Research Article

Rolling Bearing Fault Diagnosis Using Modified Neighborhood Preserving Embedding and Maximal Overlap Discrete Wavelet Packet Transform with Sensitive Features Selection

Fei Dong,1,2 Xiao Yu,1,2,3 Enjie Ding,1,2 Shoupeng Wu,1,2 Chunyang Fan,1,2 and Yanqiu Huang4

1School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China
2IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China
3School of Medicine Information, Xuzhou Medical University, Xuzhou 221000, China
4Institute of ElectroDynamics and Microelectronics, University of Bremen, 28359 Bremen, Germany

Correspondence should be addressed to Enjie Ding; enjied@cumt.edu.cn

Received 20 December 2017; Accepted 6 February 2018; Published 26 March 2018

Academic Editor: Simone Cinquemani

Copyright © 2018 Fei Dong 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.

In order to enhance the performance of bearing fault diagnosis and classification, features extraction and features dimensionality reduction have become more important. The original statistical feature set was calculated from single branch reconstruction vibration signals obtained by using maximal overlap discrete wavelet packet transform (MODWPT). In order to reduce redundancy information of original statistical feature set, features selection by adjusted rand index and sum of within-class mean deviations (FSASD) was proposed to select fault sensitive features. Furthermore, a modified features dimensionality reduction method, supervised neighborhood preserving embedding with label information (SNPEL), was proposed to realize low-dimensional representations for high-dimensional feature space. Finally, vibration signals collected from two experimental test rigs were employed to evaluate the performance of the proposed procedure. The results show that the effectiveness, adaptability, and superiority of the proposed procedure can serve as an intelligent bearing fault diagnosis system.

1. Introduction

Bearings are one of the most crucial elements of rotating machinery [1, 2] and bearing faults can seriously affect safe and stable operations of the rotary mechanical equipment [3, 4]. If no effective actions are taken, device faults will inevitably occur, and such faults may lead to serious casualties and enormous pecuniary loss [5]. Thus, it is of significance to identity bearing faults to maintain safety of the device and reduce maintenance cost. Vibration signals collected from rolling bearings usually carry rich information on machine operation conditions [6]. In recent years, with the rapid development of signal processing, data mining, and artificial intelligence technology, the data-driven methods are becoming more important in the fault diagnosis of rolling bearings. Four main steps are necessary for these methods based on vibration signals analysis: signal processing, features extraction, features reduction, and patterns recognition [7, 8]. The first three steps are the foundation of patterns recognition.

In the phase of signal processing and features extraction, due to the complexity of equipment structure and variety of operation conditions [5], the signals collected from rolling bearings often exhibit strong nonlinearity and nonstationarity. Therefore, the time-domain and frequency-domain analysis approaches cannot have essential effects [9]. For these signals, time-frequency analysis can provide an effective way for features extraction. There are representative and commonly used time-frequency analysis methods, such as empirical mode decomposition (EMD), short-time Fourier transform (STFT), Wigner-Ville distribution (WVD), and wavelet transform (WT) [10].

In recent years, various intelligent fault diagnosis systems based on EMD [11–15], STFT [16–18], and WVD [19–21]
have been widely developed for monitoring the condition of bearings in rotating machines with varying degrees of success. However, for these time-frequency methods, some challenges exist in the application. EMD has some problems such as over envelope, end effects, and mode mixing [22–24]. The effectiveness of STFT is still hampered by the limitation of single triangular basis [25, 26]. WVD can produce interference terms on the time-frequency domain in a critical condition and high computational complexity [27]. Wavelet analysis is another important time-frequency analysis method, and it is outstanding in rotary machine diagnosis because its multiresolution merit is suitable for analyzing nonlinear and nonstationary signals [28]. Continuous wavelet transform (CWT) and discrete wavelet transform (DWT) are two categories of WT. They have perfect local properties in both time and frequency spaces and can be used as an effective method to preserve signal characteristics [27]. In [29], wavelet filtering to detect periodical impulse components from vibration signals was presented. In [30], the DWT for extracting the rotor bar faults feature was studied. In [31], the CWT and the wavelet coefficients of signals are used to process vibration signals. However, both CWT and DWT have drawbacks. CWT can generate redundant data. Therefore, it has a huge operand and requires a long time to use [31, 32]. Although DWT can overcome this drawback by using the decomposition of the original complex signal to several resolutions [33, 34], DWT requires the sample size to be exactly a power of 2 for the full transform because of the downsampling and has very poor frequency resolution at high frequencies [35–37]. In order to overcome these drawbacks, a new wavelet-based algorithm is developed, namely, maximal overlap discrete wavelet packet transform (MODWPT) [38]. It not only provides better frequency resolution, but also has no restriction about sample size [36, 38]. In [36], simulation signals and gear fault vibration signals collected form a test stand are decomposed into a set of monocomponent signals by MODWPT; then the corresponding Hilbert spectrum is applied for gear fault diagnosis; the simulation and practical application examples show that the Hilbert spectrum based on MODWPT is superior to EMD. However, the time-frequency analysis methods mentioned above can cause a high-dimensional feature vector that can be a primary reason for fault classification accuracy degradation [39]. Thus features selection or dimensionality reduction is needed to find the most useful fault features that can keep intrinsic information about the defects.

Generally, the statistical properties of the signal in time, frequency, and time-frequency domain are extracted to represent features information, such as peak value (PV), root mean square (RMS), variance (V), skewness (Sw), and kurtosis (K). In [40], 21 time-domain statistical characteristics are extracted from different IMFs obtained by EMD as the feature vectors. Then, principle component analysis (PCA) was employed to extract the dominant components from statistical characteristics for gear faults detection. In [41], two time-domain and two frequency-spectrum statistical characteristics are selected as the features to train the SVM with a novel hybrid parameter optimization algorithm for fault diagnosis of the rolling element bearings. In [31], the statistical parameters of the wavelet coefficients in 1–64 scales were calculated for the vibration signal. In [42], 40 statistical features of wavelet packet coefficients were calculated for a single sample for each state of bearing condition. In [43], for each wavelet packet node, 10 statistical features are extracted from its associated wavelet packet coefficients and 10 statistical features are extracted from frequency spectra of its associated wavelet packet coefficients. However, considering the complex mapping relations between some bearing faults and their signs, it is often difficult to determine which statistical property is worthy of reflecting the fault nature from the feature space. If unsuitable features are used for fault diagnosis, it may lead to a decline in accuracy and efficiency of fault diagnosis [10, 44]. Therefore, how to select the fault sensitive statistical characteristics as the basis of subsequent fault analysis garners considerable attention and is further studied. In this paper, a features extraction method, features selection by adjusted rand index and sum of within-class mean deviations (FSASD), is proposed. FSASD combines the K-means method and sum of within-class mean deviations (SWD) of feature data, which can select the sensitive statistical characteristics for fault analysis.

For the high-dimensional statistical characteristics data, if these data are used directly in fault classification, it will lead to the very high computational complexity and fault classification accuracy degradation. Therefore, features dimensionality reduction is another crucial stage in the fault diagnosis process [21]. Up to now, dimension reduction algorithms for machinery fault diagnosis have been intensively investigated [46, 47] and many classical methods have been proposed [48]. Principal component analysis (PCA) and linear discriminant analysis (LDA), as two classical linear dimensionality reduction methods, have been widely used for linear data; when the distribution of a dataset is nonlinear, PCA and LDA may be invalid [49]. Therefore, recently, some nonlinear dimensionality reduction methods, kernel principal components analysis (KPCA), Isomap, Laplacian Eigenmaps (LE), and Local Linear Embedding (LLE), and so on, are presented to provide a valid solution for the dimensionality reduction of nonlinear data [12]. Although nonlinear dimensionality reduction methods have been successfully applied in many fields, they also have some problems in practical applications, such as the problem of “out-of-sample” that has no explicit mapping matrix [50], the problem of overlearning of locality [51], and high computational complexity. Inspired by nonlinear dimensionality reduction methods, a lot of linear unsupervised dimensionality reduction methods based on manifold learning are proposed [52], such as neighborhood preserving embedding (NPE) [53], orthogonal neighborhood preserving projection (ONPP) [54], and locality preserving projections (LPP) [55]. They are the representative ones, which preserve the local geometric structure on the data manifold using linear approximation to the nonlinear mappings [52]. In recent years, some other manifold learning-based dimensionality reduction methods are presented to provide valid solutions for dimensionality reduction. In [56], a novel supervised method, called locality preserving embedding (LPE), is proposed and gives a low-dimensional embedding for discriminative multiclass
submanifolds and preserves principal structure information of the local submanifolds. In [57], maximal local interclass embedding (MLIE) is proposed. MLIE can be viewed as a linear method of a multimanifold-based learning framework, in which the information of neighborhood is integrated with the local interclass relationships [57]. In [52], a general sparse subspace learning framework, called sparse linear embedding (SLE), is proposed and can integrate the local geometric structure to obtain sparse projections. And the ONPP is taken as an example to design a novel sparse subspace learning framework [52]. In [58–62], some supervised and semisupervised dimensionality reduction methods based on NPE are proposed. NPE, as a manifold learning method, is a kind of linear approximation of LLE by replacing the nonlinear mapping relation to achieve dimensionality reduction [53, 63]. NPE aims at preserving the local neighborhood structure on the data manifold, and it can work well with multimodal data. In [63], the NPE is applied for bearing fault identification and classification and performs well in feature extraction. However, NPE could not utilize the label information in dimensionality reduction [57]. LDA is a supervised dimensionality reduction method and takes the label information into account in features reduction. Based on the respective attributes of NPE and LDA, supervised neighborhood preserving embedding with label information (SNPEL), a modified NPE, is proposed in this paper, where the fault label information is considered.

The contribution of this paper is the development of an intelligent fault diagnosis system of rolling bearings based on multidomain features, systematically combining statistical analysis methods with artificial intelligence techniques. FSASD, a novel features extraction method, was proposed to select the fault sensitive statistical characteristics as the basis of subsequent fault analysis. A modified features reduction method, SNPEL, was proposed to excavate abundant and valuable information with low dimensionality. The execution of the proposed bearing fault diagnosis method is divided into four steps: signal processing, features extraction, features reduction, and fault patterns identification. In the first step, vibration signals collected from bearings are decomposed into different terminal nodes by MODWPT, and multidomain features were calculated from the reconstructed signal. In the second step, the adjusted rand index (ARI) criterion of the clustering method and SWD of samples were used to select fault sensitive statistical characteristics, which can represent the fault peculiarity under different working conditions. Furthermore, due to information redundancy and a high-dimensional dataset, the third step, SNPEL was applied to obtain a new lower-dimensional space in which the new constructed features were obtained by transformations of the original higher-dimensional features such that certain properties were preserved. Finally, vibration signals collected from two test rigs were conducted to validate the effectiveness, adaptability, and superiority of the proposed method for the identification and classification of bearing faults. The first test rig is from Case Western Reserve University; four cases with 12 working conditions were employed to verify the performance of the proposed method. The second test rig is SQI-MFS test rig; two cases with 10 working conditions were employed to verify the performance of the proposed method. The analysis results for the vibration signals of roller bearing under different working conditions show the effectiveness, adaptability, and superiority of the proposed fault diagnosis approach.

The rest of this paper is organized as follows. In Section 2, a theoretical background of the LDA technique, NPE technique, and SVM is summarized. In Section 3, a description of the proposed diagnosis technique is given, and the system framework of the proposed method is illustrated. In Section 4, bearing faulty vibration signals collected from two experimental test rigs are employed to verify the proposed fault diagnosis method. Finally, some conclusions are drawn in Section 5.

2. Theoretical Background

2.1. Bearing Fault Effects on the Vibration in Frequency Domain. For the bearing, the inner race, outer race, ball, and cage which are placed in the space between the rings make rotating possible. However, due to the inappropriate lubrication of the bearing rolling elements, inadequate bearing selection improper mounting, indirect failure and material defects, and manufacturing errors, various defects can occur [21], such as surface fatigue damage, bonding, and wear. The most common of these faults is the surface fatigue damage, which is further categorized as spalling, crack, or other abnormal conditions [64]. When a fault appears on the surface of bearing, the cyclical impulsive vibration emerges. The frequency of the impulsive vibration is known as the fault symptoms, of which the value depends on the fault size, rotational speed, and damage location [65].

For different bearing components (i.e., outer race, inner race, and ball, as shown in Figure 1), main fault frequencies are the cage fault frequency (CF), the inner raceway fault frequency (IRFF), the outer raceway fault frequency (ORFF), and the ball/roller fault frequency (BRFF). When the outer ring is fixed, the aforementioned fault frequencies are mathematically described as

\[
\begin{align*}
\text{CFF} & = \frac{f_r}{2} \left( 1 - \frac{d}{D_m} \cos \alpha \right) = \frac{f_r}{2N_b}, \\
\text{IRFF} & = \frac{f_r}{2} N_b \left( 1 + \frac{d}{D_m} \cos \alpha \right), \\
\text{ORFF} & = \frac{f_r}{2} N_b \left( 1 - \frac{d}{D_m} \cos \alpha \right), \\
\text{BRFF} & = \frac{f_r D_m}{2d} \left( 1 - \left( \frac{d}{D_m} \cos \alpha \right)^2 \right),
\end{align*}
\]  

(1)

where \( f_r \) is the motor driving frequency or rotational frequency of shaft, \( d \) is the ball/roller diameter, \( D_m \) is the pitch diameter, \( N_b \) is the number of rolling elements, and \( \alpha \) is the ball contact angle (zero for rollers) [21]. Therefore, a lot of research work has been carried out based on vibration signal for bearings fault analysis.
2.2. Maximal Overlap Discrete Wavelet Packet Transform (MODWPT).

WT can be treated as a fast-evolving mathematical and signal processing tool in dealing with nonstationary signals [66] and has been widely applied in many engineering fields for decomposing, denoising, and signal analysis over nonstationary signals [26, 42]. Continuous wavelet transform (CWT) and discrete wavelet transform (DWT) are two categories of WT. The CWT have some drawbacks; one of these is that CWT generates redundant data. Therefore, it has a huge operand and requires a long time to use [31, 32]. The DWT can overcome this drawback by using the decomposition of the original complex signal to several resolutions [33, 34]. Let $X$ be a column vector containing a sequence $X_0, X_1, \ldots, X_{N-1}$, namely, $\{X_t, t = 0, 1, \ldots, N-1\}$, and $N$ is a power of 2. The even-length scaling (low-pass) filter can be denoted by $\{g_l, l = 0, 1, \ldots, L-1\}$ and the wavelet (high-pass) filter can be denoted by $\{h_l, l = 0, 1, \ldots, L-1\}$. These low-pass filters satisfy

$$\sum_{l=0}^{L-1} g_l^2 = 1,$$

$$\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{+\infty} g_l g_{l+2n} = 0 \quad (2)$$

for all nonzero integers $n$. These high-pass filters are also required to satisfy (2). In addition, both low-pass filters and high-pass filters are chosen to be quadrature mirror filters satisfying

$$h_l = (-1)^l g_{L-1-l},$$

$$g_l = (-1)^{l+1} h_{L-1-l} \quad \text{for } l = 0, 1, \ldots, L-1. \quad (3)$$

With $V_{0,t} = X_t$, for $t = 0, 1, \ldots, N-1$, the $j$th level wavelet and scaling coefficients are given by

$$V_{j,t} = \sum_{l=0}^{L-1} h_l V_{j-1,(2l+1-t)} \mod N_j^{-1},$$

$$W_{j,t} = \sum_{l=0}^{L-1} h_l V_{j-1,(2l+1-t)} \mod N_j^{-1} \quad (t = 0, 1, \ldots, N_j - 1), \quad (4)$$

where mod means modulus after division [35–37].

Although the DWT has been developed to improve the drawback mentioned above of CWT [33, 34], it requires the sample size to be exactly a power of 2 for the full transform because of the downsampling step in the DWT [35]. In order to overcome these drawbacks, maximal overlap discrete wavelet transform (MODWT) is developed [36]. MODWT could be considered as a revised version DWT. While the DWT of level $j$ restricts the sample size to an integer multiple of $2^j$, the MODWT of level $j$ is well defined for any sample size $N$ [35–37]. A scaling of the defining filters is required to conserve energy and filters are given by

$$\tilde{g}_l = \frac{g_l}{\sqrt{2}},$$

$$\tilde{h}_l = \frac{h_l}{\sqrt{2}}. \quad (5)$$

Thus, (2) becomes

$$\sum_{l=0}^{L-1} \tilde{g}_l^2 = \frac{1}{2},$$

$$\sum_{l=0}^{L-1} \tilde{g}_l \tilde{g}_{l+2n} = \sum_{l=-\infty}^{+\infty} \tilde{g}_l \tilde{g}_{l+2n} = 0 \quad (6)$$

and the filters are still quadrature mirror filters satisfying

$$\tilde{h}_l = (-1)^l \tilde{g}_{L-1-l},$$

or $\tilde{g}_l = (-1)^{l+1} \tilde{h}_{L-1-l}$ for $l = 0, 1, \ldots, L-1. \quad (7)$

In order to avoid downsampling, the MODWT creates appropriate new filters at each stage by inserting $2^j - 1$ zeros between the elements of $[\tilde{g}_l]$ and $[\tilde{h}_l]$

$$\tilde{g}_0, 0, \ldots, 0, \tilde{g}_1, 0, \ldots, 0, \ldots, \tilde{g}_{L-2}, 0, \ldots, 0, \tilde{g}_{L-1},$$

$$\tilde{h}_0, 0, \ldots, 0, \tilde{h}_1, 0, \ldots, 0, \ldots, \tilde{h}_{L-2}, 0, \ldots, 0, \tilde{h}_{L-1}. \quad (8)$$

With $\tilde{V}_{0,t} = X_t$, for $t = 0, 1, \ldots, N-1$, the $j$th level scaling coefficients $[\tilde{V}_{j,t}]$ and wavelet coefficients $[\tilde{W}_{j,t}]$ are given by

$$\tilde{V}_{j,t} = \sum_{l=0}^{L-1} \tilde{g}_l \tilde{V}_{j-1,(2l+1-t)} \mod N_j^{-1} \quad (t = 0, 1, \ldots, N_j - 1), \quad (9)$$

$$\tilde{W}_{j,t} = \sum_{l=0}^{L-1} \tilde{h}_l \tilde{V}_{j-1,(2l+1-t)} \mod N_j^{-1} \quad (t = 0, 1, \ldots, N_j - 1).$$
However, both the DWT and the MODWT have very poor frequency resolution at low frequencies [36]. For this drawback, the maximal overlap discrete wavelet packet transform (MODWPT) can further decompose the high frequency band, which is not decomposed in the DWT and the MODWT. Let $W_{j,n}^{(M)} = \{W_{j,n,t}^{(M)} : t = 0, 1, \ldots, N - 1\}$ be the sequence of MODWPT coefficients at $j$th level and the frequency-index $n$. With $W_{0,0}^{(M)} = X$, given the series $\{W_{j-1,n/2}^{(M)}\}$ of length $N$, the $\{W_{j,n}^{(M)}\}$ can be obtained by using

$$W_{j,n}^{(M)} = \sum_{l=0}^{L-1} r_{j,n} W_{j-1,n/2,(l-2t)} \mod N'$$

when $n \mod 4 = 0$ or 3, then $r_{j,n} = [\tilde{g}_l]$; when $n \mod 4 = 1$ or 2, then $r_{j,n} = [\tilde{h}_l]$.

Therefore, with the suitable decomposition scale and disjoint dyadic decomposition, the complicated signal could be decomposed into a number of components whose instantaneous amplitude and instantaneous frequency attain physical meaning [36, 37].

2.3. Linear Discriminant Analysis (LDA) and Neighborhood Preserving Embedding (NPE)

2.3.1. Linear Discriminant Analysis (LDA). The LDA was proposed by Fisher [67] for dimension reduction, which finds an embedding transformation such that the between-class scatter is maximized and within-class scatter is minimized [68–70]. The objective of the original Fisher’s LDA, namely, Fisher’s criterion, is to maximize the ratio of between-class scatter variance and within-class scatter variance. Let $\mu$ be the associated class labels, where $\mu_i \in \{1, 2, \ldots, c\}$ be the number of samples and $c$ is the total number of classes. Let $n_t$ be the number of samples in class $i$. When $r > 1$, where $r = c - 1$, a projection matrix $U$ is needed. Both $U^T S_b U$ and $U^T S_w U$ are $r \times r$ matrices, and the ratio of them cannot be computed directly. The determinant ratio is used:

$$J (U) = \max_U \left[ \frac{|U^T S_b U|}{|U^T S_w U|} \right].$$

where $S_b = \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^T$, $S_w = \sum_{i=1}^{c} \sum_{y \neq y_i} (x_i - \mu_i)(x_i - \mu_i)^T$, where $\mu_t$ is the mean of the samples in class $i$ and $\mu$ is the mean of all samples:

$$\mu_i = \frac{1}{n_t} \sum_{x_i \in y_i} x_i,$$

$$\mu = \frac{1}{n} \sum x_i.$$

The between-class scatter matrix $S_b$ and within-class scatter matrix $S_w$ also have equivalent form [71, 72]:

$$S_b = \frac{1}{2} n \sum_{i,j=1}^{n} p_{ij}^b (x_i - x_j)(x_i - x_j)^T = X (D^b - P^b) X^T,$$

$$S_w = \frac{1}{2} n \sum_{i,j=1}^{n} p_{ij}^W (x_i - x_j)(x_i - x_j)^T = X (D^W - P^W) X^T,$$

where

$$p_{ij}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_t}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

$$p_{ij}^W = \begin{cases} \frac{1}{n}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

and the within-class scatter matrix $S_w$ also have equivalent form [71, 72]:

$$S_b = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^b (x_i - x_j)(x_i - x_j)^T = X (D^b - P^b) X^T,$$

$$S_w = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^W (x_i - x_j)(x_i - x_j)^T = X (D^W - P^W) X^T,$$

where

$$p_{ij}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_t}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

$$p_{ij}^W = \begin{cases} \frac{1}{n}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

and the within-class scatter matrix $S_w$ also have equivalent form [71, 72]:

$$S_b = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^b (x_i - x_j)(x_i - x_j)^T = X (D^b - P^b) X^T,$$

$$S_w = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^W (x_i - x_j)(x_i - x_j)^T = X (D^W - P^W) X^T,$$

where

$$p_{ij}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_t}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

$$p_{ij}^W = \begin{cases} \frac{1}{n}, & y_i = y_j = l \\ \frac{1}{n_t}, & y_i \neq y_j \end{cases},$$

and the within-class scatter matrix $S_w$ also have equivalent form [71, 72]:

$$S_b = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^b (x_i - x_j)(x_i - x_j)^T = X (D^b - P^b) X^T,$$

$$S_w = \frac{1}{2} \sum_{i,j=1}^{n} p_{ij}^W (x_i - x_j)(x_i - x_j)^T = X (D^W - P^W) X^T,$$
Given a dataset of \( N \) samples assembled in a matrix \( X = [x_1, x_2, \ldots, x_N] \), the dimension of each sample is \( M \), and a transformation matrix \( A \) can be found that maps these \( N \) samples to a dataset of \( N \) samples assembled in \( Y = [y_1, y_2, \ldots, y_N] \). The dimension of each sample is \( D (D \ll M) \), where the \( r \)th column vector of \( Y \) corresponds to that of \( X \). Thus, the transformation can be expressed by \( Y = A^T X \).

The specific procedure can be presented as follows [53, 63]:

1. Constructing an adjacency graph: calculate the Euclidean distance between samples \( x_i \) and \( x_j \). The \( k \)-nearest neighbors (knn) are used to construct the adjacency graph \( G \). The distance \( d(x_i, x_j) \) represents the edge connecting \( x_i \) and \( x_j \), as
   \[
   d(x_i, x_j) = \| x_i - x_j \|_2. \tag{17}
   \]

2. Computing the weights: in this step, the weights of the edges can be computed. Let \( W \) denote the weight matrix with \( W_{i,j} \) having the weight of the edge from node \( i \) to node \( j \) and let it be 0 if there is no such edge. The weights of the edges can be computed by the minimizing weighted objective function which is presented as follows:
   \[
   O(W) = \min \sum_i \left[ x_i - \sum_j W_{i,j} x_j \right]^2, \tag{18}
   \]
   with constraints
   \[
   \sum_i W_{i,j} = 1, \quad j = 1, 2, \ldots, N. \tag{19}
   \]
   A reasonable criterion for choosing an expected map is to minimize that cost function which is presented as follows [72]:
   \[
   P(W) = \min \sum_i \left[ y_i - \sum_j W_{i,j} y_j \right]^2. \tag{20}
   \]
   This optimization problem can be converted to the following expression:
   \[
   P(W) = \min \left\{ \operatorname{tr} \left( A^T X Z X^T A \right) \right\}, \tag{21}
   \]
   where \( Z = (I-W)(I-W), I = \text{diag}(1,\ldots,1) \), and \( \operatorname{tr} \) is the trace of \( X Z X^T \). \( Z \) is symmetric and semi-positive definite. The specific procedure of how to solve the above minimization problem can be seen in [72].

2.4. Support Vector Machine (SVM). The key concept of SVM [73], which is originally developed for binary classification problems, uses a hyperplane to define decision boundaries between data points with different class. The idea behind SVM is that it can seek to construct optimal separating hyperplane to separate the two patterns, where the hyperplane minimizes the upper bound of the generalization error by maximizing the margin between the separating hyperplane and the nearest sample points [24]. SVM is able to handle both simple linear classification tasks and the classification of complex and nonlinear multiclass data [12].

Considering that a dataset \( [x_i, y_i]_{i=1}^n \) consists of \( D \)-dimensional sample, \( x_i \in \mathbb{R}^d (i = 1, 2, \ldots, n) \) presents the attribute and the corresponding label \( y_i \in \{-1, +1\} \) defines the type. In order to acquire a hyperplane to separate the two types of samples, a linear decision boundary, \( w \phi(x) + b \), can be learned from the training samples, where \( w \) is the normal direction of a separation plane and \( b \) is the bias [12, 24]. Samples of each type can be classified through the following constraints:

\[
\begin{align*}
wx_i + b &\geq +1 \quad \text{for } y_i = +1, \\
wx_i + b &\leq -1 \quad \text{for } y_i = -1.
\end{align*}
\]

The optimal hyperplane can be obtained by solving the following optimization problem:

\[
\min \frac{1}{2} \| w \|^2, \tag{23}
\]

Subject to \( y_i [wx_i + b] - 1 \geq 0, \quad i = 1, 2, \ldots, n. \)

When the data are linearly separable, the formulations presented above can work accurately. However, they will be ineffective when the investigated sample is overlapping or nonlinear [12]. Thus, a parameter \( \xi \) is adopted to make the classifier more robust, which allows a certain degree of misclassification for some points around the decision boundary. Furthermore, a penalty parameter \( C \), imposing a trade-off between training error and generalization [24], is introduced to control the number of misclassified points and adjust the margin between different classes [12]. Therefore, the optimization problem to find the optimal decision can be described as follows:

\[
\min \frac{1}{2} \| w \|^2 + C \sum_{i=1}^n \xi_i, \tag{24}
\]

Subject to \( y_i [wx_i + b] - 1 + \xi_i \geq 0, \quad i = 1, 2, \ldots, n. \)

For the constrained optimization problem, by using the duality theory of optimization, the final decision function can be presented as [24]

\[
\begin{align*}
f(x) \; \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i \langle \phi(x_i), \phi(x) \rangle + b \right) \\
&= \text{sgn} \left( \sum_{i=1}^n y_i \alpha_i K(x, x_i) + b \right),
\end{align*}
\]

where \( \alpha_i \) symbolizes Lagrange multipliers and \( K(x, x_i) \) is a kernel function, which is positive definite. Typical examples of kernel function [10] offer these choices: linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel.
For roller element bearings, the fault detection is a multiclass pattern recognition task, which can be generally solved by decomposing the multiclass problem into several binary class problems [74]. In [75], the multiclass patterns recognition was handled by the “one-against-one” approach. In this paper, we select the polynomial kernel to solve the multiclass pattern recognition task.

3. Proposed Method and the System Framework

3.1. Features Extraction Method FSASD (Features Selection by Adjusted Rand Index and Sum of Within-Class Mean Deviations). In this paper, we suggest that the most sensitive statistical characteristics should be selected before the implementation of the fault patterns recognition technique. For this reason, the K-means method and SWD are applied to a dataset that includes different statistical characteristics for the case of bearing conditions. In FSASD, each kind of statistical characteristic is clustered by K-means method, from which the clustering result adjusted rand index (ARI) becomes an evaluation index of each statistical characteristic. For each kind of statistical characteristic, we compute SWD of characteristic samples in each bearings condition. The sum of SWD in all bearing conditions can be obtained. For each statistical characteristic, the higher the value of ARI is, the greater the class discriminative degree will be. The lower the value of SWD is, the greater the class cohesion of the characteristic will be. Therefore, the ratio of ARI and SWD is selected to indicate the sensitivity of statistical characteristic. The description of FSASD is summarized in the following steps.

Step 1. In the training samples, there are M kinds of bearing fault types, N vibration signals samples in each type of bearings fault pattern, and K kinds of statistical characteristics. By vibration signals processing, we can obtain original feature sets, \[FS^1, FS^2, \ldots, FS^K],\] where \( FS^K \) can be expressed by

\[
FS^K = \begin{bmatrix}
S^k_{11} & S^k_{12} & \cdots & S^k_{1N} \\
S^k_{21} & S^k_{22} & \cdots & S^k_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
S^k_{M1} & S^k_{M2} & \cdots & S^k_{MN}
\end{bmatrix},
\]

where \( S^k_{ij} \) is the \( k \)-th statistical characteristic of the \( j \)-th sample in the \( i \)-th kind of bearings fault type.

Next, \( FS^K \) can be classified into \( M \) clustering partitions using the K-means method. The ARI of the clustering partitions can be calculated to judge the accuracy of clustering results [76, 77].

Consider a set of \( n \) objects \( X = \{x_1, x_2, \ldots, x_n\} \) and suppose \( U = \{u_1, u_2, \ldots, u_R\} \) and \( V = \{v_1, v_2, \ldots, v_C\} \) represent two different partitions of the objects in \( X \) such that \( \bigcup_{i=1}^{R} u_i = X = \bigcup_{j=1}^{C} v_j \) and \( u_i \cap u_{i'} = \emptyset = v_j \cap v_{j'} \) for \( 1 \leq i \neq i' \leq R \) and \( 1 \leq j \neq j' \leq C \), where \( R \) and \( C \) represent subsets. The ARI is then defined as [78, 79]

\[
ARI = \frac{(a - (a + c)(a + b)/d)}{((a + c) + (a + b))/2 - (a + c)(a + b)/d},
\]

where

\( a_0, a_1, \ldots, a_{D-1} \) are arranged according to their corresponding eigenvalues \( \lambda_i \leq \lambda_1 \leq \cdots \leq \lambda_{D-1} \). Thus, the embedding is as follows:

\[
x_i \rightarrow y_i = A^T x_i, \quad A = (a_0, a_1, \ldots, a_{D-1}),
\]

where \( y_i \) is a \( D \)-dimensional vector and \( A \) is an \( M \times D \) matrix.

\( a \) is number of objects in a pair being placed in the same class in \( U \) and in the same class in \( V \),

\( b \) is number of objects in a pair being placed in the same class in \( U \) and in different classes in \( V \),

\( c \) is number of objects in a pair being placed in different classes in \( U \) and in the same class in \( V \),

\( d \) is number of objects in a pair being placed in different classes in \( U \) and in different classes in \( V \).

ARI can give a measure of the agreement between partitions and in classification problems [79]. When the ARI value is 1 (maximum), it indicates that the algorithm is doing the correct distinction between classes [79]. Necessarily, the greater the value of ARI is, the better the clustering performance will be. Therefore, the ARI can give us the characteristic’s discriminant power [79].

Once clustering analysis is performed for the characteristics sets, \([FS^1, FS^2, \ldots, FS^K]\), \( ARI = \{ARI(1), ARI(2), \ldots, ARI(K)\} \) can be obtained. In this paper, we presume that the greater the value of ARI(\( k \)) is, the greater the characteristic class discriminative degree will be.

Step 2. The SWD of characteristic samples of a kind of statistical characteristic in each type of bearings conditions can be obtained. In this paper, we select the polynomial kernel to solve the following generalized eigen-vector problem:

\[
XZ X^T a = \lambda XX^T a.
\]

(28)

where

\( x_i \) is a D-dimensional vector and \( A \) is an \( M \times D \) matrix.

\( a \) is number of objects in a pair being placed in the same class in \( U \) and in the same class in \( V \),

\( b \) is number of objects in a pair being placed in the same class in \( U \) and in different classes in \( V \),

\( c \) is number of objects in a pair being placed in different classes in \( U \) and in the same class in \( V \),

\( d \) is number of objects in a pair being placed in different classes in \( U \) and in different classes in \( V \).

\( a_0, a_1, \ldots, a_{D-1} \) are arranged according to their corresponding eigenvalues \( \lambda_i \leq \lambda_1 \leq \cdots \leq \lambda_{D-1} \). Thus, the embedding is as follows:

\[
x_i \rightarrow y_i = A^T x_i, \quad A = (a_0, a_1, \ldots, a_{D-1}),
\]

where \( y_i \) is a D-dimensional vector and \( A \) is an \( M \times D \) matrix.

\( a \) is number of objects in a pair being placed in the same class in \( U \) and in the same class in \( V \),

\( b \) is number of objects in a pair being placed in the same class in \( U \) and in different classes in \( V \),

\( c \) is number of objects in a pair being placed in different classes in \( U \) and in the same class in \( V \),

\( d \) is number of objects in a pair being placed in different classes in \( U \) and in different classes in \( V \).

(27)

where

\[
ARI = \frac{(a - (a + c)(a + b)/d)}{((a + c) + (a + b))/2 - (a + c)(a + b)/d},
\]

(29)

where

\( a_0, a_1, \ldots, a_{D-1} \) are arranged according to their corresponding eigenvalues \( \lambda_i \leq \lambda_1 \leq \cdots \leq \lambda_{D-1} \). Thus, the embedding is as follows:

\[
x_i \rightarrow y_i = A^T x_i, \quad A = (a_0, a_1, \ldots, a_{D-1}),
\]

where \( y_i \) is a D-dimensional vector and \( A \) is an \( M \times D \) matrix.

(30)

where

\[\overline{S^k_j} = \frac{1}{N} \sum_{j=1}^{N} S^k_{ij}\]
Next, we can obtain SSWD(k), which is the sum of the SWD of characteristic samples of the kth statistical characteristic for all cases of bearing conditions, where SSWD(k) can be expressed by

$$SSWD(k) = \sum_{i=1}^{M} SWD_i^k.$$ (32)

In this paper, we presume that the SWD can be used to express the cohesion of data. Thus, there is the standard deviation sequence SSSWD = \{SSWD(1), SSWD(2), ..., SSWD(K)\}, which becomes another evaluation index for features extraction. In this paper, we presume that the lower the value of SSWD(k), the greater the classification accuracy. Therefore, a novel dimensionality reduction method, SNPEL, was proposed. SNPEL naturally combines the optimization objectives of LDA and NPE. The optimization objective of SNPEL can be obtained by combining the optimization objectives of SNPEL and LDA. The underlying idea of solving the problem mentioned above is that the optimization objective of LDA can be integrated into NPE; that is, the between-class scatter is maximized and the within-class scatter is minimized.

Based on the description of NPE and LDA in Section 2, the optimization objective of SNPEL can be obtained by combining the optimization objectives of LDA and NPE. The objective function can be defined as follows:

$$\max_{A} \frac{|A^T S_b A|}{|A^T S_W A| + \sum_i \|y_i - \sum_j W_{i,j} y_j\|^2}.$$ (34)

According to (15), the above objective function can be expressed as follows:

$$\max_{A} \frac{|A^T X (D^b - P^b) X^T A|}{|A^T X (D^W - P^W) X^T A| + \sum_i \|y_i - \sum_j W_{i,j} y_j\|^2}.$$ (35)

The above optimization problem can be converted to the trace ratio optimization problem, and according to (21), the objective function (35) can be simplified as follows:

$$\max_{A} \frac{\text{tr}[A^T X (D^b - P^b) X^T A]}{\text{tr}[A^T X (D^W - P^W) X^T A] + \text{tr}(A^T X Z X^T A)},$$ (36)

where the matrix Z and matrix $D^W - P^W$ need to be normalized. Thus, the final optimization objective function is presented as follows:

$$\max_{A} \frac{\text{tr}[A^T X (D^b - P^b) X^T A]}{\text{tr}[A^T X (D^W - P^W)_{\text{nor}} X^T A] + \text{tr}(A^T X Z_{\text{nor}} X^T A)},$$ (37)

where $(D^W - P^W)_{\text{nor}}$ and $Z_{\text{nor}}$ represent the normalized matrix $(D^W - P^W)$ and the normalized matrix $Z$, respectively.

Finally, the dimensionality reduction projection matrix $A$ can be formed by solving a generalized eigenvalue problem:

$$XLX^T a = AXX^T a,$$ (38)

where $L = (D^W - P^W)_{\text{nor}}, S = (D^W - P^W)_{\text{nor}} + Z_{\text{nor}}$, and $a_0, a_1, ..., a_M$ are arranged according to their corresponding eigenvalues $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_M$. The projection matrix $A$ is composed of the first $D$ eigenvectors of $L$; that is, $A = [a_0, a_1, ..., a_{D-1}]$. Therefore, given $x_i \in R^M (i = 1, 2, ..., N)$, the corresponding embedding projection $y_i \in R^D : x_i \rightarrow y_i = A^T x_i$ can be obtained.

The detailed procedures of SNPEL are listed as follows.

Step 1. Compute Euclidean distance between samples $x_i$ and $x_j$, and the $k$-nearest neighbors (knn) are used to construct the adjacency graph $G$.

Step 2. Compute the weights on the edges. Let $W$ denote the weight matrix with $W_{i,j}$ having the weight of the edge from node $i$ to node $j$, and let it be 0 if there is no such edge. The weights of the edges can be computed by the minimizing weighted objective equation (18).

Step 3. Compute the $M$-dimensional mean vectors for the different classes of the dataset.

Step 4. Compute between-class scatter matrix $S_b = X(D^b - P^b) X^T$ and within-class scatter matrix $S_w = X(D^W - P^W) X^T$.

Step 5. Compute the eigenvectors and corresponding eigenvalues for the matrix $XLX^T$ and the matrix $XX^T$. Thus eigenvectors $a_0, a_1, ..., a_M$ and corresponding eigenvalues $\lambda_0, \lambda_1, ..., \lambda_M$ are obtained.

Step 6. Sort the eigenvectors by decreasing eigenvalues and choose $D$ eigenvectors with the largest eigenvalues to form the $M \times D$-dimensional projection matrix $A$.

Step 7. Compute the equation $Y = A^T X$. The $M$-dimensional samples can be transformed to the $D$-dimensional samples and procedures of dimensionality reduction have been completed.
Finally, with the utility of SNPEL, the low-dimensional feature matrices of the training and testing dataset can be obtained with more sensitive and less redundant information for the bearings fault identification and classification.

3.3. System Framework. The implementation of the proposed method is shown in Figure 2, where the statistical analysis and the artificial intelligence approaches are systematically blended to detect and diagnose rolling element bearing faults. The whole fault diagnosis procedure is divided into four steps: signal processing, feature extraction, features reduction, and patterns recognition.

In the first step, vibration signals collected from bearings are decomposed into different wavelet packet nodes by MODWPT. The single branch reconstruction signals of terminal nodes will be applied to generate statistical characteristics. With the utility of the proposed FSASD, the most sensitive statistical characteristics can be selected to construct feature vectors for the training classifier. The most sensitive statistical characteristics will be directly applied to extracting features for testing samples. Then, for the feature reduction, the low-dimensional training feature space is obtained by the proposed SNPEL, which generates a projection that can be used for dimensionality reduction of the testing feature space. The low-dimensional testing feature space can be obtained. SASD and projection matrix are obtained by processing the training set, which can be directly used by testing set. In the last step, the low-dimensional training feature set is employed as the input of the fault type to train the classifier. The trained classifier will be employed to conduct fault patterns recognition using the low-dimensional testing feature set. The procedure of this proposed method outputs the fault identification and classification accuracy.

4. Experiments and Analysis Results

4.1. Experiments Based on Test Rig 1

4.1.1. Experimental Setup and Cases. The vibration dataset is freely provided by the Bearing Data Center of Case Western Reserve University (CWRU) [45]. Figure 3 shows the system used for measuring the data that includes an electric motor (left), a torque transducer/encoder (center), a dynamometer (right), and control circuitry (not shown). The bearings used in this work are deep groove ball bearings of the type 6205-2RS JEM SKF at DE. The single fault (including ball fault, inner race fault, and outer race fault) was separately seeded on the normal bearing with different defect sizes (0.007 in, 0.014 in, 0.021 in, and 0.028 in) using electro-discharge machining [12]. The vibration signals were collected using accelerometers under different motor loads of 0–3 hp (motor speeds of 1730 to 1797 rpm).

In order to evaluate the effectiveness, adaptability, and robustness of the proposed bearing fault diagnosis method, the vibration signals of different fault types and degrees were employed. The detailed information of the used dataset is
Table 1: The detailed information of the used vibration dataset.

| Condition of the bearings | Defect Size (mm) | Number of training samples | Number of testing samples | Class |
|---------------------------|------------------|----------------------------|---------------------------|-------|
|                           |                  | 2 hp (case 1)              | 3 hp (case 2)             |       |
| Healthy ball              | 0                | 20                         | 40                        | 40    | 1    |
| Ball fault                | 0.007            | 20                         | 40                        | 40    | 2    |
|                           | 0.014            | 20                         | 40                        | 40    | 3    |
|                           | 0.021            | 20                         | 40                        | 40    | 4    |
|                           | 0.028            | 20                         | 40                        | 40    | 5    |
| Inner race fault          | 0.007            | 20                         | 40                        | 40    | 6    |
|                           | 0.014            | 20                         | 40                        | 40    | 7    |
|                           | 0.021            | 20                         | 40                        | 40    | 8    |
|                           | 0.028            | 20                         | 40                        | 40    | 9    |
| Outer race fault          | 0.007            | 20                         | 40                        | 40    | 10   |
|                           | 0.014            | 20                         | 40                        | 40    | 11   |
|                           | 0.021            | 20                         | 40                        | 40    | 12   |
| Number of samples         |                  | 240                        | 480                       | 480   |

presented in Table 1, where ball and inner race faults have four fault degrees, respectively, while the outer race fault has three fault degrees. Furthermore, there is also a normal condition. Therefore, there are 12 working conditions, and these conditions correspond to 12 fault patterns. In each fault pattern, 60 samples are acquired from vibration signals, while each sample contains 2000 continuous data points. The 60 samples of each fault pattern were collected from the bearing installed at the drive end of the motor housing, where the sampling frequency is 12 kHz. In order to verify the adaptability of the proposed diagnosis method, the samples of a kind of motor load are selected as the training samples and the samples of different motor loads are selected as the testing samples. This experimental setup is different from other setups employed in previous research [5, 21, 80]. Therefore, two cases are employed in experiments. Cases 1 and 2 are comparative cases. In case 1, 40 random samples of 2 hp are selected as the testing samples. In case 2, 40 random samples of 3 hp are selected as the testing samples. For the training samples, two cases use the same remaining 20 samples of 2 hp.

4.1.2. Analysis Results. According to the system framework shown in Figure 2, the first step is signal processing, in which vibration signals collected from bearings are decomposed into different wavelet packet nodes by MODWPT. In this paper, the decomposition level is 4 and the "dmev" is selected as mother wavelet. One ball fault vibration signal sample from the training set of 2 hp and the corresponding single branch reconstruction signals of terminal nodes are presented in Figure 4.

According to the decomposition of vibration signals, 16 terminal nodes and the corresponding coefficients can be obtained. Then, we obtain 16 single branch reconstruction signals of terminal nodes and 16 corresponding Hilbert envelope spectra (HES), which can generate 192 statistical characteristics using 6 statistical parameters shown in Table 2. For 192 statistical characteristics of each sample, the class discriminative degree of each characteristic is different, which is reflected in Figures 5 and 6. In this paper, we provide four examples, of which two are time-domain characteristics (energy and energy entropy) and two are HES statistical characteristics (standard deviation and kurtosis).

The original feature set is composed of 192 statistical characteristics. Then, the FSASD is employed to select the sensitive statistical characteristics as the input feature vectors for the training classifier. The ARI, SSWD, and ASD of 192 statistical characteristics of the training samples are presented in Figures 7, 8, and 9, respectively. In Figure 7, the horizontal axis represents the number of statistical characteristics. 1–6, 7–12, ..., 85–90, and 91–96 represent time-domain characteristics of single branch reconstruction signals of terminal wavelet packet nodes 1–16, respectively. 97–102, 103–108, ..., 181–186, and 187–192 represent HES characteristics of single branch reconstruction signals of terminal wavelet packet nodes 1–16, respectively.

In order to verify the effectiveness and adaptability of the proposed bearing fault diagnosis method, a series of comparative experiments are divided into two groups. The detailed descriptions of them are presented below. Furthermore, in order to verify the superiority of MODWPT, WPT is also applied for fault diagnosis, and the results are compared with those of MODWPT.

In the first group, the FSASD is not applied. The original feature set contains 192 statistical characteristics which are directly processed by some dimensionality reduction methods. OFS-SVM is a SVM-based diagnosis model, in which the OFS is the input of SVM. OFS-PCA/NPE/LDA/SNPEL-SVM are SVM-based diagnosis models with the use of PCA, NPE, LDA, and SNPEL, respectively. According to Tables 3–7, the performance of each model using MODWPT is better than that of the model using WPT.

The detailed results of all models using MODWPT are presented below. For the testing set of case 1, all models can achieve preferable performance. The accuracies of each model can reach over 96%, and the highest accuracy can reach 100%. For the testing set of case 2, compared with OFS-SVM, OFS-PCA-SVM and OFS-NPE-SVM have improvement in diagnosis accuracy. But the performance of
Figure 4: A vibration signal sample and the corresponding single branch reconstruction signals of terminal nodes.
Table 2: Statistical parameters.

| Number | Feature          | Expression                                      |
|--------|------------------|-------------------------------------------------|
| (1)    | Range            | $T_1 = \max(|x(i)|) - \min(|x(i)|)$             |
| (2)    | Mean value       | $T_2 = \frac{1}{n} \sum_{i=1}^{n} x(i)$         |
| (3)    | Standard deviation| $T_3 = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x(i) - T_1)^2}$ |
| (4)    | Kurtosis         | $T_4 = \sum_{i=1}^{n} (x(i) - T_1)^3$            |
| (5)    | Energy           | $T_5 = \sum_{i=1}^{n} |x(i)|^2$                |
| (6)    | Energy entropy   | $T_6 = -\sum_{i=1}^{n} \varepsilon(i) \cdot \log \varepsilon(i)$ |

Here $x(i)$ is a series of a dataset for $i = 1, 2, \ldots, n$, $n$ is the number of data points, and $\varepsilon(i)$ is the energy distribution of the signal $x(i)$.

Table 3: Bearing fault diagnosis results obtained by OFS-SVM.

|                  | MODWPT Case 1 testing accuracy (%) | MODWPT Case 2 testing accuracy (%) | WPT Case 1 testing accuracy (%) | WPT Case 2 testing accuracy (%) |
|------------------|------------------------------------|-----------------------------------|---------------------------------|---------------------------------|
| MODWPT           | 98.54                              | 83.54                             | 95.21                           | 76.25                           |

Figure 5: Two time-domain statistical characteristics of the training samples.

Figure 6: Two HES statistical characteristics of the training samples.
OFS-LDA-SVM and OFS-SNPEL-SVM is better than that of OFS-SVM, OFS-PCA-SVM, and OFS-NPE-SVM, and the highest accuracy of OFS-SNPEL-SVM can reach 94.58%. In the experiments mentioned above, two cases are tested in various approaches. According to the experimental results, it is evident that the fault diagnosis model using SNPEL can achieve preferable performance.

In the second group, the FSASD is applied to select the sensitive statistical characteristics before the implementation of features reduction and fault diagnosis. OFS-FSASD-SVM is a SVM-based diagnosis model, in which the sensitive characteristics can be selected from OFS by FSASD. OFS-FSASD-PCA/NPE/LDA/SNPEL-SVM are SVM-based diagnosis models with the use of PCA, NPE, LDA, and SNPEL, respectively. According to Tables 8–12 and Figures 10–21, the performance of each model using MODWPT is better than that of the model using WPT. The detailed results of all models using MODWPT are presented below.

The sfn is the number of selected characteristics. For the testing set of case 1, all models can achieve preferable performance. For the testing set of case 2, compared with the experimental results of the first group, diagnosis accuracies...
### Table 4: Bearing fault diagnosis results obtained by OFS-PCA-SVM.

| Dimension size | MODWPT | WPT |
|----------------|--------|-----|
|                | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
| 5              | 94.79  | 72.08 | 93.54  | 82.92  |
| 10             | 98.75  | 81.45 | 93.54  | 73.96  |
| 15             | 98.33  | 84.58 | 93.54  | 77.50  |
| 20             | 99.17  | 84.79 | 95.63  | 78.13  |
| 30             | 99.37  | 85.00 | 96.87  | 77.71  |

### Table 5: Bearing fault diagnosis results obtained by OFS-NPE-SVM.

| Dimension size | MODWPT | WPT |
|----------------|--------|-----|
|                | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
| 5              | 69.17  | 68.75 | 44.38  | 34.17  |
| 10             | 83.54  | 74.17 | 77.71  | 61.25  |
| 15             | 90.00  | 76.46 | 83.33  | 67.29  |
| 20             | 95.21  | 88.75 | 78.96  | 66.46  |
| 30             | 96.46  | 74.38 | 83.13  | 53.54  |

### Table 6: Bearing fault diagnosis results obtained by OFS-LDA-SVM.

| Dimension size | MODWPT | WPT |
|----------------|--------|-----|
|                | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
| 5              | 99.79  | 74.91 | 96.67  | 66.04  |
| 7              | 99.79  | 83.54 | 98.96  | 76.46  |
| 9              | 99.79  | 83.12 | 99.38  | 77.92  |
| 11             | 100.00 | 93.75 | 99.58  | 79.79  |

### Table 7: Bearing fault diagnosis results obtained by OFS-SNPEL-SVM.

| Dimension size | MODWPT | WPT |
|----------------|--------|-----|
|                | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
| 5              | 83.75  | 72.92 | 66.67  | 54.17  |
| 10             | 98.54  | 91.67 | 94.58  | 81.67  |
| 15             | 98.75  | 94.58 | 98.13  | 89.58  |
| 20             | 98.75  | 94.58 | 98.13  | 89.79  |
| 30             | 98.75  | 94.58 | 97.91  | 90.00  |

### Table 8: Bearing fault diagnosis results obtained by OFS-FSASD-SVM.

| sfn | MODWPT | WPT |
|-----|--------|-----|
|     | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
| 20  | 87.91  | 63.54 | 99.79  | 44.17  |
| 30  | 99.79  | 45.21 | 98.13  | 75.63  |
| 40  | 97.79  | 54.38 | 98.13  | 75.63  |
| 50  | 97.29  | 81.88 | 97.29  | 83.54  |
| 70  | 98.13  | 87.08 | 98.13  | 97.71  |
| 90  | 99.58  | 95.63 | 97.50  | 97.29  |
| 120 | 98.96  | 75.21 | 98.54  | 68.13  |
| 140 | 98.75  | 79.38 | 97.50  | 70.00  |
| 160 | 98.54  | 79.58 | 98.13  | 73.13  |
| 180 | 98.95  | 82.71 | 95.83  | 77.08  |
The performance of OFS-FSASD-SNPEL-SVM and OFS-FSASD-LDA-SVM is better than that of OFS-FSASD-SVM, OFS-FSASD-PCA-SVM, and OFS-FSASD-NPE-SVM. For OFS-FSASD-SNPEL-SVM and OFS-FSASD-LDA-SVM, the performance of OFS-FSASD-SNPEL-SVM is better. For the testing set of case 1, both the diagnosis accuracies of OFS-FSASD-SNPEL-SVM and OFS-FSASD-LDA-SVM can reach 100%. For the testing set of case 2, the maximum diagnosis accuracy of OFS-FSASD-SNPEL-SVM can reach 100%, but the maximum diagnosis accuracy of OFS-FSASD-LDA-SVM can only reach 97.92%. According to the experimental results of the second group, when a suitable parameter sfn is selected, it can achieve a desirable improvement on the diagnosis accuracy. According to Figures 12–21, we find that the fault diagnosis can attain better performance when the parameter sfn is in a relatively wide range; for example, for the performance of OFS-FSASD-SNPEL-SVM, the highest diagnosis accuracy can reach 100%. Therefore, on the one hand, the validity of the design of the correlation parameter can be
Table 12: Bearing fault diagnosis results obtained by OFS-FSASD-SNPEL-SVM (dimension size is 20).

| sfn | MODWPT Case 1 testing accuracy (%) | MODWPT Case 2 testing accuracy (%) | WPT Case 1 testing accuracy (%) | WPT Case 2 testing accuracy (%) |
|-----|------------------------------------|------------------------------------|--------------------------------|--------------------------------|
| 20  | 17.08                              | 16.46                              | 99.79                          | 75.00                          |
| 30  | 100.00                             | 66.67                              | 99.38                          | 62.08                          |
| 40  | 99.79                              | 69.79                              | 100.00                         | 74.17                          |
| 50  | 100.00                             | 74.38                              | 100.00                         | 75.00                          |
| 60  | 100.00                             | 75.00                              | 100.00                         | 83.75                          |
| 70  | 100.00                             | 100.00                             | 99.79                          | 99.17                          |
| 90  | 100.00                             | 100.00                             | 99.79                          | 99.17                          |
| 120 | 99.38                              | 76.88                              | 99.17                          | 83.13                          |
| 140 | 98.33                              | 84.58                              | 99.17                          | 89.17                          |
| 160 | 99.79                              | 90.83                              | 99.17                          | 85.21                          |
| 180 | 99.58                              | 92.92                              | 98.13                          | 89.79                          |

Figure 12: The diagnosis results obtained by OFS-FSASD-PCA-SVM using MODWPT with different dimension sizes for PCA. The PC represents the number of dimension sizes.

verified. On the other hand, it can verify that the proposed bearing fault diagnosis algorithm has great adaptability.

4.2. Experiments Based on the Test Rig 2

4.2.1. Experimental Setup and Cases. In order to validate the adaptability of the proposed bearing fault diagnosis method, we collected vibration signals from SQI-MFS test rig to conduct some experiments. Figure 22 shows the test rig and Figure 23 shows that the bearings used in this work are the type SER205. The single fault (including ball fault, inner race fault, and outer race fault) was separately seeded on the normal bearing with different defect sizes (0.05 mm, 0.1 mm, and 0.2 mm) using laser machining. The vibration signals were collected from the bearings using accelerometers under different motor speeds of 1200 rmp and 1800 rmp, where the sampling frequency is 16 kHz.

The detailed information of the used vibration dataset is presented in Table 13, where ball, inner race, and outer race faults have three fault degrees, respectively, and there is also a normal condition. Therefore, there are 10 working conditions, corresponding to 10 fault patterns. In each fault pattern, 60 samples are acquired from vibration signals, while each sample contains 5000 continuous data points. Two cases are employed in the experiments, the same as test rig 1. The samples of a kind of motor speed are selected as the training samples and the samples of different motor speeds are selected as the testing samples. In case 1, 40 random samples of 1800 rmp are selected as the training samples. In case 2, 40 random samples of 1200 rmp are selected as testing samples. For training samples, two cases use the same remaining 20 samples of 1800 rmp.

4.2.2. Analysis Results. The procedure of bearing fault diagnosis for SQI-MFS test rig is the same as that for the test rig 1. In the experiments, MODWPT is applied for vibration signals processing. For 192 statistical characteristics, the class discriminative degree of each characteristic is reflected in
When the original feature set has been obtained, the FSASD is employed to select the sensitive statistical characteristics as the input feature vectors for the bearing fault diagnosis. Then, ARI, SSWD, and ASD of 192 statistical characteristics of training samples can be obtained. They are presented in Figures 26, 27, and 28, respectively.

In order to verify the effectiveness and adaptability of the proposed fault diagnosis method for SQI-MFS test rig, a series of comparative experiments are divided into two groups. In the first group, the FSASD is not applied. The fault diagnosis results of OFS-SVM, OFS-PCA-SVM, OFS-NPE-SVM, OFS-LDA-SVM, and OFS-SNPEL-SVM are presented in Tables 14–18. For the testing set of case 1, all models can achieve preferable performance. The accuracies of each model can reach over 95%, and the highest accuracy can reach 99.67%. For the testing set of case 2, all models have
no desirable diagnosis accuracies. According to Tables 15–18, the performance of OFS-PCA-SVM and OFS-SNPEL-SVM is better than that of OFS-NPE-SVM and OFS-LDA-SVM, which indicates that the diagnosis model using different dimensionality reduction methods has different impacts on diagnosis accuracy.

In OFS, different statistical characteristics have different fault sensitivity; some are beneficial to fault identification and classification, but some are not. The FSASD can evaluate the fault sensitivity of each statistical characteristic and select the sensitive statistical characteristics. For the second group, the FSASD is applied before the implementation of features reduction and fault diagnosis. The fault diagnosis results of OFS-FSASD-SVM, OFS-FSASD-PCA-SVM, OFS-FSASD-NPE-SVM, OFS-FSASD-LDA-SVM, and OFS-FSASD-SNPEL-SVM are presented in Tables 19–23. The corresponding curve representations are presented in Figures 29–33. For the testing set of case 1, all models can achieve preferable performance. The accuracies of each model can reach over 97.50%, and the highest accuracy can reach 100%.

Figure 15: The diagnosis results obtained by OFS-FSASD-NPE-SVM model using WPT with different dimension sizes for NPE. The NPE(5) represents that the number of dimension sizes is 5.

Figure 16: The diagnosis results obtained by OFS-FSASD-LDA-SVM using MODWPT with different dimension sizes for LDA. The LDA(5) represents that the number of dimension sizes is 5.
And the performance of OFS-FSASD-LDA-SVM and OFS-FSASD-SNPEL-SVM is better than that of OFS-FSASD-SVM, OFS-FSASD-PCA-SVM, and OFS-FSASD-NPE-SVM. For the testing set of case 2, the maximum diagnosis accuracy of OFS-FSASD-SNPEL-SVM can reach 89.83%, but the maximum diagnosis accuracy of OFS-FSASD-LDA-SVM can only reach 83.17%. For comparison, the diagnosis results of OFS-FSASD-SVM, OFS-FSASD-PCA-SVM, OFS-FSASD-NPE-SVM, and OFS-FSASD-SNPEL-SVM are also presented in Figure 34. According to the experimental results of the second group, compared with the first group, on the one hand, the performance of the diagnosis model using the FSASD can have an improvement, which indicates that the different numbers of sensitive features have an effect on fault diagnosis accuracy. According to Figures 29–34, we find that the fault diagnosis can attain better performance when the parameter sfn is in a range; for example, for the performance of OFS-FSASD-SNPEL-SVM, the highest...
diagnosis accuracy can reach 89.83%, which can verify that a desirable improvement on the diagnosis accuracy can be achieved when a suitable parameter sfn is selected. On the other hand, the performance of the diagnosis model using different dimensionality reduction methods can also lead to different impacts on fault diagnosis accuracy, especially the fault diagnosis accuracy of the testing set of case 2. Because the proposed SNPEL can preserve the local geometry of the data and work well with multimodal data, at the same time, it can also take the label information into account in dimensionality reduction. Therefore, the low-dimensional feature space obtained by SNPEL is more beneficial to fault identification and classification. Through a series of comparative experiments, the effectiveness and adaptability of the proposed bearing fault diagnosis procedure for SQI-MFS test rig can be verified.
Figure 21: The diagnosis results of models using WPT for the testing sets of two cases with the use of FSASD and different dimensionality reduction methods. The output dimension sizes of PCA, LDA, and SNPEL are 20, 11, and 20, respectively. The "NO" represents the model without using dimensionality reduction method.

Figure 22: Experimental test rig 2.

Figure 23: SER205 bearings.
Figure 24: Two time-domain statistical characteristics of training samples.

Figure 25: Two HES statistical characteristics of training samples.

Table 13: The detailed information of the used vibration dataset.

| Condition of the bearings | Defect size (mm) | Number of training samples | Number of testing samples | Class |
|---------------------------|------------------|-----------------------------|---------------------------|-------|
|                           |                  | 1800 tr/min                 | 1800 tr/min (case 1)      | 1200 tr/min (case 2) |       |
| Healthy ball              | 0                | 20                          | 40                        | 40    | 1     |
| Ball fault                | 0.05             | 20                          | 40                        | 40    | 2     |
|                           | 0.1              | 20                          | 40                        | 40    | 3     |
|                           | 0.2              | 20                          | 40                        | 40    | 4     |
| Inner race fault          | 0.05             | 20                          | 40                        | 40    | 5     |
|                           | 0.1              | 20                          | 40                        | 40    | 6     |
|                           | 0.2              | 20                          | 40                        | 40    | 7     |
| Outer race fault          | 0.05             | 20                          | 40                        | 40    | 8     |
|                           | 0.1              | 20                          | 40                        | 40    | 9     |
|                           | 0.2              | 20                          | 40                        | 40    | 10    |
| Number of samples         |                  | 240                         | 480                       | 480   |       |
Figure 26: The ARI of 192 statistical characteristics of training samples.

Figure 27: The SSWD of 192 statistical characteristics of training samples.

Figure 28: The ASD of 192 statistical characteristics of training samples.

Figure 29: The diagnosis results of the OFS-FSASD-SVM model with different sfn.

Table 14: Bearing fault diagnosis results obtained by the OFS-SVM.

|            | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|------------|-----------------------------|-----------------------------|
|            | 98.17                       | 77.17                       |

Table 15: Bearing fault diagnosis results obtained by the OFS-PCA-SVM.

| Dimension size | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|----------------|-----------------------------|-----------------------------|
| 5              | 89.00                       | 68.83                       |
| 10             | 95.17                       | 70.83                       |
| 15             | 96.67                       | 74.50                       |
| 20             | 97.00                       | 74.00                       |
| 30             | 97.50                       | 75.50                       |

Table 16: Bearing fault diagnosis results obtained by the OFS-NPE-SVM.

| Dimension size | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|----------------|-----------------------------|-----------------------------|
| 5              | 67.67                       | 57.33                       |
| 10             | 84.00                       | 61.33                       |
| 15             | 93.83                       | 53.17                       |
| 20             | 94.50                       | 56.33                       |
| 30             | 95.50                       | 25.83                       |

5. Conclusions

This paper proposed a novel procedure in order to identify and classify different bearing fault conditions. The proposed procedure, systematically blending statistical analysis with
Figure 30: The diagnosis results obtained by OFS-FSASD-PCA-SVM with different number of dimension sizes for PCA. The PC represents the number of dimension sizes.

Figure 31: The diagnosis results obtained by OFS-FSASD-NPE-SVM with different number of dimension sizes for NPE. The NPE(5) represents that the number of dimension sizes is 5.

artificial intelligence, is developed using MODWPT as multidomain features generation approach. Using the proposed FSASD as the most sensitive features extraction method, the modified NPE (SNPEL) as a feature dimensionality reduction technique, and SVM as an automated fault patterns recognition system, the experimental data collected from two experimental test rigs contain different bearing fault conditions such as ball fault, inner race fault, and outer race fault at different defect sizes.

According to the experimental results, the proposed bearing fault diagnosis method has great potential to be an effective and adaptable tool for precise identification and classification of bearing faults for a variety of bearing conditions. For the experimental test rig 1, in the proposed...
procedure, two cases are employed in experiments. Cases 1 and 2 are a set of comparative cases. They use the testing samples with different motor loads, which are 2 hp and 3 hp, respectively. They use samples with the same motor load (2 hp) as the training samples. Experimental results indicate that the maximum diagnosis accuracy of case 1 can reach 100%. The diagnosis accuracies of case 2 can reach over 99% when the parameter sfn is in a relatively wide range. In order to verify the adaptability of the proposed procedure, vibration

**Table I7:** Bearing fault diagnosis results obtained by the OFS-LDA-SVM.

| Dimension size | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|----------------|-----------------------------|-----------------------------|
| 5              | 91.67                       | 60.00                       |
| 7              | 98.83                       | 59.83                       |
| 9              | 99.67                       | 59.17                       |
| 11             | 99.67                       | 59.50                       |
Table 18: Bearing fault diagnosis results obtained by the OFS-SNPEL-SVM.

| Dimension size | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|----------------|-----------------------------|----------------------------|
| 5              | 78.50                       | 64.33                      |
| 10             | 99.17                       | 65.33                      |
| 15             | 99.17                       | 65.17                      |
| 20             | 99.17                       | 66.33                      |
| 30             | 99.17                       | 66.00                      |

Table 19: Bearing fault diagnosis results obtained by the OFS-FSASD-SVM.

| sfn | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|-----|-----------------------------|----------------------------|
| 10  | 60.67                       | 54.67                      |
| 20  | 62.50                       | 67.67                      |
| 30  | 62.83                       | 50.67                      |
| 50  | 64.67                       | 57.00                      |
| 70  | 73.00                       | 60.17                      |
| 90  | 96.33                       | 75.33                      |
| 110 | 97.00                       | 76.83                      |
| 130 | 97.17                       | 76.83                      |
| 150 | 97.33                       | 75.00                      |
| 170 | 98.33                       | 74.50                      |
| 190 | 98.00                       | 76.50                      |

Table 20: Bearing fault diagnosis results obtained by the OFS-FSASD-PCA-SVM (dimension size is 20).

| sfn | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|-----|-----------------------------|----------------------------|
| 20  | 70.50                       | 71.67                      |
| 30  | 71.50                       | 69.00                      |
| 50  | 78.00                       | 63.00                      |
| 70  | 82.67                       | 64.50                      |
| 90  | 86.50                       | 68.83                      |
| 110 | 93.33                       | 72.33                      |
| 130 | 93.00                       | 75.50                      |
| 150 | 95.17                       | 74.83                      |
| 170 | 96.17                       | 73.67                      |
| 190 | 97.50                       | 74.33                      |

Table 21: Bearing fault diagnosis results obtained by the OFS-FSASD-NPE-SVM (dimension size is 20).

| sfn | Case 1 testing accuracy (%) | Case 2 testing accuracy (%) |
|-----|-----------------------------|----------------------------|
| 20  | 91.50                       | 39.00                      |
| 30  | 92.50                       | 44.83                      |
| 50  | 94.67                       | 54.50                      |
| 70  | 98.17                       | 47.67                      |
| 90  | 97.17                       | 50.00                      |
| 110 | 97.50                       | 48.67                      |
| 130 | 96.50                       | 56.50                      |
| 150 | 96.83                       | 44.00                      |
| 170 | 97.33                       | 58.83                      |
| 190 | 94.50                       | 56.33                      |
datasets collected from the experimental test rig 2 (SQI-MFS) are employed. Cases 1 and 2 use the testing samples with different motor speeds, which are 1200 rmp and 1800 rmp, respectively. They use samples with the same motor speed (1800 rmp) as the training samples. The experimental results can also indicate that the diagnosis model using the proposed methods can achieve preferable performance.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (no. 2017YFC0804400, no. 2017YFC0804401) and the National Key Basic Research Program of China (973 Program, no. 2014CB046300).

References

[1] L. Zhang, G. Xiong, H. Liu, H. Zou, and W. 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.
[2] N. Gebrareel, M. Lawley, R. Liu, and V. Parmeshwaran, “Residual life predictions from vibration-based degradation signals: a neural network approach,” IEEE Transactions on Industrial Electronics, vol. 51, no. 3, pp. 694–700, 2004.
[3] Y. Lei, J. Lin, Z. He, and M. J. Zuo, “A review on empirical mode decomposition in fault diagnosis of rotating machinery,” Mechanical Systems and Signal Processing, vol. 35, no. 1-2, pp. 108–126, 2013.
[4] N. A. Othman, N. S. Damanhuri, and V. Kadirkamanathan, “The study of fault diagnosis in rotating machinery,” in Proceedings of the 2009 5th International Colloquium on Signal Processing and Its Applications, CSPA 2009, pp. 69–74, mws, March 2009.
[5] X. Xue and J. Zhou, “A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery,” ISA Transactions®, vol. 66, pp. 284–295, 2017.
[6] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, “A multidimensional hybrid intelligent method for gear fault diagnosis,” Expert Systems with Applications, vol. 37, no. 2, pp. 1419–1430, 2010.
[7] Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques—part I: fault diagnosis with model-based and signal-based approaches,” IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3757–3767, 2015.
[8] Z. Gao, C. Cecati, and S. X. Ding, “A survey of fault diagnosis and fault-tolerant techniques—part II: fault diagnosis with knowledge-based and hybrid/active approaches,” IEEE Transactions on Industrial Electronics, vol. 62, no. 6, pp. 3768–3774, 2015.
[9] M. Kang, J. Kim, L. M. Wills, and J.-M. Kim, “Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis,” IEEE Transactions on Industrial Electronics, vol. 62, no. 12, pp. 7749–7761, 2015.
[10] Z. Feng, M. Liang, and F. Chu, “Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples,” Mechanical Systems and Signal Processing, vol. 38, no. 1, pp. 165–205, 2013.
[11] L. Saidi, J. Ben Ali, F. Fnaiech, and B. Morello, “Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis,” in Proceedings of the 6th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2014, pp. 25–30, Tunisia, August 2014.
[12] L. Lu, J. Yan, and C. W. De Silva, “Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition,” Journal of Sound and Vibration, vol. 344, pp. 464–483, 2015.
[13] Z. Yang, B. W.-K. Ling, and C. Bingham, “Trend extraction based on separations of consecutive empirical mode decomposition components in Hilbert marginal spectrum,” Measurement, vol. 46, no. 8, pp. 2481–2491, 2013.
[14] X. Liu, L. Bo, and H. Luo, “Bearing faults diagnostics based on hybrid LS-SVM and EMD method,” Measurement, vol. 59, pp. 145–166, 2015.
[15] T. Guo and Z. Deng, “An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing,” Applied Acoustics, vol. 127, pp. 46–62, 2017.
[16] X. Jiao, B. Jing, Y. Huang, J. Li, and G. Xu, “Research on fault diagnosis of airborne fuel pump based on EMD and probabilistic neural networks,” Microelectronics Reliability, vol. 75, pp. 296–308, 2017.
[17] J. Zhong and Y. Huang, "Time-frequency representation based on an adaptive short-time Fourier transform," IEEE Transactions on Signal Processing, vol. 58, no. 10, pp. 5118–5128, 2010.

[18] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," IEEE Transactions on Industrial Informatics, vol. 9, no. 2, pp. 760–771, 2013.

[19] H. Wang and P. Chen, "Fuzzy diagnosis method for rotating machinery in variable rotating speed," IEEE Sensors Journal, vol. 11, no. 1, pp. 23–34, 2011.

[20] C. K. E. Nizwan, S. A. Ong, M. F. M. Yusof, and M. Z. Baharom, "A wavelet decomposition analysis of vibration signal for bearing fault detection," in Proceedings of the 2nd International Conference on Mechanical Engineering Research, ICMER 2013, Malaysia, July 2013.

[21] I. Attou, N. Fergani, N. Boutassetta, B. Oudjani, and A. Deliou, "A new time–frequency method for identification and classification of ball bearing faults," Journal of Sound and Vibration, vol. 397, pp. 241–265, 2017.

[22] J. 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.

[23] J.-H. Ahn, D.-H. Kwak, and B.-H. Koh, "Fault detection of a roller-bearing system through the EMD of a wavelet denoised signal," Sensors, vol. 14, no. 8, pp. 15022–15038, 2014.

[24] X. Yu, E. Ding, C. Chen, X. Liu, and L. Li, "A novel characteristic frequency bands extraction method for automatic bearing fault diagnosis based on Hilbert Huang transform," Sensors, vol. 15, no. 11, pp. 27869–27893, 2015.

[25] Y. Yang, X. J. Dong, Z. K. Peng, W. M. Zhang, and G. Meng, "Vibration signal analysis using parameterized time-frequency method for features extraction of varying-speed rotary machinery," Journal of Sound and Vibration, vol. 335, pp. 350–366, 2015.

[26] S. Luo, Y. Yang, J. Chen et al., "Tanshinol stimulates bone formation and attenuates dexamethasone-induced inhibition of osteogenesis in larval zebrafish," Journal of Orthopaedic Translation, vol. 4, pp. 35–45, 2016.

[27] Y. Wang, G. Xu, L. Liang, and K. Jiang, "Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis," Mechanical Systems and Signal Processing, vol. 54–55, pp. 259–276, 2015.

[28] C. Wang, M. Gan, and C. Zhu, "Intelligent fault diagnosis of rolling element bearings using sparse wavelet energy based on overcomplete DWT and basis pursuit," Journal of Intelligent Manufacturing, vol. 28, no. 6, pp. 1377–1391, 2017.

[29] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," Journal of Sound and Vibration, vol. 289, no. 4–5, pp. 1066–1090, 2006.

[30] H. Talhaoui, A. Menacer, A. Kessal, and R. Kechida, "Fast Fourier and discrete wavelet transforms applied to sensorless vector control induction motor for rotor bar faults diagnosis," ISA Transactions, vol. 53, no. 5, pp. 1639–1649, 2014.

[31] H. Bafroui and A. Ohadi, "Application of wavelet energy and Shannon entropy for feature extraction in gearbox fault detection under varying speed conditions," Neurocomputing, vol. 133, pp. 437–445, 2014.

[32] J.-D. Wu and C.-H. Liu, "An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network," Expert Systems with Applications, vol. 36, no. 3, pp. 4278–4286, 2009.

[33] S. Prabhakar, A. R. Mohanty, and A. S. Sekhar, "Application of discrete wavelet transform for detection of ball bearing race faults," Tribology International, vol. 35, no. 12, pp. 793–800, 2002.

[34] S. Serhat and A. Emine, "Feature extraction related to bearing damage in electric motors by wavelet analysis," Journal of The Franklin Institute, vol. 340, no. 2, pp. 125–134, 2003.

[35] E. Tsakiroglou and A. T. Walden, "From Blackman - Tukey pilot estimators to wavelet packet estimators: A modern perspective on an old spectrum estimation idea," Signal Processing, vol. 82, no. 10, pp. 1425–1441, 2002.

[36] Y. Yang, Y. He, J. Cheng, and D. Yu, "A gear fault diagnosis using Hilbert spectrum based on MODWPT and a comparison with EMD approach," Measurement, vol. 42, no. 4, pp. 542–551, 2009.

[37] P.-W. Shan and M. Li, "Nonlinear time-varying spectral analysis: HHT and MODWPT," Mathematical Problems in Engineering, vol. 2010, Article ID 618231, 2010.

[38] A. T. Walden and A. Contreras Cristian, "The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events," Proceedings of the Royal Society A Mathematical, Physical and Engineering Sciences, vol. 454, no. 1976, pp. 2243–2266, 1998.

[39] M. Kang, J. Kim, A. C. Tan, E. Y. Kim, and B. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," IEEE Transactions on Power Electronics, vol. 30, no. 5, pp. 2786–2797, 2015.

[40] C. Y. Yang and T. Y. Wu, "Diagnostics of gear deterioration using EEMD approach and PCA process," Measurement, vol. 61, pp. 75–87, 2015.

[41] X. Zhang, D. Qiu, and F. Chen, "Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis," Neurocomputing, vol. 149, pp. 641–651, 2015.

[42] C. Rajeswari, B. Sathiyabharma, S. Devendiran, and K. Manivannan, "Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine," Procedia Engineering, vol. 97, pp. 1772–1783, 2014.

[43] X. Wan, D. Wang, P. W. Tse, G. Xu, and Q. Zhang, "A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions," Measurement, vol. 78, pp. 138–150, 2016.

[44] H. Sun, Z. He, Y. Zi et al., "Multiwavelet transform and its applications in mechanical fault diagnosis – A review," Mechanical Systems & Signal Processing, vol. 43, no. 1–2, pp. 1–24, 2014.

[45] "C.W.R.University," http://csegroups.case.edu/bearingdatacenter/pages/download-data-file.

[46] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti, and S. Longhi, "Electric motor defects diagnosis based on kernel density estimation and Kullback-Leibler divergence in quality control scenario," Engineering Applications of Artificial Intelligence, vol. 44, no. 2335, pp. 25–32, 2015.

[47] S. Yin, X. Zhu, and O. Kaynak, "Improved PLS focused on key performance indicator related fault diagnosis," IEEE Transactions on Industrial Electronics, 2014.

[48] F. Nie, D. Xu, I. W. Tsang, and C. Zhang, "Flexible manifold embedding: a framework for semi-supervised and unsupervised dimension reduction," IEEE Transactions on Image Processing, vol. 19, no. 7, pp. 1921–1932, 2010.
Y. Liu, Y. Zhang, Z. Yu, and M. Zeng, “Incremental supervised locally linear embedding for machinery fault diagnosis,” *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 60–70, 2016.

Y. Bengio, P. Vincent, O. Delalleau et al., “Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering,” in *International Conference on Neural Information Processing Systems*, pp. 177–184, MIT Press, 2003.

N. Vlassis, Y. Motomura, and B. Kröse, “Supervised dimension reduction of intrinsically low-dimensional data,” *Neural Computation*, vol. 14, no. 1, pp. 191–215, 2002.

Z. Lai, W. Wong K. Y. Xu et al., “Approximate Orthogonal Sparse Embedding for Dimensionality Reduction,” *IEEE Transactions on Neural Networks & Learning Systems*, vol. 27, no. 4, p. 723, 2016.

X. He, D. Cai, S. Yan, and H. Zhang, “Neighborhood preserving embedding,” in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV ’05)*, pp. 1208–1213, Beijing, China, October 2005.

E. Kokioxioupolou and Y. Saad, “Orthogonal neighborhood preserving projections,” in *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM 2005)*, pp. 234–241, USA, November 2005.

X. He and P. Niyogi, “Locality preserving projections,” *Advances in Neural Information Processing Systems*, vol. 45, no. 1, pp. 186–197, 2005.

Z. Lai, M. Wan, and Z. Jin, “Locality preserving embedding for face and handwriting digital recognition,” *Neural Computing and Applications*, vol. 20, no. 4, pp. 565–573, 2011.

Z. Lai, C. Zhao, Y. Chen, and Z. Jin, “Maximal local interclass embedding with application to face recognition,” *Machine Vision and Applications*, vol. 22, no. 4, pp. 619–627, 2011.

X. Bao, L. Zhang, B. Wang, and J. Yang, “A supervised neighborhood preserving embedding for face recognition,” in *Proceedings of the 2014 International Joint Conference on Neural Networks*, ICNN 2014, pp. 278–284, China, July 2014.

C. Tan and G. Ji, “Semisupervised local preserving embedding algorithm based on maximum margin criterion for large-scale data streams,” *Concurrency & Computation Practice & Experience*, no. 11, e2426 pages, 2017.

Y. Cui, Z. Jin, and J. Jiang, “A novel supervised feature extraction and classification fusion algorithm for land cover recognition of the off-land scenario,” *Neurocomputing*, vol. 140, pp. 77–83, 2014.

X. Yuan, Z. Ge, L. Ye, and Z. Song, “Supervised neighborhood preserving embedding for feature extraction and its application for soft sensor modeling,” *Journal of Chemometrics*, vol. 30, no. 8, pp. 430–441, 2016.

J. Wen, W. Yan, and W. Lin, “Supervised linear manifold learning feature extraction for hyperspectral image classification,” in *Proceedings of the Joint 2014 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2014 and the 35th Canadian Symposium on Remote Sensing, CSRS 2014*, pp. 3710–3713, Canada, July 2014.

S. Zhang and W. Li, “Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM,” *Mathematical Problems in Engineering*, vol. 2014, Article ID 781583, 10 pages, 2014.

Q. Liu, F. Chen, Z. Zhou, and Q. Wei, “Fault diagnosis of rolling bearing based on wavelet packet transform and ensemble empirical mode decomposition,” *Advances in Mechanical Engineering*, vol. 2013, Article ID 792584, 6 pages, 2013.

M. M. Tahir, A. Q. Khan, N. Iqbal, A. Hussain, and S. Badshah, “Enhancing Fault Classification Accuracy of Ball Bearing Using Central Tendency Based Time Domain Features,” *IEEE Access*, vol. 5, pp. 72–83, 2017.

C. Kar and A. R. Mohanty, “Monitoring gear vibrations through motor current signature analysis and wavelet transform,” *Mechanical Systems and Signal Processing*, vol. 20, no. 1, pp. 158–187, 2006.

R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

C. Radhakrishna Rao, “The utilization of multiple measurements in problems of biological classification,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, pp. 159–193; discussion, 194–203, 1948.

C. Yao, Z. Lu, J. Li, Y. Xu, and J. Han, “A subset method for improving Linear Discriminant Analysis,” *Neurocomputing*, vol. 138, pp. 310–315, 2014.

A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical Pattern Recognition: A Review,” *Optica Acta: International Journal of Optics*, vol. 27, no. 11, pp. 1502–1502, 2000.

M. Sugiyama, “Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis,” *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

V. Cherkassky, “The Nature Of Statistical Learning Theory,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 8, no. 6, pp. 1564–1564, 1997.

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.

C. W. Hsu and C. J. Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 13, no. 2, pp. 415–425, 2002.

W. M. Rand, “Objective criteria for the evaluation of clustering methods,” *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

R. J. G. B. Campello, “A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment,” *Pattern Recognition Letters*, vol. 28, no. 7, pp. 833–841, 2007.

L. Hubert and P. Arabie, “Comparing partitions,” *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

J. M. Santos and M. Embrechts, “On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification,” in *Artificial Neural Networks – ICANN 2009*, vol. 5769 of Lecture Notes in Computer Science, pp. 175–184, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

M. Van and H.-J. Kang, “Bearing defect classification based on individual wavelet local fisher discriminant analysis with particle swarm optimization,” *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 124–135, 2016.
