where $A \in [0, 1]$ denotes the adjustment coefficient to adjust the weight of the signal amplitude mean and the deviation between the amplitudes. Thus, the probability of $p(\pi_j^{m,\tau})$ appearing in the entire time series is $\pi_j^{m,\tau}$.

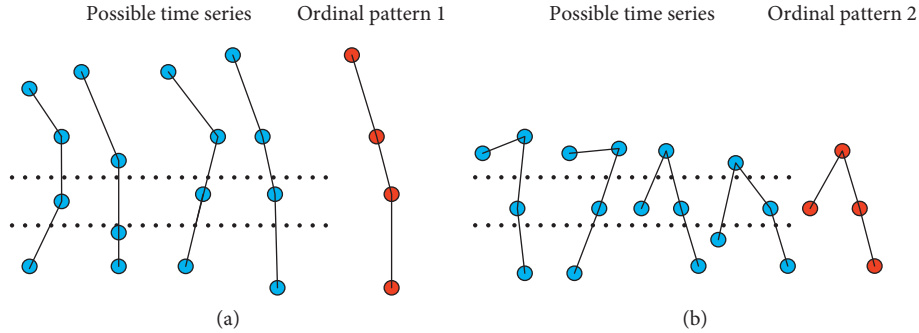

Figure 2: Two examples of diverse time series with the same sort mode.

$$p\left(\pi_j^{m,\tau}\right) = \frac{p\left(\pi_j^{m,\tau}\right)}{\sum_{t=1}^{N-m+1}\left((A/m)\sum_{k=1}^{m}\left|x_{t+(k-1)\tau}\right| + ((1-A)/(m-1))\sum_{k=2}^{m}\left|x_{t+(k-1)\tau} - x_{t+(k-2)\tau}\right|\right)}. \tag{6}$$

The AAPE value is computed as

$$\mathrm{AAPE}(m,\tau,n) = -\sum_{\pi_k=1}^{\pi_k=m!} p(\pi_k)\ln p(\pi_k). \tag{7}$$

2.2.2. GCMAAPE. The fault information contained in the vibration signals of rolling bearings usually appears on multiple scales, and a large amount of fault information will be lost if only single-scale analysis is performed. Thus, it is necessary to perform multiscale analysis to adequately extract the fault feature, and a multiscale AAPE (MAAPE) is accordingly put forward. Nevertheless, the coarse-grained approach adopted by MAAPE has the following defects. The coarse-graining process divides a time series into equal-length nonoverlapping segments and calculates the average of all data points in each segment. Therefore, the sequence of different scales of the original signal obtained by using only a single feature of the data mean will inevitably cause the loss of many potentially useful information. Consequently, a generalized composite multiscale method to address the defects of MAAPE is adopted to resolve the deficiencies of the traditional coarse-grained method. Specific steps are described as follows [16]:

(1) For the time series $x = \{x_1, x_2, \ldots, x_N\}$, the defined generalized coarse-grained time series $y_k^{(s)} = \left\{y_{k,j_1}^{(s)}, y_{k,j_2}^{(s)}, \ldots, y_{k,j_s}^{(s)}\right\}$ for scale factor s can be calculated as follows.

$$y_{k,j}^{(s)} = \frac{1}{s}\sum_{i=(j-1)s+k}^{js+k-1}\left(x_i - \overline{x_i}\right)^2,$$
$$1 \le j \le \frac{N}{S}, 2 \le k \le s, \ \overline{x_i} = \frac{1}{s}\sum_{k=0}^{s-1} x_{i+k}. \tag{8}$$

(2) For the scale factor s, the AAPE values of s generalized coarse-grained time series $y_k^{(s)}$ $(k = 1, 2, \ldots, s)$ are computed.

(3) The average of the s PE values is taken as the GCMAAPE value of the raw time series at the scale factor s, calculated as follows:

$$\mathrm{GCMAAPE}(x,m,A,\tau,s) = \frac{1}{s}\sum_{k=1}^{s}\mathrm{AAPE}\left(y_k^{(s)},m,A,\tau\right). \tag{9}$$

The PE value obtained by formula (9) is drawn as a function of the scale factor, called generalized composite multiscale amplitude-aware permutation entropy analysis. GCMAAPE not only synthesizes the information of multiple coarse-grained time series at the same scale but also generalizes the first-order moment (mean) to the second-order moment (variance). Theoretically, the performance of GCMAAPE is better than that of MAAPE method. Different from AAPE with single-scale analysis, GCMAAPE and MAAPE analyze time series from multiple scales. If GCMAAPE value of one time series is larger than another at most scales, indicating that the former is more random than the latter and has a higher probability of dynamic mutation.

2.2.3. The Parameter Choice Analysis for GCMAAPE. In the GCMAAPE, four vital parameters are required to be selected beforehand: embedding dimension m, adjustment coefficient A, time delay t, and scale factor s. Specifically, if the value of embedding dimension m is too small, the reconstructed vector includes too few states, and the algorithm loses its significance and effectiveness. However, the phase space reconstruction will homogenize the time series when m is too large; this not only spends much time to calculate but also fails to reflect the subtle transformation in the time series. Therefore, the embedding dimension is generally 3–7.

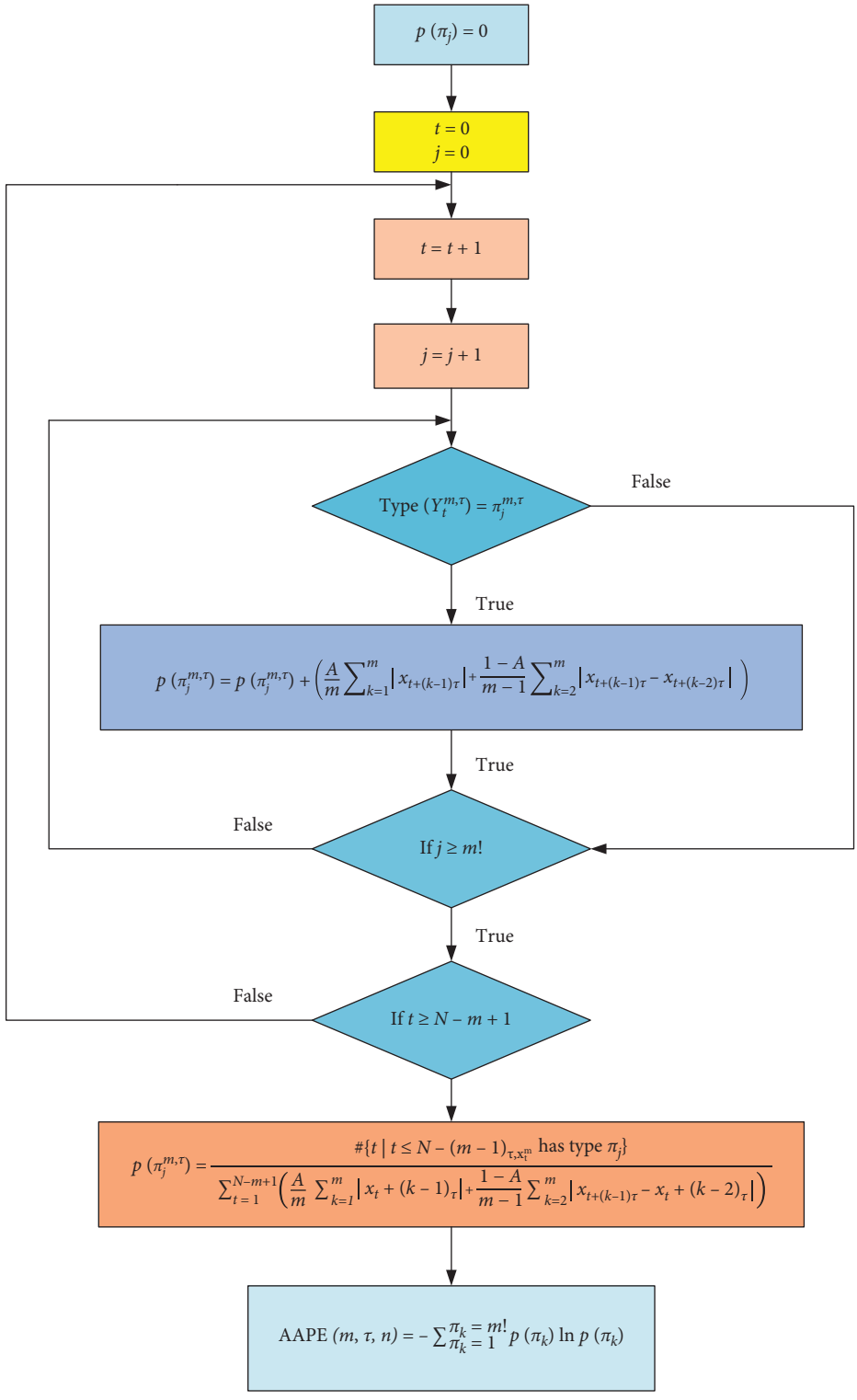

Figure 3: The flowchart of the AAPE algorithm.

Besides, the adjustment factor A is set to 0.5 according to the literature [12]. The time delay t has a small effect on the performance of GCMAAPE and is normally set to 1. Moreover, scale factor s is generally set to be larger than 10; there are no specific selection criteria while too small scale factor will lead to insufficient feature extraction and make it hard to effectively quantify the fault feature. However, too large scale factor will cause a large increase in the amount of calculation, as well as redundancy of features.

The vibration signals of rolling bearing under normal condition and the slight inner race fault condition were analyzed to understand the effect of the embedding

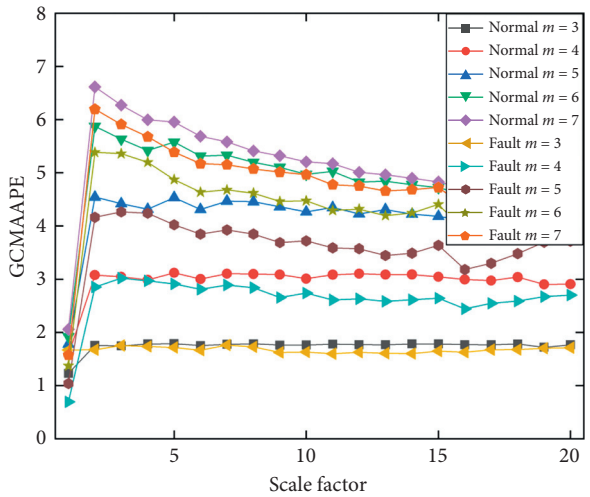

Figure 4: Effect of embedding dimension on the performance of GCMAAPE.

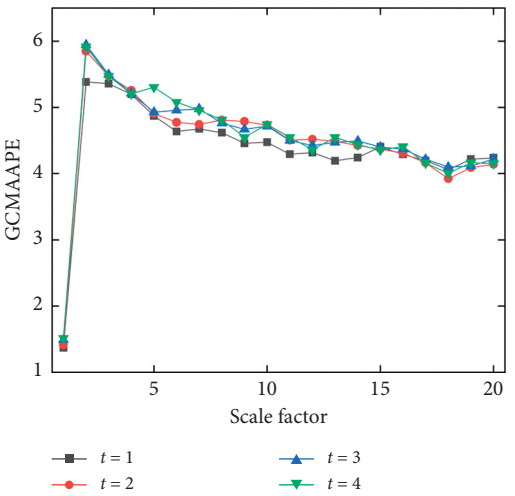

Figure 5: Effect of time delay on the performance of GCMAAPE.

dimension on GCMAAPE; the sampling points were 2400. The effect of embedding dimension ($m = 3, 4, 5, 6, 7$) on GCMAAPE performance when the time delay $t = 1$ and adjustment coefficient $A = 0.5$ is presented in Figure 4. It can be observed that the performance of GCMAAPE is the best when the embedding dimension $m = 6$. At this time, the entropy value of each scale factor has the largest difference; besides, the fault and normal states can be clearly distinguished. Consequently, the embedding dimension m is set to 6.

Secondly, the time delay $t = 1, 2, 3, 4$ when $m = 6$ is selected to test the effect of time delay t on the performance stability of GCMAAPE. The test results of the normal state vibration signal of the bearing under different time delay are illustrated in Figure 5. The different curves almost overlap together, and the difference in entropy value is very small. This indicates that the time delay t has little effect on GCMAAPE. Thus, t is set to 1 in this research according to the recommendations [13].

Finally, depending on the above analysis and the suggestions of the literature [13], the parameter settings are set as $m = 6$, $t = 1$, $A = 0.5$, and $s = 10$ in the subsequent experiments of this article.

2.3. Deep Belief Network (DBN). The deep belief network (DBN) is a probabilistic generation model consisting of multiple layers of restricted Boltzmann machines (RBMs), each of which can be simply considered independent. The layers are connected to each other, and each layer is an abstract representation of the visual layer data. A DBN network model that can be used is obtained through pre-training and fine-tuning. The detailed training procedure is divided into two steps [35].

Step 1: each layer of RBM is trained separately to allow each layer to contain as many features of the input data as possible. Specifically, the input vector is first mapped to the output through the weight. Then, the output is obtained, and the input vector is reconstructed in turn. The reconstructed deviation is used as the basis for updating the weight. This process is repeated until the deviation between the input vector and the output vector is tiny. The procedure of forward and backward is the learning process of RBM.

Step 2: in the former process, each RBM network only optimizes the mapping relationship between the input and output of its own layer, while it does not make the entire network structure reach the optimal. Therefore, it is necessary to establish a softmax classifier in the ultimate layer of DBN; then, the output feature vector of RBM is taken as the input feature vector of the softmax classifier to train the softmax classifier with supervision; next, the error between output and input is propagated to RBM of each layer from top to bottom; finally, all parameters of the network can be finely turned. The structure of the DBN classifier is presented in Figure 6.

In the learning process of DBN, the training of RBM is the core. The network parameters are initialized using the layer-by-layer learning of RBM. Although the resulting initialized network parameters are not optimal parameters, they are generally in the vicinity of the optimal parameters, avoiding the BP algorithm to easily fall into the local optimal and training time and other defects caused by the random initialization of network parameters during training DBN.

3. The Proposed Approach

This paper proposes an integrated fault identification technology for rolling bearings based on the advantages of DTCWPT, GCMAAPE, and t-SNE in rolling bearing fault feature extraction and combining DBN models that can handle high-dimensional data classification problems, this paper proposes an integrated fault identification technology for rolling bearings. The integrated technology includes the following two parts.

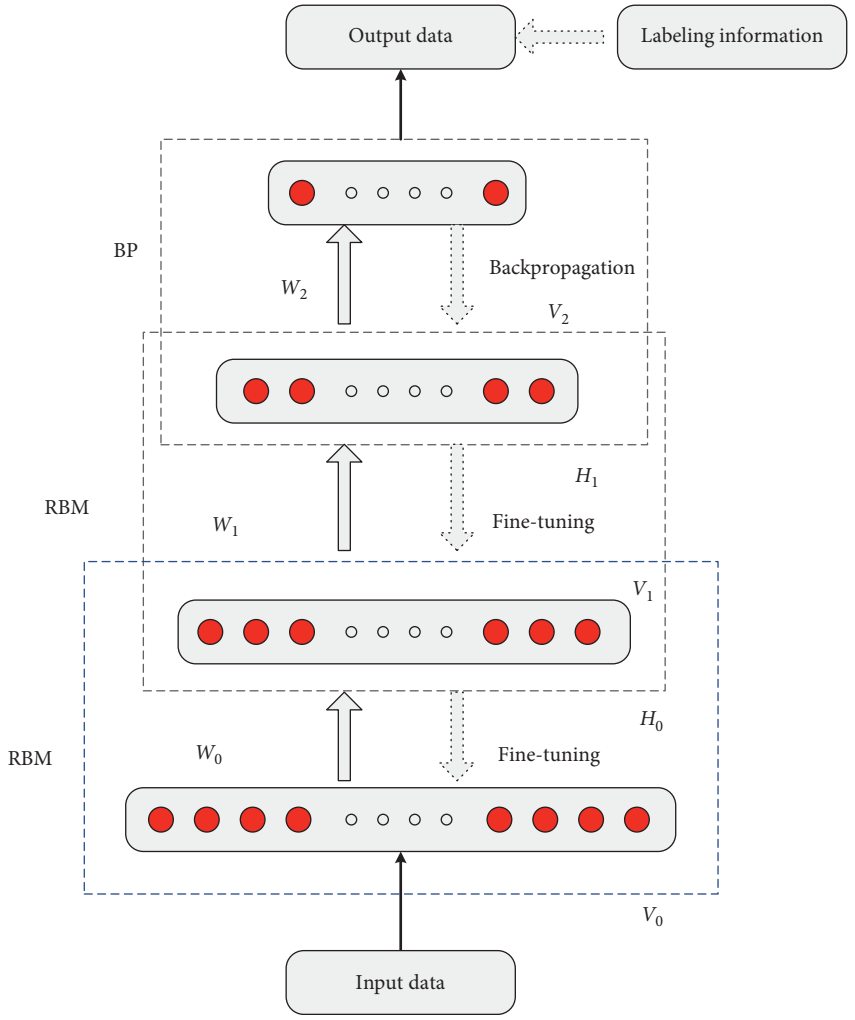

Figure 6: Structure diagram of double hidden layer DBN network.

3.1. Fault Predetection. As an improvement on the permutation entropy, AAPE has similar functions to PE. AAPE can detect the failure of equipment, indicating the ability to detect the failure. The sensitivity of AAPE to normal and the fault was used to screen the normal and fault of bearing.

When the scale factor is 1, the GCMAAPE value of the normal vibration signal is smaller than that of the fault vibration signal, with an obvious difference value. Therefore, the GCMAAPE value when the scale factor is 1 is used to differentiate the normal and fault states. Therefore, a threshold value is designed to detect the current health condition of the bearing to screen more intuitively.

3.2. Fault Classification. After the predetection, it is necessary to make further analysis to judge the bearing fault type and severity if the bearing fault is detected. A novel time-frequency multiscale feature extraction approach based on DTCWPT, GCMAAPE, and t-SNE was proposed. There are two traditional multiscale feature extraction methods: (1) the nonlinear time-frequency algorithm is adopted to decompose the signal into multiple components, and then the single entropy of multiple components is acquired; (2) the MSE of a single component is computed. Compared with the traditional method, the new time-frequency multiscale feature extraction method avoids the problem of insufficient feature extraction by extracting the multiscale entropy of multiple components to highlight the impact and resonant components in the fault vibration signal. The basic principle is to decompose the fault vibration signal into several components of different frequency bands and then use GCMAAPE to extract the fault characteristics of each component. Next, t-SNE is utilized to select sensitive features to acquire low-dimensional final feature vector. Finally, the DBN classification model is trained and tested with the final feature vector to classify different fault states. The technical route of the presented means is illustrated in Figure 7. The process of implementing the integrated fault diagnosis method includes the following six steps.

Step 1. Vibration data acquisition: collect vibration signals of running rolling bearings under different work conditions with sensors. Divide the collected experimental data into multiple samples of length N, which have no overlap between the sequences.

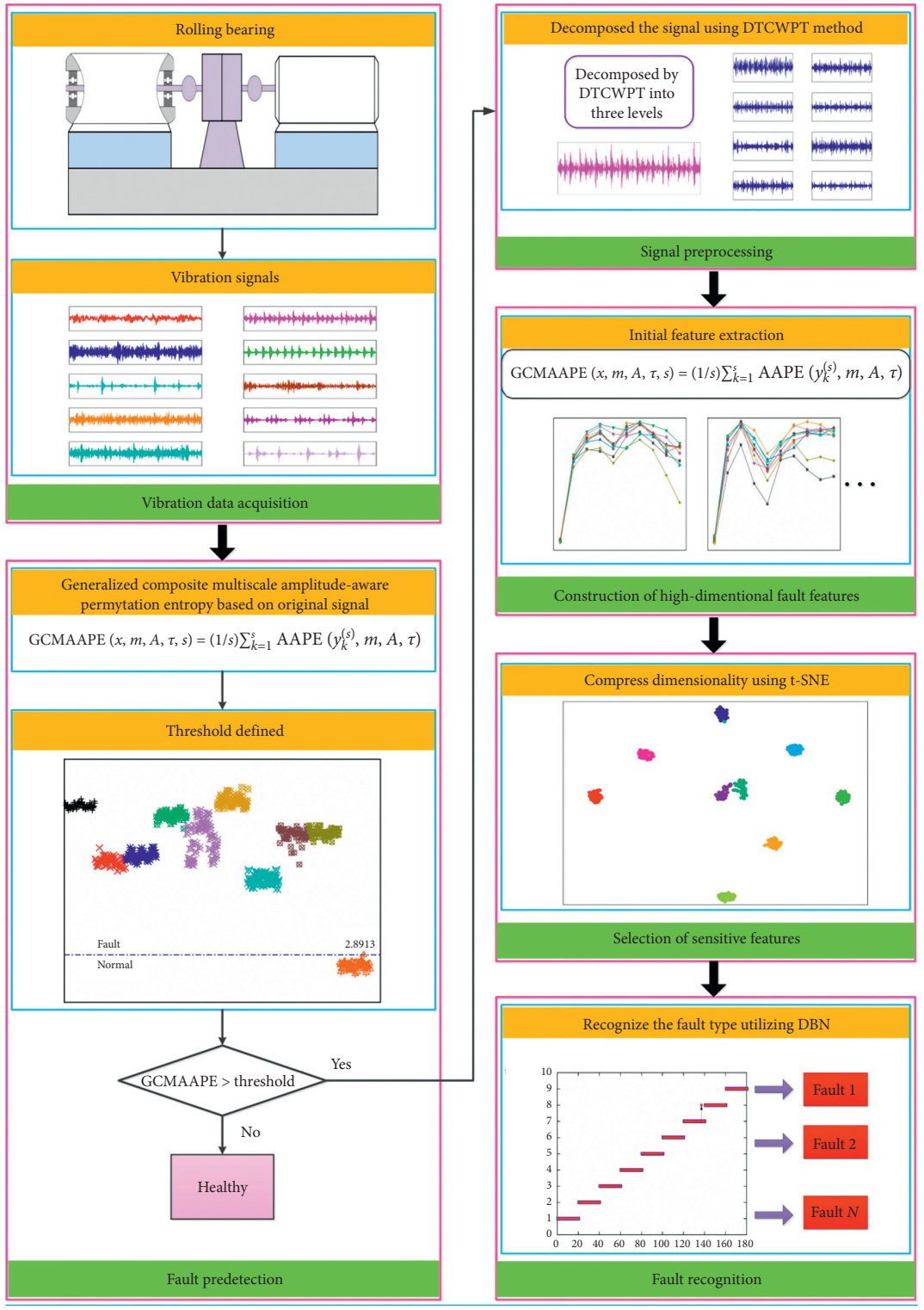

Figure 7: The technical route of the presented approach.

Step 2. Fault predetection: calculate the GCMAAPE value when the scale factor is 1 and set a threshold value based on the GCMAAPE value to judge whether the bearing is healthy. If the GCMAAPE value of the bearing vibration signal to be detected is less than the threshold value, the bearing is healthy, the output is normal, and the diagnosis ends. Otherwise, proceed to the next step to judge the type and severity of the bearing failure.

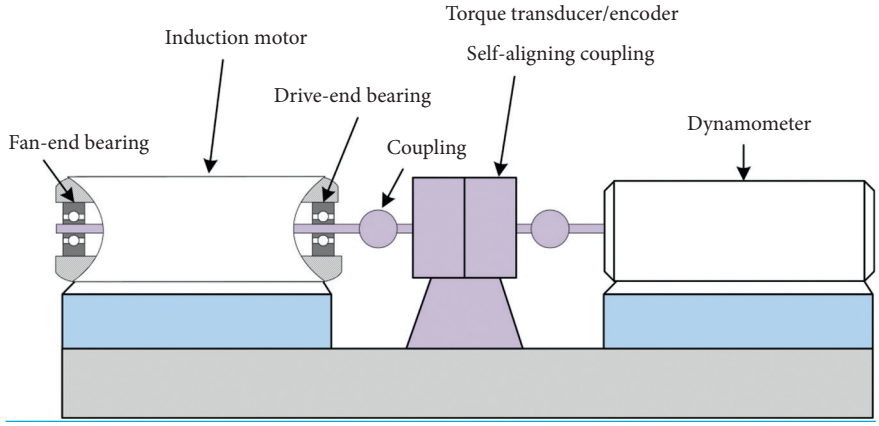

Figure 8: Bearing failure simulation test bench.

Step 3. Signal preprocessing: decompose the fault vibration signal with DTCWPT to obtain several subbands of different frequency bands.

Step 4. Construction of high-dimensional fault features: calculate the GCMAAPE value of each subband component to form the initial candidate feature vector set and perform normalization processing as input for the next step.

Step 5. Selection of sensitive features: choose sensitive features from the normalized initial features using the t-SNE to construct a final feature vector.

Step 6. Fault recognition: divide the normalized feature vector into a training sample and a test sample and establish an optimal DBN classification model through the training and test.

4. Experimental Process and Results

4.1. Experimental Data. The bearing vibration signal is adopted to verify the performance of the proposed DTCWPT-GCMAAPE-t-SNE-DBN model. The rolling bearing vibration signal data used in the experiment are collected from the rolling bearing failure simulation test bench of the Electrical Engineering Laboratory of Case Western Reserve University in the United States [36]. The rolling bearing model used for the test is 6205-2RS-SKF deep groove ball bearings. A bearing failure simulation test bench is exhibited in Figure 8.

The bearing data used include vibration data of the drive end bearing under 10 operating conditions, which are normal working conditions (labeled NM), inner race fault conditions (labeled IRF1, IRF2, and IRF3), outer race fault conditions (labeled ORF1, ORF2, and ORF3), and ball fault conditions (labeled BF1, BF2, and BF3). The fault diameters of the three fault types are 0.1778 mm, 0.3556 mm, and 0.5334 mm. The different fault diameters indicate the severity of bearing damage. In this test, the sampling frequency is 12 kHz, the rotating speed of the motor is 1797 rpm, and the load is 0 HP. The details of the data used in the experiment are listed in Table 1. Each group of signals is divided into multiple groups of nonoverlapping samples. Since each sample consists of 2400 sampling points, each state contains 50 samples. Therefore, the experimental data used are composed of 10 working conditions, each of which contains 50 sets of samples. Among them, the samples of each state use 30 groups as the training set for the DBN classification model and the remaining 20 groups are used as the testing set.

4.2. Results and Analysis. The time-domain waveforms of vibration signals of bearing with different fault types and severity are illustrated in Figure 9. The waveform of the vibration signal lacks regularity, making it difficult to determine the working condition of the bearing directly from the time-domain waveform. Therefore, further measures need to be taken to determine the working state of the bearing. Similar to PE, AAPE can detect the state of the bearing to avoid secondary damage to the bearing, according to the previous theoretical analysis. The AAPE value of all samples is presented in Figure 10. It can be observed that the AAPE value of the bearing in the fault state is generally large, and the AAPE value of the bearing in the normal state is small; this is significantly different from the AAPE value of the fault state. Therefore, this method can be used to detect the normal state of the bearing. The value at the red dotted line is defined as the AAPE threshold (2.8913). Besides, the normal and fault states can be clearly distinguished by comparing the AAPE value of the vibration signal with the threshold. Moreover, the two indicators of detection accuracy (DTA) and missed detection rate (MDR) are used to make it more intuitive to evaluate the efficiency of the approach in the predetection. As demonstrated in Figure 10, the entropy values of all fault samples are distributed above the threshold, and all normal samples are distributed below the threshold. According to mathematical statistical analysis, the indicator DTA has reached 100%, and the MDR is 0%. Depending on the definition, the larger the value of DTA, the higher the accuracy of detecting normal samples, and the smaller the MDR, the lower the probability of misdiagnosis. In summary, the larger the DTA and the smaller the MDR, the more effective the method. Furthermore, it can be verified

Table 1: The brief description of experimental data.

Health state	Fault diameter (mm)	Fault severity	Abbreviation	Training sample number	Testing sample number	Classification label
Normal	None		NM	30	20	0
Inner race fault	0.1778	Minor	IRF1	30	20	1
	0.3556	Medium	IRF2	30	20	2
	0.5334	Severe	IRF3	30	20	3
Outer race fault	0.1778	Minor	ORF1	30	20	4
	0.3556	Medium	ORF2	30	20	5
	0.5334	Severe	ORF3	30	20	6
Ball fault	0.1778	Minor	BF1	30	20	7
	0.3556	Medium	BF2	30	20	8
	0.5334	Severe	BF3	30	20	9

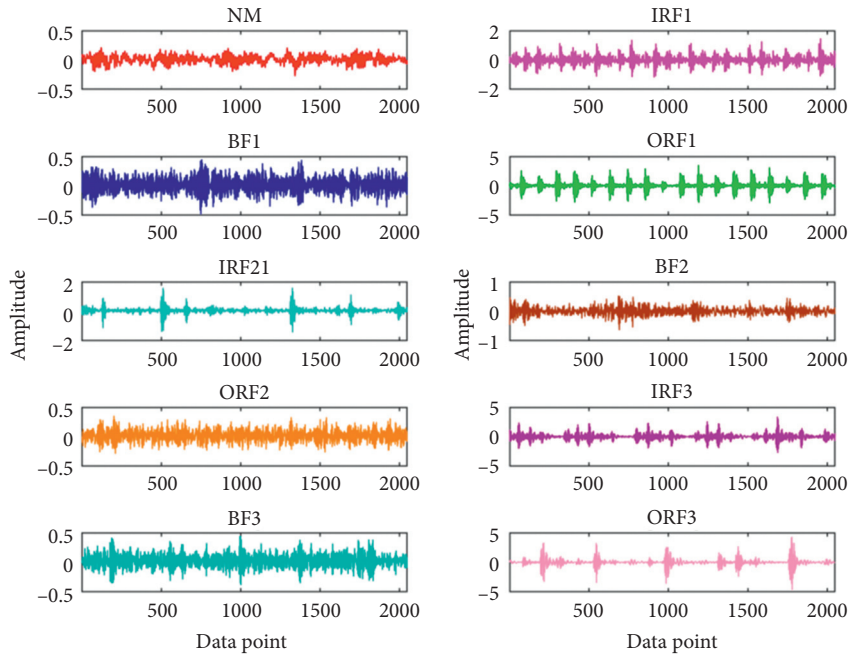

Figure 9: The time-domain waveforms of bearings with different types and fault severity.

by the experimental results that the proposed approach has an excellent performance in the bearing predetection stage.

After predetecting the working condition of the bearing, the fault recognition approach is used to diagnose the type and severity of the fault if the bearing does have a fault. First, each sample is decomposed to three levels adopting DTCWPT to highlight the impact components in the fault vibration signal and reduce the interference between the components. Next, eight subband components including diverse frequency band information can be acquired. Besides, only the DTCWPT decomposition results of minor outer race faults (ORF1) are used as representatives to reduce the space footprint. The results of the decomposition are exhibited in Figure 11.

After decomposing the signal, the GCMAAPE algorithm is used to fetch features of diverse scales from each subband signal. Considering space limitations, the GCMAAPE values of only subband signal 1, subband signal 2, subband signal 3, and subband signal 4 over scale factor of 10 under 9 working conditions are illustrated in Figure 12. It can be observed from Figure 12 that the GCMAAPE curves of the four subbands almost overlap when the scale factor is 1–4, demonstrating a poor separability of each state at this time. Moreover, the curve of each state has a significant difference when the scale factor is 4–10, indicating a strong separability. However, it is not possible to directly determine the fault condition based solely on the curve distribution in the figure, and further analysis is required to make the features obvious enough so as to recognize the diverse type of the bearing.

After completing the initial feature extraction, the R^{450*80} dimensional feature can be obtained. Apparently, the original feature is high-dimensional and redundant. It will not only reduce the efficiency but also guarantee the recognition accuracy if used directly for classification. Thus, it is indispensable to reduce the dimension of the fault feature. The t-SNE is adopted to select sensitive features for the

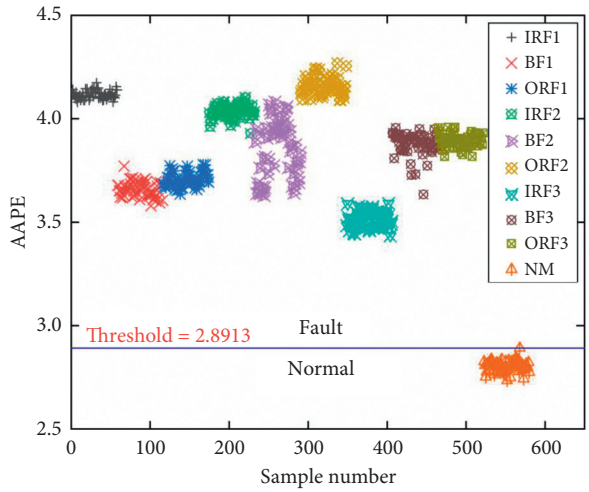

Figure 10: The amplitude-aware permutation entropy distribution of all samples.

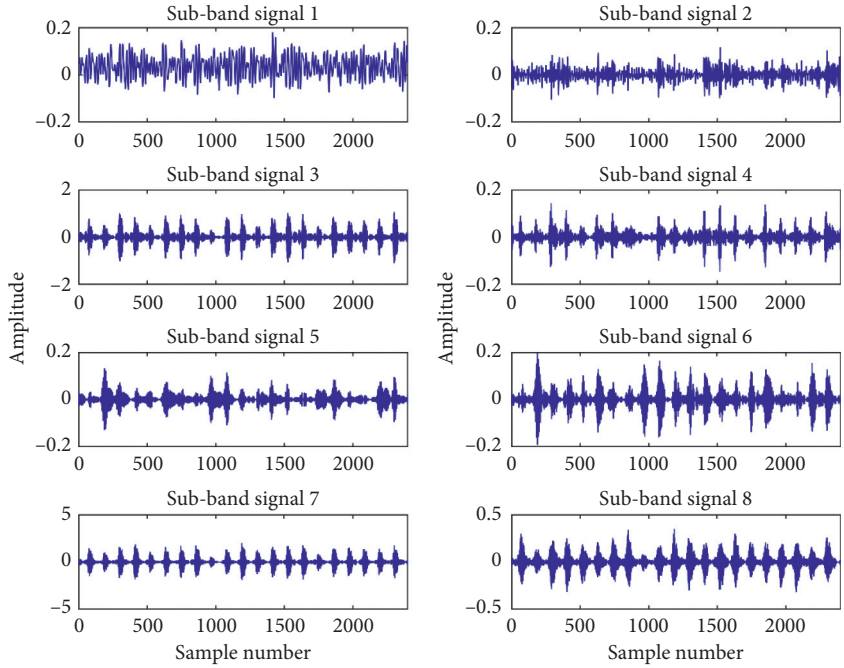

Figure 11: DTCWPT results of the sample with the minor outer race fault.

original high-dimensional features. The feature vector after dimensionality reduction is R^{450*3}. Then, the DBN classification model is trained using 270 groups of samples, and the property of the trained model is verified employing the remaining as test samples. The classification results of the proposed diagnostic model for nine fault conditions under one trial are illustrated in Figure 13. In the 180 sets of test samples, one BF1 sample was misclassified as BF2 sample. It can be revealed that the classification accuracy of minor outer race fault (BF2) conditions is 95% while the classification accuracy of the other eight states reaches 100%. Besides, the recognition rate of the proposed diagnosis model reaches as high as 99.44% for all working conditions, fully demonstrating that the solution possesses the best diagnostic performance and good robustness for diverse working conditions and fault severity.

The proposed approach is compared with the other three approaches used on the bearing dataset to verify its excellent performance. The other three methods are DTCWPT and GCMPE (DTCWPT-GCMPE), DTCWPT and MAAPE (DTCWPT-MAAPE), and DTCWPT and MSE (DTCWPT-MSE). For DTCWPT-GCMPE [37], $r = 3$, $m = 6$, $t = 1$, and $s_{\max} = 10$. For DTCWPT-MAAPE, $r = 3$, $m = 6$, $t = 1$, $A = 0.5$, and $s_{\max} = 10$. For DTCWPT-MSE, $r = 3$, $m = 2$, $t = 1$, rd = 0.15, and $s_{\max} = 10$. Among them, r denotes the number of decomposition layers of DTCWPT, m refers to the embedding dimension, t is the time delay, rd represents the tolerance of the signal, A is the adjustment coefficient, and

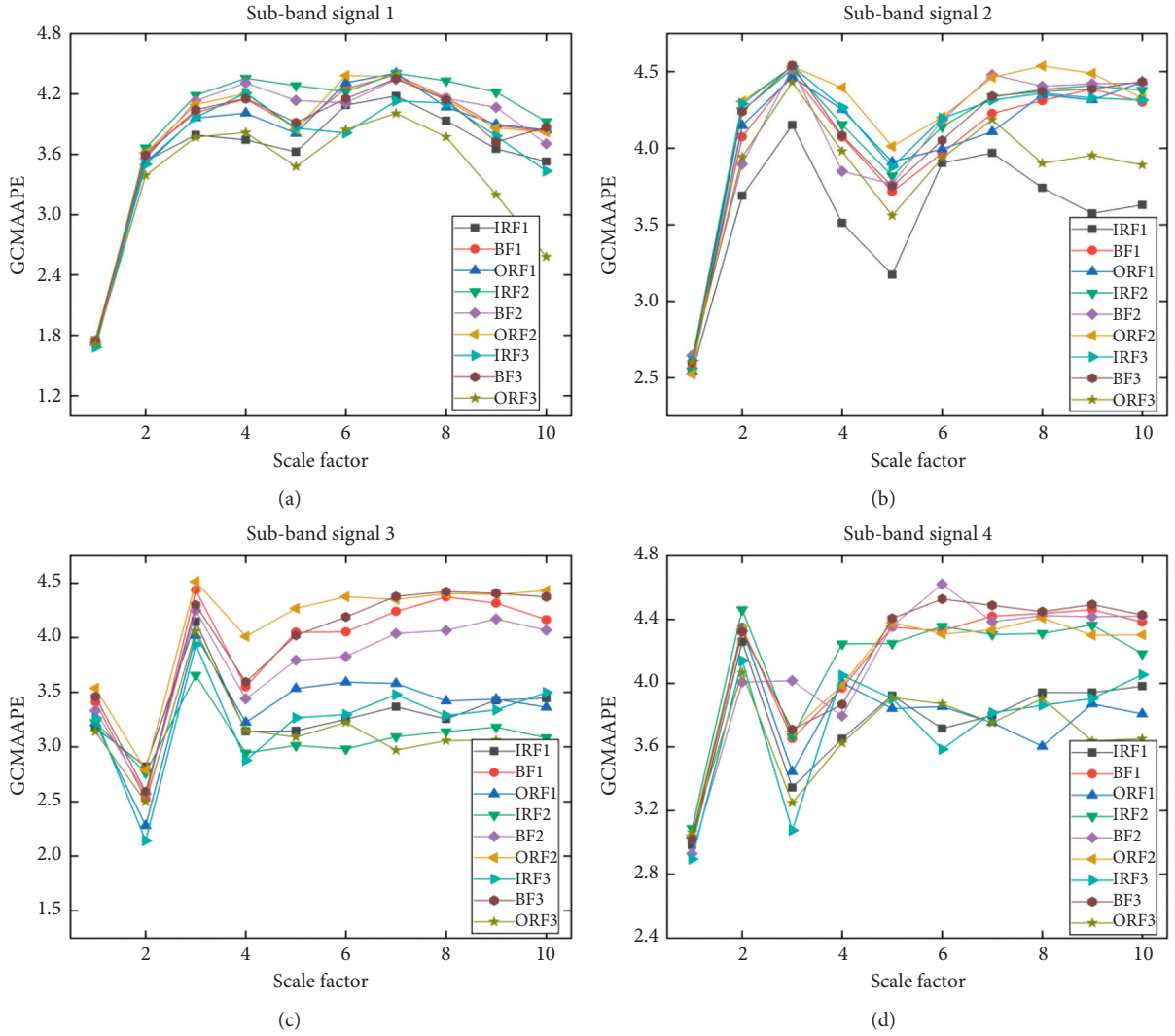

FIGURE 12: GCMAAPE value of four subband signals over 10 scales under nine different working conditions. (a) Subband signal 1. (b) Subband signal 2. (c) Subband signal 3. (d) Subband signal 4.

s_{max} indicates the scale factor. The classifier used in each method is DBN. Each method is repeated 25 times to avoid errors caused by random factors such as accidental factors. The experimental results of diverse approaches are presented in Figure 14 and Table 2. It can be clearly observed from Figure 14 and Table 2 that the classification effect of the proposed method is obviously superior to several other methods. The proposed method exhibits a maximum recognition rate of 100% and an average classification accuracy of 98.82%. After replacing GCMAAPE with GCMPE, the classification accuracy reaches 98.33%, and the average classification accuracy is 96.58%, lower than that of GCMAAPE. The main reason is that GCMAAPE introduces the amplitude information of the vibration signal into the calculation process, contributing to improving the utilization rate of fault information and obtaining higher-quality features. Besides, the average correct rate of MAAPE is lower than that of GCMAAPE, verifying that the generalized coarse-grained method adopted is superior to the traditional coarse-grained method used by MAAPE. Generally, the GCMAAPE method used is superior to several common entropy-based methods in performance, which is directly reflected in the higher recognition rate. Thus, the robustness of the approach to fault classification problems is fully demonstrated.

The advantages of this method in feature extraction are verified by comparing the performance of the DTCWPT-GCMAAPE feature extraction method with the following feature extraction methods. GCMAAPE acts on the raw vibration signal, the approach based on EEMD and GCMAAPE (EEMD-GCMAAPE) which can be found in the literature [38, 39] and the approach based on WT and GCMAAPE (WT-GCMAAPE). The parameters of these three methods are set as follows. For EEMD-GCMAAPE, $M = 100$, $sd = 0.2$, $m = 6$, $A = 0.5$, $t = 1$, and $s_{max} = 10$. For WT-GCMAAPE, $r = 3$, wavelet basis function is db4, $m = 6$, $A = 0.5$, $t = 1$, and $s_{max} = 10$. Among them, M is the ensemble number of EEMD, and sd is the standard deviation of added white noise in EEMD. The fault features extracted by the above three methods are finally inputted to the DBN

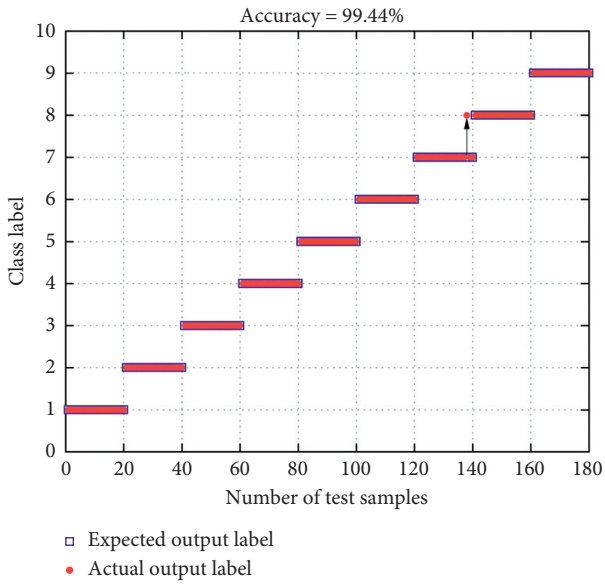

Figure 13: Classification result of the proposed model.

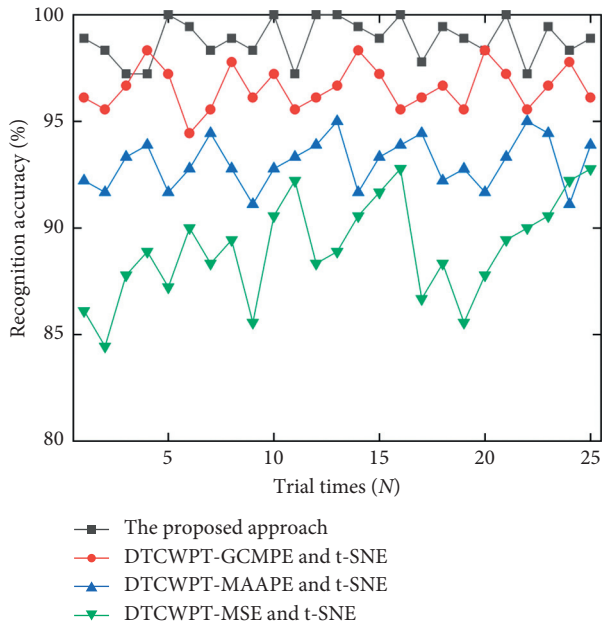

Figure 14: The classification results of five approaches during 25 trials.

Table 2: The classification results of different approaches.

Approaches	Accuracy (%)		
	Max	Min	Mean
The proposed approach	100	97.22	98.82
DTCWPT-GCMPE	98.33	94.44	96.58
DTCWPT-MAAPE	95	91.11	93.07
DTCWPT-MSE	92.78	84.44	89.04

classifier to classify the fault type. The selection of parameters and the proportion of samples are consistent with the proposed method. The fault classification results of the three methods are presented in Figure 15, and the classification accuracy rates reach 93.33%, 95%, and 96.11%, respectively, lower than the proposed method in Figure 13. This is because when the GCMAAPE method is directly applied to the original vibration signal, the information such as the fault frequency in the fault vibration signal cannot be extracted, and the fault information is not sufficiently analyzed, resulting in a reduction in the quality of the feature. Both WT-GCMAAPE and EEMD-GCMAAPE are a multiscale analysis method based on time-frequency analysis. These two methods also have some problems limiting the quality of feature extraction. Besides, the WT method cannot effectively decompose the high-frequency part of the signal when analyzing the signal. EEMD has a serious mode aliasing effect, making the decomposed IMF have a large interference component. Compared with these three methods, the proposed DTCWPT-GCMAAPE feature extraction method is a multiscale analysis method based on time-frequency preprocessing and can reflect more fault information by highlighting different frequency components of vibration signals and by multiscale analysis, contributing to improving the quality of extracted features for better classification.

The classification experiments of the four methods in two conditions are compared to explore the necessity of dimensionality reduction and validate the performance of t-SNE in feature dimensionality reduction. The two conditions are without dimension reduction and with LDA dimension reduction. The experimental results of the four methods under different dimensionality reduction conditions are provided in Table 3. It can be observed from Table 3 that DTCWPT-GCMAAPE has achieved the best classification effect in both conditions while the accuracy rate of using LDA dimensionality reduction and nondimensionality reduction is lower than that presented in Table 2. It indicates that dimensionality reduction is necessary, and the dimensionality reduction performance of t-SNE is better than that of LDA. Without loss of the generality, two features are selected from the original features, as illustrated in Figure 16(a), and the two-dimensional visualization after LDA dimensionality reduction is presented in Figure 16(b). Moreover, the two-dimensional visualization after t-SNE dimensionality reduction is exhibited in Figure 16(c). Apparently, the different fault states in Figure 16(a) are well separated compared to Figures 16(a) and 16(b). Simultaneously, a BF2 sample is erroneously divided into IRF3 samples, confirming the classification result of Figure 13. Through dimension reduction, the subsequent classification becomes faster and efficiency and accuracy are improved.

The feature samples obtained by the proposed method are sent to different classifications for comparison (namely, support vector machine (SVM) and random forest (RF)) to verify the performance of the selected DBN classifier. The classification results and average running time are presented in Table 4. Besides, each method runs for 30 times to ensure that it is not affected by random factors. It can be revealed from Table 3 that the DBN classifier achieves the best classification effect, and the RF classifier requires the least running time. Thus, DBN has the highest classification accuracy, even though the DBN classifier has more running

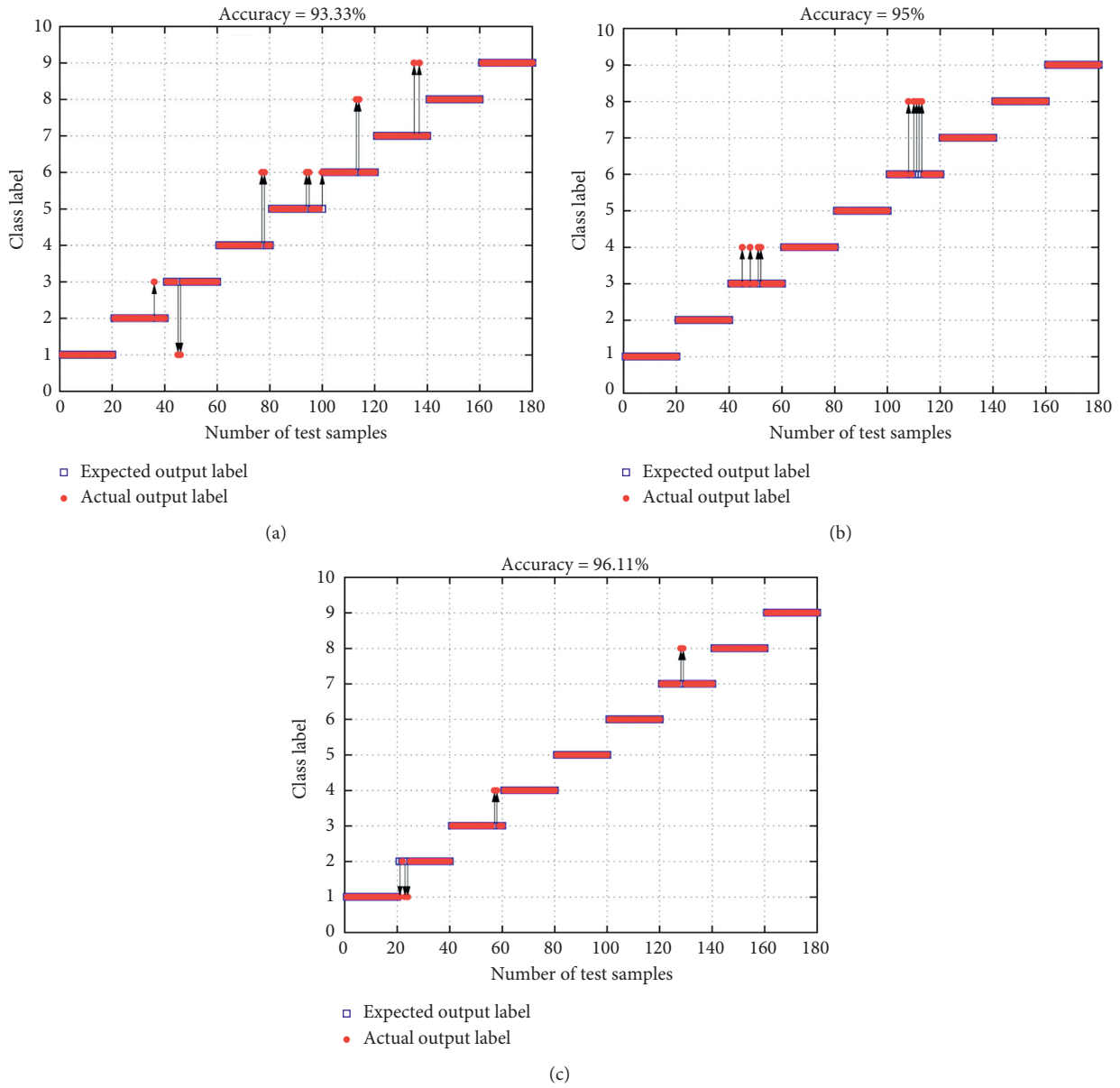

Figure 15: Outputs of GCMAAPE, WT-GCMAAPE, and EEMD-GCMAAPE based fault diagnosis approaches. (a) GCMAAPE. (b) WT-GCMAAPE. (c) EEMD-GCMAAPE.

Table 3: The recognition rate of four methods under different conditions.

Different conditions	Different methods	Accuracy (%)		
		Max	Min	Mean
LDA	DTCWPT-GCMAAPE	96.67	92.78	94.51
	DTCWPT-GCMPE	95.56	91.11	92.74
	DTCWPT-MAAPE	91.67	87.22	89.86
	DTCWPT-MSE	86.11	81.67	83.72
Without t-SNE	DTCWPT-GCMAAPE	93.33	90	91.58
	DTCWPT-GCMPE	92.22	88.89	90.14
	DTCWPT-MAAPE	89.44	85	86.95
	DTCWPT-MSE	82.78	78.89	81.29

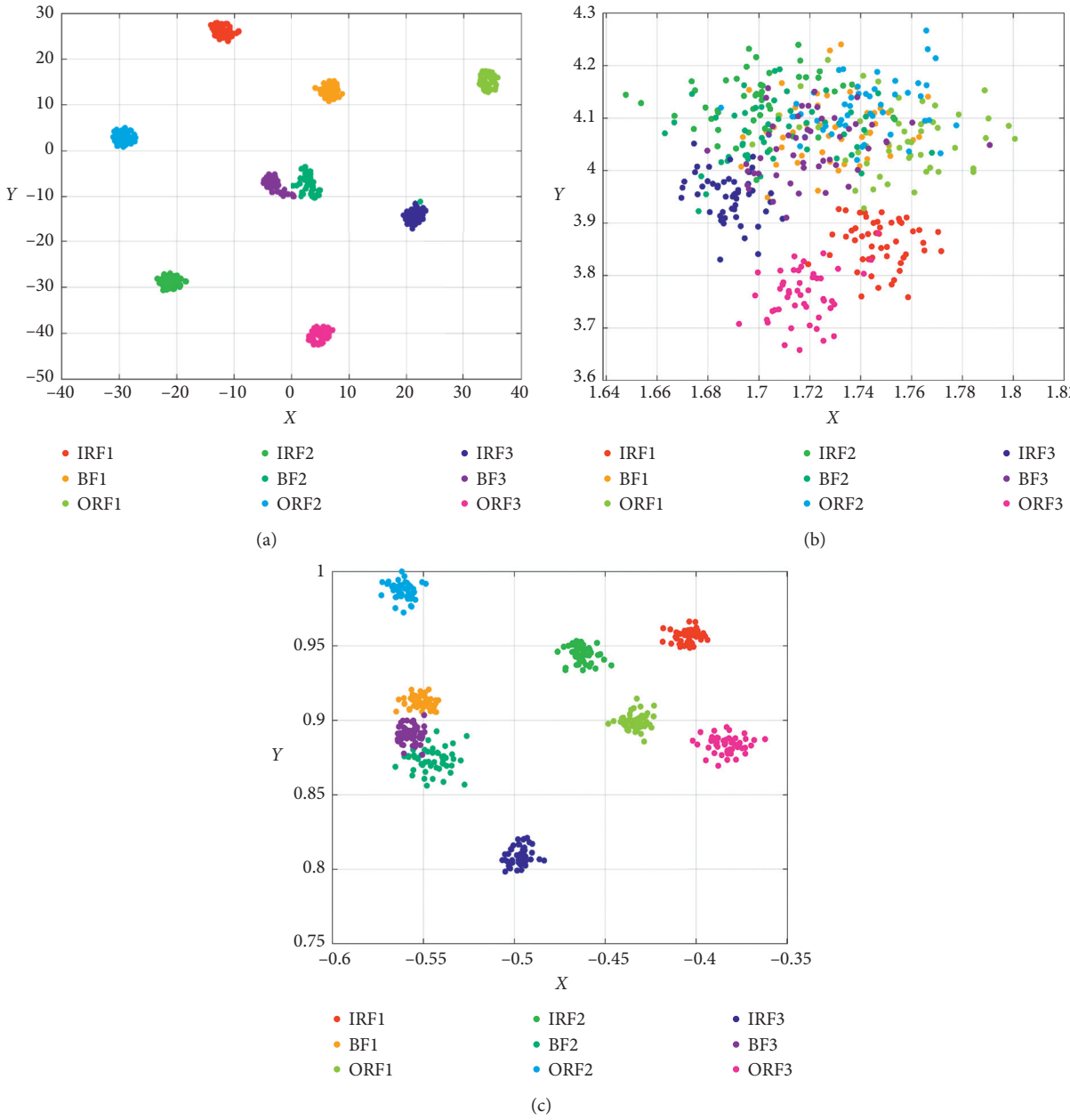

Figure 16: The two-dimensional view of the first two features by DTCWPT-GCMAAPE: (a) adopting t-SNE; (b) without adopting t-SNE; (c) adopting LDA.

Table 4: The classification results of different classification models for 30 trials.

Classifier	Accuracy (%)			Time (s)
	Max	Min	Mean	
DBN	100	96	98.47	7.98
RF	100	93	96.23	5.61
SVM	95	88	91.25	11.36

time than RF. DBN achieves better performance by sacrificing part of the running time, which is acceptable to a certain extent. The running time of SVM is the longest among the three methods while its classification effect is not ideal. This is mainly because SVM is suitable for processing small sample data and cannot achieve the best performance when processing large batches of high-dimensional data. Generally, the DBN classification model not only has a shorter running time but also exhibits better performance.

5. Conclusion

In this research, a synthesized fault identification technology including bearing predetection and fault classification is proposed to detect and identify the status of the bearing. At the stage of predetection, an AAPE threshold value that can differentiate between normal and fault conditions is defined by calculating the AAPE value of the bearing vibration signal under different working conditions so as to screen out the bearings with normal working conditions. Specifically, the time-frequency multiscale method DTCWPT-GCMAAPE-t-SNE is utilized to extract the fault feature and generate a fault feature sample if the bearing is detected to be faulty. Finally, a DBN classifier with powerful classification performance is used to classify the acquired high-dimensional features. The classification effects of WT-GCMAAPE, EEMD-GCMAAPE, and GCMAAPE are compared. The results indicate that the proposed approach can accurately highlight the fault information in the vibration signal and improve the quality of the features extracted subsequently. Simultaneously, it is compared with MAAPE, GCMPE, and MSE, demonstrating that GCMAAPE can effectively extract fault features from the DTCWPT processed signal and has better robustness. Generally, compared with other common fault diagnosis methods, this paper introduces bearing predetection, which avoids the subsequent model classification with uncertainty and improves the diagnosis efficiency. It has practical engineering significance and is more suitable for practical engineering application.

Data Availability

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

Conflicts of Interest

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

References

[1] G. Tang, X. Wang, and Y. He, "A novel method of fault diagnosis for rolling bearing based on dual tree complex wavelet packet transform and improved multiscale permutation entropy," *Mathematical Problems in Engineering*, vol. 2016, Article ID 5432648, 13 pages, 2016.

[2] 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.

[3] W. Caesarendra, M. Pratama, B. Kosasih et al., "Parsimonious network based on a fuzzy inference system (PANFIS) for time series feature prediction of low speed slew bearing prognosis," *Applied Sciences*, vol. 8, no. 12, 2018.

[4] A. Glowacz, "Acoustic fault analysis of three commutator motors," *Mechanical Systems and Signal Processing*, vol. 133, 2019.

[5] A. Glowacz, "Fault detection of electric impact drills and coffee grinders using acoustic signals," *Sensors*, vol. 19, no. 2, 2019.

[6] J. Zhang, Y. Zhao, X. Li, and M. Liu, "Bearing Fault diagnosis with kernel sparse representation classification based on adaptive local iterative filtering-enhanced multiscale entropy features," *Mathematical Problems in Engineering*, vol. 2019, Article ID 7905674, 17 pages, 2019.

[7] M.-a. Li, H.-n. Liu, W. Zhu et al., "Applying improved multiscale fuzzy entropy for feature extraction of MI-EEG," *Applied Sciences*, vol. 7, no. 1, 2017.

[8] R. Yan and R. X. Gao, "Approximate entropy as a diagnostic tool for machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 824–839, 2007.

[9] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, Article ID H2039, 2000.

[10] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, pp. 1741021–1741024, 2002.

[11] J. Xingmeng, W. Li, P. Liwu et al., "Rolling Bearing Fault diagnosis based on ELCD permutation entropy and RVM," *Journal of Engineering*, vol. 2016, Article ID 1308108, 7 pages, 2016.

[12] 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.

[13] Y. S. Chen, T. H. Zhang, W. J. zhao et al., "Fault Diagnosis of rolling bearing using multiscale Amplitude-aware permutation entropy and random forest," *Algorithms*, vol. 12, no. 9, p. 184, 2019.

[14] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, no. 6, 2002.

[15] Y. Li, G. Li, Y. Wei, B Liu, and X Liang, "Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy," *ISA Transactions*, vol. 81, pp. 329–341, 2018.

[16] Y. Wei, Y. Li, M. Xu, and W. Huang, "Intelligent Fault Diagnosis of rotating machinery using ICD and generalized composite multi-scale fuzzy entropy," *IEEE Access*, vol. 7, pp. 38983–38995, 2019.

[17] 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, no. 20, p. 4542, 2019.

[18] W. B. Zhang and J. Z. 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.

[19] S. Wan and X. Zhang, "Teager energy entropy ratio of wavelet packet transform and its application in Bearing Fault diagnosis," *Entropy*, vol. 20, no. 5, p. 388, 2018.

[20] J. S. Chen, D. J. Yu, J. S. Tang, and Y. Yang, "Application of SVM and SVD technique based on EMD to The fault diagnosis of the rotating machinery," *Shock and Vibration*, vol. 16, no. 1, pp. 89–98, 2009.

[21] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters," in *Proceedings of the 8th IEEE Digital Signal Processing Workshop*, 1998.

[22] I. Bayram and I. W. Selesnick, "On the dual-tree complex wavelet packet and *M*-Band transforms," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2298–2310, 2008.

[23] M.-F. Ge, Z. Ge, H. Pan et al., "A deep condition feature learning approach for rotating machinery based on MMSDE and optimized SAEs," *Measurement Science and Technology*, 2020, In press.

[24] Y. Yu and C. Junsheng, "A roller bearing fault diagnosis method based on EMD energy entropy and ANN," *Journal of Sound and Vibration*, vol. 294, no. 1-2, pp. 269–277, 2006.

[25] H. Ao, J. Cheng, K. Li, and T. K. Truong, "A roller Bearing Fault diagnosis method based on LCD energy entropy and ACROA-SVM," *Shock and Vibration*, vol. 2014, Article ID 825825, 12 pages, 2014.

[26] G. E. Hinton, S. Osindero, and Y. W. The, "A fast learning algorithm for deep belief nets," *Neural Comput," Neural Computation*, vol. 18, no. 7, pp. 1257–2544, 2006.

[27] W. Deng, H. Liu, J. Xu et al., "An improved quantum-inspired differential evolution algorithm for deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319–7327, 2020.

[28] W. Deng, J. Xu, and H. Zhao, "An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem," *IEEE Access*, vol. 7, pp. 20281–20292, 2019.

[29] H. Zhao, H. Liu, J. Xu, and W. Deng, "Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine," *Ieee Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 4165–4172, 2020.

[30] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 3, pp. 983–994, 2020.

[31] J. Gai, J. Shen, H. Wang, and Y. Hu, "A parameter-optimized DBN using GOA and its application in fault diagnosis of gearbox," *Shock and Vibration*, vol. 2020, Article ID 4294095, 11 pages, 2020.

[32] Q. Tong, J. Cao, B. Han et al., "A fault diagnosis approach for rolling element bearings based on dual-tree complex wavelet packet transform-improved intrinsic time-scale decomposition, singular value decomposition, and online sequential extreme learning machine," *Advances in Mechanical Engineering*, vol. 9, no. 12, 2017.

[33] S. Zhou, S. Qian, W. Chang et al., "A novel bearing multi-Fault Diagnosis approach based on weighted permutation entropy and an improved SVM ensemble classifier," *Sensors*, vol. 18, no. 6, 2018.

[34] B. Fadlallah, B. Chen, A. Keil et al., "Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 87, no. 2, Article ID 022911, 2013.

[35] G. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum*, vol. 9, pp. 1–21, 2010.

[36] Case Western Reserve University, *Bearing Data Center*, Case Western Reserve University, Cleveland, OH, USA, 2019, http://csegroups.case.edu/bearingdatacenter/pages/download-data-fifile.

[37] J. D. Zheng, T. Liu, R. Meng et al., "Generalized composite multiscale permutation entropy and PCA based fault diagnosis of rolling bearings," *Journal of Vibration and Shock*, vol. 37, no. 20, pp. 61–66, 2018, in Chinese.

[38] Y. Ye, Y. Zhang, Q. Wang, Z. Wang, Z. Teng, and H. Zhang, "Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106565, 2020.

[39] X. Zhang, Y. Liang, J. Zhou, and Y. zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.

[40] Q. Tong, J. Cao, B. Han et al., "a fault diagnosis approach for rolling element bearings based on RSGWPT-LCD bilayer screening and extreme learning machine," *IEEE Access*, vol. 5, pp. 5515–5530, 2017.