

Research Article

A Novel Method of Fault Diagnosis for Rolling Bearing Based on Dual Tree Complex Wavelet Packet Transform and Improved Multiscale Permutation Entropy

Guiji Tang, Xiaolong Wang, and Yuling He

School of Energy, Power and Mechanical Engineering, North China Electric Power University, Baoding 071000, China

Correspondence should be addressed to Yuling He; heyuling1@163.com

Received 8 February 2016; Accepted 7 April 2016

Academic Editor: Wen Chen

Copyright © 2016 Guiji Tang 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.

A novel method of fault diagnosis for rolling bearing, which combines the dual tree complex wavelet packet transform (DTCWPT), the improved multiscale permutation entropy (IMPE), and the linear local tangent space alignment (LLTSA) with the extreme learning machine (ELM), is put forward in this paper. In this method, in order to effectively discover the underlying feature information, DTCWPT, which has the attractive properties as nearly shift invariance and reduced aliasing, is firstly utilized to decompose the original signal into a set of subband signals. Then, IMPE, which is designed to reduce the variability of entropy measures, is applied to characterize the properties of each obtained subband signal at different scales. Furthermore, the feature vectors are constructed by combining IMPE of each subband signal. After the feature vectors construction, LLTSA is employed to compress the high dimensional vectors of the training and the testing samples into the low dimensional vectors with better distinguishability. Finally, the ELM classifier is used to automatically accomplish the condition identification with the low dimensional feature vectors. The experimental data analysis results validate the effectiveness of the presented diagnosis method and demonstrate that this method can be applied to distinguish the different fault types and fault degrees of rolling bearings.

1. Introduction

Rolling bearings are one of the most widely used parts in rotating machineries because they affect the operation reliability, the performance precision, and the service life of the entire equipment. The failures of rolling bearings may cause catastrophic accidents and result in great loss. Therefore, condition monitoring and fault diagnosis for rolling bearings are of great significance in engineering application [1, 2].

Due to the factors such as friction, strike, and structure transmutation, the vibration signals of bearings are often characterized by nonlinearity and nonstationarity. And the major challenge for bearing condition monitoring and fault diagnosis is to acquire the reliable and sensitive features from the vibration signals [3]. In recent years, with the development of nonlinear dynamic theories, a series of nonlinear parameter estimation techniques have been investigated and introduced to the field of bearing condition monitoring and fault diagnosis. For example, the correlation dimension

was chosen as a tool for discovering the fault features of bearings by Kang et al. [4]. Unfortunately, the estimation of correlation dimension usually requires sufficient data, which prevents this technology from being widely used. Yan and Gao [5] applied approximate entropy (AE) to monitor the bearing condition. However, AE depends heavily on the signal length and the calculated value is uniformly smaller than the expected one when processing short term signals [6]. Later, the sample entropy (SE) was proposed by Richman and Moorman [7] to overcome the drawback of AE. The collected signals from bearing systems usually consist of multiple temporal scale structures. But AE and SE both evaluate the complexity of signal at a single scale. Hence, these two approaches have limited performance in analyzing the bearing signal. Considering the disadvantage of the single scale analysis, the multiscale entropy (MSE) was developed by Costa et al. [8] to estimate the complexity of time series over a range of scales, and this technology was used by Zhang et al. [9] to extract the features of bearing signal. However, the

estimation of MSE is easily affected by the outliers of signal, and the computational efficiency of MSE is very low for long term signals.

In literature [10], a new kind of entropy named permutation entropy (PE) was proposed to measure the complexity and detect the dynamic changes of signal. Compared with AE and SE, the calculation of PE is simple and immune to noises. But similar to AE and SE, PE also conducts the entropy measure in a single scale. Then the multiscale permutation entropy (MPE) method based on PE was further proposed by Aziz and Arif [11] to depict the multiple temporal scale structures of the signal. And this method was, respectively, applied by Li and Zheng to bearing fault diagnosis [12, 13]. Nevertheless, the analysis results of MPE are usually unstable for short term signals. Recently, a novel method called improved multiscale permutation entropy (IMPE) was proposed by Azami and Escudero [14] to remedy the weakness of MPE and the effectiveness of this method has been verified by the simulated signal and the real biomedical signal. In view of the advantage of IMPE in digging the inherent features of signal, this method is introduced to the field of fault diagnosis and utilized to identify the condition of rolling bearing in this paper.

Usually, the collected bearing signals are more or less contaminated by external environmental noises, and the interference between the components in the complicated signal is inevitable. These factors lead to the difficulty of feature information extraction using IMPE method directly. And it will be of benefit for the subsequent analysis procedure if the original signals are processed in advance.

Up to now, many signal processing techniques such as empirical mode decomposition (EMD), local mean decomposition (LMD), and discrete wavelet transform (DWT) have been developed and applied in different research fields. As a kind of adaptive signal processing method, EMD could decompose a signal into a set of intrinsic mode functions. However, EMD lacks a forceful mathematical framework and has the problems of mode mixture and end effects. Similar to EMD, the LMD algorithm also involves these drawbacks which have not been fundamentally addressed [15]. DWT is a classic signal processing tool and has been widely used to analyze the mechanical fault signals, but its disadvantages of shift variance and frequency aliasing may cause the loss of useful information. Then Kingsbury [16] proposed the dual tree complex wavelet transform (DTCWT) method, which possesses many excellent properties such as nearly shift invariance, good directional selectivity, and reduced aliasing in comparison with DWT. However, DTCWT cannot achieve multiresolution analysis in the high frequency region where the useful feature information usually exists. As a kind of expansion of DTCWT, dual tree complex wavelet packet transform (DTCWPT) is developed to offset this shortcoming [17]. After performing DTCWPT on the collected signal, precise frequency band partitions over the whole analyzed frequency domain could be achieved and the corresponding subband signals could be obtained. And it will be more effective to discover the feature information of the original signal by analyzing the subband signals using IMPE. Based

on the above analysis, DTCWPT is combined with IMPE for the first time in this study for bearing fault diagnosis.

After feature extraction using DTCWPT and IMPE, the acquired feature vectors need to be fed into the classifier to achieve condition recognition. However, the acquired vectors with high dimension inevitably contain redundancy information. It is time-consuming and may lead to a decline in the diagnostic accuracy if the entire vectors are adopted as the inputs of classifier. Then the manifold learning algorithm named linear local tangent space alignment (LLTSA) [18] is employed in this paper to reduce the dimension of vectors. By using LLTSA, the high dimensional feature vectors are automatically compressed and sensitive feature vectors with lower dimension could be obtained, which will not only reduce the computational burden but also improve the diagnostic precision.

Naturally, an intelligent classifier is needed to automatically distinguish the bearing condition based on the obtained sensitive feature vectors. Extreme learning machine (ELM) [19] is a novel powerful intelligent machine learning approach based on single hidden layer feed-forward networks. Compared with some classic machine learning methods such as support vector machines (SVM), artificial neural network (ANN), and K nearest neighbor classifier (KNNC), the main advantages of ELM lie in better generalization ability on the small samples, faster learning speed, and less human intervention. Thus, in this paper, ELM is utilized to distinguish the bearing condition.

The rest of this paper is organized as follows. Section 2 proposes the feature extraction method based on DTCWPT and IMPE. Section 3 presents the feature dimension reduction method based on LLTSA. Section 4 briefly introduces the ELM classifier. Section 5 illustrates the detailed procedures of the proposed diagnosis method. In Section 6, the proposed method is applied to rolling bearing experimental data and some comparisons are made. Finally, conclusions are drawn in Section 7.

2. Feature Extraction Based on DTCWPT and IMPE

Fault diagnosis for rolling bearings is comprised of feature extraction and pattern recognition. Feature extraction is the most important part during the fault diagnosis, because the bearing conditions are identified according to the extracted features. Aiming to take the advantages of DTCWPT in processing the nonstationary and nonlinear signal and meanwhile utilize the capability of IMPE in characterizing the property of signal, these two methods are combined together to extract the feature information from the bearing signal.

2.1. A Brief View of DTCWPT. DTCWPT is an enhancement to the traditional discrete wavelet packet transform (WPT). In the decomposition and the reconstruction process of DTCWPT, two parallel WPTs with different low pass and high pass filters in each level are utilized. These can be, respectively, regarded as the real tree and the imaginary tree in DTCWPT algorithm. And information complementation

can be achieved in the process of signal processing [20]. The decomposition process of DTCWPT is implemented through a set of low pass and high pass filters recursively as follows.

Real tree decomposition is as follows:

$$\begin{aligned} c_{l+1,2N}^{\text{Re}}(k) &= \sum_m h_0(m-2k) c_{l,N}^{\text{Re}}(m), \\ c_{l+1,2N+1}^{\text{Re}}(k) &= \sum_m h_1(m-2k) c_{l,N}^{\text{Re}}(m). \end{aligned} \quad (1)$$

Imaginary tree decomposition is as follows:

$$\begin{aligned} c_{l+1,2N}^{\text{Im}}(k) &= \sum_n g_0(n-2k) c_{l,N}^{\text{Im}}(n), \\ c_{l+1,2N+1}^{\text{Im}}(k) &= \sum_n g_1(n-2k) c_{l,N}^{\text{Im}}(n), \end{aligned} \quad (2)$$

where h_0 and h_1 , respectively, represent the low pass and the high pass filters used by WPT of the real tree, while g_0 and g_1 are the low pass and the high pass filters used by WPT of the imaginary tree. $c_{l,N}^{\text{Re}}$ and $c_{l,N}^{\text{Im}}$, respectively, denote the coefficients in the real tree and the imaginary tree at the l th level, N th node. When level $l = 0$, coefficients $c_{0,0}^{\text{Re}}$ and $c_{0,0}^{\text{Im}}$ are the original signal $x(t)$; namely, $c_{0,0}^{\text{Re}} = c_{0,0}^{\text{Im}} = x(t)$. The decomposition process of DTCWPT is illustrated in Figure 1.

The corresponding reconstruction operation of DTCWPT is as follows.

Real tree reconstruction is as follows:

$$\begin{aligned} c_{l,N}^{\text{Re}}(k) &= \sum_m \tilde{h}_0(k-2m) c_{l+1,2N}^{\text{Re}}(m) \\ &+ \sum_m \tilde{h}_1(k-2m) c_{l+1,2N+1}^{\text{Re}}(m). \end{aligned} \quad (3)$$

Imaginary tree reconstruction is as follows:

$$\begin{aligned} c_{l,N}^{\text{Im}}(k) &= \sum_n \tilde{g}_0(k-2n) c_{l+1,2N}^{\text{Im}}(n) \\ &+ \sum_n \tilde{g}_1(k-2n) c_{l+1,2N+1}^{\text{Im}}(n), \end{aligned} \quad (4)$$

where \tilde{h}_0 and \tilde{h}_1 , respectively, represent the low pass and the high pass reconstruction filters used by WPT of the real tree, while \tilde{g}_0 and \tilde{g}_1 denote the low pass and the high pass reconstruction filters used by WPT of the imaginary tree.

2.2. Background of PE, MPE, and IMPE

2.2.1. Permutation Entropy. The permutation entropy (PE) was proposed to detect the dynamic changes of time series based on comparison of neighboring values of time series [13]. The calculation steps of PE are described as follows.

Given a time series $Y = \{y_1, y_2, \dots, y_N\}$ with the length of N , h dimensional delay embedding vector at the moment t can be constructed as $Y_t^{h,q} = \{y_t, y_{t+q}, \dots, y_{t+(h-1)q}\}$ ($t = 1, 2, \dots, N - (h-1)q$), where h represents the embedding dimension and q is the time delay. It is thought that $Y_t^{h,q}$ has a permutation type $\pi_{m_0 m_1 \dots m_{h-1}}$ if it satisfies

$$y_{t+m_0 q} \leq y_{t+m_1 q} \leq \dots \leq y_{t+m_{h-1} q}, \quad (5)$$

where $0 \leq m_i \leq h-1$ and $m_i \neq m_j$.

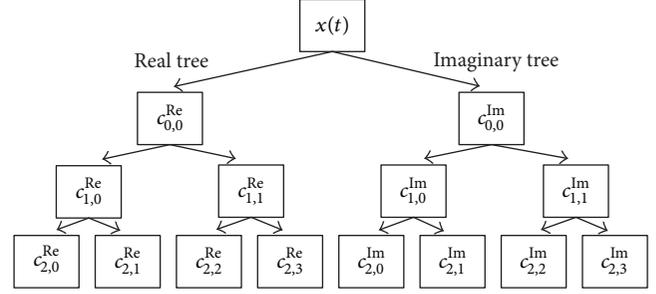


FIGURE 1: Decomposition process of DTCWPT.

There are $h!$ kinds of different permutation types for h dimensional vector. For each permutation type $\pi \in T$ (T denotes the set of all permutation types), $p(\pi)$ demonstrates the relative frequency as follows:

$$\begin{aligned} p(\pi) &= \frac{\text{Number} \{t \mid 1 \leq t \leq N - (h-1)q, Y_t^{h,q} \text{ has type } \pi\}}{N - (h-1)q}. \end{aligned} \quad (6)$$

Then PE of time series Y is calculated as follows:

$$\text{PE}(Y, h, q) = - \sum_{\pi \in T} p(\pi) \ln(p(\pi)). \quad (7)$$

2.2.2. Multiscale Permutation Entropy. The multiscale permutation entropy (MPE) is defined as the PE set of time series at different scales. Considering the time series $X = \{x_1, x_2, \dots, x_N\}$ with the length of N , the computational procedures of MPE are as follows.

(1) The original time series is firstly divided into several coarse-grained series $y_j^{(\tau)}$ according to (8) and the schematic of this procedure is shown in Figure 2:

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

where τ denotes the scale factor.

(2) The PE of each coarse-grained series is calculated based on (6) and (7) and then plotted as a function of the scale factor τ , which can be expressed as follows:

$$\text{MPE}(X, h, q, \tau) = \text{PE}(y_j^{(\tau)}, h, q). \quad (9)$$

2.2.3. Improved Multiscale Permutation Entropy. From Figure 2, it can be found that the coarse-grained procedure in MPE method can be considered as the procedure of averaging the original time series within a τ -length window and then downsampling by a scale factor of τ . However, the imprecise and unreliable results may occur in the process of downsampling at a certain scale [14]. To overcome the drawback of MPE, IMPE algorithm is proposed, and the calculation steps are as follows.

(1) For a defined scale factor τ , the original time series is divided into τ different coarse-grained series $z_i^{(\tau)} =$

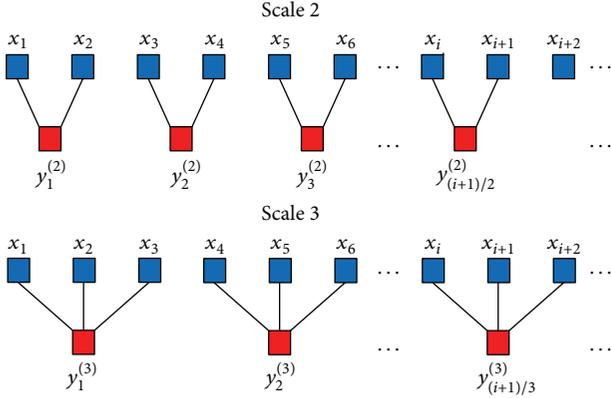


FIGURE 2: The schematic of the coarse-grained procedure for scale factors $\tau = 2$ and $\tau = 3$.

$\{y_{i,1}^{(\tau)}, y_{i,2}^{(\tau)}, \dots\}$ ($i = 1, 2, \dots, \tau$) based on the following equation:

$$y_{i,j}^{(\tau)} = \frac{\sum_{f=0}^{\tau-1} x_{f+i+\tau(j-1)}}{\tau}. \quad (10)$$

Then the τ different coarse-grained series $z_i^{(\tau)}$ ($i = 1, \dots, \tau$) corresponding to the scale factor τ are considered in IMPE algorithm, while, in MPE algorithm, only $z_1^{(\tau)}$ is taken into account.

(2) Calculate the PE of each coarse-grained series $z_i^{(\tau)}$ ($i = 1, \dots, \tau$) corresponding to the scale factor τ separately. Then, IMPE could be obtained based on the average value of PE:

$$\text{IMPE}(X, h, q, \tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} \text{PE}(z_i^{(\tau)}, h, q). \quad (11)$$

2.3. Feature Extraction. Usually, the collected signals of the bearings with local defect are complicated and the interference between the components in the signal is inevitable. Besides, the differences among the original signals of the bearings in various operating conditions may be subtle. These factors will result in the difficulties of feature information extraction. Then the signal processing procedure combining DTCWPT with IMPE is presented to address this issue.

As a useful tool for signal processing, DTCWPT is suitable for analyzing the complicated bearing signals. The original signal could be decomposed into several subband signals using DTCWPT, and the subband signal will be simpler than the original signal. Then the interference between the components in each subband signal will be slighter than that in the original signal. And the hidden features in the original signal will be easier to be discovered by analyzing the subband signals. Therefore, DTCWPT is regarded as a preprocessing technology to analyze the original signal. And the IMPE algorithm, which can effectively evaluate the complexity and detect the dynamic changes of the signal, is used in the subsequent analysis process.

After performing DTCWPT on the original signal, each node of the wavelet packet coefficients is reconstructed at

a single level and the corresponding subband signals could be obtained. Then IMPE is further used to calculate the PE values of each subband signal at different scales. If the decomposition level of DTCWPT is l and the scale factor of IMPE is τ , then the number of the obtained subband signal is 2^l and the number of the calculated PE values of each subband signal is τ . Therefore, $2^l \times \tau$ PE values could be obtained for every original signal, and the constructed feature vectors based on these PE values could be used to comprehensively reflect the differences of the signals under different bearing conditions.

3. LLTSA for Dimension Reduction

For the classifier, the large amount of features will not only increase the computational complexity but also lead to a decline in the classification accuracy. Therefore, the dimension of the obtained feature vectors needs to be reduced. The objective of the dimension reduction in fault diagnosis mainly contains two aspects: (1) removing the disturbed and redundant information within the high dimensional feature vectors; (2) increasing the separability of the samples, namely, making different-class samples far from each other while making same-class samples close to each other.

Based on the previous analysis, in this paper, the LLTSA algorithm is utilized to compress the original vectors into the new vectors with a lower dimension. The basic idea of LLTSA is to use the tangent space in the neighborhood of a data point to represent the local geometry of the feature. Then the local manifold structures of space are lined up to construct the global coordinates [21].

Given a dataset $X_{\text{ORG}} = [x_{\text{org}1}, x_{\text{org}2}, \dots, x_{\text{org}N}]$ from Euclidean space R^m , generally, X_{ORG} , an underlying d dimensional nonlinear manifold M^d ($M^d \subset R^d$) embedded in R^m ($d < m$) exists. Then the target problem for LLTSA is to find transformation matrix A which can map the original set $X_{\text{ORG}} = [x_{\text{org}1}, x_{\text{org}2}, \dots, x_{\text{org}N}]$ in R^m to the set $Y = [y_1, y_2, \dots, y_N]$ in R^d ; that is,

$$Y = A^T X_{\text{ORG}} H_N, \quad (12)$$

where $H_N = I - ee^T/N$ represents the centering matrix, I is the identifying matrix, e is N dimensional column vector of all ones, and N denotes the number of the data.

The LLTSA algorithm procedures are described as follows.

(1) *PCA Projection.* Project the raw dataset X_{ORG} into the PCA subspace by throwing away the minor components. In order to make it clear, X is used to represent the dataset in the PCA subspace in the following steps and A_{pca} is applied to denote the transformation matrix of PCA.

(2) *Determining the Neighborhood Size.* The Euclidean distance matrix for all data points is constructed, and the k nearest neighbors x_{i_j} ($j = 1, 2, \dots, k$) belonging to the same class of point x_i ($i = 1, 2, \dots, N$) are obtained by analyzing the distance matrix.

(3) *Extracting Local Information.* Compute the tangent space matrix V_i composed of d eigenvectors of $X_i H_k$ ($X_i = [x_{i_1}, x_{i_2}, \dots, x_{i_k}]$) corresponding to its d largest eigenvalues, and $H_k = I - ee^T/k$.

(4) *Constructing Alignment Matrix.* Form matrix B by locally summing as follows:

$$B(I_i, I_i) \leftarrow B(I_i, I_i) + W_i W_i^T \quad (i = 1, 2, \dots, N), \quad (13)$$

where the initialization $B = 0$, and $I_i = \{i_1, i_2, \dots, i_k\}$ denotes the set of indices for the k nearest neighbors of x_i and $W_i = H_k(I - V_i V_i^T)$ ($i = 1, 2, \dots, N$).

(5) *Computing the Maps.* Compute the eigenvectors and the eigenvalues for the generalized eigenvalue problem as

$$X H_N B H_N X^T a = \lambda X H_N X^T a. \quad (14)$$

Then the eigenvector $\alpha_1, \alpha_2, \dots, \alpha_d$ ordered according to the eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_d$ could be obtained, and $A_{LLTSA} = (\alpha_1, \alpha_2, \dots, \alpha_d)$. Thus, the ultimate transformation matrix is as follows: $A = A_{pca} A_{LLTSA}$ and $X \rightarrow Y = A^T X_{ORG} H_N$.

Due to the good clustering performance of LLTSA, the d dimensional eigenvector set Y outputted by LLTSA can be served as the input vectors of the classifier for the pattern recognition.

4. ELM Classifier

The ELM proposed by Huang et al. [22] is a new and fast machine learning technique based on single layer feed-forward networks. A brief description of ELM is as follows.

Given a training dataset with N samples $\{x_i, y_i\}_{i=1}^N$, where $x_i \in R^d$ is the input vector and $y_i \in R^s$ stands for the target vector, the output of ELM with L hidden neurons can be represented as

$$\sum_{i=1}^L b_i g(\omega_i \cdot x_j + \beta_i) = o_j \quad (j = 1, 2, \dots, N), \quad (15)$$

where $g(\cdot)$ is the activation function, ω_i is the vector of the link weights between the i th hidden neuron and the input layer, b_i is the vector of the link weights between the i th hidden neuron and the output layer, β_i indicates the bias of the i th hidden neuron, and o_j is the output vector of the j th input sample. If ELM can approximate these samples without error, then

$$\sum_{i=1}^L b_i g(\omega_i \cdot x_j + \beta_i) = y_j \quad (j = 1, 2, \dots, N). \quad (16)$$

And (16) can be rewritten as

$$HB = Y, \quad (17)$$

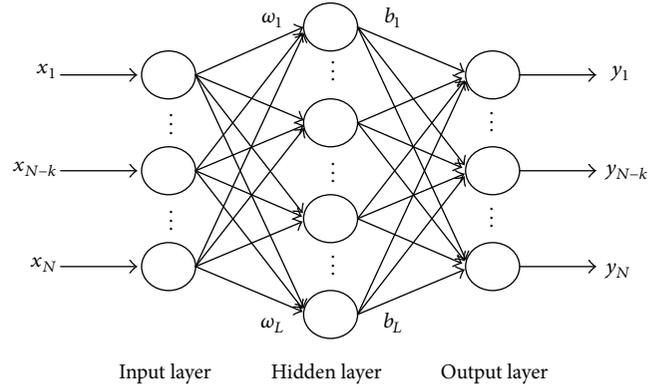


FIGURE 3: The structure of ELM.

where H denotes the output matrix of the hidden layer and can be expressed as

$$H = \begin{pmatrix} g(\omega_1 \cdot x_1 + \beta_1) & \cdots & g(\omega_L \cdot x_1 + \beta_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 \cdot x_N + \beta_1) & \cdots & g(\omega_L \cdot x_N + \beta_L) \end{pmatrix}_{N \times L} \quad (18)$$

and $B = [b_1, b_2, \dots, b_k]^T$ is the matrix of the link weights from the hidden layer to the output layer, while $Y = [y_1, y_2, \dots, y_N]^T$ is the matrix of the target vectors. Typically, B can be determined by the Moore-Penrose (MP) generalized inverse of H :

$$B = H^* T. \quad (19)$$

Then, utilizing the MP inverse method, the ELM generalization performance can be achieved. The structure of ELM is displayed in Figure 3.

5. The Proposed Fault Diagnosis Method

Based on the advantages of DTCWPT, IMPE, LLTSA, and ELM, a novel bearing fault diagnosis method is proposed in this paper, and the flow chart of this method is shown in Figure 4. The detailed procedures are described as follows.

(1) Process the collected samples using DTCWPT and acquire the corresponding subband signals. Considering the tradeoff between the classification accuracy and the computational burden, without loss of generality, the decomposition level of DTCWPT is set to 2 in this study. Then each sample is decomposed into four subband signals after performing DTCWPT.

(2) Apply IMPE algorithm to calculate the PE values of the obtained subband signals at different scales. Before using IMPE, four parameters including the embedding dimension h , the length of signal N , the time delay q , and the scale factor τ need to be set. Since h determines the number of accessible states $h!$, the estimation of PE relies heavily on the selected embedding dimension. If the dimension is too small, the scheme will not work because there are too few distinct states. When the dimension is too large, it will lead to being

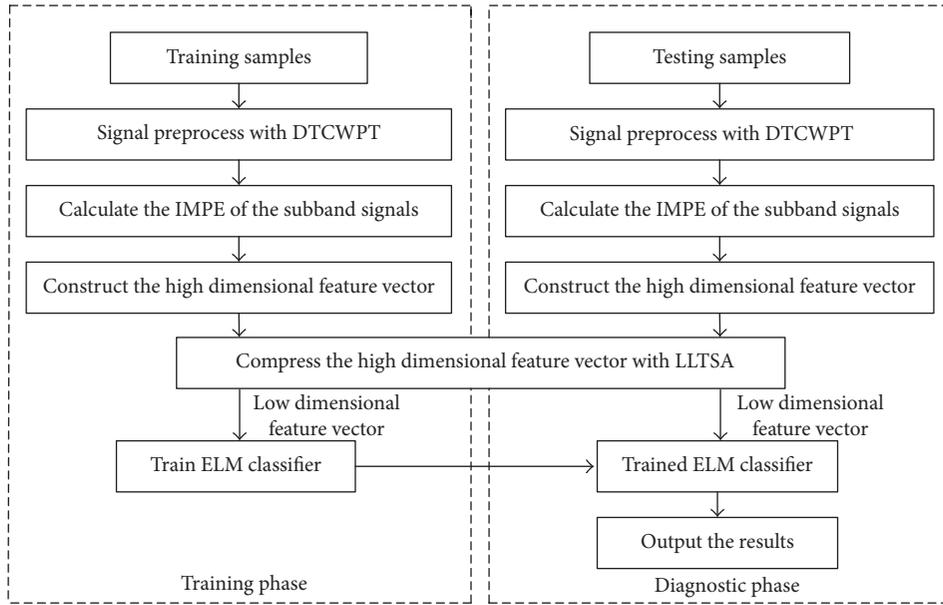


FIGURE 4: Flow chart of the proposed method.

time-consuming. To evaluate the complexity of the signal, the embedding dimension h is often chosen by tradeoff between the information loss and the computational burden. In this paper, h is set to 4. The signal length N also influences the estimation of PE. It is noticeable that N should satisfy the criterion $N \geq 5h!$ which is recommended in literature [23] to obtain a reliable statistics. However, a too large value of N will decrease the computational efficiency. The signal with 1024 points is enough to obtain a reliable result. Therefore, we set $N = 1024$ in this study. The time delay q has little effect on the calculated result; here we set $q = 1$. As for the selection of the scale factor τ , when τ is too small, the acquired feature information from the signal will be insufficient. On the other hand, if the scale factor τ is too large, the obtained PE values in large scales will be unstable. Taking these constraints into consideration, based on the criterion $\tau \leq N/(h+1)!$ proposed in literature [14] and the selected N and h , the scale factor τ is set to 8 in this paper.

(3) Combine the calculated IMPE of each subband signal and construct the feature vector for each sample. Since each sample is decomposed into four subband signals, and the number of calculated PE is 8 for each subband signal, the dimension of the constructed feature vector is 32; namely, 32 features are extracted for each sample.

(4) Utilize LLTSA algorithm to compress the dimension of the constructed feature vectors and acquire the new feature vectors with lower dimension. In LLTSA algorithm, two parameters including the neighborhood size k and the intrinsic dimension d need to be adjusted. If parameter k is too small, LLTSA cannot well discover the intrinsic structure information of the high dimensional feature vectors. Contrarily, LLTSA will lose the ability of nonlinear dimension reduction. As for the intrinsic dimension d , if this parameter is chosen larger than what it really is, much redundant information will be preserved. When it is selected

smaller, useful information of the feature vectors will be thrown out during the dimension reduction. For LLTSA, there is an approximate linear relation between the optimal neighborhood size k and the intrinsic dimension d [24]. Hence, we choose $k = d$ according to this linear relationship in this paper, using the cross validation method to determine the intrinsic dimension d .

(5) Feed the acquired new feature vectors into the ELM classifier for training and testing and distinguish the bearing condition automatically. Compared with some classic classifiers, ELM requires less human interventions. Only the number of the hidden neurons needs to be selected. Generally, as long as the number of hidden neurons is larger than 20, the classification accuracy of ELM will remain stable [25]. Therefore, the number of the hidden neurons is set to 20 in this paper.

6. Analysis on Experimental Data

6.1. Experimental Data Description. The experimental data from Case Western Reserve University are applied to verify the proposed method [26]. Figure 5 displays the experimental system, which consists of an electric motor, a torque transducer/encoder, and a dynamometer. The SKF6205-2RS deep groove ball bearing supporting the shaft at the drive end was used in the test. The rolling bearings were seeded with single point defects whose diameters were 0.1778 mm, 0.3556 mm, and 0.5334 mm, respectively, using the electric discharge machining technology. The defects were set on the inner race, the outer race, and the rolling element, respectively. An accelerometer was mounted on the motor housing to collect the vibration signals of the bearings under three different kinds of fault types as well as normal condition. The rotating speed of the motor was 1797 r/min and the sampling frequency was 12000 Hz. Every fault type contains

TABLE 1: The detailed description of the experimental datasets.

Fault type	Fault diameter (mm)	Fault degree	Number of training samples	Number of testing samples	Class label
Normal	0		10	40	1
Inner race	0.1778	Slight	10	40	2
	0.3556	Medium	10	40	3
	0.5334	Severe	10	40	4
Outer race	0.1778	Slight	10	40	5
	0.3556	Medium	10	40	6
	0.5334	Severe	10	40	7
Rolling element	0.1778	Slight	10	40	8
	0.3556	Medium	10	40	9
	0.5334	Severe	10	40	10

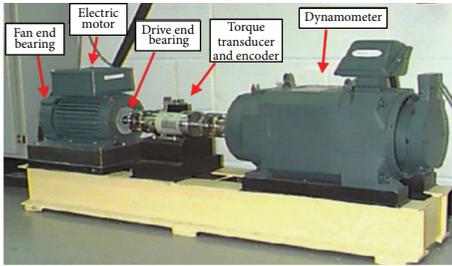


FIGURE 5: The rolling bearing experimental system.

three kinds of fault degrees with respect to different defect diameters. Therefore, a ten-condition classification problem for the bearings is investigated in this paper. The collected vibration signals are divided into several nonoverlapping segments with the length of 1024 points, and each segment is a sample. Each bearing condition includes 50 samples, from which 10 samples are randomly selected to train the classifier, while the residual 40 samples are used for testing. The detailed description of the experimental datasets is displayed in Table 1. The time domain waveforms of the samples of the bearings, respectively, with the slight inner race fault (Slight-IRF), the medium inner race fault (Medium-IRF), the severe inner fault (Severe-IRF), the slight outer race fault (Slight-ORF), the medium outer race fault (Medium-ORF), the severe outer race fault (Severe-ORF), the slight rolling element fault (Slight-REF), the medium rolling element fault (Medium-REF), and the severe rolling element fault (Severe-REF), as well as the normal condition, are shown in Figure 6, respectively.

6.2. Results and Discussions. Since the measured vibration signals of the bearings under different conditions represent the nonlinear and nonstationary characteristics, it is difficult to distinguish the different fault types and fault degrees only using the time domain waveforms in Figure 6. Therefore, it is very essential to perform an effective method to identify different operating conditions accurately. Then the proposed diagnosis method is applied.

Firstly, in order to reduce the interference among the components in the original sample and discover the hidden feature information more effectively, each sample is decomposed to 2 levels using DTCWPT. Then four subband signals containing different frequency band information could be obtained. For the sake of space, only the decomposition results of the samples with the slight inner race fault (Slight-IRF) are shown in Figure 7 as a representative.

According to the flow chart of the proposed diagnosis method indicated in Figure 1, after completing the signal decomposition and reconstruction, the IMPE algorithm is then utilized to extract the features at different scales from each subband signal for each sample. Figure 8 illustrates the IMPE values of subband signal 1, subband signal 2, subband signal 3, and subband signal 4 over 8 scales under 10 conditions. As shown in Figure 8, for each subband signal under different conditions, the divisibility among the PE values is high at some scales, while the differences of the PE values are not obvious at some scales. It is still unable to distinguish the different fault types with various fault degrees from the IMPE curves in Figure 8. Then the feature vector is constructed based on the acquired PE values of four subband signals for each sample. And a multifault classifier is applied to recognize the different bearing conditions. In this paper, the ELM classifier is used to achieve this purpose.

If the constructed feature vectors containing 32 PE values are directly taken as the inputs of the classifier, it will be time-consuming. Even worse, ELM cannot effectively distinguish the conditions of samples since feature vectors inevitably contain certain interference and redundancy information. Then LLTSA is further employed to compress the high dimensional feature vectors.

Before using LLTSA, an important problem about selecting the intrinsic dimension d of the original feature vectors needs to be addressed. In this paper, this parameter is determined using a fivefold cross validation method [27, 28]. That is, the 100 training samples are randomly divided into five equal-sized subsets. Each subset is validated on the ELM classifier that was trained using the other four subsets. The process was repeated 5 times; the accuracy rate of the classifier is then obtained by means of averaging the recorded accuracy

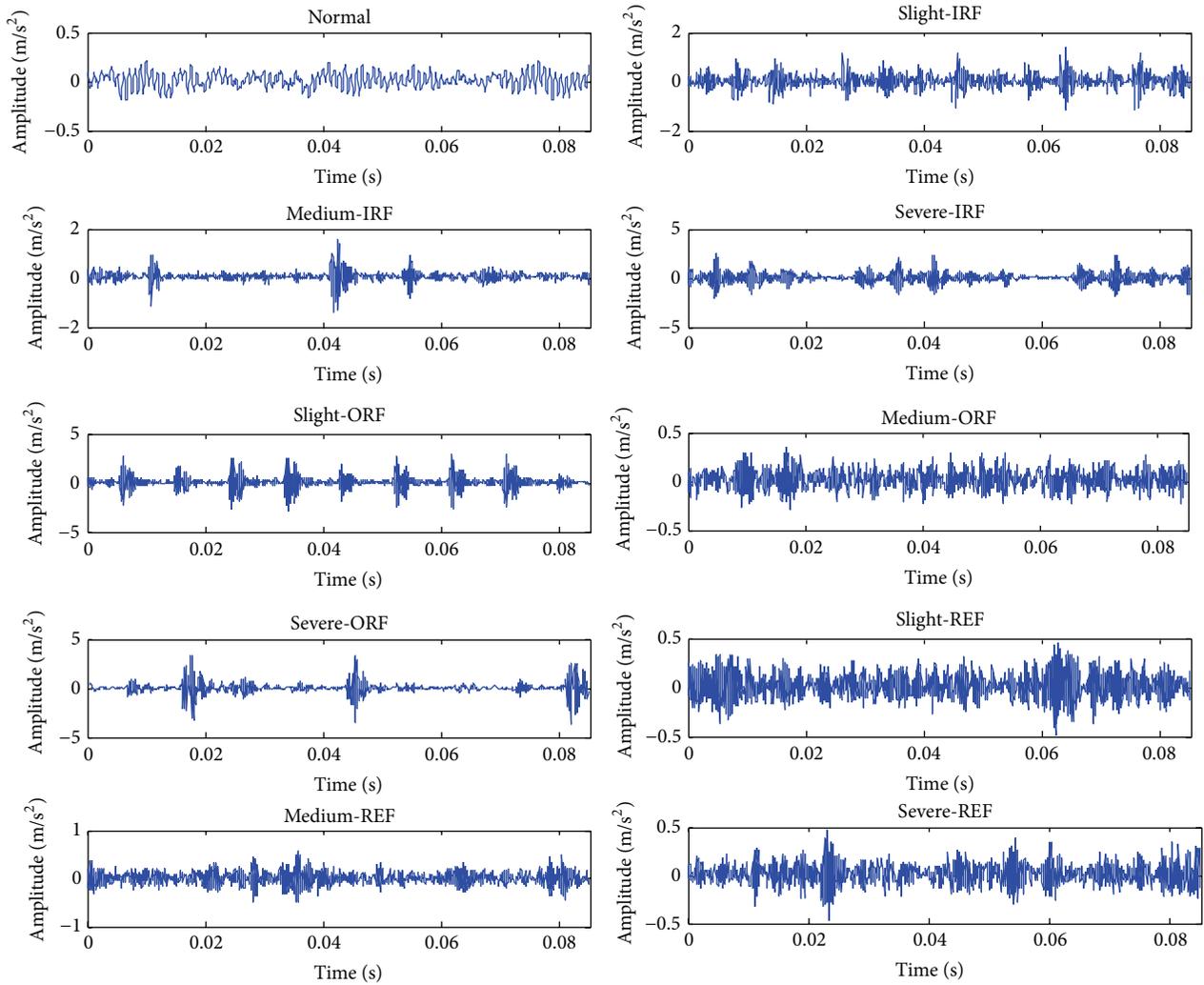


FIGURE 6: The time domain waveforms of the samples of the bearings under ten kinds of different conditions.

rate in each testing fold. Finally, choose parameter d which provides the best classification accuracy. In this paper, the intrinsic dimension d in the five-fold validation varies in the interval $[3, m/2]$ with an incremental step size of 1, where $m = 32$ denotes the dimension of the original feature vectors. Figure 9 shows the curve of the classification accuracy versus the intrinsic dimension. As indicated in Figure 9, the accuracy reaches 100% when the intrinsic dimension is larger than 10. In order to avoid information redundancy as far as possible, we select the intrinsic dimension $d = 10$; then the neighborhood size k is set to 10 according to the approximate linear relationship between the intrinsic dimension and the neighborhood size.

After parameter selections, LLTSA is performed on the constructed feature vectors. Then the original high dimensional feature vectors are projected into a low dimensional space, based on which the new 10 dimensional feature vectors could be obtained. Then feed them into ELM for training and testing. After training the classifier with the 100 feature vectors of the training samples, the remaining 400 feature vectors of the testing samples are used to test the ELM

classifier. The classification results of the classifier are shown in Figure 10, where the red asterisks denote the ELM actual output classifications of the samples, while the blue squares represent the desired output classifications. The 100 samples on the left side and the 400 samples on the right side of the dotted line are, respectively, the training samples and the testing samples. It is suggested that, for each sample, the actual ELM output classification is consistent with the desired one. There is no misclassified sample, and the recognition accuracy achieves a perfect level of 100%. The proposed method obtains perfect classification results, which means that this method is exactly suitable and effective in bearing fault diagnosis.

In order to verify the advantage of IMPE, as a representative, a comparison is taken between IMPE and MPE by analyzing the 50 independent subband signals of the bearing with the slight inner race fault. The selected parameters in MPE algorithm are the same as those in IMPE algorithm. Figure 11 represents the mean values and the standard deviations of the PE values, using the IMPE and the MPE algorithms. The following conclusions can be drawn from

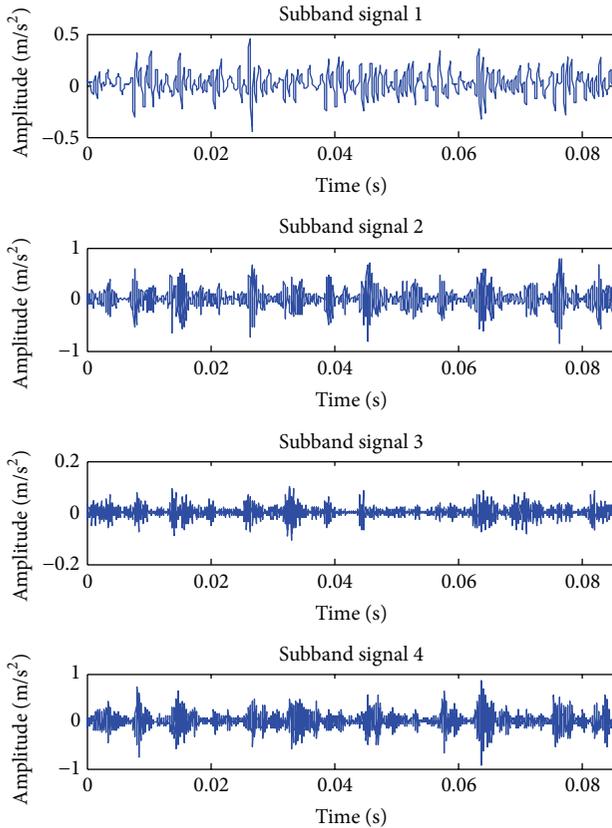


FIGURE 7: DTCWPT results of the sample with the slight inner race fault.

Figure 11. Firstly, the mean curves of the PE values derived from IMPE are really close to those derived from MPE. Secondly, compared with the MPE algorithm, the IMPE algorithm is able to get smaller standard deviations of the PE values. These conclusions can also be drawn through analyzing the subband signals of the bearings under the other conditions. It is indicated that the IMPE algorithm is more stable than the traditional MPE algorithm, which means that IMPE can provide a more accurate PE estimation on the nonlinear and nonstationary signals.

To further illustrate the advantage of IMPE, the feature vectors extracted by the processing method based on DTCWPT, MPE, and LLTSA are also fed into the ELM classifier to distinguish the bearing conditions. The actual ELM output classifications and the desired output classifications of the training and the testing samples are shown in Figure 12. On the right side of the dotted line, the locations of eight red asterisks are inconsistent with those of the blue squares. It is indicated that eight testing samples are misclassified and the classification accuracy is 98%. It can be easily observed from Figure 12 that two testing samples with Slight-IRF are misclassified as Medium-IRF, two testing samples with Slight-ORF are misclassified as Medium-IRF and Medium-ORF, a testing sample with Severe-ORF is misclassified as Medium-IRF, and three testing samples with Medium-REF are misclassified as Slight-REF and Medium-IRF, respectively.

The comparison results displayed in Figures 10 and 12 provide compelling evidence that IMPE can provide more accurate estimation of entropy values with higher distinguishability than MPE. These analysis results can be explained by the fact that when the MPE algorithm is used to analyze the short term series, the calculation points will be decreased exponentially as the scale factor is increased. It can not only give rise to the questionable and uncertain estimations of the entropy values but also increase the standard deviations of the features. However, the IMPE algorithm is able to avoid the drawbacks of MPE effectively and result in better classification accuracy.

To validate the necessity of the dimension reduction using LLTSA, the constructed original feature vectors without dimension reduction are adopted as the inputs of the ELM classifier for a comparison. The classification results are displayed in Figure 13, from which it can be seen that three testing samples with the rolling element fault are misclassified into the wrong fault degrees. The accuracy is 99.25%, which is lower than that of the method with dimension reduction. In the process of the dimension reduction, LLTSA can get the low dimensional sensitive feature vectors from the high dimensional feature vectors with interference and redundancies. Therefore, the recognition precision of ELM could be improved. It is indicated that the dimension reduction using LLTSA is of benefit for the bearing condition classification. Also, the necessity of this procedure is demonstrated at the same time.

In addition, in order to verify the superiority of the proposed feature extraction method based on DTCWPT, IMPE, and LLTSA, the calculated IMPE values of the original samples are directly taken as the input feature vectors of the ELM classifier. The 100 training samples and the 400 testing samples, as well as the selected parameters, remain the same as mentioned previously. The actual ELM output classification and the desired output classification of all the samples are shown in Figure 14, where 28 testing samples on the right side of the dotted line are misclassified. The recognition accuracy is 93%. It is shown that the extracted features of the samples directly using IMPE cannot completely reflect the distinctions of different bearing conditions. Thus, the obtained classification results of the ELM classifier are unsatisfied. This comparison demonstrates the superiority of the proposed feature extraction method which combines IMPE with DTCWPT and LLTSA due to the abilities of DTCWPT and LLTSA in restraining the interference among the components and highlighting the feature information of the samples.

Finally, the recognition accuracies of ELM, SVM, ANN, and KNNC using different feature extraction methods are compared. The training and the testing samples are the same for each comparison. And the feature vectors taken as the inputs of these classifiers are extracted by four different methods, respectively. The first method is the proposed method used in this paper, that is, the combination of DTCWPT, IMPE, and LLTSA (DTCWPT + IMPE + LLTSA). The second method utilizes MPE instead of IMPE and obtains the feature vectors based on DTCWPT, MPE, and LLTSA (DTCWPT + MPE + LLTSA). The third method extracts the feature vectors

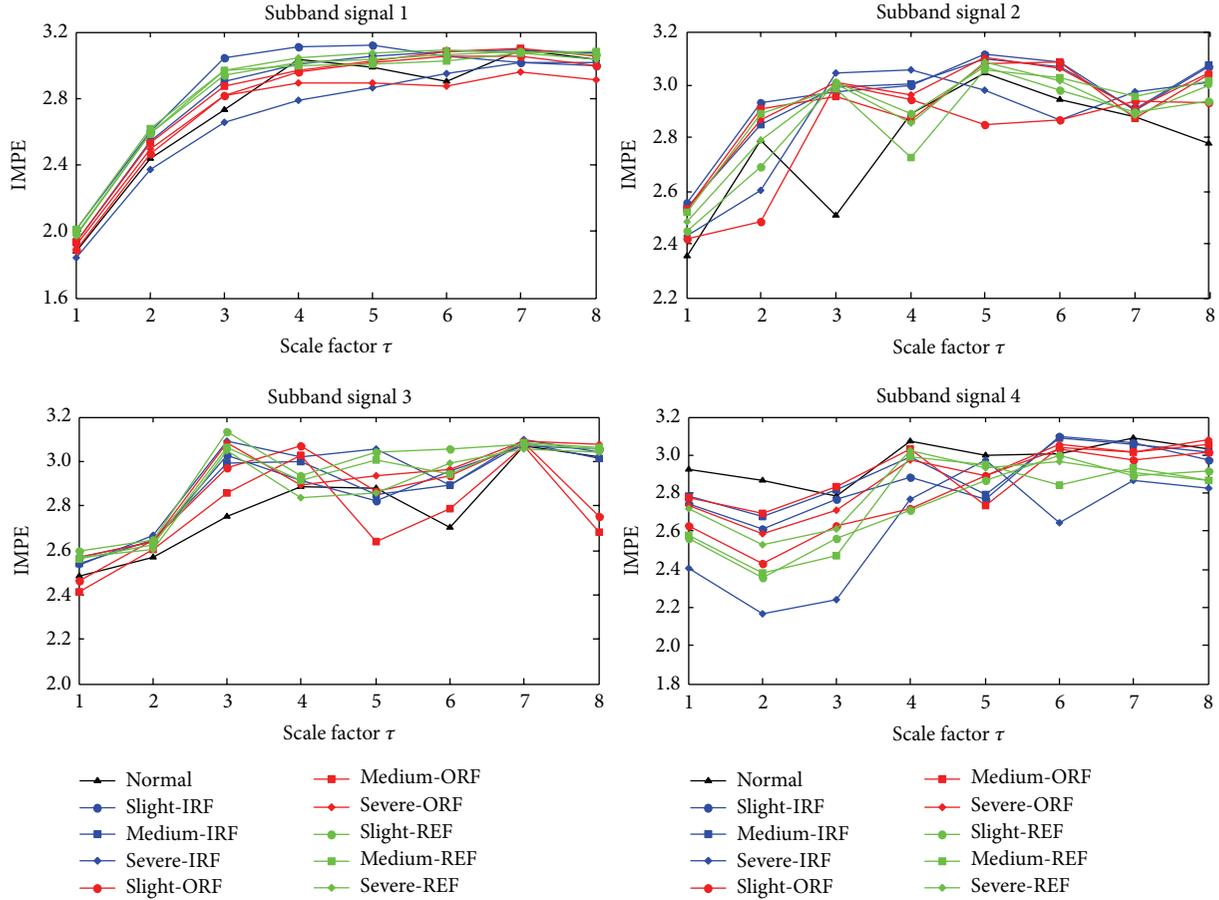


FIGURE 8: IMPE of each subband signal over 8 scales under 10 different conditions.

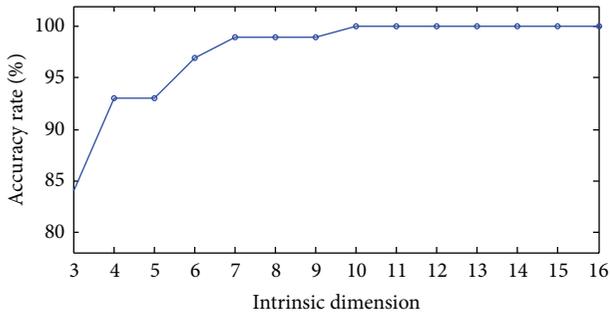


FIGURE 9: Curve of classification accuracy versus intrinsic dimension.

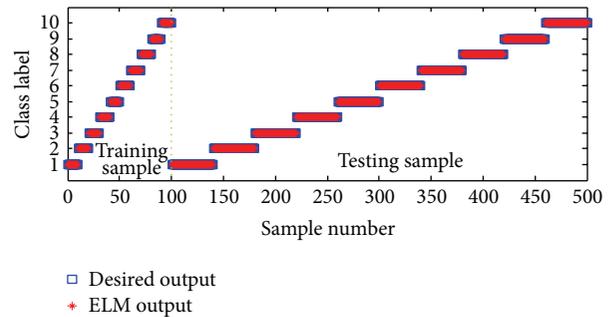


FIGURE 10: Classification results of the proposed method.

using DTCWPT and IMPE without LLTSA (DTCWPT + IMPE). The last method treats the calculated IMPE of the samples as the feature vectors (IMPE). The parameters of SVM are chosen as follows: the penalty factor C is set to 100 and the RBF kernel parameter γ is set to 0.01 [29]. The parameters of ANN are selected as follows: the number of the hidden neurons $N = 20$, the maximum number of the iterations $I = 500$, the learning rate $\alpha = 0.1$, and the training error $e = 0.001$ [30]. The neighborhood number k of KNNC is set to 7 [31]. The classification results of ELM, SVM, ANN,

and KNNC using different feature extraction methods are shown in Table 2 and Figure 15. No matter what kind of method, it can be noted that the classification accuracy of ELM is higher than that of the other three classifiers. This verifies the advantage of ELM in classification performance.

It is suggested from Table 2 and Figure 15 that, using the feature vectors extracted through the first method, the testing accuracies of the classifiers are all higher than those of the classifiers using the feature vectors extracted via the other three kinds of methods. On one hand, for the average testing accuracies of the four classifiers, the first feature extraction

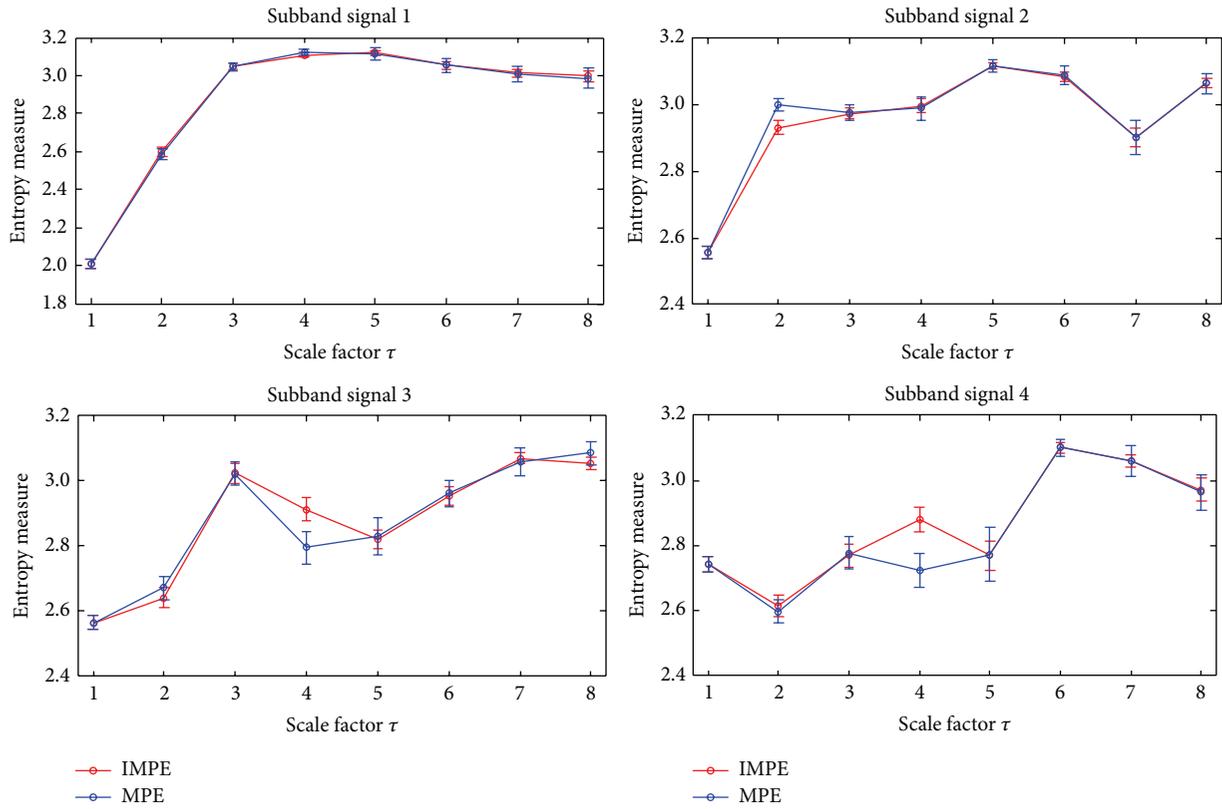


FIGURE 11: MPE and IMPE comparison results of the samples of the bearing with slight inner race fault.

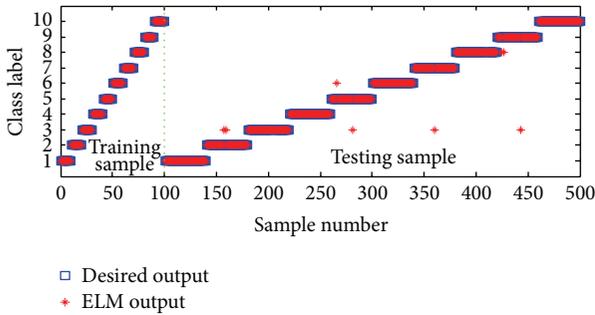


FIGURE 12: Classification results of the method based on DTCWPT, MPE, and LLTSA.

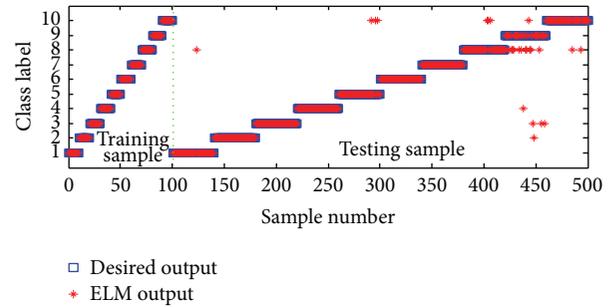


FIGURE 14: Classification results of the method based on IMPE.

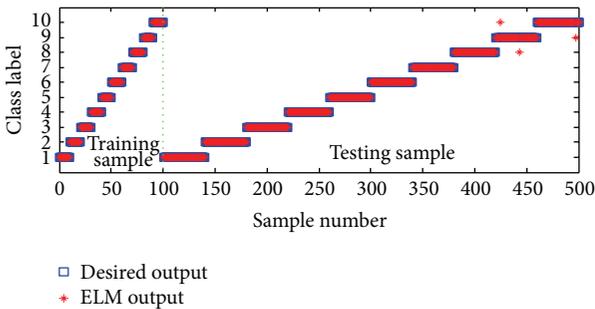


FIGURE 13: Classification results of the method based on DTCWPT and IMPE.

method is 2.19%, 0.62%, and 7.93%, better than the second, the third, and the last method, respectively, which in turn verifies the advantage of the presented feature extraction method based on DTCWPT, IMPE, and LLTSA.

7. Conclusions

IMPE is a recently proposed novel technique for evaluating the complexity and detecting the dynamic changes of time series. Its application in bearing fault diagnosis is firstly investigated in this work. And a novel fault diagnosis method for rolling bearings combining IMPE with DTCWPT, LLTSA, and ELM is proposed in this paper. Focusing on the nonlinear and nonstationary characteristics of the bearing vibration

TABLE 2: Classification results of different classifiers with feature vectors extracted by different methods.

Classifier	Testing accuracies with feature vectors extracted by different methods			
	DTCWPT + IMPE + LLTSA	DTCWPT + MPE + LLTSA	DTCWPT + IMPE	IMPE
ELM	100%	98%	99.25%	93%
SVM	99.75%	97.75%	99.25%	91.5%
ANN	99%	96%	98.25%	90.5%
KNNC	99.5%	97.75%	99%	91.5%
Average of four classifiers	99.56%	97.37%	98.94%	91.63%

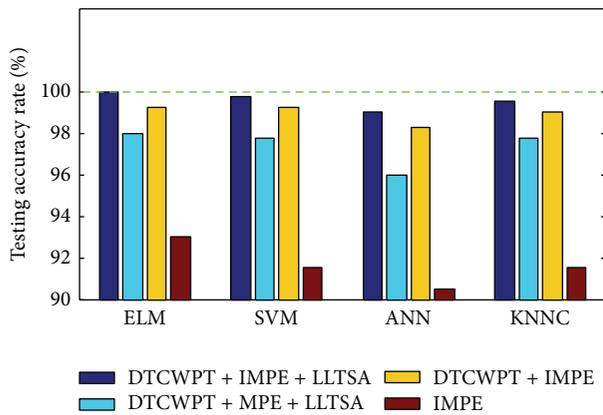


FIGURE 15: Classification results of different classifiers with feature vectors extracted by different methods.

signals, DTCWPT is employed to preprocess the signal and obtain the corresponding subband signals. IMPE is then taken as the feature extractor to calculate the PE values of each subband signal at different scales. To solve the dimension reduction problem of the constructed feature vectors, LLTSA is applied to compress the high dimensional vectors and sift out the principal sensitive features used to construct the new low dimensional vectors. Besides, the ELM classifier is adopted to implement the condition identification. For comparison purpose, the presented feature extraction method is compared with other methods. The comparison results indicate that the presented method is able to obtain the feature vectors with a higher divisibility. Also, the classification performance of the ELM classifier is also compared with other widely used classifiers, and the advantage of ELM is verified by the comparison results. The experimental data analysis results demonstrate that the proposed fault diagnosis method in this paper is suitable and effective in recognizing the different fault types and fault degrees of rolling bearings.

In the highly automated industry, since the proposed diagnosis method is data-driven without operators' experiences, it is much easier to be widely used. It is mentioned that the proposed method is a promising approach, which is not limited to rolling bearing fault diagnosis but also could be applied in fault diagnosis of other mechanical equipment.

To some extent, limited by the consumption of computer resources, the proposed diagnosis method may not be satisfactory enough in real time. In addition, only the constant working load is discussed in this paper. If the working load is dramatically changed, the accuracy and the efficiency of the proposed method may be influenced. Consequently, further studies will be focused on solving this problem.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (no. 51307058 and no. 51475164), the Natural Science Foundation of Hebei Province, China (no. 2014502052 and no. 2015502013), and the Fundamental Research Funds for the Central Universities (no. 2015ZD27 and no. 2015XS120).

References

- [1] G. J. Tang, X. L. Wang, and Y. L. He, "Diagnosis of compound faults of rolling bearings through adaptive maximum correlated kurtosis deconvolution," *Journal of Mechanical Science and Technology*, vol. 30, no. 1, pp. 43–54, 2016.
- [2] X. Ding, Q. He, and N. Luo, "A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification," *Journal of Sound and Vibration*, vol. 335, pp. 367–383, 2015.
- [3] J. Zhou, J. Xiao, H. Xiao, W. Zhang, W. Zhu, and C. Li, "Multi-fault diagnosis for rolling element bearings based on intrinsic mode permutation entropy and ensemble optimal extreme learning machine," *Advances in Mechanical Engineering*, vol. 6, Article ID 803919, 10 pages, 2014.
- [4] J. Q. Kang, F. Sun, and Y. B. Liu, "Fault diagnosis of the rolling bearing based on correlation dimension improved algorithm," *Chinese Journal of Scientific Instrument*, vol. 27, no. 6, pp. 404–406, 2006.
- [5] 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.
- [6] Y.-H. Pan, Y.-H. Wang, S.-F. Liang, and K.-T. Lee, "Fast computation of sample entropy and approximate entropy in

- biomedicine,” *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 382–396, 2011.
- [7] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate and sample entropy,” *American Journal of Physiology—Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [8] 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.
- [9] L. Zhang, G. L. Xiong, H. S. Liu, H. J. Zou, and W. Z. Guo, “Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference,” *Expert Systems with Applications*, vol. 37, no. 8, pp. 6077–6085, 2010.
- [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] W. Aziz and M. Arif, “Multiscale permutation entropy of physiological time series,” in *Proceedings of the 9th International Multitopic Conference (INMIC '05)*, pp. 1–6, IEEE, Karachi, Pakistan, December 2005.
- [12] Y. B. Li, M. Q. Xu, Y. Wei, and W. H. Huang, “A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree,” *Measurement*, vol. 77, pp. 80–94, 2016.
- [13] J. Zheng, J. Cheng, and Y. Yang, “Multiscale permutation entropy based rolling bearing fault diagnosis,” *Shock and Vibration*, vol. 2014, Article ID 154291, 8 pages, 2014.
- [14] H. Azami and J. Escudero, “Improved multiscale permutation entropy for biomedical signal analysis: interpretation and application to electroencephalogram recordings,” *Biomedical Signal Processing and Control*, vol. 23, pp. 28–41, 2016.
- [15] J. D. Zheng, J. Cheng, and Y. Yang, “A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy,” *Mechanism and Machine Theory*, vol. 70, pp. 441–453, 2013.
- [16] 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.
- [17] 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.
- [18] F. Li, B. P. Tang, and R. S. Yang, “Rotating machine fault diagnosis using dimension reduction with linear local tangent space alignment,” *Measurement*, vol. 46, no. 8, pp. 2525–2539, 2013.
- [19] H. L. Chen, B. Yang, D. Y. Liu et al., “Using blood indexes to predict overweight statuses: an extreme learning machine based approach,” *PLOS ONE*, vol. 10, no. 11, Article ID e0143003, 2015.
- [20] J. X. Qu, Z. S. Zhang, and T. Gong, “A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion,” *Neurocomputing*, vol. 171, pp. 837–853, 2016.
- [21] T. H. Zhang, J. Yang, D. L. Zhao, and X. L. Ge, “Linear local tangent space alignment and application to face recognition,” *Neurocomputing*, vol. 70, no. 7–9, pp. 1547–1553, 2007.
- [22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [23] M. Matilla-García, “A non-parametric test for independence based on symbolic dynamics,” *Journal of Economic Dynamics & Control*, vol. 31, no. 12, pp. 3889–3903, 2007.
- [24] X. Q. Lu and Y. Yuan, “Hybrid structure for robust dimensionality reduction,” *Neurocomputing*, vol. 124, pp. 131–138, 2014.
- [25] X. He, H. L. Wang, J. H. Lu, and W. Jiang, “Analog circuit fault diagnosis method based on preferred wavelet packet and ELM,” *Chinese Journal of Scientific Instrument*, vol. 34, no. 11, pp. 2614–2619, 2013.
- [26] Bearing Data Center, Case Western Reserve University, 2006, <http://cseggroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [27] Y. B. Li, M. Q. Xu, R. X. Wang, and W. H. Huang, “A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy,” *Journal of Sound and Vibration*, vol. 360, Article ID 12638, pp. 277–299, 2016.
- [28] Y. Zhang, B. Li, W. Wang, T. Sun, X. Yang, and L. Wang, “Supervised locally tangent space alignment for machine fault diagnosis,” *Journal of Mechanical Science and Technology*, vol. 28, no. 8, pp. 2971–2977, 2014.
- [29] S. J. Dong, L. L. Chen, B. P. Tang, X. Y. Xu, Z. Y. Gao, and J. Liu, “Rotating machine fault diagnosis based on optimal morphological filter and local tangent space alignment,” *Shock and Vibration*, vol. 2015, Article ID 893504, 9 pages, 2015.
- [30] K. Zhao, C. Y. Wang, J. Hu et al., “Prostate cancer identification: quantitative analysis of T2-weighted MR images based on a back propagation artificial neural network model,” *Science China Life Sciences*, vol. 58, no. 7, pp. 666–673, 2015.
- [31] F. Li, J. X. Wang, B. P. Tang, and D. Q. Tian, “Life grade recognition method based on supervised uncorrelated orthogonal locality preserving projection and K-nearest neighbor classifier,” *Neurocomputing*, vol. 138, pp. 271–282, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

