

Research Article

Rolling Bearing Fault Diagnosis Based on Sensitive Feature Transfer Learning and Local Maximum Margin Criterion under Variable Working Condition

Shiyuan Liu,^{1,2} Xiao Yu,³ Xu Qian ,² and Fei Dong³

¹College of Applied Science and Technology, Beijing Union University, Beijing 100083, China

²School of Mechanical Electronic & Information Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

³IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China

Correspondence should be addressed to Xu Qian; xuqian@cumtb.edu.cn

Received 20 February 2020; Revised 19 May 2020; Accepted 15 June 2020; Published 1 September 2020

Academic Editor: Cristina Castejón

Copyright © 2020 Shiyuan Liu 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 real industrial scenarios, the working conditions of bearings are variable, and it is therefore difficult for data-driven diagnosis methods based on conventional machine-learning techniques to guarantee the desirable performance of diagnosis models, as the models assume that the distributions of both the training and testing data are the same. To enhance the performance of the fault diagnosis of bearings under different working conditions, a novel diagnosis framework inspired by feature extraction, transfer learning (TL), and feature dimensionality reduction is proposed in this work, and dual-tree complex wavelet packet transform (DTCWPT) is used for signal processing. Additionally, transferable sensitive feature selection by ReliefF and the sum of mean deviation (TSFSR) is proposed to reduce the redundant information of the original feature set, to select sensitive features for fault diagnosis, and to reduce the difference between the marginal distributions of the training and testing feature sets. Furthermore, a modified feature reduction method, the local maximum margin criterion (LMMC), is proposed to acquire low-dimensional mapping for high-dimensional feature spaces. Finally, bearing vibration signals collected from two test rigs are analyzed to demonstrate the adaptability, effectiveness, and practicability of the proposed diagnosis framework. The experimental results show that the proposed method can achieve high diagnosis accuracy and has significant potential benefits in industrial applications.

1. Introduction

Rolling element bearings (REBs) are one of the most common machine elements of rotating machinery equipment in modern industry and smart manufacturing [1, 2], and the health state of REBs can seriously affect the safe and stable operation of rotary mechanical equipment [2]. REBs often operate in harsh working environments, and their failure probability is therefore higher than that of other components [3, 4]. Thus, REB fault diagnosis is of great significance for the guarantee of equipment safety and the reduction of maintenance costs [4]. In the past decade, because the vibration signals usually carry rich information about the machine operating conditions, the vibration signals collected from REBs have been commonly used as the

analytical signals in many intelligent machine fault diagnosis systems [5]. In recent years, with the rapid development of signal processing, data mining, and artificial intelligence technology, data-driven fault diagnosis has become a popular research topic [5]. Data-driven fault diagnosis consists of four steps, namely, signal collection and processing, feature extraction, feature reduction, and pattern recognition [5–8], among which feature extraction is the crucial step for the extraction of more useful information of the original vibration signals for fault pattern recognition. However, most existing data-driven intelligent diagnosis methods have two main limitations that hinder their applicability in real industrial scenarios [5, 6, 9]: (1) most existing feature extraction and fault classification models assume that the training and testing data have the same distributions. Due to

the harsh working environment and different working requirements in an industry, the working conditions are not consistent; this can therefore lead to differences between the distributions of the training and testing data [6]. (2) The variability of working conditions and the diversity of the types of failure of rotating machines often lead to insufficient labeled target fault data. Therefore, diagnostic models based on conventional machine-learning techniques that are learned with training data do not guarantee the preferred diagnosis performance of the use of testing data collected from industrial scenarios. To overcome these two limitations, it is necessary to use an improved fault diagnosis framework in which the training data are the labeled data under one working condition, and the resulting model can be applied to the unlabeled data under other working conditions.

Signal processing is the first step in data-driven REB fault diagnosis methods and has been carried out in many previous investigations by numerous scholars. Because vibration signals collected from REBs generally have nonlinear and nonstationary features, a time-frequency analysis can be effective for feature extraction [10]. Some commonly used and representative conventional time-frequency domain analysis approaches include empirical mode decomposition (EMD), short-time Fourier transform (STFT), the Wigner–Ville distribution (WVD), and wavelet transform (WT) [11]. In addition, parameterized time-frequency transform (PTFT) methods [12, 13] have been proposed to achieve a more accurate extraction of the instantaneous rotation frequency (IRF) from strong nonstationary vibration signals. In the work by Wang and Xiang [14], spline-kernelled chirplet transform (SCT), one of the PTFT methods, was employed to calculate the time-frequency distribution and extract the instantaneous rotation frequency for REB fault diagnosis under varying speed conditions. In the work by Wang et al. [15], polynomial chirplet transform (PCT), another PTFT method, was employed to estimate the IRF of REBs from the vibration signals for fault diagnosis.

EMD is a common and effective time-frequency approach in the fault diagnosis of rotating machinery and can automatically decompose nonstationary and nonlinear signals into multiple modal compositions [16, 17]. In some previous studies [18–21], EMD was employed to process the original signals and extract features for REB fault diagnosis. However, there are some limitations of EMD, such as overenveloping, end effects, and mode mixing [17, 22]. STFT is also an effective time-frequency analysis approach that can be used to divide the entire time domain into numerous segments of the same length, and each time period is an approximately stationary process [23, 24]. Some researchers [25–28] have used STFT for fault diagnosis, but its effectiveness is still hampered by the limitation of its single triangular basis [17, 29]. The WVD is also a widely used nonlinear time-frequency distribution for signal processing due to its excellent resolution and localization in the time-frequency domain [30]; however, the presence of cross-terms when they are applied to multicomponent signals can result in misleading interpretations [31, 32]. WT, including

continuous wavelet transform (CWT) and discrete wavelet transform (DWT), is an outstanding and powerful method in rotary machine diagnosis because its multiresolution capability is suitable for the analysis of nonlinear and nonstationary signals [33]. However, CWT can generate redundant data, has a huge operand, and is very time consuming [34, 35]. DWT can overcome these drawbacks of CWT, but its limitations of shift variance and frequency aliasing may lead to the loss of useful information [36]. To address these shortcomings of DWT, dual-tree complex wavelet transform (DTCWT) was proposed by Tang et al. [37, 38] and further investigated by Selesnick et al. [39] in the dyadic case. DTCWT possesses some advantageous properties [36, 37, 40], including its (1) near shift-invariance and reduced aliasing, (2) good directional selectivity, which can overcome the lack of the directional selection of DWT, (3) limited redundancy and efficient order, and (4) ability to acquire amplitude information and achieve perfect reconstruction. These properties are all beneficial for feature extraction in the task of mechanical fault diagnosis [31]. Dual-tree complex wavelet packet transform (DTCWPT) is an extension of DTCWT and can overcome the foremost limitation of DTCWT, namely, that it cannot realize multiresolution analysis in the high-frequency band. In previous research [36, 37, 40, 41], DTCWPT has been employed to process signals and extract features for REB fault diagnosis. In this paper, DTCWPT is introduced to process the original vibration signals collected from REBs.

For the construction of a feature set for fault pattern recognition, the statistical properties of the signals in the time, frequency, and time-frequency domains can be extracted to represent feature information [9, 42–44], such as the peak value (PV), root mean square (RMS), variance (V), skewness (Sw), kurtosis (K), energy, and energy entropy. In [42], after the vibration signals were processed by wavelet analysis, single-branch reconstruction signals and the corresponding HHT envelope spectrum (HES) were used to generate 192 statistical features using 6 statistical parameters for bearing fault diagnosis. In [43], vibration signals collected from REBs were decomposed into several different IMFs by EMD. The first four IMFs were selected to obtain the HHT marginal spectrum and HES, which were then used to calculate the original statistical properties. In [9], 29 statistical parameters were selected to extract 29 statistical features, which formed a high-dimensional original feature dataset for REB fault diagnosis. In [44], more than 30 feature indicators of vibration signals were calculated for axle bearings under different conditions, and the features that could more effectively and representatively reflect the fault features were selected for fault detection. In the research [45], the RMS and K were used to calculate fault features for wind turbine bearing fault diagnosis. It is often difficult to determine which statistical property can best reflect the nature of a fault from the feature space because of the complex mapping relations between some bearing faults and their signals [42, 43]. Thus, when unsuitable statistical features are chosen for fault pattern recognition, it may lead to a decline in the accuracy and efficiency of fault diagnosis. According to some previous studies [11, 42, 43], the selection

of a feature subset that is formed by fault-sensitive features is a crucial step for the achievement of the expected diagnostic accuracy.

As discussed previously, the two foremost limitations of most existing data-driven intelligent diagnosis methods [5, 6] are that (1) they assume the uniform distributions of the training and testing data are the same and (2) the variability of working conditions and the diversity of failures often lead to insufficient labeled target fault data. Recently, these two problems have garnered considerable attention and have been further investigated by some researchers. An et al. [5] proposed a novel three-layer model inspired by a recurrent neural network (RNN) and TL for REB fault diagnosis under different working conditions. Ma et al. [9], aiming at overcoming the first limitation, proposed a transfer diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains. In the work by Gao et al. [46], the finite element method (FEM) was employed to simulate samples with different faults to overcome the missing sufficient and complete fault samples. In a study by Liu et al. [47], aiming at the problem of the faulty samples of real-world running mechanical systems being difficult to obtain, a personalized fault diagnosis method for the detection of bearing faults was proposed for the activation of smart-sensor networks using FEM simulations. Some existing research has shown that TL [48] has broad application prospects and wide applicability in various fields [49–51]. Feature-based transfer, a mainstream branch of TL technology, has been used in image classification [52–54] and has inspired a novel idea for overcoming the two limitations of data-driven intelligent fault diagnosis. In this paper, the use of a novel feature extraction procedure, namely, transferable sensitive feature selection via ReliefF and the sum of standard deviation ratio (TSFRS), is proposed. TSFRS has the following two aspects: (1) it is characterized by the selection of fault-sensitive features and combines the ReliefF algorithm and the sum of within-class mean deviations (SMD) of feature data; (2) a feature-based TL method, namely, transfer component analysis (TCA), is used to reduce the differences between the marginal distributions of the training and testing data.

After the steps of signal processing and feature extraction, a high-dimensional feature set can usually be generated; if this feature set is used directly in fault pattern recognition, it will lead to very high computational complexity and the degradation of fault diagnosis accuracy [42]. Hence, dimensionality reduction is another key step that must be taken before fault pattern recognition. In fact, the dimensionality reduction of features can not only limit storage requirements and increase the algorithm speed, but can also improve the predictive accuracy of the classifier model by removing noisy and redundant features while retaining the most useful information regarding diverse bearing failures [55]. Dimensionality reduction methods can be classified into either linear or nonlinear methods. Principal component analysis (PCA) and linear discriminant analysis (LDA), as two classical linear dimensionality reduction methods, have been extensively used for linear data, but they may be invalid for nonlinear data [56]. Therefore,

some nonlinear dimensionality reduction methods, namely, kernel principal components analysis (KPCA), Isomap, Laplacian eigenmaps (LE), and local linear embedding (LLE), among others, have presented valid solutions for the dimensionality reduction of nonlinear data [56]. However, nonlinear dimensionality reduction methods have some limitations in practical applications, such as the problem of “out-of-sample” that has no explicit mapping matrix [57], the problem of the overlearning of locality [58], and high computational complexity. In recent years, some unsupervised manifold learning methods that preserve the local geometric structure on the data manifold using the linear approximation of the nonlinear mappings have been proposed, and some representative methods include locality-preserving projections (LPP) [59], neighborhood-preserving embedding (NPE) [60], and orthogonal neighborhood-preserving projection [61]. Among these manifold learning methods, LPP has attracted attention in the fault diagnosis field [62–64], but it does not utilize the label information in dimensionality reduction. LDA is a supervised dimensionality reduction method that considers the label information in feature reduction, but it cannot be directly applied when the within-class and between-class scatter matrixes are singular because of the small sample size (SSS) problem [65]. Based on the respective dominant attributes of LPP and LDA, a novel dimensionality reduction method, namely, local Fisher discriminant analysis (LFDA), was proposed by Sugiyama [66]. LFDA takes into account the label information of data while simultaneously preserving the local geometric structures of the feature data. However, LFDA only considers the neighbor relationships between samples of the same class while ignoring those between samples of different classes. Aiming at the alleviation of the SSS problem of LDA, the maximum margin criterion (MMC), a supervised dimensionality reduction method, was proposed [65]. Inspired by the attributes of LFDA and MMC, this paper proposes a novel feature reduction method, namely, the local maximum margin criterion (LMMC), an improved MMC in which both the neighbor relationships between samples of the same class and those between samples of different classes are considered.

Therefore, the contributions of this paper are summarized as follows. To solve the problem of fault diagnosis via vibration data that are variably distributed under different working conditions, a novel intelligent fault diagnosis framework of REBs based on multidomain features that systematically combine statistical feature extraction, feature-based TL, feature reduction, and pattern recognition is proposed. TSFRS, a novel feature extraction procedure, is proposed for the selection of the transferable fault-sensitive statistical features as the basis of the subsequent fault analysis. LMMC, an improved feature reduction method, is proposed for the excavation of abundant and valuable information with low dimensionality, which is beneficial for fault diagnosis. The execution of the proposed fault diagnosis framework of REBs is divided into four steps, namely, signal processing, feature extraction, feature reduction, and fault pattern recognition. First, DTCWPT is performed on raw vibration signals collected from REBs, and different terminal

nodes can be obtained. Multidomain statistical features are then extracted from the reconstructed signals of the terminal nodes to construct the original feature set. Secondly, based on the Relief algorithm and mean deviation, a new evaluation index, namely, the ratio of the feature weight value and the SMD, is employed to indicate the sensitivity of statistical features; the most sensitive features can be selected to form a feature subset that represents the fault peculiarity of REBs. Additionally, TCA is used to reduce the differences between the marginal distributions of feature datasets under different working conditions. Thirdly, LMMC is performed on the original high-dimensional feature set to acquire a new lower-dimensional projection of it. Finally, vibration signals collected from two test rigs under different working conditions are employed to validate the effectiveness, adaptability, and superiority of the proposed method for the identification and classification of REB faults. The first test rig is from Case Western Reserve University, on which two cases with 12 fault types under different motor loads of 2 hp and 3 hp are employed for validation experiments. The second test rig is an SQI-MFS test rig, on which two cases with 10 fault types under different motor speeds of 1200 rpm and 1800 rpm are employed to further verify the adaptability of the proposed method.

The remainder of this paper is organized as follows. In Section 2, the theoretical backgrounds of the DTCWPT technique, TCA technique, and MMC are summarized. In Section 3, a description of the proposed diagnosis technique is provided, and the fault diagnosis framework of REBs is illustrated. In Section 4, REB fault vibration signals collected from two experimental test rigs are investigated to verify the performance of the proposed method. Finally, the conclusion of this work is presented in Section 5. Some acronyms used in this paper are presented in Table 1.

2. Theoretical Background

2.1. Dual-Tree Complex Wavelet Packet Transform (DTCWPT). DTCWT, an enhancement of DWT, is characterized by some important properties including near shift-invariance and the inhibition of frequency aliasing components [36]. However, DTCWT cannot be used for multiresolution analysis in the high-frequency band where useful fault feature information usually exists [67]. To address this limitation, DTCWPT, which is composed of two parallel discrete wavelet packet transforms with different low- and high-pass filters, can present a more precise frequency band partition over the entire analyzed frequency band [36, 40]. DTCWPT is divided into real- and imaginary-part wavelet packet transforms, which can be, respectively, regarded as the real and imaginary trees. The real tree decomposition and the corresponding coefficients can be expressed as follows [37]:

$$\begin{cases} c_{l+1,2N}^{R_e}(k) = \sum_m h_0(m-2k)c_{l,N}^{R_e}(m), \\ c_{l+1,2N+1}^{R_e}(k) = \sum_m h_1(m-2k)c_{l,N}^{R_e}(m), \end{cases} \quad (1)$$

where $c_{l,N}^{R_e}$ is the coefficients in the real tree at a scale l and node N and h_0 and h_1 are the low-pass filter and the high-

TABLE 1: A list of acronyms used in this paper.

Acronym	Full form
DTCWPT	Dual-tree complex wavelet packet transform
TSFSR	Transferable sensitive feature selection by the Relief and sum of mean deviation
LMMC	Local maximum margin criterion
REBs	Rolling element bearings
EMD	Empirical mode decomposition
STFT	Short-time Fourier transform
WVD	Wigner-Ville distribution
WT	Wavelet transform
PTFT	Parameterized time-frequency transform
IRF	Instantaneous rotation frequency
SCT	Spline-kernelled chirplet transform
PCT	Polynomial chirplet transform
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
DTCWT	Dual-tree complex wavelet transform
PV	Peak value
RMS	Root mean square
V	Variance
Sw	Skewness
K	Kurtosis
HES	HHT envelope spectrum
RNN	Recurrent neural network
FEM	Finite element method
TL	Transfer learning
SMD	Sum of within-class mean deviations
TCA	Transfer component analysis
PCA	Principal component analysis
LDA	Linear discriminant analysis
KPCA	Kernel principal component analysis
LE	Laplacian eigenmaps
LLE	Local linear embedding
LPP	Locality-preserving projections
NPE	Neighborhood-preserving embedding
SSS	Small size sample
LFDA	Local Fisher discriminant analysis
MMC	Maximum margin criterion
MMD	Maximum mean discrepancy
MD	Mean deviations
WV	Weight value
SFS	Sensitive feature set
CWRU	Case Western Reserve University
OFS	Original feature set

pass filter, respectively. The imaginary tree decomposition and the corresponding coefficients can be expressed as follows:

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

where $c_{l,N}^{I_m}$ is the coefficients in the imaginary tree at a scale l and node N and g_0 and g_1 are the low-pass filter and the high-pass filter, respectively. When the scale l is 0, the coefficients are both equal to the original signal $x(t)$, namely, $c_{0,0}^{R_e} = c_{0,0}^{I_m} = x(t)$. The decomposition coefficients of DTCWPT are composed of $c_{l,N}^{R_e}$ and $c_{l,N}^{I_m}$, and it can be expressed as follows:

$$c_{l,N}(k) = c_{l,N}^{R_e}(k) + ic_{l,N}^{I_m}(k). \quad (3)$$

The reconstruction procedure of DTCWPT is as follows:

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

where \tilde{h}_0 and \tilde{h}_1 are the wavelet packet reconstruction filters of the real and imaginary trees, respectively. DTCWPT is characterized by two prominent advantages: (1) it is beneficial to the detection of multiple harmonic signals and (2) it can help to extract the periodic impact features of signals. Therefore, in this work, DTCWPT is used to process the original vibration signals, and the corresponding single-branch reconstruction signals of the terminal nodes are used to extract original features.

2.2. Maximum Margin Criterion (MMC) and Local Fisher Discriminant Analysis (LFDA). Linear discriminant analysis (LDA), one of the most popular methods for dimension reduction in statistics research fields [68, 69], was proposed by Fisher [70] for the dimension reduction of binary classification problems and was further extended to multiclass cases by Rao [71]. However, LDA cannot be directly applied when the within-class and between-class scatter matrixes are singular because of the SSS problem [65]. To address this drawback, Li et al. [65] and Song et al. [72] used the difference of the between-class and within-class scatter matrixes as a discriminant criterion called the maximum margin criterion (MMC), which can make the inverse matrix not to be constructed. Thus, the SSS problem in traditional LDA is alleviated.

Let $X = \{x_1, x_2, \dots, x_N\} \in R^M$ be the input data set and $l_i \in L = \{C_1, C_2, \dots, C_c\}$ be the associated class label set, where $x_i (i = 1, \dots, N)$ is an M -dimensional sample, N is the number of samples, and c is the total number of classes. To reduce the dimensionality of a sample $x \in R^M$, some measures are needed to be employed to assess the similarity or dissimilarity. We want to find a linear transformation W ,

transforming x from R^M to R^D , where $D \leq M$. After the dimensionality reduction, the similarity or dissimilarity information is preserved as much as possible. In the work by Li et al. [65], the Euclidean distance was applied to measure the dissimilarity, and the objective of the MMC is for a sample to be close to those in the same class but far from those in different classes. Thus, the MMC can be presented as follows:

$$J = \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j d(C_i, C_j), \quad (5)$$

where p_i and p_j are the prior probability of the class C_i and C_j , respectively. The $d(C_i, C_j)$ is defined as the distance between mean vectors, that is,

$$d(C_i, C_j) = d(m_i, m_j), \quad (6)$$

where m_i and m_j are the mean vectors of the classes C_i and C_j , respectively. However, due to the fact that (6) neglects the scatter of classes, (6) is not suitable. Though $d(m_i, m_j)$ is large, it is not easy to separate two classes that have the large spread and overlap with each other. For this problem, considering the scatter of classes, the between-class distance can be redefined as follows:

$$d(C_i, C_j) = d(m_i, m_j) - s(C_i) - s(C_j), \quad (7)$$

where $s(C_i)$ is some measure of the scatter of the class C_i . The generalized variance $|S_i|$ or overall variance trace S_i ($\text{tr}(S_i)$) is usually used to measure the scatter of data, where S_i is the covariance matrix of the class C_i . Thus, based on (5) and (7), two new parts can be obtained by decomposing (5):

$$\begin{aligned} J &= \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j (d(m_i, m_j) - s(C_i) - s(C_j)) \\ &= \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j (d(m_i, m_j)) - \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j (\text{tr}(S_i) + \text{tr}(S_j)). \end{aligned} \quad (8)$$

By employing the Euclidean distance, the first part in (8) can be simplified as

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j d(m_i, m_j) &= \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j (m_i - m + m - m_j)^T (m_i - m + m - m_j) \\ &= \frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j \left[(m_i - m)^T (m_i - m) + (m - m_j)^T (m_i - m) + (m_i - m)^T (m - m_j) + (m - m_j)^T (m - m_j) \right]. \end{aligned} \quad (9)$$

Because $\sum_{j=1}^c p_j = 1$ and $\sum_{j=1}^c p_j m_j = m$, (9) can be further simplified as

$$\frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j d(m_i, m_j) = \text{tr} \left(\sum_{i=1}^c p_i (m_i - m) (m_i - m)^T \right) = \text{tr}(S_b). \quad (10)$$

The second part in (8) can be simplified as

$$\frac{1}{2} \sum_{i=1}^c \sum_{j=1}^c p_i p_j (\text{tr}(S_i) + \text{tr}(S_j)) = \sum_{i=1}^c p_i \text{tr}(S_i) = \text{tr} \left(\sum_{i=1}^c p_i S_i \right) = \text{tr}(S_w). \quad (11)$$

Equation (5) can be transformed to

$$J = \text{tr}(S_b - S_w), \quad (12)$$

where S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, and $\text{tr}(S_b)$ measures the between-class separation, while $\text{tr}(S_w)$ measures the within-class cohesion.

Local Fisher discriminant analysis (LFDA), a linear supervised dimensionality reduction method, was proposed by Sugiyama [66]. LFDA can not only maximize between-class separability and preserve the within-class local manifold structure at the same time in a reduced dimensional space, but also inherits an excellent property from LDA; that is, it has an analytic form of the embedding matrix, and the solution can be easily computed by solving a generalized eigenvalue problem [66]. LFDA and LDA have the same optimization framework $J(U)$. Furthermore, LFDA incorporates local information into the definition of weight. The objective of LDA is to maximize the ratio of the between-class scatter matrix S_b to the within-class scatter matrix S_w :

$$J(U) = \max_U \frac{|U^T S_b U|}{|U^T S_w U|}, \quad (13)$$

where U is a projection matrix, and the definitions of S_b and S_w are as follows:

$$S_b = \sum_{l=1}^c n_l (\mu_l - \mu)(\mu_l - \mu)^T, \quad (14)$$

$$S_w = \sum_{l=1}^c \sum_{i: y_i=l} (x_i - \mu_l)(x_i - \mu_l)^T,$$

where n_l is the number of samples in class l , μ_l is the mean of the samples in class l , and μ is the mean of all samples:

$$\mu_l = \frac{1}{n_l} \sum_{i: y_i=l} x_i, \quad (15)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{l=1}^c n_l \mu_l,$$

where n is the number of samples. According to the literature [66, 73], S_b and S_w also have equivalent form [66, 73]:

$$S_b = \frac{1}{2} \sum_{i,j=1}^n p_{ij}^b (x_i - x_j)(x_i - x_j)^T = X(D^b - P^b)X^T, \quad (16)$$

$$S_w = \frac{1}{2} \sum_{i,j=1}^n p_{ij}^w (x_i - x_j)(x_i - x_j)^T = X(D^w - P^w)X^T, \quad (17)$$

where

$$p_{ij}^b = \begin{cases} \frac{1}{n} - \frac{1}{n_l}, & y_i = y_j = l, \\ \frac{1}{n}, & y_i \neq y_j, \end{cases} \quad (18)$$

$$p_{ij}^w = \begin{cases} \frac{1}{n_l}, & y_i = y_j = l, \\ 0, & y_i \neq y_j, \end{cases}$$

$P^b = [p_{ij}^b]_{n \times n}$ and $P^w = [p_{ij}^w]_{n \times n}$ are the weight matrices, D^b and D^w are the diagonal matrices, d_i^b is the i th diagonal samples of D^b and the sum of elements of the i th row of P^b , and d_i^w is the i th diagonal samples of D^w and the sum of elements of the i th row of P^w .

LFDA incorporates local information into the definition of weight. Thus, S_b has been replaced by \tilde{S}_b and S_w has been replaced by \tilde{S}_w . \tilde{S}_b and \tilde{S}_w are presented as follows [66]:

$$\tilde{S}_b = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^b (x_i - x_j)(x_i - x_j)^T = X(\tilde{D}^b - \tilde{P}^b)X^T,$$

$$\tilde{S}_w = \frac{1}{2} \sum_{i,j=1}^n \tilde{p}_{ij}^w (x_i - x_j)(x_i - x_j)^T = X(\tilde{D}^w - \tilde{P}^w)X^T, \quad (19)$$

where

$$p_{ij}^b = \begin{cases} A_{ij} \left(\frac{1}{n} - \frac{1}{n_l} \right), & y_i = y_j = l, \\ \frac{1}{n}, & y_i \neq y_j, \end{cases} \quad (20)$$

$$p_{ij}^w = \begin{cases} \frac{A_{ij}}{n_l}, & y_i = y_j = l, \\ 0, & y_i \neq y_j, \end{cases}$$

where $A_{ij} \in [0, 1]$ can be defined as follows:

$$A_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\gamma_i \gamma_j}\right), \quad (21)$$

where γ_i is the local scaling around x_i , defined by $\|x_i - x_i^k\|$ and x_i^k is the k -th nearest neighbor of x_i . If x_i and x_j are close to each other in the feature space, A_{ij} is large; otherwise, it is small [66]. According to (21), for the far apart sample pairs in the same class, it can be weighted and have less influence on \tilde{S}_b and \tilde{S}_w . Furthermore, the sample pairs in different classes cannot be weighted [74].

2.3. Transfer Component Analysis (TCA). Transfer component analysis (TCA) [75] is a typical feature-based TL method. Given the source domain data that are the training dataset with corresponding labels and the target domain data

that are the dataset without corresponding labels, TCA aims to reduce the difference between the marginal distributions of the different datasets by leveraging the transferable features or knowledge from the source domain [75].

A domain D consists of a D -dimensional feature space X , whose marginal probability distribution is $P(X)$, where $X = \{x_1, \dots, x_n\}$ is a training dataset, and the representation of D can be $\{X, P(X)\}$. $T = \{Y, f(X)\}$ consists of a label space Y and a predictive function $f(X)$, where $Y = \{y_1, \dots, y_n\}$ is a training dataset label and $f(X) = Q(Y|X)$ represents the conditional probability distribution. There are two learning tasks, namely, task T_S of D_S and task T_T of D_T . Feature transfer is employed to facilitate the learning process of the target predictive function $f_T(X)$ in D_T by using the knowledge and information in D_S and T_S , where $D_S \neq D_T$ or $T_S \neq T_T$ [75]. Given two datasets X_S and X_T , $P_S(X_S) \neq P_T(X_T)$, and a transformation ϕ exists such that $P_S(\phi(X_S)) \approx P_T(\phi(X_T))$ and $P_S(Y_S|\phi(X_S)) \approx P_T(Y_T|\phi(X_T))$, where ϕ is a nonlinear mapping function in a reproducing kernel Hilbert space H . The learning objective of TCA is to find a domain-invariant feature space in which the marginal distribution distance between the source domain and the target domain is minimized. The distribution distance is measured using the maximum mean discrepancy (MMD) criterion, which is defined as follows [76]:

$$\text{MMD}(X_S, X_T) = \left\| \frac{1}{n_S} \sum_{x_i \in X_S} \phi(x_i) - \frac{1}{n_T} \sum_{x_j \in X_T} \phi(x_j) \right\|_H^2 = \text{tr}(\text{KL}), \quad (22)$$

where

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix}. \quad (23)$$

In equation (22), n_S and n_T represent the numbers of source domain samples and target domain samples, respectively, $\text{tr}(\cdot)$ represents the trace of the matrix, K is a kernel matrix, and $K_{S,S}$, $K_{S,T}$, and $K_{T,T}$ are the kernel matrices in the source domain, cross domain, and target domain, respectively. L can be calculated as

$$L = \begin{cases} \frac{1}{n_S^2}, & x_i, x_j \in X_S, \\ \frac{1}{n_T^2}, & x_i, x_j \notin X_S, \\ \frac{1}{n_S n_T}, & \text{other.} \end{cases} \quad (24)$$

TCA maps the features of two domain datasets into the same kernel space through the unified kernel function. The resultant kernel matrix can be calculated as follows:

$$\tilde{K} = (KK^{-(1/2)}\tilde{W})(\tilde{W}KK^{-(1/2)}) = KWW^TK, \quad (25)$$

where $W = K^{-(1/2)}\tilde{W}$, and the distribution distance between the different domain datasets can be defined as

$$\text{MMD}(X_S, X_T) = \text{tr}((KWW^TK)L) = \text{tr}(W^TKLKW). \quad (26)$$

The complexity of W needs to be controlled by a regularization term $\text{tr}(W^TW)$, which is employed to avoid the rank deficiency of the denominator. Thus, the objective function of TCA can be rewritten as

$$\begin{aligned} \min_W \quad & \text{tr}(W^TKLKW) + \mu \text{tr}(W^TW) \\ \text{s.t.} \quad & W^TKHKW = I, \end{aligned} \quad (27)$$

where μ is a trade-off parameter and can be used to guarantee that the optimization objective can be well defined. $I \in R^{m \times m}$ represents an identity matrix. H is a centering matrix. $W^TKHKW = I$ can avoid the trivial solution $W = 0$.

According to the introduction of TCA, the optimization objective of TCA is that the latent space spanned by the learned samples preserves the variance of the data and minimizes the marginal distributions between the different domain datasets as much as possible. The optimization problem of equation (27) can be efficiently solved by the trace optimization problem.

3. Proposed Method and System Framework

3.1. Feature Extraction Procedure TSFRS (Transferable Sensitive Feature Selection by ReliefF and the Sum of Mean Deviation). TSFRS has two components: (1) the selection of fault-sensitive features, which combines the ReliefF algorithm and the sum of within-class mean deviations (SMD) of feature data, and (2) the feature-based TL method, in which TCA is used to reduce the difference in the marginal distributions between the training and testing data.

In this paper, it is suggested that the sensitive statistical features be selected before the implementation of fault pattern recognition. Thus, the ReliefF algorithm [77] and MD are employed for a dataset that includes different statistical features for the case of REB conditions. Each type of statistical feature is evaluated by the ReliefF algorithm to determine its weight value (WV). ReliefF, a supervised algorithm for feature ranking, is usually applied in data preprocessing as a feature subset selection method. The basic concept of ReliefF is to compute instances at random, compute their nearest neighbors, and adjust a feature weighting vector to give more weight to features that discriminate the instances from the neighbors of different classes. For each kind of the statistical feature, the MD of the feature data samples in each REB condition can be calculated, and the sum of MD in all REB conditions can be further calculated. Aiming at the evaluation of each statistical feature, the higher the WV, the greater the discriminative degree of the feature class. The lower the value of MD, the greater the class cohesion of the characteristic. Therefore, the ratio of WV and SMD is selected to indicate the sensitivity of a statistical feature, based on which the sensitive feature subset can be selected from the original feature set.

Furthermore, the variable working conditions of REBs in industry scenarios can lead to distribution differences between the training and testing data [6]. Therefore, after the construction of the sensitive feature subset, TCA is employed to reduce the difference between the distributions

of the sensitive feature training and testing subsets. The description of TSFRS is summarized as the following steps.

Step 1. In the training samples, there are M types of REB faults, N vibration signal samples in each type of REB fault pattern, and K types of statistical features. Via the processing of the vibration signals, original feature sets $[FS^1, FS^2, \dots, FS^K]$ can be obtained, where FS^k can be expressed by

$$FS^k = \begin{bmatrix} S_{11}^k & S_{12}^k & \cdots & S_{1N}^k \\ S_{21}^k & S_{22}^k & \cdots & S_{2N}^k \\ \vdots & \vdots & \ddots & \vdots \\ S_{M1}^k & S_{M2}^k & \cdots & S_{MN}^k \end{bmatrix}, \quad (28)$$

where S_{ij}^k is the k -th statistical feature of the j -th sample in the i -th type of REB fault. Next, FS^k can be evaluated to obtain the corresponding feature WV using the ReliefF algorithm, and the WV of each statistical feature can be used to evaluate the distinguishability of the feature. The higher the WV, the greater the discriminative degree of the feature class.

Step 2. The MD of feature samples of a type of statistical feature in each type of REB condition is calculated, i.e., the MD of the elements of row of FS^k . Therefore, an MD set, $[MD_1^k, MD_2^k, \dots, MD_M^k]$, can be obtained, where MD_i^k can be expressed by

$$MD = \frac{\sum_{i=1}^N |x_i - \bar{x}|}{N}, \quad (29)$$

where

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N}. \quad (30)$$

Next, $SMD(k)$ can be obtained, that is, the sum of the MD of feature samples of the k -th statistical feature for all cases of REB conditions; $SMD(k)$ can be expressed by

$$SMD(k) = \sum_{i=1}^M MD_i^k. \quad (31)$$

In this paper, it is presumed that the MD can be used to express the cohesion of data. Thus, there is a mean deviation sequence $SMD = \{SMD(1), SMD(2), \dots, SMD(K)\}$, which becomes another evaluation index for sensitive feature selection. The lower the value of $SMD(k)$, the greater the class cohesion of the feature.

Step 3. A new sequence $WSD = \{WSD(1), WSD(2), \dots, WSD(K)\}$ is obtained, where the definition of $WSD(k)$ is as follows:

$$WSD(k) = \frac{WV(k)}{SMD(k)}. \quad (32)$$

In this paper, it is presumed that the greater the value of $WSD(k)$, the better the fault sensitivity of the corresponding feature elements. Therefore, the sorted ratio sequence of WV

and SMD can be obtained by sorting the WSD in the descending mode.

Step 4. For the labeled training data under one working condition and unlabeled testing data under another working condition, based on the training data, the sorted sequence WSD in the descending mode is acquired and is used to select the most sensitive statistical features that can construct a sensitive feature set (SFS). The most sensitive statistical features will be directly applied to the extraction of features for testing data. Thus, two sensitive feature sets can be obtained; the first is the SFS of training data, called SFS_{train} , and the other is the SFS of testing data, called SFS_{test} . Furthermore, SFS_{train} and SFS_{test} are used as the input of TCA, and a new feature set $TSFS_{test}$, in which the difference in marginal distributions between SFS_{train} and $TSFS_{test}$ is minimized, can be generated.

3.2. Local Maximum Margin Criterion (LMMC).

Although the MMC can avoid the SSS problem of LDA, it may be invalid for nonlinear datasets due to the lack of consideration of the local structure of the dataset. LFDA considers the neighbor relationships between samples of the same class while ignoring those between samples of different classes. Aiming at this problem and inspired by the attributes of the MMC and LFDA, this paper proposes a novel feature reduction method, LMMC, which is an improved MMC. The LMMC naturally inherits the merits of the MMC and LFDA, and the underlying idea of the solution to the problem mentioned previously is that the optimization objective of LFDA can be integrated into the MMC; in addition, the neighbor relationships between samples of different classes are taken into consideration.

Based on the descriptions of the MMC and LFDA provided in Section 2, the optimization objective of the LMMC can be obtained by combining the optimization objectives of the MMC and LFDA. In addition, the LMMC has an improvement on the local information of the definition of the weight. The LMMC and MMC have the same optimization framework, but S_b and S_w are, respectively, replaced by S_b^L and S_w^L . The objective function can be presented as follows:

$$J = \text{tr}(S_b^L - S_w^L). \quad (33)$$

According to equations (16) and (17), the local S_b^L and the local S_w^L are defined as follows:

$$S_b^L = \frac{1}{2} \sum_{i,j=1}^n \tilde{P}_{ij}^{Lb} (x_i - x_j)(x_i - x_j)^T = X(\tilde{D}^{Lb} - \tilde{P}^{Lb})X^T, \quad (34)$$

$$S_w^L = \frac{1}{2} \sum_{i,j=1}^n \tilde{P}_{ij}^{LW} (x_i - x_j)(x_i - x_j)^T = X(\tilde{D}^{LW} - \tilde{P}^{LW})X^T, \quad (35)$$

where

$$\tilde{P}_{ij}^{Lb} = \begin{cases} A_{ij} \left(\frac{1}{n} - \frac{1}{n_i} \right), & y_i = y_j = l, \\ \frac{n_j}{n}, & y_i \neq y_j (j \in Nst(i)), \\ \frac{1}{n}, & \text{else,} \end{cases} \quad (36)$$

$$\tilde{P}_{ij}^{LW} = \begin{cases} \frac{A_{ij}}{n_l}, & y_i = y_j = l, \\ 0, & y_i \neq y_j, \end{cases} \quad (37)$$

where \tilde{P}^{Lb} and \tilde{P}^{LW} are the weight matrices and \tilde{D}^{Lb} and \tilde{D}^{LW} are the diagonal matrices. In \tilde{P}^{Lb} , the means of $y_i \neq y_j (j \in Nst(i))$ is that j is the nearest neighbor of i , and they belong to different classes.

According to equations (30)–(37), the local structure of the dataset, including the neighbor relationships between samples of the same class and the neighbor relationships between samples of different classes, can be considered into the dimensionality reduction by changing the WV.

Let $W \in R^{M \times L}$ be a linear transformation that transforms the high-dimensional dataset from R^M to R^D , where $D \leq M$. Thus, in the lower-dimensional space, the scatter matrices, respectively, become $S_b^{LL} = W^T S_b^L W$ and $S_w^{LL} = W^T S_w^L W$, where W can be determined by maximizing

$$J(W) = \text{tr}(S_b^{LL} - S_w^{LL}) = \text{tr}(W^T (S_b^L - S_w^L) W). \quad (38)$$

It is assumed that W is constituted by the unit vectors, that is, $W = [w_1, \dots, w_L]$ and $w_k^T w_k = 1$, $k = 1, 2, \dots, L$. Thus, W can be obtained by solving the following constrained optimization:

$$\begin{aligned} \max \quad & \sum_{k=1}^L w_k^T (S_b^L - S_w^L) w_k \\ \text{s.t.} \quad & w_k^T w_k = 1, k = 1, 2, \dots, L. \end{aligned} \quad (39)$$

The above constrained optimization can be transformed to the eigenvalue problem:

$$(S_b^L - S_w^L) w_k = \lambda_k w_k. \quad (40)$$

Thus,

$$J(W) = \sum_{k=1}^L w_k^T (S_b^L - S_w^L) w_k = \sum_{k=1}^L \lambda_k w_k^T w_k = \sum_{k=1}^L \lambda_k. \quad (41)$$

According to equation (41), W is composed of the eigenvectors of $(S_b^L - S_w^L)$ corresponding to the first L largest nonnegative eigenvalues. Finally, with the utility of the LMMC, the low-dimensional feature matrices of the training and testing datasets can be obtained with more sensitive and less redundant information for REB fault diagnosis.

3.3. System Framework. The implementation of the proposed fault diagnosis framework is presented in Figure 1, in which the statistical analysis and artificial intelligence

approaches are systematically blended to detect and diagnose REB faults under different working conditions. The entire fault diagnosis procedure is divided into four steps, namely, signal processing, feature extraction, feature reduction, and fault pattern recognition.

In the signal processing step, vibration signals collected from REBs under different working conditions are decomposed into different wavelet packet nodes by DTCWPT. The single-branch reconstruction signals of terminal nodes will be employed to extract statistical features. In the feature extraction step, with the utilization of the proposed TSFRS, the most sensitive statistical features can be selected based on the training dataset to construct a sensitive feature subset for the training classifier, and these most sensitive statistical features will be directly applied to the extraction of features for the testing dataset. The sensitive feature subsets of the training and testing datasets are, respectively, used as the source domain data and the target domain data. TCA is used to reduce the difference in marginal distributions of the source domain data and the target domain data. For the feature reduction, the low-dimensional training feature space is acquired by the proposed LMMC, which generates a projection that can be directly used for the dimensionality reduction of the testing feature dataset; thus, the low-dimensional testing feature dataset can be obtained. The WSD and projection matrix W are obtained by processing the training dataset and can be directly used for the testing dataset. In the final step, the low-dimensional training feature dataset is employed as the input of the fault type to train the classifier. In this paper, support vector machine (SVM) is used as the fault pattern recognition classifier. The trained classifier will be employed to conduct fault pattern recognition using the low-dimensional testing feature dataset. Finally, the procedure of the proposed method outputs the fault identification and classification accuracy.

4. Experiments and Analysis Results

4.1. Experiments Based on Experimental Test Rig 1

4.1.1. Experimental Setup and Cases. The REB vibration data from Case Western Reserve University (CWRU) [78], which reproduces several fault scenarios, were used to verify the effectiveness of the proposed methods. The experimental test rig is presented in Figure 2; the test rig was composed of an electric motor (left), a torque transducer/encoder (center), a dynamometer (right), and control circuitry (not shown). An SKF6205-2RS deep-groove REB was used in the test rig, and electro-discharge machining was employed to set single-point defects with different fault diameters, namely, 0.007, 0.014, 0.021, and 0.028 inches. The collected vibration signals of the REBs consisted of inner race fault signals, ball fault signals, outer race fault signals, and normal signals. The test rig supported a motor load of 0–3 horsepower (hp), and the corresponding motor speeds were 1730 to 1797 rpm. Three accelerometers were, respectively, placed at the 12 o'clock position. The sampling frequency was 12 kHz for the drive-end and fan-end bearings.

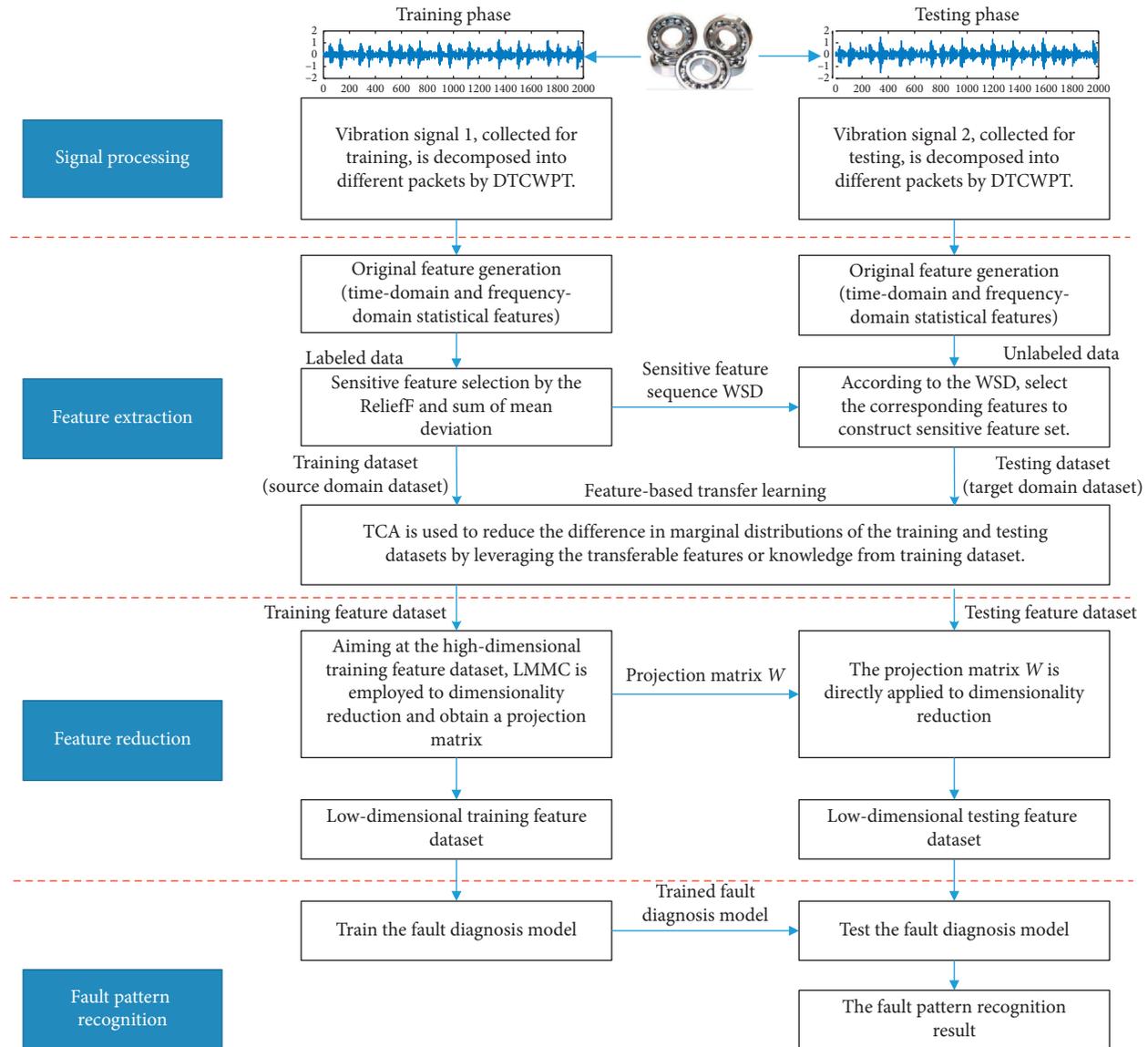


FIGURE 1: Implementation of the proposed fault diagnosis framework.

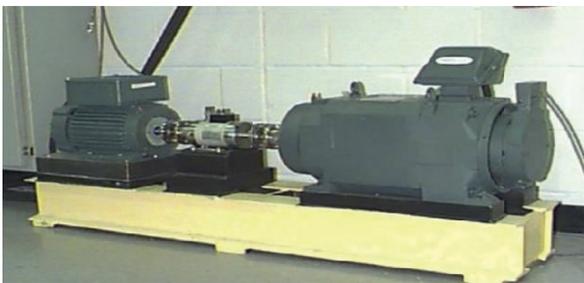


FIGURE 2: Experimental test rig 1.

In order to verify the effectiveness, adaptability, and the practical value of the proposed bearing fault diagnosis framework under different working conditions, the vibration signals of different fault types and diameters under different motor loads are employed. The signal samples of

2 hp and 3 hp are applied for experiments, and there are four bearing conditions (normal, ball fault, inner race fault, and outer race fault). Ball fault and inner race fault have four fault diameters (0.007 inches, 0.014 inches, 0.021 inches, and 0.028 inches, respectively). Outer race fault has three fault diameters (0.007 inches, 0.014 inches, and 0.021 inches). Therefore, there are 12 bearing conditions which can correspond to 12 patterns for fault diagnosis. The bearing vibration signals are divided into several data segments, and each segment which is used as a sample has 2000 data points. Each bearing condition contains 60 samples, among which 40 random samples are selected as the testing samples and 20 random samples are selected as the training samples. Based on these samples, two group datasets are used in experiments. The first group dataset includes two cases (cases 1 and 2), and the samples of 2 hp are used as training samples. In case 1, the samples of 2 hp are used as testing samples. In case 2, the samples of 3 hp are used as testing samples. The second

TABLE 2: The detailed information of the first group dataset.

Bearing condition	Defect diameter (mm)	Number of training samples	Number of testing samples		Class label
		2 hp	2 hp (case 1)	3 hp (case 2)	
Normal	0	20	40	40	1
	0.007	20	40	40	2
	0.014	20	40	40	3
	0.021	20	40	40	4
	0.028	20	40	40	5
Ball 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
Inner 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	

TABLE 3: The detailed information of the second group dataset.

Bearing condition	Defect diameter (mm)	Number of training samples	Number of testing samples		Class label
		3 hp	3 hp (case 3)	2 hp (case 4)	
Normal	0	20	40	40	1
	0.007	20	40	40	2
	0.014	20	40	40	3
	0.021	20	40	40	4
	0.028	20	40	40	5
Ball 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
Inner 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	

group dataset includes two cases (cases 3 and 4), the samples of 3 hp are used as training samples. In case 3, the samples of 3 hp are used as testing samples. In case 4, the samples of 2 hp are used as testing samples. The detailed information of the two group experimental datasets is shown in Tables 2 and 3, respectively.

4.1.2. Analysis Results. According to the diagnosis framework shown in Figure 1, each sample is decomposed into different wavelet packet nodes by DTCWPT, and the decomposition level is 4. Thus, 16 terminal nodes, namely, subband signals, can be obtained. Then, 16 single-branch reconstruction signals of terminal nodes can be obtained, and 16 single-branch reconstruction signals are selected to generate 16 Hilbert envelope spectra (HES). By using the 6 statistical parameters shown in Table 4, each single-branch reconstruction signal can generate 6 statistical features by calculating 6 statistical parameters, and each HES can generate 6 statistical features by calculating 6 statistical parameters. Thus, 16 single-branch reconstruction signals and 16 HES can generate 192 statistical features which compose the original feature set (OFS). Then, the TSFRS is performed to select the sensitive statistical

TABLE 4: Statistical parameters.

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

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

features and reduce the difference of distribution between the training sensitive feature subset and the testing sensitive feature subset. The WV, SSMD, and WSD of 192 statistical features of the training samples (2 hp) are presented in Figures 3–5, respectively. In Figure 4, the horizontal axis represents the number of statistical features. The 1–6, 7–12, \dots , 85–90, and 91–96 represent time domain features of single-branch reconstruction signals of terminal wavelet packet nodes 1–16, respectively. The 97–102, 103–108, \dots , 181–186, and 187–192 represent the HES features. After the procedure of TSFRS, the

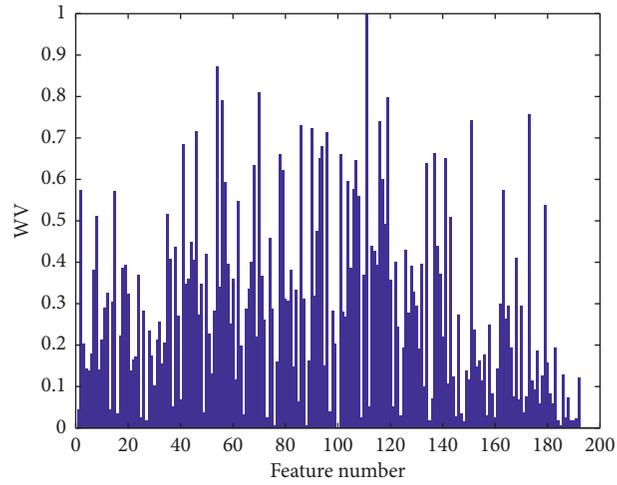


FIGURE 3: The WV of 192 statistical features of the training samples.

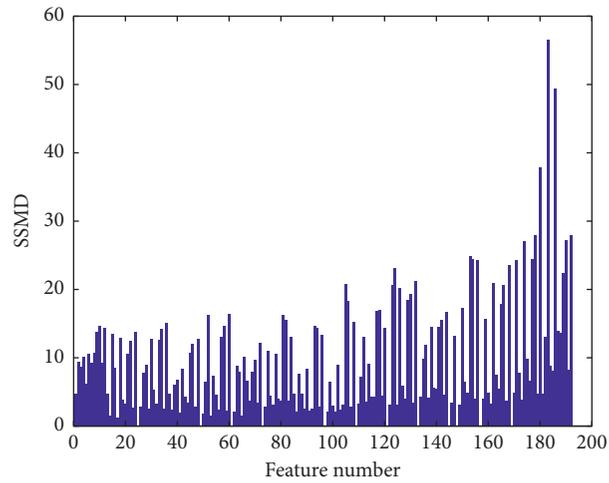


FIGURE 4: The SSMD of 192 statistical features of the training samples.

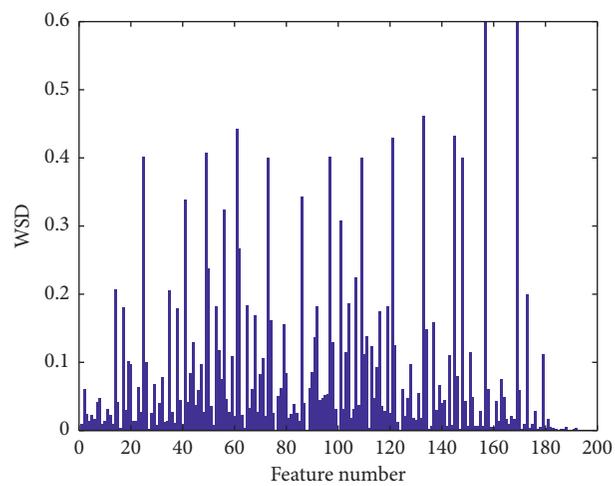


FIGURE 5: The WSD of 192 statistical features of the training samples.

TABLE 5: Bearing fault diagnosis results obtained by OFS-SVM.

DTCWPT				WPT			
Testing accuracy (%)				Testing accuracy (%)			
Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
98.13	83.33	98.13	78.54	95.00	73.13	97.71	76.67

TABLE 6: Bearing fault diagnosis results obtained by OFS-PCA-SVM.

Dimension size	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	95.21	72.08	92.92	73.33	89.58	59.79	94.17	70.21
10	98.33	81.46	96.67	81.04	96.25	75.00	98.33	69.58
15	98.75	84.58	98.33	85.00	97.70	73.33	98.54	81.88
20	99.38	85.00	97.71	83.75	96.88	76.46	99.17	83.54
30	99.38	85.00	97.92	86.04	97.71	78.75	98.96	84.58

TABLE 7: Bearing fault diagnosis results obtained by OFS-LDA-SVM.

Dimension size	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	96.25	80.62	95.00	82.08	92.50	77.71	95.00	76.88
7	98.13	86.25	95.83	88.33	94.58	80.42	97.71	71.67
9	98.75	86.04	96.25	88.75	95.21	80.42	97.71	76.67
11	99.17	85.83	96.46	88.54	95.63	80.21	98.13	77.08

TABLE 8: Bearing fault diagnosis results obtained by OFS-LFDA-SVM.

Dimension size	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	98.75	81.67	94.79	72.71	81.46	54.79	86.88	52.08
7	99.58	86.45	95.83	73.33	82.71	55.42	92.08	53.75
9	99.58	86.67	96.46	73.54	82.92	56.25	92.92	53.75
11	99.79	86.88	96.46	73.54	83.13	56.25	92.92	53.75

TABLE 9: Bearing fault diagnosis results obtained by OFS-MMC-SVM.

Dimension size	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	92.71	79.79	96.46	71.67	95.00	75.00	96.67	62.71
7	97.29	82.92	96.25	80.42	95.21	72.08	99.17	80.21
9	98.75	86.04	95.42	82.08	96.25	72.50	99.17	82.71
11	99.17	86.46	97.08	85.21	96.25	72.50	98.96	83.54

TABLE 10: Bearing fault diagnosis results obtained by OFS-LMMC-SVM.

Dimension size	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	99.79	74.79	99.79	88.13	96.04	63.54	97.71	69.17
7	99.79	83.54	99.38	90.00	97.92	75.83	98.54	80.00
9	99.79	83.13	100.00	92.08	98.33	75.21	99.38	80.42
11	100.00	93.75	100.00	91.88	98.13	82.29	99.58	90.00

TABLE 11: Testing accuracies obtained by OFS-TSFRS-SVM.

sf _n	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
10	93.54	82.70	94.58	72.08	96.04	52.92	75.63	53.13
20	95.00	88.54	94.17	68.33	98.96	66.04	95.83	89.58
30	94.38	86.46	93.54	69.38	98.96	66.04	97.29	82.08
40	94.38	86.46	93.75	69.38	95.63	79.79	97.92	99.38
50	96.88	86.25	96.04	73.13	96.46	84.17	99.38	94.38
70	97.71	86.04	96.25	73.96	95.21	59.58	98.54	94.38
90	97.92	83.13	96.67	75.00	95.21	59.58	98.96	94.79
120	98.75	81.88	97.08	79.38	98.13	67.08	97.50	74.79
140	98.75	83.75	97.92	77.92	97.29	69.79	97.92	74.58
160	98.75	83.96	97.92	77.92	97.71	70.42	98.13	77.29
180	98.96	83.54	98.13	78.96	96.67	74.79	96.88	76.46

TABLE 12: Testing accuracies obtained by OFS-TSFRS-PCA-SVM.

sf _n	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
10	91.88	77.29	91.04	66.88	96.25	53.75	93.96	37.08
20	97.71	83.75	94.58	72.92	93.38	66.04	93.54	58.54
30	97.50	85.63	94.79	74.17	99.38	67.29	91.67	62.29
40	97.50	85.63	94.79	74.17	98.96	73.13	91.67	38.13
50	97.92	88.75	95.63	84.38	98.54	81.88	90.83	43.33
70	97.71	87.92	97.71	82.29	96.88	67.50	91.04	39.58
90	97.29	88.54	97.92	76.25	96.88	77.92	90.21	40.00
120	96.67	79.79	97.29	82.92	98.75	75.83	96.04	57.50
140	99.17	83.33	96.88	80.63	98.33	77.50	95.21	58.96
160	99.38	83.33	96.88	80.63	97.92	78.33	97.71	72.92
180	99.38	85.00	97.08	80.83	97.29	76.88	97.92	68.13

Dimension size is 10.

TABLE 13: Testing accuracies obtained by OFS-TSFRS-LDA-SVM.

sf _n	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
20	95.42	90.83	93.96	76.25	97.29	73.54	95.83	63.13
30	95.41	85.42	92.50	68.96	97.71	73.54	98.13	72.92
40	95.42	85.42	92.50	68.96	95.63	76.25	93.54	51.25
50	98.96	89.17	95.63	73.54	95.42	77.71	95.21	57.71
70	98.75	89.58	95.83	74.79	96.67	81.25	94.79	58.54
90	98.96	91.25	95.42	78.75	97.71	81.25	93.54	60.42
120	98.33	83.33	96.04	80.63	98.13	80.83	97.50	67.08
140	99.17	85.83	96.04	88.33	97.29	81.25	95.00	64.38
160	99.17	85.83	96.04	88.33	97.29	81.25	98.13	80.83
180	99.17	85.83	96.67	87.50	96.88	82.08	97.92	77.08

Dimension size is 11.

feature reduction method LMMC is further performed to obtain a low-dimensional feature set which is used as the input of the SVM.

In order to verify the effectiveness of the proposed TSFRS and LMMC, two group comparative experiments are

performed. In addition, WPT is also used in experiments, and the results of experiments using WPT are compared with those of DTCWPT, which can help to verify the superiority of DTCWPT. In this paper, the training dataset is employed to train the fault diagnosis model, the testing

TABLE 14: Testing accuracies obtained by OFS-TSFRS-LFDA-SVM.

sfn	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
20	99.79	94.58	99.38	85.83	97.71	74.79	97.50	72.29
30	99.17	98.75	99.58	90.63	99.58	66.46	98.54	67.50
40	99.58	98.96	99.79	97.50	99.17	65.63	98.96	76.88
50	90.79	66.25	100.00	67.08	99.58	64.79	99.58	66.46
70	100.00	66.67	100.00	73.33	99.79	61.46	99.38	67.71
90	100.00	72.08	100.00	66.88	99.58	59.79	99.79	58.54
120	100.00	80.42	100.00	83.33	99.79	63.75	99.38	61.46
140	100.00	89.38	99.79	90.83	99.17	64.38	98.54	71.25
160	100.00	97.29	99.38	83.54	96.04	62.50	95.42	66.25
180	99.17	88.75	98.96	76.46	96.04	62.08	91.25	71.04

Dimension size is 11.

TABLE 15: Testing accuracies obtained by OFS-TSFRS-MMC-SVM.

sfn	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
20	97.92	92.08	94.79	75.63	94.79	71.67	93.75	68.54
30	96.04	86.04	93.75	73.54	96.04	71.88	96.88	72.71
40	96.04	86.04	93.75	73.54	96.25	75.83	93.96	53.33
50	97.71	78.54	94.79	71.46	97.71	82.50	94.38	54.38
70	97.71	77.50	96.25	76.46	96.88	61.67	95.42	56.46
90	98.33	82.29	97.29	77.08	96.88	68.54	93.96	60.83
120	99.17	80.21	97.92	84.79	97.50	75.83	97.92	71.46
140	99.58	86.46	96.46	84.79	98.54	79.17	98.75	71.46
160	99.58	86.46	96.46	85.42	98.54	76.88	98.96	79.38
180	99.58	86.46	96.88	85.83	96.46	73.54	98.96	82.08

Dimension size is 11.

TABLE 16: Testing accuracies obtained by OFS-TSFRS-LMMC-SVM.

sfn	DTCWPT				WPT			
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
20	98.33	95.00	99.79	85.21	99.38	72.08	96.67	77.08
30	99.58	99.17	100.00	90.63	99.79	66.67	99.38	88.75
40	100.00	100.00	100.00	93.75	100.00	67.08	99.79	91.25
50	100.00	83.33	100.00	91.46	100.00	69.58	99.79	90.83
70	100.00	83.13	100.00	91.88	100.00	75.63	100.00	91.25
90	100.00	83.33	100.00	94.38	100.00	76.46	99.79	96.67
120	100.00	91.04	100.00	99.58	99.79	82.50	99.79	99.58
140	100.00	90.42	100.00	99.79	99.38	66.67	99.79	96.25
160	100.00	91.67	100.00	91.25	98.96	80.42	99.58	91.88
180	100.00	96.67	100.00	91.46	99.17	81.04	99.79	90.83

Dimension size is 11.

dataset is employed to test the fault diagnosis model, and the accuracy results presented in a series of tables and figures are the average diagnostic accuracy of 12 bearing conditions. Thus, we use the average diagnostic accuracy results for experimental analysis, and the detailed experimental analysis is described as follows.

In the first group of experiments, the TSFRS is not applied. The OFS containing 192 statistical features is

directly processed by some dimensionality reduction methods, namely, PCA, LDA, MMC, LFDA, and LMMC. OFS-SVM is a diagnosis model based on SVM, in which the OFS is used as the input of the SVM. OFS-PCA/LFDA/LFADA/MMC/LMMC-SVM are also SVM-based diagnosis models that, respectively, use PCA, LDA, LFDA, MMC, and LMMC. According to results shown in Tables 5–10, for the cases 1 and 2, the performance of each model using DTCWPT is

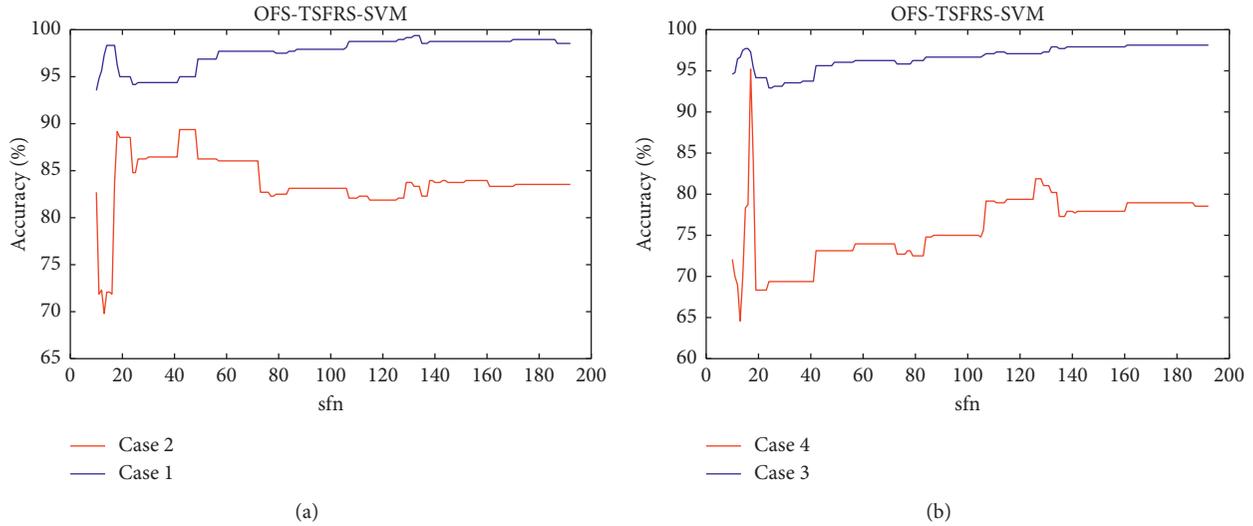


FIGURE 6: The curve representation of the diagnosis results of OFS-TSFRS-SVM using DTCWPT with different sfn.

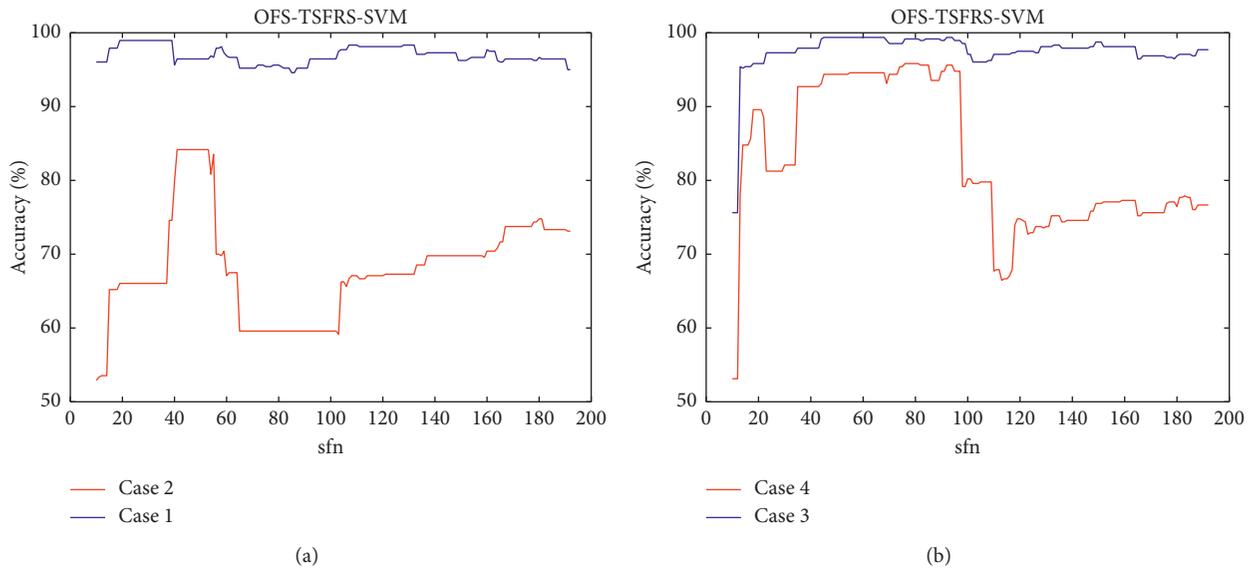


FIGURE 7: The curve representation of diagnosis results of OFS-TSFRS-SVM using WPT with different sfn.

better than that of the model using WPT; for the cases 3 and 4, the performance of OFS-SVM, OFS-LFDA-SVM, and OFS-LMMC-SVM models using DTCWPT is better than that of these models using WPT. For the OFS-PCA-SVM, OFS-LDA-SVM, and OFS-MMC-SVM models using DTCWPT, the diagnosis results of case 4 are better than that of models using WPT. In general, the DTCWPT has more advantages than WPT.

The detailed experimental results of all models using DTCWPT are presented below. For the testing set of case 1, all models can obtain preferable diagnosis accuracy. The maximum accuracy of each model can attain over 98%, and the highest accuracy can attain 100%, which is obtained by

OFS-LMMC-SVM. For the testing set of case 2, the working condition is different from the training set. The diagnosis accuracy of OFS-SVM can only attain 83.33%, compared with OFS-SVM, and all models have enhancement in diagnosis accuracy. But, the performance of OFS-LMMC-SVM is better than that of other models, and the highest accuracy of OFS-LMMC-SVM can attain 93.75% when the dimension size is 11. For the testing set of case 3, the maximum accuracy of each model can attain over 96%, and the highest accuracy can attain 100%, which is obtained by OFS-LMMC-SVM. For the testing set of case 4, the diagnosis accuracy of OFS-SVM can only attain 78.54%, and the model using PCA, LDA, MMC, and LMMC has an obvious

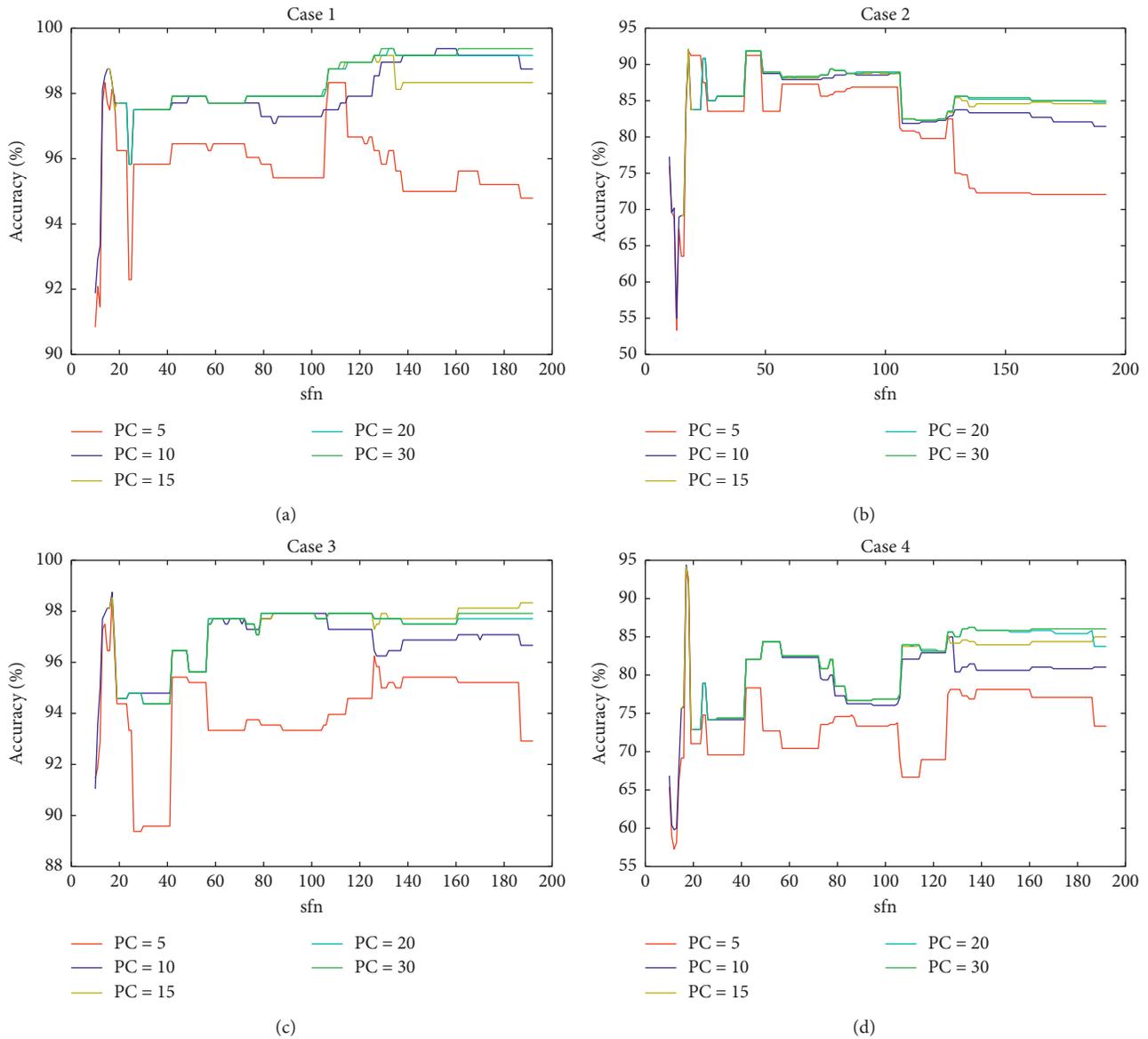


FIGURE 8: The curve representation of diagnosis results obtained by OFS-TSFRS-PCA-SVM using DTCWPT with different dimension sizes for PCA. The PC represents the number of dimension size.

enhancement in diagnosis accuracy, respectively. The highest diagnosis accuracy can attain 92.08%, which is obtained by OFS-LMMC-SVM. According to the experimental results of the four testing cases under various models, it is evident that the fault diagnosis model using LMMC can achieve preferable diagnosis performance.

In the second group of experiments, the TSFRS is applied before the implementation of feature reduction and fault pattern recognition. OFS-TSFRS-SVM is an SVM-based diagnosis model, in which the most sensitive features can be selected from OFS according to WSD; in addition, the TCA is employed to reduce difference of distribution between the training sensitive feature subset and the testing sensitive feature subset. OFS-TSFRS-PCA/LDA/LFDA/MMC/LMMC-SVM are also SVM-based diagnosis models that, respectively, use PCA, LDA, LFDA, MMC, and LMMC.

According to Tables 11–16 and Figures 6–19, the detailed experimental results of all models using DTCWPT are presented in the following.

For the testing set of case 1, all models can achieve preferable performance, which is reflected on diagnosis accuracy. The maximum diagnosis accuracy of each models can attain over 98%, and the highest diagnosis accuracy can attain 100%, which is obtained by both OFS-TSFRS-LFDA-SVM and OFS-TSFRS-LMMC-SVM. For the testing set of case 2, as compared with the experimental results of the first group, diagnosis accuracies of all models using TSFRS appear with an enhancement. Among the models mentioned above, the performance of OFS-TSFRS-LFDA-SVM and OFS-TSFRS-LMMC-SVM is better than that of OFS-TSFRS-SVM, OFS-TSFRS-PCA-SVM, OFS-TSFRS-LDA-SVM, and OFS-TSFRS-MMC-SVM. The maximum diagnosis

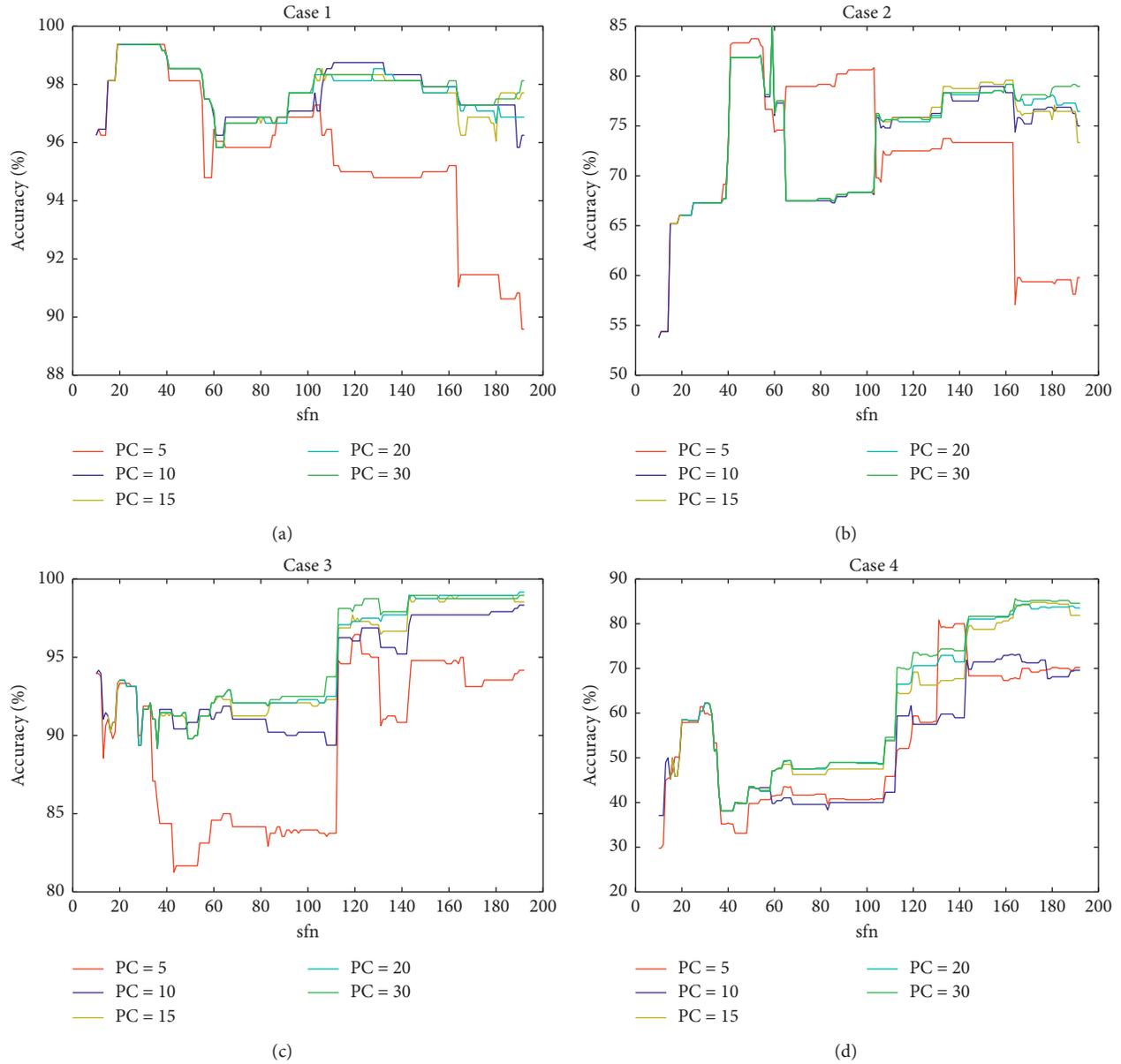


FIGURE 9: The curve representation of diagnosis results obtained by OFS-TSFRS-PCA-SVM using WPT with different dimension sizes for PCA. The PC represents the number of dimension size.

accuracies of OFS-TSFRS-LMMC-SVM and OFS-TSFRS-LFDA-SVM can attain over 98%, but the highest diagnosis accuracy of OFS-TSFRS-LMMC-SVM can attain 100%. For the testing set of case 3, the maximum accuracy of each model can attain over 96%, and OFS-LMMC-SVM can achieve 100% fault diagnosis accuracy, which is higher than that of the other models. For the testing set of case 4, as compared with the experimental results of the first group, diagnosis accuracies of all models using TSFRS appear with an enhancement. The performance of OFS-TSFRS-LMMC-SVM is better than that of other models, and the highest diagnosis accuracy can attain 99.79% (when the sfn (selected feature number) is 140).

According to the experimental results of the second group, when a suitable parameter sfn is selected, it can achieve a desirable improvement on the performance of the fault diagnosis model, which can attain a preferable diagnosis accuracy. According to Figures 6–19, it is evident that the fault diagnosis model can attain better diagnosis performance when a suitable sfn is selected. For example, for the testing sets of cases 1, 2, and 3, the diagnosis accuracy of OFS-TSFRS-LMMC-SVM can attain 100% when the sfn is between 35 and 42, and for the testing set of case 4, the diagnosis accuracy of OFS-TSFRS-LMMC-SVM can attain 99.79% when the sfn is between 137 and 146. In summary, the effectiveness of the proposed TSFRS and LMMC can be

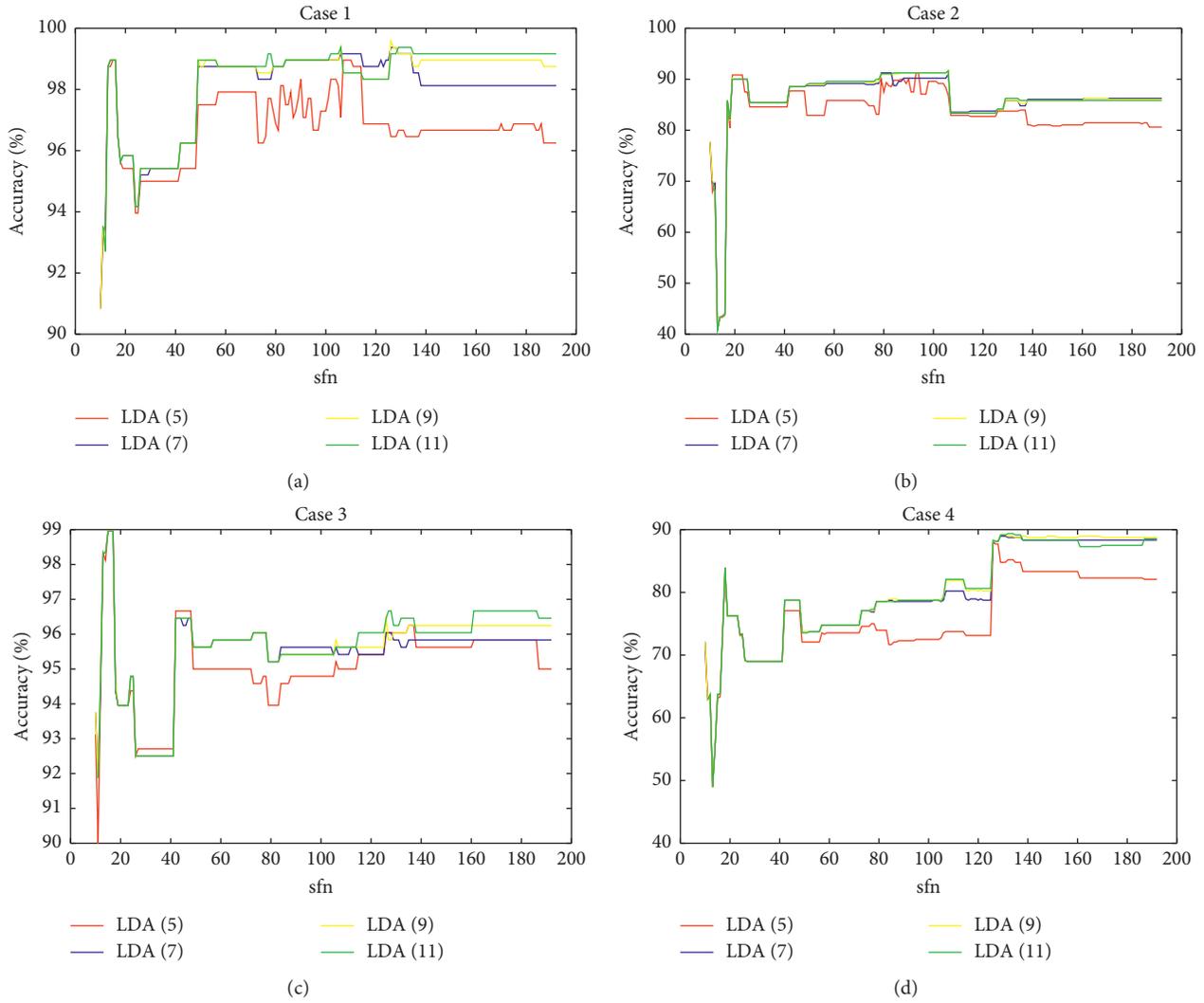


FIGURE 10: The curve representation of diagnosis results obtained by OFS-TSFRS-LDA-SVM using DTCWPT with different dimension sizes for LDA. The LDA (5) represents that the number of dimension size is 5.

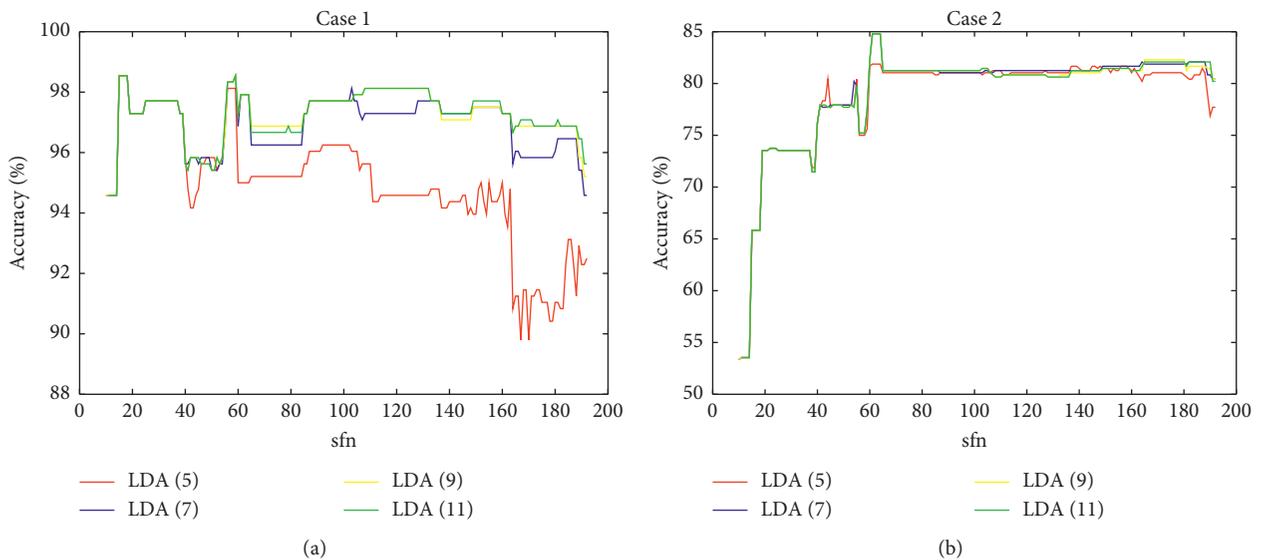


FIGURE 11: Continued.

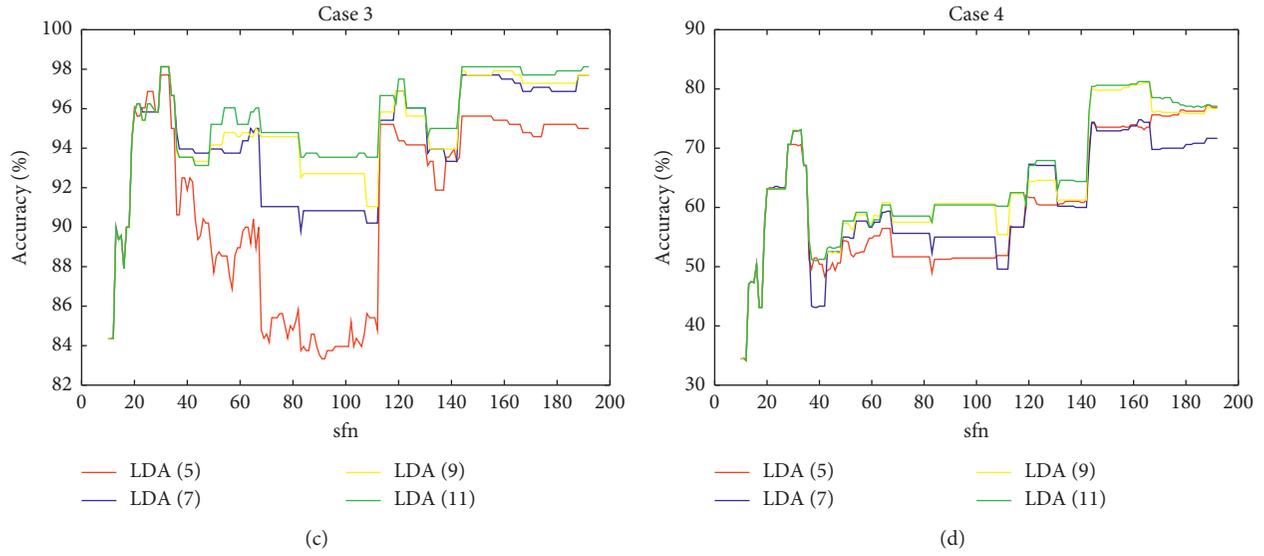


FIGURE 11: The curve representation of diagnosis results obtained by the OFS-TSFRS-LDA-SVM model using WPT with different dimension sizes for LDA. The LDA (5) represents that the number of dimension size is 5.

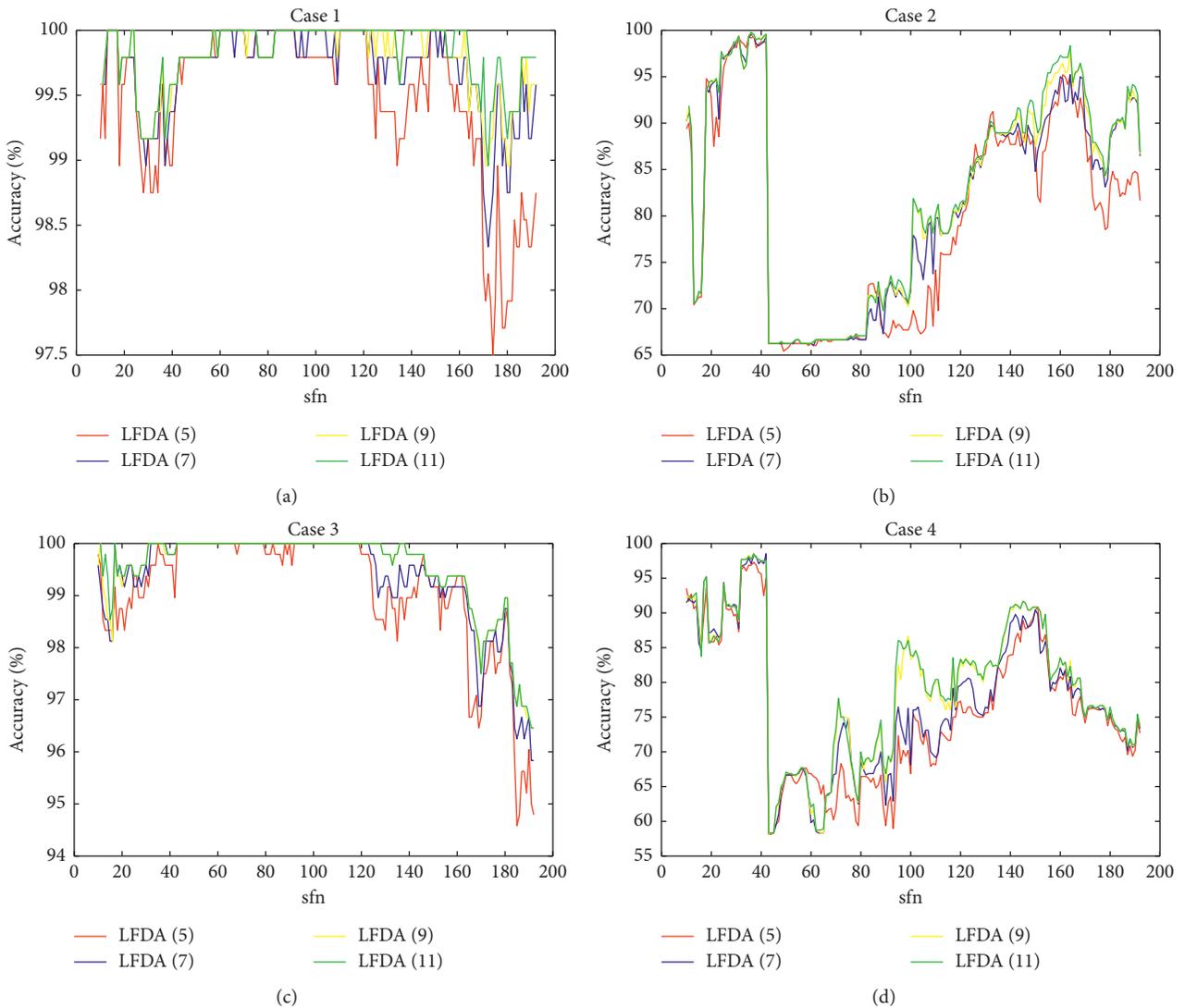


FIGURE 12: The curve representation of diagnosis results obtained by OFS-TSFRS-LFDA-SVM using DTCWPT with different dimension sizes for LFDA. The LFDA (5) represents that the number of dimension size is 5.

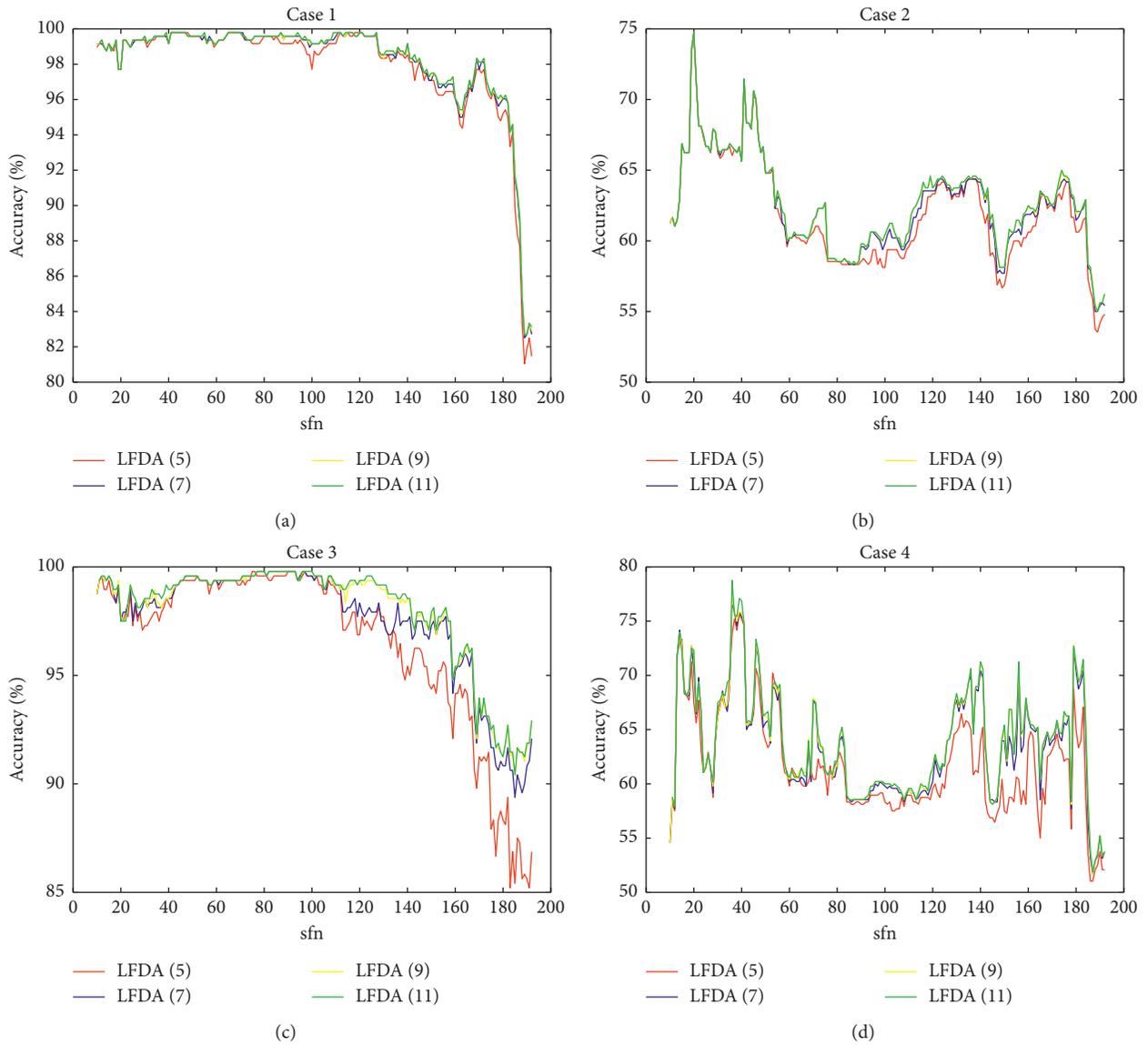


FIGURE 13: The curve representation of diagnosis results obtained by OFS-TSFRS-LFDA-SVM using WPT with different dimension sizes for LFDA. The LFDA (5) represents that the number of dimension size is 5.

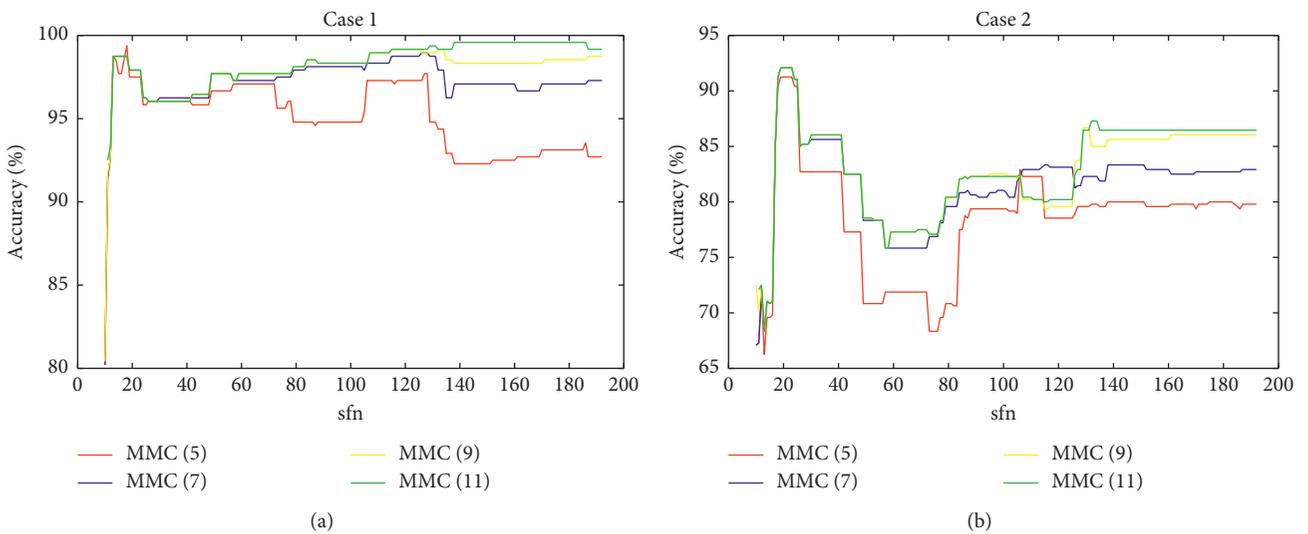


FIGURE 14: Continued.

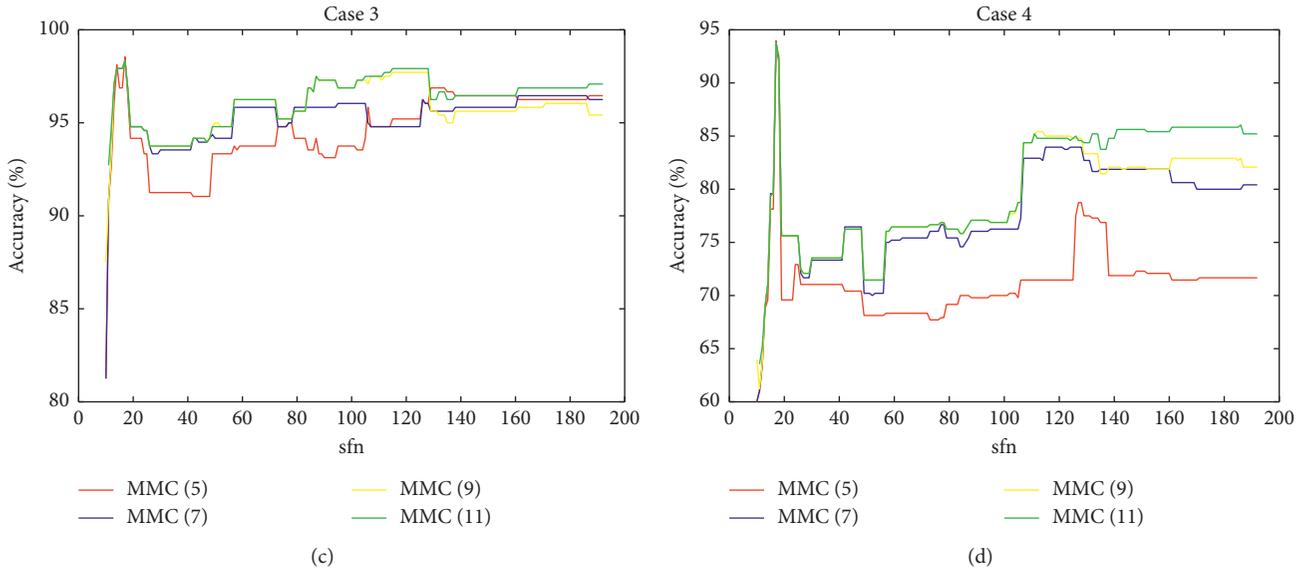


FIGURE 14: The curve representation of diagnosis results obtained by OFS-TSFRS-MMC-SVM using DTCWPT with different dimension sizes for MMC. The MMC (5) represents that the number of dimension size is 5.

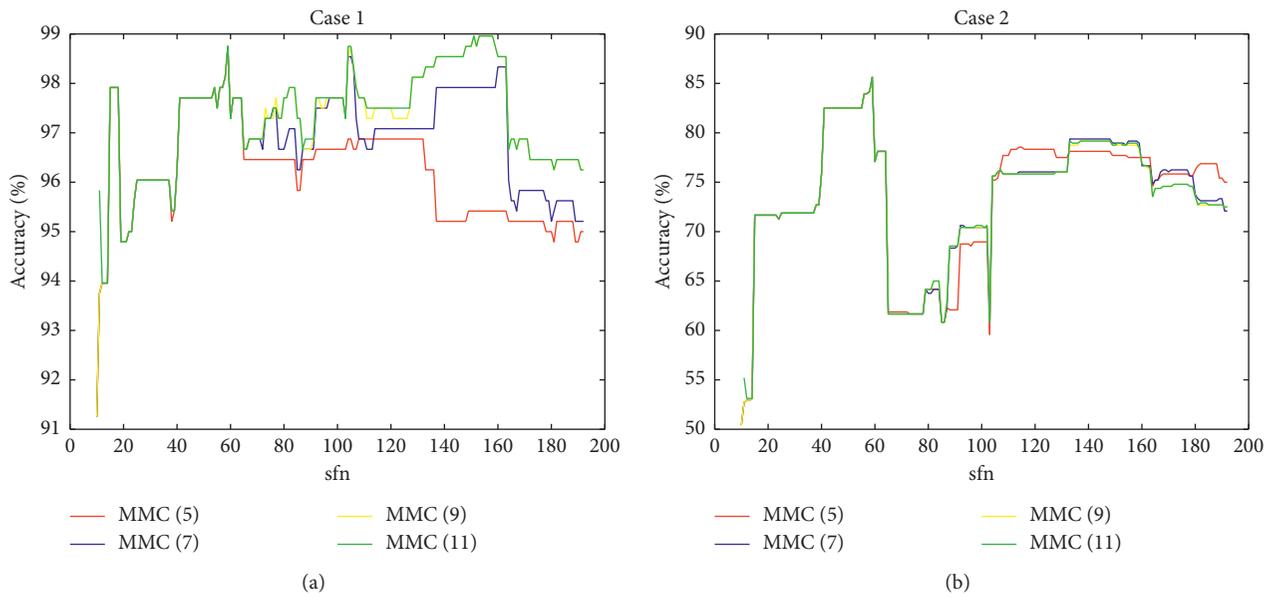


FIGURE 15: Continued.

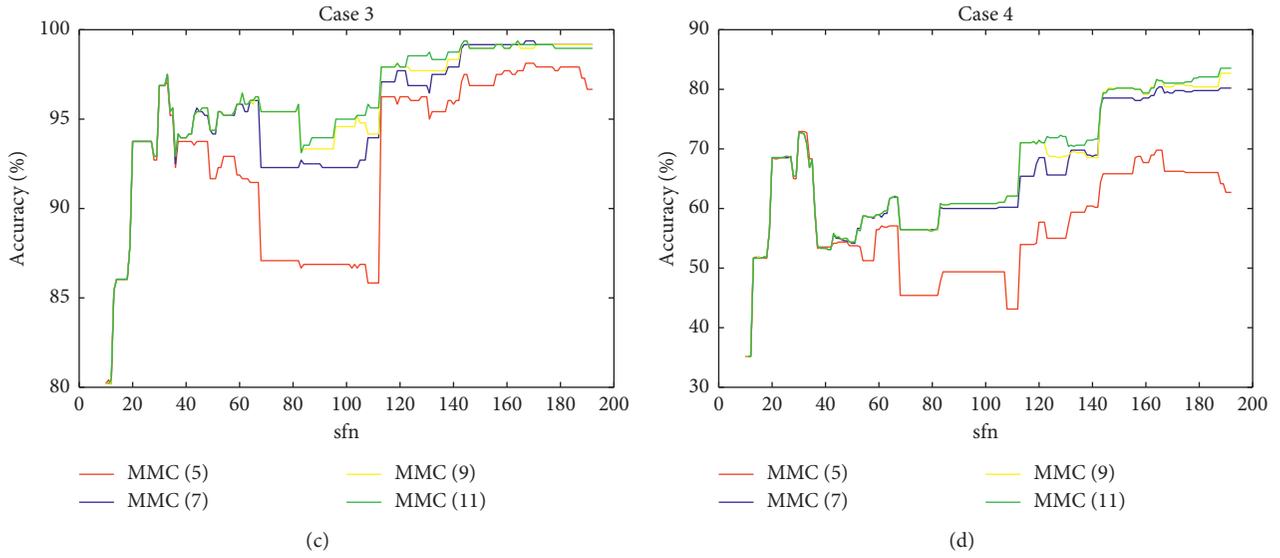


FIGURE 15: The curve representation of diagnosis results obtained by OFS-TSFRS-MMC-SVM using WPT with different number of dimension sizes for MMC. The MMC (5) represents that the number of dimension size is 5.

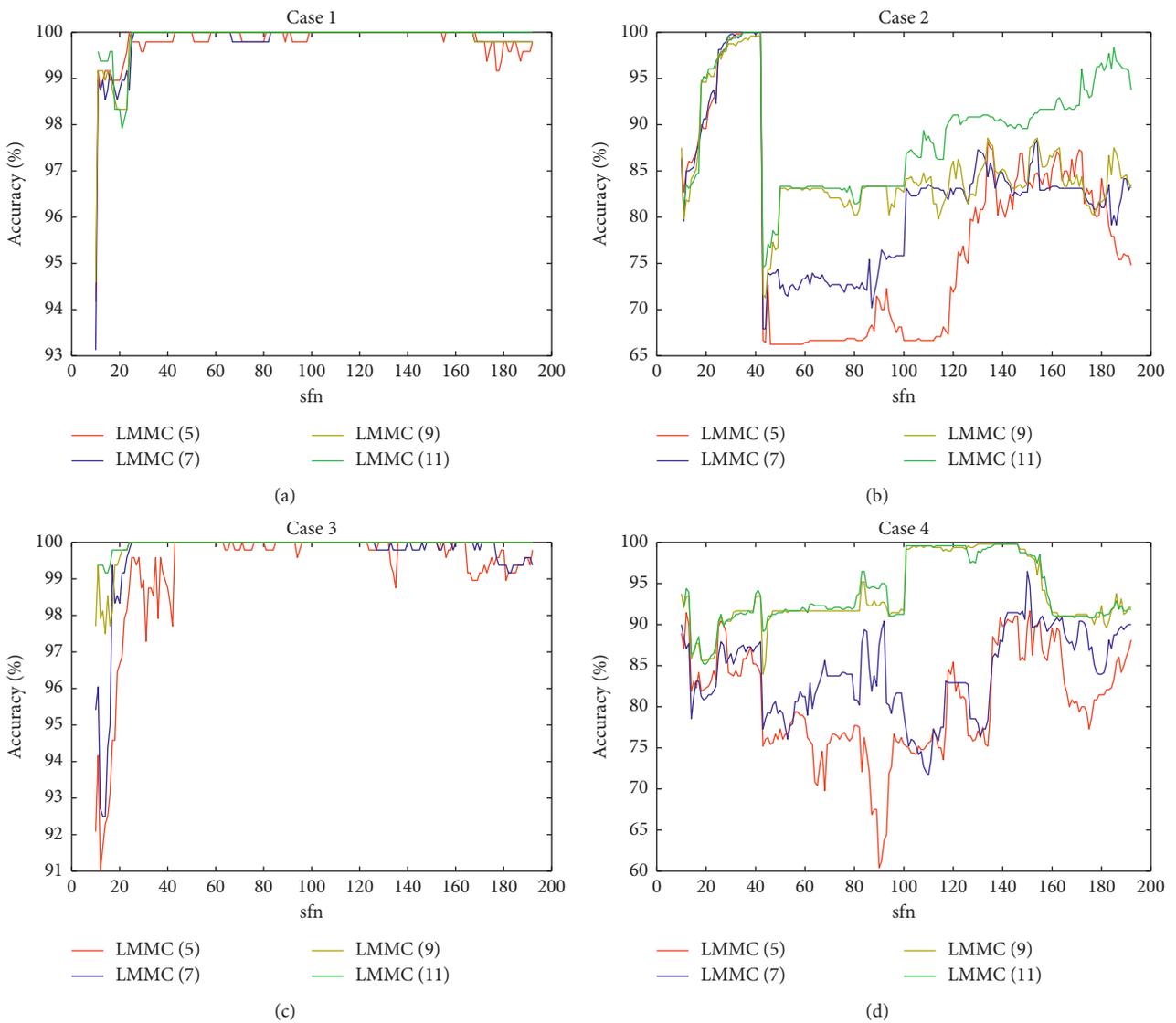


FIGURE 16: The curve representation of diagnosis results obtained by OFS-TSFRS-LMMC-SVM using DTCWPT with different dimension sizes for LMMC. The LMMC (5) represents that the number of dimension size is 5.

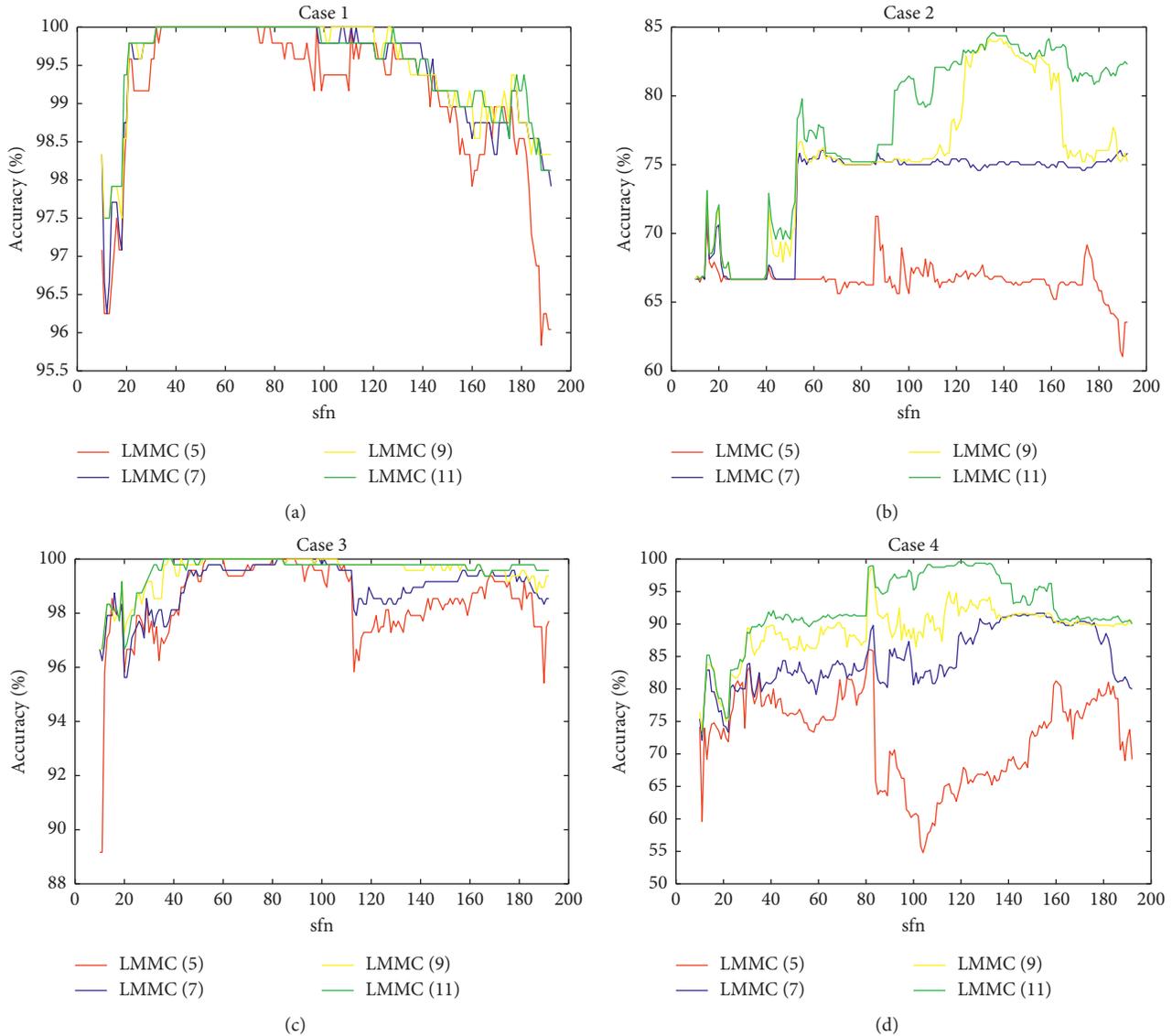


FIGURE 17: The curve representation of diagnosis results obtained by OFS-TSFRS-LMMC-SVM using WPT with different number of dimension sizes for LMMC. The LMMC (5) represents that the number of dimension size is 5.

demonstrated, and a diagnosis model trained by the proposed diagnosis framework using the training data of a single working condition can achieve a desirable accuracy for a testing set collected from bearings under different working conditions.

4.2. Experiments Based on the Experimental Test Rig 2

4.2.1. Experimental Setup and Cases. In order to further validate the effectiveness and adaptability of the proposed methods, the SQI-MFS test rig [42] is used to collect bearing vibration signals for fault diagnosis. The experimental test rig is shown in Figure 20, and the corresponding bearings with different fault types are shown in Figure 21. The laser machining is employed to set single-point defects with the different fault diameters, which contains 0.05 mm, 0.1 mm,

and 0.2 mm. The collected vibration signals of the bearing consist of the inner race fault signals, ball fault signals, outer race fault signals, and normal signals. The test rig supports the motor speeds of 1200 rpm and 1800 rpm. Two accelerometers are used to collect vibration signals. The sampling frequency is 16 kHz.

For the experimental data of the SQI-MFS test rig, the signal samples of 1200 rpm and 1800 rpm are applied for experiments, and there are four bearing conditions (normal, ball fault, inner race fault, and outer race fault). Ball fault, inner race fault, and outer race fault have three fault diameters (0.05 mm, 0.1 mm, and 0.2 mm), respectively. Therefore, there are 10 bearing conditions which can correspond to 10 patterns for fault diagnosis. The bearing vibration signals are divided into several data segments, and each segment which is used as a sample has 5000 data points. Each bearing condition contains 60 samples, among which

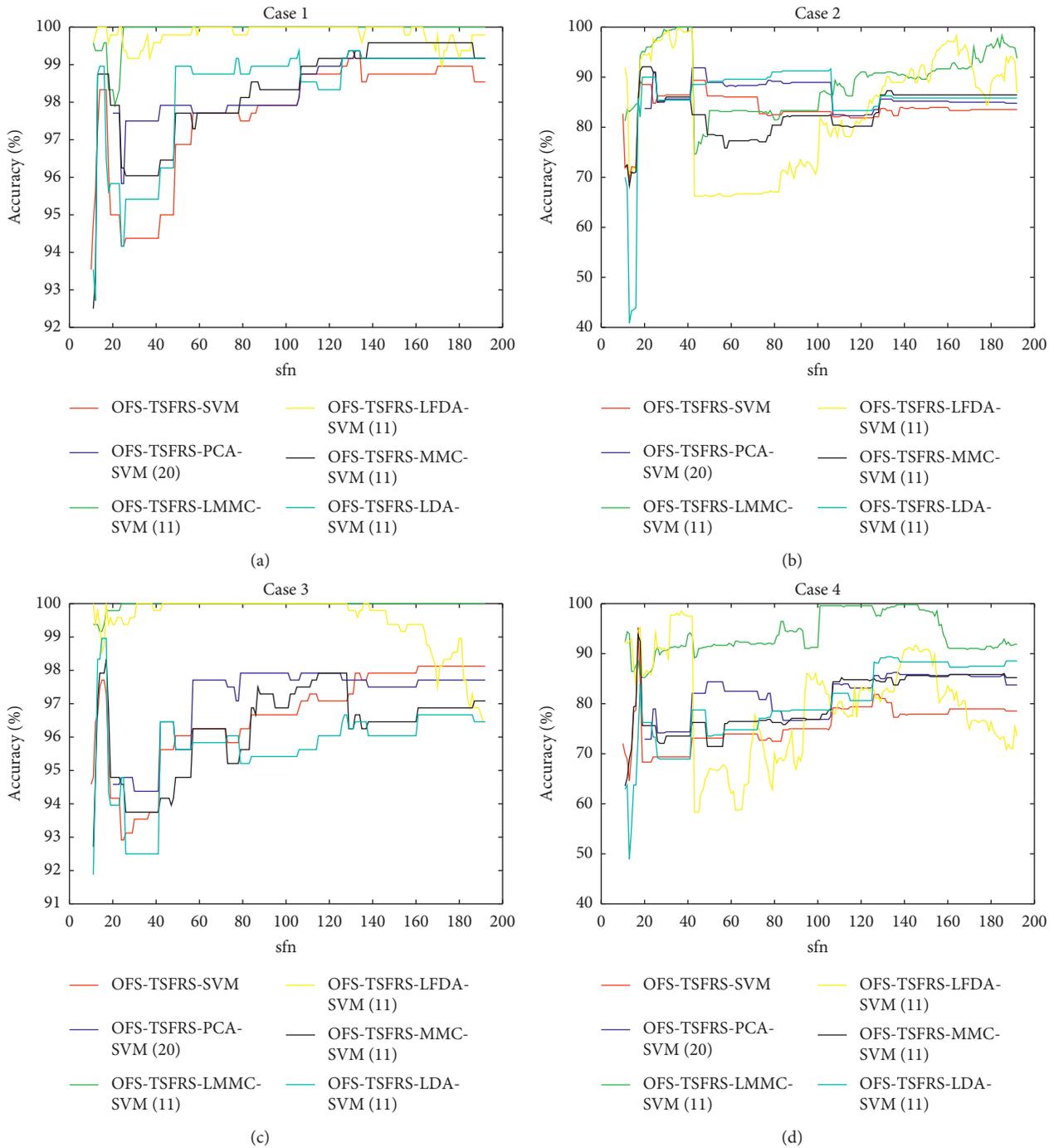


FIGURE 18: The curve representation of diagnosis results of models using DTCWPT for the testing sets of two cases with the use of TSFRS and different dimensionality reduction methods. The output dimension sizes of PCA, LDA, LFDA, MMC, and LMMC are 20, 11, 11, 11, and 11, respectively.

40 random samples are selected as the testing samples and 20 random samples are selected as the training samples. Based on these samples, two group datasets are used in experiments. The first group dataset includes cases 1 and 2. The samples of 1800 rpm are used as training samples. In case 1, the samples of 1800 rpm are used as testing samples. In case 2, the samples of 1200 rpm are used as testing samples. The second group dataset includes cases 3 and 4. The samples of

1200 rpm are used as training samples. In case 3, the samples of 1200 rpm are used as testing samples. In case 4, the samples of 1800 rpm are used as testing samples. The detailed information of the two group experimental datasets is shown in Tables 17 and 18, respectively.

4.2.2. *Analysis Results.* The experimental procedure is the same as that of the test rig 1, and each sample is decomposed

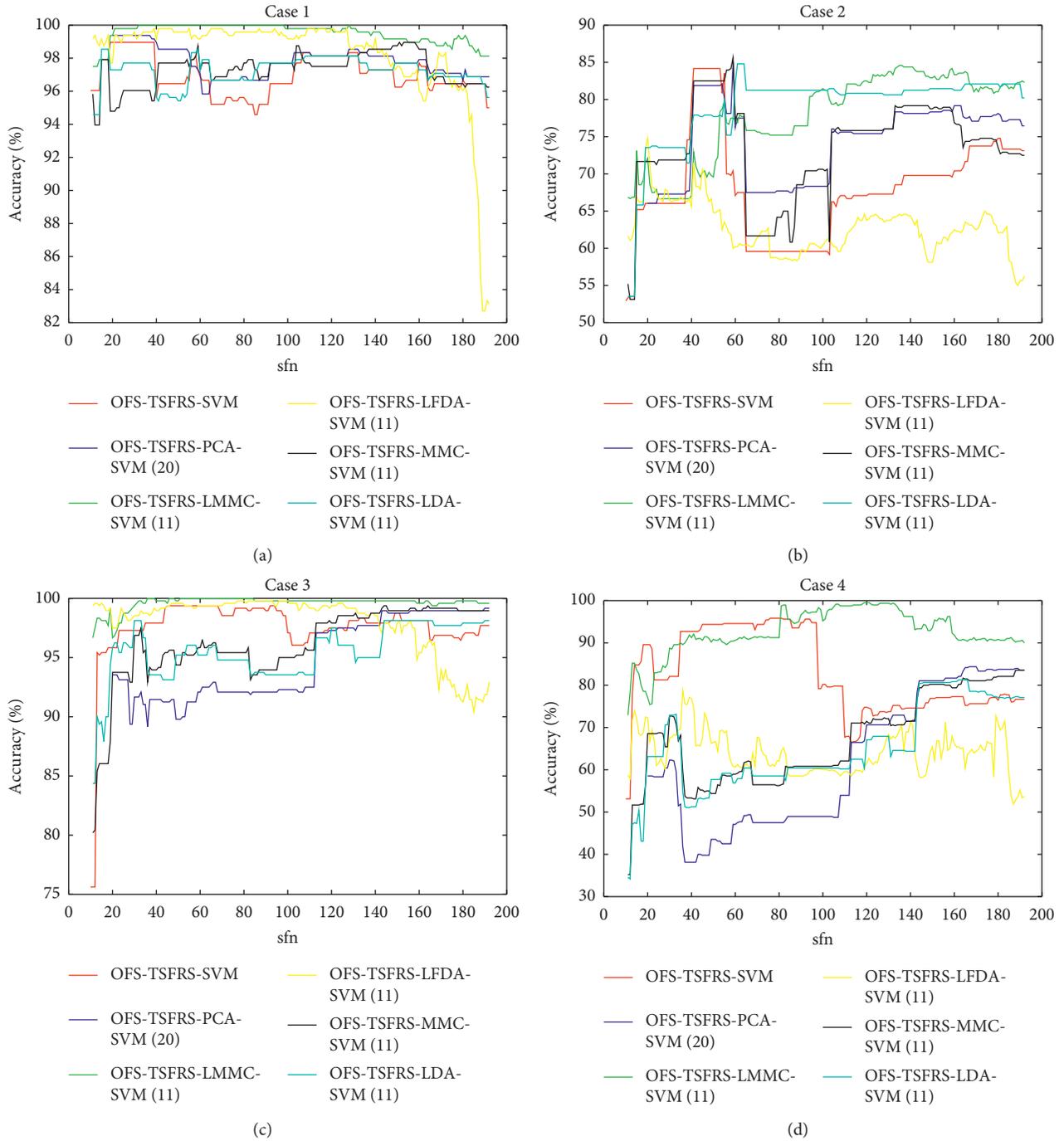


FIGURE 19: The curve representation of diagnosis results of models using WPT for the testing sets of two cases with the use of TSFRS and different dimensionality reduction methods. The output dimension sizes of PCA, LDA, LFDA, MMC, and LMMC are 20, 11, 11, 11, and 11, respectively.

into different wavelet packet nodes by DTCWPT; the decomposition level is 4. Thus, 16 terminal nodes, namely, subband signals, can be obtained. Then, 16 single-branch reconstruction signals of terminal nodes can be obtained, and 16 single-branch reconstruction signals are selected to generate 16 HES. By using 6 statistical parameters shown in Table 4, each single-branch reconstruction signal can generate 6 statistical features by calculating 6 statistical

parameters, and each HES can generate 6 statistical features by calculating 6 statistical parameters. Thus, 16 single-branch reconstruction signals and 16 HES can generate 192 statistical features which compose the OFS. Then, the TSFRS is performed to select the sensitive statistical features and reduce the difference of distribution between the training sensitive feature subset and the testing sensitive feature subset. The WV, SSMD, and WSD of the 192 statistical



FIGURE 20: Experimental test rig 2.

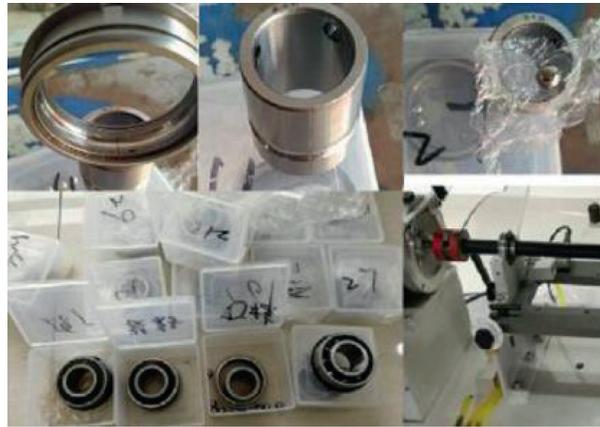


FIGURE 21: SER205 bearings.

TABLE 17: The detailed information of the first group dataset.

Bearing condition	Defect diameter (mm)	Number of training samples		Number of testing samples		Class label
		1800 tr/min		1800 tr/min (case 1)	1200 tr/min (case 2)	
Normal	0	20	20	40	40	1
	0.05	20	20	40	40	2
Ball fault	0.1	20	20	40	40	3
	0.2	20	20	40	40	4
Inner race fault	0.05	20	20	40	40	5
	0.1	20	20	40	40	6
	0.2	20	20	40	40	7
Outer race fault	0.05	20	20	40	40	8
	0.1	20	20	40	40	9
	0.2	20	20	40	40	10
Number of samples		240	240	480	480	

features of the training samples are presented in Figures 22–24, respectively.

In order to verify the adaptability and effectiveness of the fault diagnosis model using proposed methods, OFS-LMMC-SVM and OFS-TSFRS-LMMC-SVM are employed for experiments. The experimental results of OFS-LMMC-SVM are presented in Table 19. For the testing set of cases 1

and 3, the maximum accuracy can attain 99.67% and 99.17%, respectively. When the testing set is from cases 2 and 4, the maximum accuracy can attain 67.67% and 59.50%. When the TSFRS is applied, according to the experimental results of OFS-TSFRS-LMMC-SVM shown in Table 20, it is evident that the diagnosis accuracies appear with an enhancement by the use of TSFRS. OFS-TSFRS-LMMC-SVM can achieve

TABLE 18: The detailed information of the second group dataset.

Bearing condition	Defect diameter (mm)	Number of training samples		Number of testing samples		Class label
		1200 tr/min	1800 tr/min	1200 tr/min (case 3)	1800 tr/min (case 4)	
Normal	0	20	20	40	40	1
Ball fault	0.05	20	20	40	40	2
	0.1	20	20	40	40	3
	0.2	20	20	40	40	4
Inner race fault	0.05	20	20	40	40	5
	0.1	20	20	40	40	6
	0.2	20	20	40	40	7
Outer race fault	0.05	20	20	40	40	8
	0.1	20	20	40	40	9
	0.2	20	20	40	40	10
Number of samples		240	240	480	480	

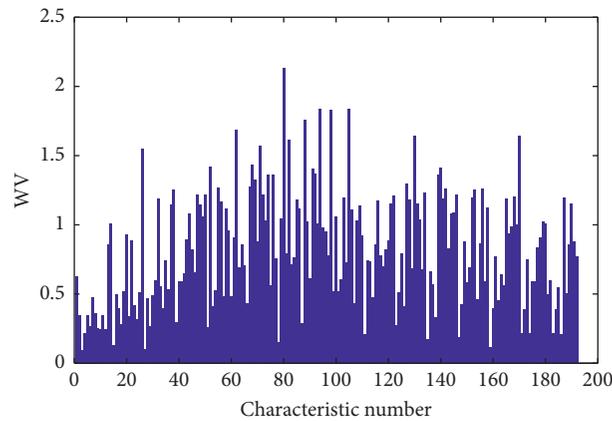


FIGURE 22: The WV of 192 statistical features of training samples.

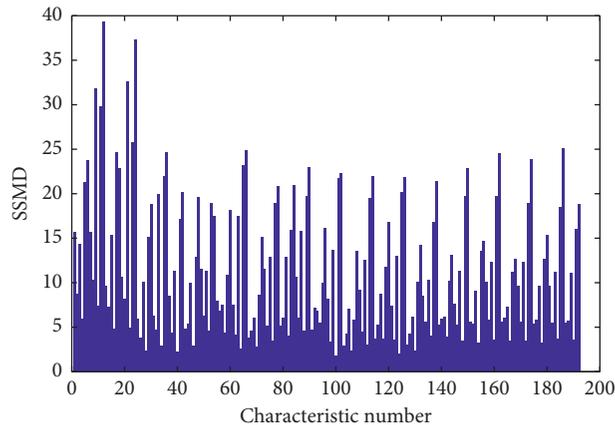


FIGURE 23: The SSMD of 192 statistical features of training samples.

desirable performance. For the testing sets of cases 1, 2, 3, and 4, the maximum accuracies can attain 100.00%, 89.00%, 100%, and 84.33%, respectively. The curve representation of the experimental results of OFS-TSFRS-LMMC-SVM is shown in Figure 25. For comparison, the diagnosis results of OFS-TSFRS-SVM, OFS-TSFRS-PCA-SVM, OFS-TSFRS-LDA-SVM, OFS-TSFRS-LFDA-SVM, and OFS-TSFRS-MMC-SVM are shown in Figure 26. According to the

experimental results, especially for the testing set of cases 2 and 4, it is evident that the performance of OFS-TSFRS-LMMC-SVM is better than that of other models when a suitable sfn is selected. In summary, the effectiveness of the proposed TSFRS and LMMC can be further demonstrated, and the adaptability of the proposed diagnosis framework using four testing sets collected from different bearings under different working conditions is also verified.

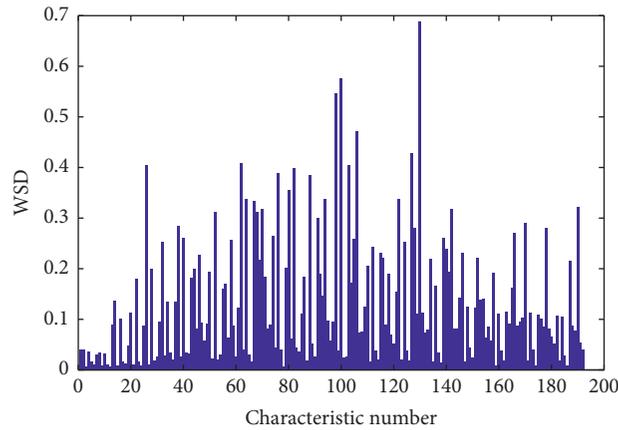


FIGURE 24: The WSD of 192 statistical features of training samples.

TABLE 19: Bearing fault diagnosis results obtained by the OFS-LMMC-SVM.

Dimension size	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
5	91.83	65.50	90.33	58.50
7	99.17	67.33	92.50	59.00
9	99.67	67.67	99.00	59.50
11	99.50	67.67	99.17	59.50

TABLE 20: Bearing fault diagnosis results obtained by the OFS-TSFRS-LMMC-SVM.

sfn	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 3 testing accuracy (%)	Case 4 testing accuracy (%)
20	73.50	61.67	64.00	70.00
30	77.17	82.83	64.17	84.33
40	81.33	82.83	77.83	80.33
50	86.17	89.00	76.83	80.67
70	99.67	88.33	100.00	72.00
100	100.00	82.17	100.00	77.50
120	99.67	82.17	100.00	77.83
140	99.83	77.83	100.00	73.00
160	99.83	66.50	99.83	66.00
180	99.33	68.33	99.83	60.33

Dimension size is 11.

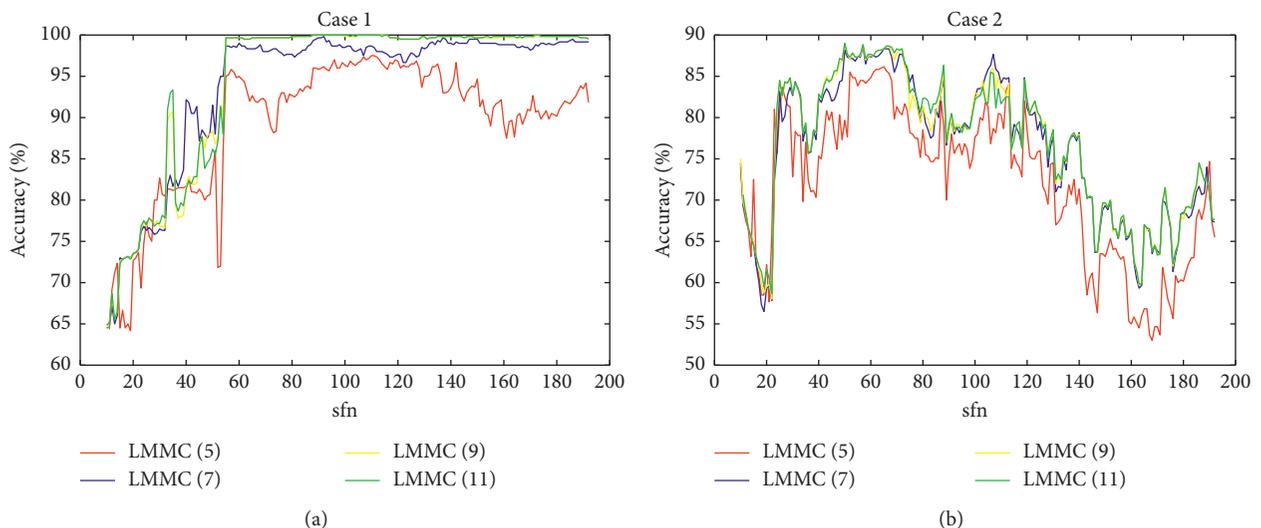


FIGURE 25: Continued.

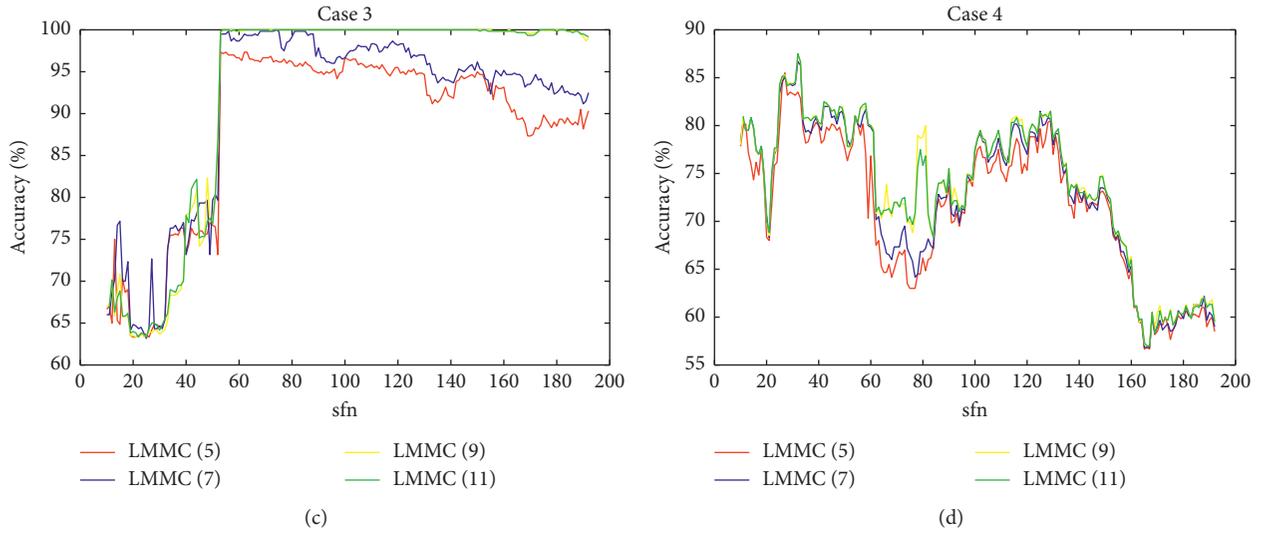


FIGURE 25: The diagnosis results obtained by OFS-TSFRS-LMMC-SVM with different number of dimension sizes for LMMC. The LMMC (5) represents that the number of dimension size is 5.

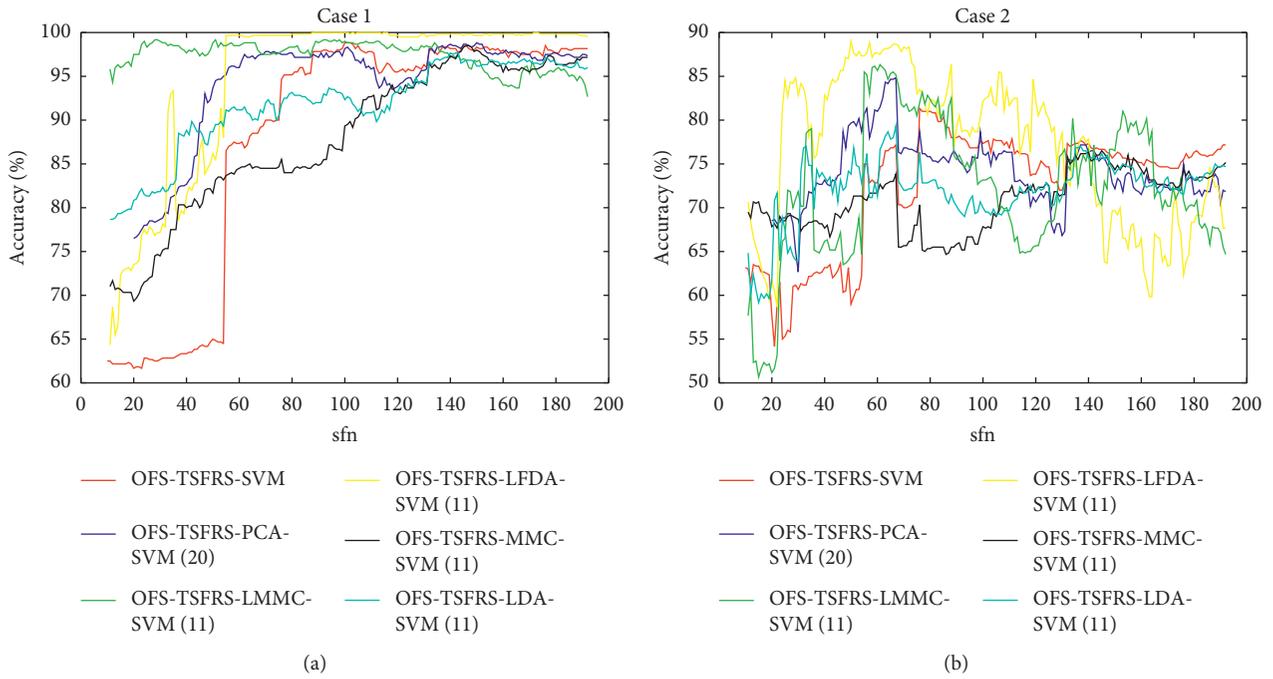


FIGURE 26: Continued.

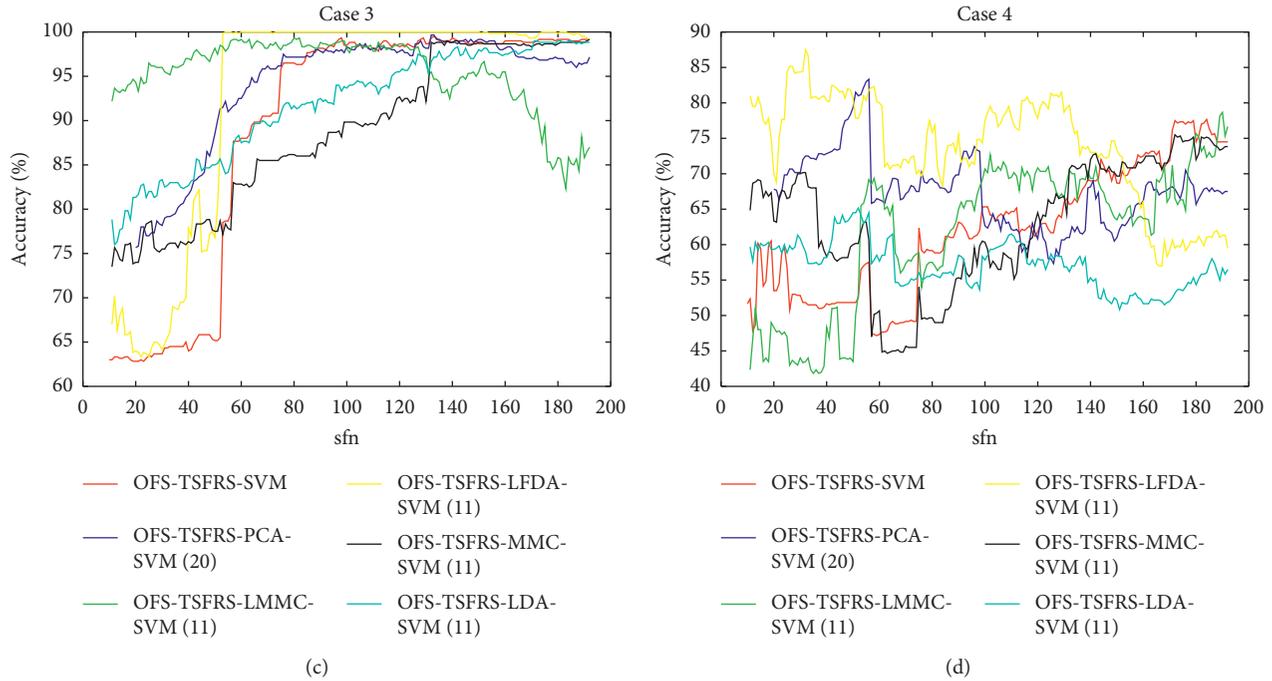


FIGURE 26: The diagnosis results of testing set of four cases with the use of TSFRS and different dimensionality reduction methods. The output dimension sizes of PCA, LDA, LFDA, MMC, and LMMC are 20, 11, 11, 11, and 11, respectively.

5. Conclusions

Due to the harsh environment and variability of the working conditions in real industrial scenarios, data-driven fault diagnosis using traditional machine-learning methods has limited the performance of the model under different working conditions, in which the distribution between the training data and testing data is different. To address this problem, this paper proposed a novel intelligent fault diagnosis framework for REBs under different conditions, with systematically blending statistical analysis with artificial intelligence. In this framework, DTCWPT is used to process raw vibration signals and extract statistical features. A new feature extraction method, TSFRS, is used to select the most sensitive features to form a sensitive feature subset and reduce the difference of distribution between the training sensitive feature subset and the testing sensitive feature subset. A modified MMC, namely, the LMMC, is used as a feature dimensionality reduction method. Compared with the other dimensionality reduction methods, the advantages of the proposed LMMC is presented. SVM is used as an automated fault pattern recognition classifier. Finally, experimental datasets collected from two experimental test rigs contain samples of different bearing fault conditions such as ball fault, inner race fault, and outer race fault at different defect diameters.

According to the experimental results, the proposed methods for REB fault diagnosis have great potential to be beneficial in industrial applications. For the experimental test rig 1, a set of comparative cases, namely, cases 1, 2, 3, and 4, are employed for experiments. The cases 1 and 2 select samples of the same motor loads of 2 hp as the training sets,

and the samples of different motor loads of 2 hp and 3 hp are selected as the testing sets, respectively. The cases 3 and 4 select samples of the same motor loads of 3 hp as the training sets, and the samples of different motor loads of 3 hp and 2 hp are selected as the testing sets, respectively. The experimental results indicate that the fault diagnosis model using the proposed methods can achieve desirable performance, when a suitable parameter sfn is selected, and the maximum diagnosis accuracies of cases 1, 2, 3, and 4 can attain 100%, 100%, 100%, and 99.79%, respectively. The experimental test rig 2 is employed to further verify the adaptability and effectiveness of the diagnosis model using the proposed methods. The experimental results show that the proposed methods can help the diagnosis model to achieve preferable diagnosis accuracy, and at the same time, the desirable adaptability of the diagnosis model using the proposed methods is demonstrated.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This work was funded by the Special Funds Project for Transforming Scientific and Technological Achievements in

Jiangsu Province (BA2016017); the National Key R&D Program of China (2017YFC0804400 and 2017YFC0804401).

References

- [1] X. Yan, Y. Liu, and M. Jia, "Research on an enhanced scale morphological-hat product filtering in incipient fault detection of rolling element bearings," *Measurement*, vol. 147, Article ID 106856, 2019.
- [2] Y. Qin, "A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2716–2726, 2017.
- [3] Z. Huo, Y. Zhang, P. Francq, L. Shu, and J. Huang, "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 5, pp. 19442–19456, 2017.
- [4] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol. 157, pp. 180–197, 2019.
- [5] Z. H. An, S. M. Li, Y. Xin, K. Xu, and H. J. Ma, "An intelligent fault diagnosis framework dealing with arbitrary length inputs under different working conditions," *Measurement Science and Technology*, vol. 30, no. 12, Article ID 125107, 2019.
- [6] H. Zheng, R. Wang, Y. Yang et al., "Cross-domain fault diagnosis using knowledge transfer strategy: a review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019.
- [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] P. Ma, H. Zhang, W. Fan, and C. Wang, "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Transactions*, vol. 99, pp. 465–478, 2020.
- [10] 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.
- [11] 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.
- [12] Z. K. Peng, G. Meng, F. L. Chu, Z. Q. Lang, W. M. Zhang, and Y. Yang, "Polynomial chirplet transform with application to instantaneous frequency estimation," *Ieee Transactions on Instrumentation and Measurement*, vol. 60, no. 9, pp. 3222–3229, 2011.
- [13] Y. Yang, Z. Peng, W. Zhang, and G. Meng, "Parameterised time-frequency analysis methods and their engineering applications: a review of recent advances," *Mechanical Systems and Signal Processing*, vol. 119, pp. 182–221, 2019.
- [14] L. Wang and J. Xiang, "A two-stage method using spline-kernelled chirplet transform and angle synchronous averaging to detect faults at variable speed," *IEEE Access*, vol. 7, pp. 22471–22485, 2019.
- [15] L. Wang, J. Xiang, and Y. Liu, "A time-frequency-based maximum correlated kurtosis deconvolution approach for detecting bearing faults under variable speed conditions," *Measurement Science and Technology*, vol. 30, no. 12, Article ID 125005, 2019.
- [16] P. Shi, S. An, P. Li, and D. Han, "Signal feature extraction based on cascaded multi-stable stochastic resonance denoising and EMD method," *Measurement*, vol. 90, pp. 318–328, 2016.
- [17] 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, no. 5, pp. 350–366, 2015.
- [18] 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.
- [19] 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.
- [20] K. Zhang, T. Lin, and X. Jin, "Low speed bearing fault diagnosis based on EMD-CIIT histogram entropy and KFCM clustering," *Journal of Shanghai Jiaotong University (Science)*, vol. 24, no. 5, pp. 616–621, 2019.
- [21] K. P. Shankar, K. L. Annamalai, and L. S. Kumar, "Selecting effective intrinsic mode functions of empirical mode decomposition and variational mode decomposition using dynamic time warping algorithm for rolling element bearing fault diagnosis," *Transactions of the Institute of Measurement and Control*, vol. 41, no. 7, pp. 1923–1932, 2018.
- [22] N. Tsakalozos, S. K. Drakakis, and S. Rickard, "A formal study of the nonlinearity and consistency of the empirical mode decomposition," *Signal Processing*, vol. 92, no. 9, pp. 1961–1969, 2012.
- [23] 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.
- [24] 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.
- [25] J. Burriel-Valencia, R. Puche-Panadero, J. Martinez-Roman, A. Sapeno-Bano, and M. Pineda-Sanchez, "Fault diagnosis of induction machines in a transient regime using current sensors with an optimized Slepian window," *Sensors*, vol. 18, no. 2, p. 146, 2018.
- [26] A. H. Boudinar, A. F. Aimer, M. E. Khodja, and N. Benouzza, "Induction motor's bearing fault diagnosis using an improved short time Fourier transform," in *Advanced Control Engineering Methods in Electrical Engineering Systems*, vol. 522, pp. 411–426, Springer, Berlin, Germany, 2019.
- [27] R. M. Ding, J. J. Shi, X. X. Jiang, and Z. K. Zhu, "Instantaneous frequency estimation via multiple ridge integration scheme for bearing fault diagnosis," in *Proceedings of the 2018 Prognostics and System Health Management Conference*, pp. 609–613, Chongqing, China, October 2018.
- [28] Y. Xin and S. M. Li, "Novel data-driven short-frequency mutual information entropy threshold filtering and its application to bearing fault diagnosis," *Measurement Science and Technology*, vol. 30, no. 11, Article ID 115006, 2019.
- [29] I. Attoui, N. Fergani, N. Boutasseta, B. Oudjani, and A. Deliou, "A new time-frequency method for identification

- and classification of ball bearing faults,” *Journal of Sound & Vibration*, vol. 397, pp. 241–265, 2017.
- [30] I. R. Quinde, J. C. Sumba, L. E. Ochoa, A. J. V. Guevara, and R. Morales-Menendez, “Bearing fault diagnosis based on optimal time-frequency representation method,” *IFAC-PapersOnLine*, vol. 52, no. 11, pp. 194–199, 2019.
- [31] 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, no. 54–55, pp. 259–276, 2015.
- [32] H. Li, Q. Zhang, X. Qin, and Y. Sun, “K-SVD-based WVD enhancement algorithm for planetary gearbox fault diagnosis under a CNN framework,” *Measurement Science and Technology*, vol. 31, no. 2, Article ID 025003, 2020.
- [33] H. 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, no. 8, pp. 437–445, 2014.
- [34] 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, 2015.
- [35] 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.
- [36] J. Qu, Z. Zhang, and T. Gong, “A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion,” *Neurocomputing*, vol. 171, no. C, pp. 837–853, 2016.
- [37] G. Tang, X. Wang, and Y. He, “A novel method of fault diagnosis for rolling bearing based on dual tree complex wavelet packet transform and improved multiscale permutation entropy,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 5432648, 13 pages, 2016.
- [38] N. G. Kingsbury, “The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters,” in *Proceedings of the 8th IEEE Digital Signal Processing Workshop*, Salt Lake City, UT, USA, 1998.
- [39] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.
- [40] Q. Niu, Q. Tong, J. Cao, F. Liu, and Y. Zhang, “Feature extraction method for condition monitoring of rolling element bearings based on dual-tree complex wavelet packet transform and VMD,” *Wireless Personal Communications*, vol. 103, no. 1, pp. 831–845, 2018.
- [41] H. Shao, H. Jiang, F. Wang, and Y. Wang, “Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet,” *ISA Transactions*, vol. 69, pp. 187–201, 2017.
- [42] D. Fei, Y. Xiao, D. Enjie, S. Wu, C. Fan, and Y. Huang, “Rolling bearing fault diagnosis using modified neighborhood preserving embedding and maximal overlap discrete wavelet packet transform with sensitive features selection,” *Shock and Vibration*, vol. 2018, Article ID 5063527, 29 pages, 2018.
- [43] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, “Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection,” *IEEE Access*, vol. 6, pp. 3715–3730, 2018.
- [44] Y. Li, X. Liang, J. Lin, Y. Chen, and J. Liu, “Train axle bearing fault detection using a feature selection scheme based multi-scale morphological filter,” *Mechanical Systems and Signal Processing*, vol. 101, pp. 435–448, 2018.
- [45] Y. Hu, S. Zhang, A. Jiang et al., “A new method of wind turbine bearing fault diagnosis based on multi-masking empirical mode decomposition and fuzzy C-means clustering,” *Chinese Journal of Mechanical Engineering*, vol. 32, no. 1, 2019.
- [46] Y. Gao, X. Liu, and J. Xiang, “FEM simulation-based generative adversarial networks to detect bearing faults,” *Ieee Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4961–4971, Jul 2020.
- [47] X. Y. Liu, H. Z. Huang, and J. W. Xiang, “A personalized diagnosis method to detect faults in a bearing based on acceleration sensors and an FEM simulation driving support vector machine,” *Sensors*, vol. 20, no. 2, p. 420, 2020.
- [48] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.
- [49] A. Khatami, M. Babaie, H. R. Tizhoosh, A. Khosravi, T. Nguyen, and S. Nahavandi, “A sequential search-space shrinking using CNN transfer learning and a radon projection pool for medical image retrieval,” *Expert Systems with Applications*, vol. 100, pp. 224–233, 2018.
- [50] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, “Wind power prediction using deep neural network based meta regression and transfer learning,” *Applied Soft Computing*, vol. 58, pp. 742–755, 2017.
- [51] S. Mun, M. Shin, S. Shon, W. Kim, D. K. Han, and H. Ko, “DNN transfer learning based non-linear feature extraction for acoustic event classification,” *IEICE Transactions on Information and Systems*, vol. E100.D, no. 9, pp. 2249–2252, 2017.
- [52] M. S. Long, J. M. Wang, G. G. Ding, J. G. Sun, and P. S. Yu, “Transfer feature learning with joint distribution adaptation,” in *Proceedings of the IEEE International Conference on Computer Vision 2013*, pp. 2200–2207, IEEE, Sydney, Australia, December 2013.
- [53] M. Ghifary, W. B. Kleijn, M. J. Zhang, D. Balduzzi, and W. Li, “Deep reconstruction classification networks for unsupervised domain adaptation,” in *Proceedings of the 14th European Conference on Computer Vision 2016*, pp. 597–613, Amsterdam, Netherlands, October 2016.
- [54] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan, and D. Krishnan, “Unsupervised pixel-level domain adaptation with generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017*, July 2017.
- [55] I. Attoui, B. Oudjani, N. Boutasseta, N. Fergani, M.-S. Bouakkaz, and A. Bouraiou, “Novel predictive features using a wrapper model for rolling bearing fault diagnosis based on vibration signal analysis,” *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 7–8, pp. 3409–3435, 2020.
- [56] 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.
- [57] Y. Bengio, P. Vincent, O. Delalleau et al., “Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering,” in *Proceedings of the International Conference on Neural Information Processing Systems*, Istanbul, Turkey, June 2003.
- [58] 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.

- [59] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 16, no. 1, pp. 186–197, 2004.
- [60] X. He, D. Cai, S. Yan et al., "Neighborhood preserving embedding," in *Proceedings of the 10th IEEE International Conference on Computer Vision 2005*, IEEE, Beijing, China, pp. 1208–1213, 2005.
- [61] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections," in *Proceedings of the IEEE International Conference on Data Mining 2005*, IEEE, New Orleans, LA, USA, November 2005.
- [62] J.-B. Yu, "Bearing performance degradation assessment using locality preserving projections," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7440–7450, 2011.
- [63] B. S. J. Costa, P. P. Angelov, and L. A. Guedes, *Fully Unsupervised Fault Detection and Identification Based on Recursive Density Estimation and Self-Evolving Cloud-Based Classifier*, Elsevier Science Publishers B. V, Amsterdam, Netherlands, 2015.
- [64] W. Guangbin, D. U. Moujun, H. Qingkai et al., "A bearing fault diagnosis method based on multi-scale sub-band sample entropy and LPP," *Journal of Vibration and Shock*, vol. 35, no. 20, pp. 71–76, 2016, in Chinese.
- [65] H. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 157–165, 2006.
- [66] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," 2007, <http://jmlr.org/>.
- [67] R. Yu and A. Baradarani, "Sampled-data design of FIR dual filter banks for dual-tree complex wavelet transforms via LMI optimization," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3369–3375, 2008.
- [68] C. Yao, Z. Lu, J. Li, Y. Xu, and J. Han, "A subset method for improving Linear Discriminant Analysis," *Neurocomputing*, vol. 138, no. 11, pp. 310–315, 2014.
- [69] A. K. Jain, P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [70] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [71] C. R. Rao, "The utilization of multiple measurements in problems of biological classification," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 10, no. 2, pp. 159–193, 1948.
- [72] F. Song, D. Zhang, D. Mei et al., "A multiple maximum scatter difference discriminant criterion for facial feature extraction," *IEEE Transaction on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 33, no. 6, pp. 1566–1599, 2007.
- [73] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [74] 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, 2017.
- [75] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.
- [76] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, Chicago, IL, USA, July 2008.
- [77] M. Robnik-Sikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1–2, pp. 23–69, 2003.
- [78] C.W.R.University, 2014, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.