

Research Article

Rolling Bearing Fault Diagnosis Based on Domain Adaptation and Preferred Feature Selection under Variable Working Conditions

Xiao Yu ^{1,2,3} Wei Chen ^{1,2} Chuanlong Wu ^{1,2} Enjie Ding ^{1,2} Yuanyuan Tian ^{1,2}
Haiwei Zuo ³ and Fei Dong ^{1,2}

¹IOT Perception Mine Research Center, China University of Mining and Technology, Xuzhou 221000, China

²School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221000, China

³School of Medicine Information and Engineering, Xuzhou Medical University, Xuzhou 221000, China

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

Received 20 August 2020; Revised 8 October 2020; Accepted 16 November 2020; Published 30 January 2021

Academic Editor: Changqing Shen

Copyright © 2021 Xiao Yu 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, with the use of conventional machine learning techniques, data-driven diagnosis models have a limitation that it is difficult to achieve the desirable fault diagnosis performance, and the reason is that the training and testing datasets are assumed to have the same feature distributions. To address this problem, a novel bearing fault diagnosis framework based on domain adaptation and preferred feature selection is proposed, in that the model trained by the labeled data collected from a working condition can be applied to diagnose a new but similar target data collected from other working conditions. In this framework, an improved domain adaptation method, transfer component analysis with preserving local manifold structure (TCAPLMS), is proposed to reduce the differences in the data distributions between different domain datasets and, at the same time, take the label information of feature dataset and the local manifold structure of feature data into consideration. Furthermore, preferred feature selection by fault sensitivity and feature correlation (PSFFC) is embedded into this framework for selecting features which are more beneficial to fault pattern recognition and reduce the redundancy of feature set. Finally, vibration datasets collected from two test platforms are used for experimental analysis. The experimental results validate that the proposed method can obviously improve diagnosis accuracy and has significant potential benefits towards actual industrial scenarios.

1. Introduction

In various parts of rotating machinery, the failure probability of rolling element bearings (REBs) is often higher than that of other components when the rotating machinery operate in harsh working environments [1, 2]. Therefore, bearing fault diagnosis features prominently in industrial applications, such as ensuring its reliability and reducing economic losses [3]. With the advent of the big data era, signal processing and data mining technologies are undergoing rapid development, and data-driven fault diagnosis methods are also developing rapidly [3–6]. However, towards its applicability in practical industrial applications, the conventional intelligent diagnosis method based on data-driven has two main disadvantages [4, 5, 7]. (1) Feature extraction and fault classification models abide by a

common premise that the training data and the test data have identical distributions. If this premise is not true, the generalization of these intelligent diagnosis methods will be greatly reduced. When the working conditions are not consistent in real industrial scenarios, the premise is not true. (2) Insufficient labeled target fault data are existed, due to that the working condition is variable and the types of failure of rotating machines are diverse. Therefore, those conventional data-driven intelligent diagnosis methods cannot establish an accurate fault diagnosis model of the target bearing in real diagnosis scenarios. In order to overcome the above limitations, it is essential to construct an advanced fault diagnosis model. This advanced model can perform accurate fault classification in a specific dataset and can show significant generalization ability in unlabeled data which is from other working conditions.

In the process of fault diagnosis, the crucial steps for fault pattern recognition are signal processing and feature extraction. In the current research on bearing fault diagnosis, the analysis signals mostly come from the vibration signal of REBs, and this signal is generally analyzed by time-frequency domain methods [5–9]. Yu et al. [10] decomposed raw signal with MODWPT and calculated the reconstructing signal to obtain the primitive statistical features representing the fault state in different frequency bands. For compound fault diagnosis of bearings, Shao et al. [11] proposed a new diagnosis method called adaptive DTCWPT based on the high-order spectrum. In order to decompose and reconstruct the vibration signal faster and more accurately, Zhang et al. [12] raised a lifting wavelet method, which is better than the traditional wavelet and combined with the morphological fractal dimension to recognize the rolling bearing's running status. By combining VMD time-frequency analysis and SVM, Zhou proposed a new method in [13] to detect the fault of rolling bearings, where the original signal is decomposed into intrinsic mode function (IMF) set through VMD. For the weak damage characteristic extraction, Xingxing et al. [8] proposed a new fault diagnosis method based on VMD, which is guided by ICF. For detection and isolation of multiple failures, a new fault detection technology is proposed for rolling bearings by Zhang in [14], which combined FAWT with time-frequency analysis. To identify the health state under varying status for the rolling bearings, Zhong et al. [15] proposed an intelligence method based on STFT and CNN. In order to combine the characteristics of WT to extract interference instantly and the adaptive ability of EMD to analyze time-varying or nonlinear signal, a new method is raised by Merainani in [16], which combined EWT with S transform (ST).

After the raw vibration signal is processed, statistical characteristics can be selected to show characteristic information, such as PV, RMS, V, Sw, K, energy, and energy entropy. In [9], the raw signal was decomposed by wavelet decomposition and the reconstructed signals were obtained, and then 192 statistical characteristics are obtained from the corresponding HHT envelope spectrum (HES) of reconstructed signals. In [17], raw signals were decomposed into many distinct IMFs through EMD, the first four IMFs were selected to obtain the HES and HHT marginal spectrum, and then they were used to calculate statistical features. Wang et al. [18] proposed a technology of the original signal preprocessing based on wavelet packet denoising (WPD) and random forest algorithm (RFs) and used the value of SNR and mean square error as the mother wavelets set of the signal preprocessing. Shuangli et al. [19] combined wavelet analysis and entropy theory to decompose and reconstruct the collected signals and then input it into GA optimized SVM (GA-SVM). Considering the modulation characteristics of the bearings fault signal and the disadvantages of selecting the resonant high-frequency band based on experience, a bearings fault diagnosis technology was proposed by Jing in [20], which combines EMD and spectral kurtosis. Considering the effectiveness of the extracted statistical features in the entire life cycle process, Xiaodong [21] proposed a feature extraction approach based on the optimal

selection of statistical indicators. Qing-feng et al. [22] used the variance contribution to eliminate false components in IMF and combined the wavelet packet decomposition to improve EMD. By using wavelet and FFT to decompose signals, Seryasat et al. [23] extracted the energy and RMS of signals in distinct frequency bands, which can recognize the bearings fault effectively.

The high-dimensional characteristic sets can usually be obtained after signal processing and feature extraction. Taking account into the compound mapping relationship between the bearing fault and its sign, it is hard to decide which statistical attribute is worth reflecting the nature of the fault from the high-dimensional characteristic space. The high-dimensional characteristic sets are prone to generate redundant characteristics and lead to reduced accuracy and efficiency of troubleshooting. The key step of the classifying process is selecting a subset of features. It is indicated in previous studies that feature selection seems to be an important premise to attain the accuracy for prospective diagnosis. Yu et al. [17] proposed a novel method of feature extraction for selecting the most sensitive characteristics, which combines the STD of characteristic data with K-means method. Fei et al. [9] raised a characteristics selection technology to determine the fault-sensitive features by combining ARI and sum of within-class MD. Considering that the stable distribution can extract features with high discriminating ability, Chouri et al. [24] combined alpha-stable distribution feature extraction with the weighted support vector machine (WSVM) to extract features efficiently. For the problem of nonsensitive characteristics in the features set, Liu et al. [25] raised a characteristic selection approach based on sensitive characteristic extraction and nonlinear feature fusion, which used CDET to choose sensitive characteristics and weighing them and then used locality preserving projections (LPP) to reduce the size of the weighted sensitive features for getting more sensitive characteristics. Liu et al. [26] used the K-SVD approach to train a sparse dictionary from sample data and apply it to sparsely decompose the feature vector for fault classification and identification. Considering that the actual working condition is complex and changeable, Chen et al. [27] raised a cross-domain features selection approach with TCA. Sun et al. [28] raised a feature extraction and diagnosis algorithm using CNN, which can automatically select a feature from time-domain vibration signals and discover distributed features of data effectively. By analyzing vibration and current signals, BR et al. [29] selected 12 statistical time-domain characteristics, among which SSC, VAR, and STD can be determined as good characteristics. Therefore, how to select the statistical features which are more beneficial to fault pattern recognition is also an important step. In this paper, a new characteristics selection approach, priority selection of features based on fault sensitivity and correlation between features (PSFFC), is proposed.

According to the above discussion, aiming at the two main limitations of conventional data-driven fault diagnosis technologies, some existing research recently has shown that transfer learning or domain adaptation methods [30] have vast application prospect and extensive applicability in every

field [31–33]. For overcoming the existing limitations, a novel idea was inspired by transfer learning or domain adaptation. For addressing the problem of insufficient fault information in significant fault samples, Chen et al. [34] presented an early fault diagnosis model, which combines DNN with transfer learning and can select the fault features of a great number of fault samples and unimportant fault features of other fault samples. With HKL and transfer learning, a robust fault diagnosis network applied to different working conditions was built by Qian in [35]. In order to overcome the problem that the training and test data have different distribution, and considering the valid and ordinary diagnostic knowledge obtained from multiple concerning source domains, Zheng et al. [3] raised a new intelligent fault recognition approach for multiple source domains. In order to overcome the problem that too few labeled samples are insufficient for accurate diagnosis, Zeng et al. [36] raised a fault diagnosis method with certain parameters unrelated to subsequent tasks for pretraining, and a great amount of unlabeled data was obtained. Zhou et al. [37] raised MSDCTL for fault diagnosis in new working condition, and it did not need new labeled data. To overcome the problem of unmarked cross-domain diagnosis of bearings, Shao et al. [38] raised a confrontation domain adaptive approach with deep transfer learning. The basic idea of the above studies is learning the fault detection knowledge from training data and transferring the knowledge to test data. It improves the capability to recognize the generalization of these models on the target dataset. Unlike the traditional machine learning techniques need to start from scratch to learn new tasks, transfer learning or domain adaptation learns the fault detection knowledge from the source domain and applies it to the target domain, which is more suitable for cross-domain learning applications [3, 6]. The transfer learning method based on characteristic is the main branch in it, and among these methods, TCA is a representative method [6]. For reducing the distance between the source domain and the target domain, features are mapped to a higher-dimensional reproducing kernel Hilbert space by means of nonlinear transformation [30]. Although it is very useful, TCA needs further optimization, which ignores the category information of samples and the local manifold structure of data. Therefore, based on TCA, transfer component analysis with preserving local manifold structure (TCAPLMS) is proposed in this paper, it reduces the distribution difference between the source domain and the target domain and improves discriminative of feature dataset.

Given the above discussion, a novel intelligent fault diagnosis method for bearings is proposed which is based on domain adaptation and preferred feature selection. The contributions of this paper are summarized as follows.

- (1) A new feature selection method, PSFFC, is raised to select features which are more beneficial to fault pattern recognition and reduce the redundancy of the feature set.
- (2) An improved feature-based transfer learning approach, TCAPLMS, is raised for domain adaptation.

TCAPLMS can reduce the differences in the marginal distributions between different domain datasets and, at the same time, take the label information of the feature dataset and the local manifold structure of feature data into consideration, and then, the domain adaptability and the discriminant performance are improved.

- (3) A new intelligent bearing fault diagnosis framework is proposed to suit variable working conditions, which address the limitations of fault diagnosis methods based on data-driven and enhance the generalization ability of fault diagnosis model in actual diagnosis scenario.

The rest of this paper is introduced as follows. Section 2 introduces the theoretical backgrounds of the MODWPT, LFDA, and TCA. Section 3 presents the fault diagnosis framework, the PSFFC and TCAPLMS. Section 4 is about the experimental analysis of the method, which is used to validate the effectiveness and adaptability. Finally, Section 5 is the conclusion. Furthermore, we present some acronyms in Table 1.

2. Theoretical Background

2.1. Maximal Overlap Discrete Wavelet Packet Transform (MODWPT). Discrete wavelet transform (DWT) is a typical time-frequency analysis method, but it also has limitations. Such as, in order to transform fully, the algorithm requires that the signal sequence length of the analyzed signal is an integer power of 2 [9]. To overcome this limitation, a highly redundant nonorthogonal wavelet transformation named MODWT is proposed, and it has no restriction on the signal sequence length. But MODWT is replaced by MODWPT, which has good frequency resolving power at high frequency and preserves the good performance of MODWT [39].

DWT can be explained as follows: $\{X_t, t = 0, K, N - 1\}$ is a real-valued time sequence and N is the signal sequence length. $\{g_l, l = 0, K, L - 1\}$ is even-length low-pass filter of DWT, which can output the low-frequency part of the input signal and filter out the high-frequency part, while the high-pass filter $\{h_l, l = 0, K, L - 1\}$ is the opposite, where L is the filter's length. For all nonzero integer n , the low-pass and high-pass filters can be simply written by the following equation [9, 39]:

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

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

Additionally, low-pass and high-pass filters are chosen to be quadrature mirror filters, so h_l and g_l are related to each other as follows:

$$h_l = (-1)^l g_{L-l-1} \text{ or } g_l = (-1)^{l+1} h_{L-l-1}. \quad (2)$$

With $V_{0,t} = X_t$, the j th level scale transformation coefficient is $\{V_{j-1,t}, t = 0, K, N_{j-1} - 1\}$, where $N_j = (N/2^j)$.

TABLE 1: A list of acronyms.

Acronym	Full form
TCAPLMS	Transfer component analysis with preserving local manifold structure
PSFFC	Preferred feature selection by fault sensitivity and feature correlation
REBs	Rolling element bearings
MODWPT	Maximum overlap discrete wavelet packet transform
SVD	Singular value decomposition
DTCWPT	Dual-tree complex wavelet packet transform
VMD	Variational mode decomposition
IMF	Intrinsic mode function
FAWT	Flexible analytical wavelet transform
TF	Time-frequency
STFT	Short-time Fourier transform
CNN	Convolutional neural network
EWT	Empirical wavelet transform
K	Kurtosis
PV	Peak value
V	Variance
RMS	Root mean square
LFDA	Local Fisher discriminant analysis
HES	HHT envelope spectrum
OFS	Original feature set
WPD	Wavelet packet denoising
CWRU	Case Western Reserve University
HES	Hilbert envelope spectra
RF	Random forest
SNR	Signal-to-noise ratio
EMD	Empirical mode decomposition
STD	Standard deviation
WSVM	Weighted support vector machine
CDET	Compensation distance evaluation technique
LPP	Locality preserving projections
TCA	Transfer component analysis
HKL	High-order Kullback–Leibler
MSDCTL	Multistage deep convolutional transfer learning
DWT	Discrete wavelet transform
MODWT	Maximal overlap discrete wavelet transform
LDA	Linear discriminant analysis
Sw	Skewness
SMD	Sum of within-class mean deviations
MMD	Maximum mean discrepancy
MD	Mean deviations
ARI	Adjusted rand index
PCC	Pearson correlation coefficient
FPSD	Feature priority selection degree
FSD	Fault sensitivity degree
RFS	Raw feature set

For the DWT pyramid algorithm, $V_{j-1,t}$ is the j th level input, so the j th level output is the j th level scaling transform coefficient and wavelet transform coefficient which are presented as follows:

$$\begin{aligned}
 V_{j,t} &= \sum_{l=0}^{l-1} g_l V_{j-1,(2t+1-l) \bmod N_{j-1}}, \quad (t = 0, K, N_j - 1), \\
 W_{j,t} &= \sum_{l=0}^{l-1} h_l V_{j-1,(2t+1-l) \bmod N_{j-1}}, \quad (t = 0, K, N_j - 1),
 \end{aligned} \tag{3}$$

where mod represents remainder after division.

DWT has the following limitations [9, 39, 40]: (1) when the length of the signal sequence to be analyzed is an integer power of 2, DWT could be fully performed; (2) when the signal is cyclically shifted, the wavelet coefficient and proportional coefficient of DWT cannot reach the identical cycle shift; (3) the wavelet coefficients and scale coefficients of DWT will be halved with the increase in the level of DWT series, which will affect the statistical analysis of the coefficients.

The sample size of the j -level DWT restricts to an integer power of 2, and MODWT does not require to consider sample size. Thus, MODWT is the optimized DWT. The filters of MODWT are redefined to conserve energy:

$$\begin{aligned}\tilde{g}_l &= \left(\frac{g_l}{\sqrt{2}} \right), \\ \tilde{h}_l &= \left(\frac{h_l}{\sqrt{2}} \right).\end{aligned}\quad (4)$$

At the same time, \tilde{g}_l and \tilde{h}_l satisfy the following equations now:

$$\begin{aligned}\sum_{l=0}^{L-1} \tilde{g}_l^2 &= \frac{1}{2}, \\ \sum_{l=0}^{L-1} \tilde{g}_l \tilde{g}_{l+2n} &= \sum_{l=-\infty}^{+\infty} \tilde{g}_l \tilde{g}_{l+2n} = 0,\end{aligned}\quad (5)$$

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

MODWT algorithm performs the weighted average of all observation starting points, which can reduce the deviation caused by cycle shift. To avoid halving the coefficient, MODWT rebuilds filters by inserting $2^{j-1} - 1$ zeros for each j th level between the elements of $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$:

$$\begin{aligned}\{\tilde{g}_0, 0, K, 0, \tilde{g}_1, 0, K, 0, K, \tilde{g}_{L-2}, 0, K, 0, \tilde{g}_{L-1}\}, \\ \{\tilde{h}_0, 0, K, 0, \tilde{h}_1, 0, K, 0, K, \tilde{h}_{L-2}, 0, K, 0, \tilde{h}_{L-1}\}.\end{aligned}\quad (6)$$

With $V_{0,t}^{(M)} = X_t$, the updated scaling coefficients $\{V_{j,t}^{(M)}\}$ and wavelet coefficients $\{W_{j,t}^{(M)}\}$ are presented as follows:

$$\begin{aligned}V_{j,t} &= \sum_{l=0}^{l-1} \tilde{g}_l V_{j-1, (t-2^{j-1}l) \bmod N}, \quad (t = 0, K, N-1), \\ W_{j,t} &= \sum_{l=0}^{l-1} \tilde{h}_l V_{j-1, (t-2^{j-1}l) \bmod N}, \quad (t = 0, K, N-1).\end{aligned}\quad (7)$$

In order to decompose high-frequency signals, MODWPT can be used to process signals. The coefficient of MODWPT is defined as follows:

$$\begin{aligned}W_{j,n,t} &= \sum_{l=0}^{L-1} \tilde{r}_{n,l} W_{j-1, [n/2], (t-2^{j-1}l) \bmod N}, \\ r_{n,l} &= \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3, \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2, \end{cases}\end{aligned}\quad (8)$$

where n is the frequency band number.

2.2. Local Fisher Discriminant Analysis (LFDA). LFDA integrates the advantages of LDA and LPP [17, 41], that is, LFDA aims to obtain the best separability between classes in the space and at the same time retain the local structure within the class. On the basis of LDA, LFDA considers the proximity relationship between the sample data of the same

classes, so that the data after dimensionality reduction are more conducive to classification [41].

LDA can be explained as follows: let $x_i \in \mathbb{R}^d$ ($i = 1, 2, K, n$) be d -dimensional samples and $y_i \in \{1, 2, K, c\}$ be associated class labels, where n is the number of samples and c is the number of classes. Let n_ℓ be the number of samples in class ℓ . Let $S^{(w)}$ be the scatter matrix within the class and $S^{(b)}$ be the scatter matrix between classes [17, 41]:

$$\begin{aligned}S^{(w)} &= \sum_{\ell=1}^c \sum_{i: y_i=\ell} (x_i - \mu_\ell)(x_i - \mu_\ell)^T, \\ S^{(b)} &= \sum_{\ell=1}^c n_\ell (\mu_\ell - \mu)(\mu_\ell - \mu)^T,\end{aligned}\quad (9)$$

where μ_ℓ is the sample average in class ℓ and μ is the average of all samples.

Assuming $S^{(w)}$ has full rank and $T^T S^{(w)} T$ is invertible, the LDA transformation matrix T_{LDA} is defined as follows:

$$T_{\text{LDA}} = \arg \max_{T \in \mathbb{R}^{d \times r}} \left[\text{tr} \left((T^T S^{(w)} T)^{-1} T^T S^{(b)} T \right) \right]. \quad (10)$$

LDA can be explained as follows: A is an affinity matrix and $A_{i,j}$ is the appetency between x_i and x_j , where $A_{i,j} \in [0, 1]$. The value of $A_{i,j}$ depends on the proximity relationship between x_i and x_j in the feature space. $A_{i,j}$ is defined as follows [40]:

$$A_{i,j} = \exp \left(- \frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j} \right), \quad (11)$$

where σ_i is the local scaling of the data samples around x_i which is defined as follows:

$$\sigma_i = \|x_i - x_i^{(K)}\|, \quad (12)$$

where $x_i^{(K)}$ is the K -th nearest neighbor of x_i and K is generally 7.

The LPP conversion matrix T_{LPP} is represented as follows:

$$T_{\text{LPP}} = \arg \min_{T \in \mathbb{R}^{d \times r}} \left(\frac{1}{2} \sum_{i,j=1}^n A_{i,j} \|T^T x_i - T^T x_j\|^2 \right), \quad (13)$$

subject to

$$T^T X D X^T T = I_r, \quad (14)$$

where D is the n -dimensional diagonal matrix and satisfies the following equation:

$$D_{i,i} = \sum_{j=1}^n A_{i,j}. \quad (15)$$

The LFDA transformation matrix T_{LFDA} is defined as follows [41]:

$$T_{\text{LFDA}} = \left[\text{tr} \left(\left(T^T \tilde{S}^{(w)} T \right)^{-1} T^T \tilde{S}^{(b)} T \right) \right], \quad (16)$$

where $\tilde{S}^{(w)}$ and $\tilde{S}^{(b)}$ are the updated scatter matrices within the class and between classes, respectively:

$$\tilde{S}^{(w)} \equiv \frac{1}{2} \sum_{i,j=1}^n \tilde{W}_{i,j}^{(w)} (x_i - x_j)(x_i - x_j)^T, \quad (17)$$

$$\tilde{S}^{(b)} \equiv \frac{1}{2} \sum_{i,j=1}^n \tilde{W}_{i,j}^{(b)} (x_i - x_j)(x_i - x_j)^T, \quad (18)$$

where $\tilde{W}_{i,j}^{(w)}$ and $\tilde{W}_{i,j}^{(b)}$ represent the weight values for the sample pair in the identical class and distinct class, respectively, which can be defined as follows:

$$\tilde{W}_{i,j}^{(w)} \equiv \begin{cases} \frac{A_{i,j}}{n_\ell}, & \text{if } y_i = y_j = \ell, \\ 0, & \text{if } y_i \neq y_j, \end{cases} \quad (19)$$

$$\tilde{W}_{i,j}^{(b)} \equiv \begin{cases} A_{i,j} \left(\frac{1}{n} - \frac{1}{n_\ell} \right) & \text{if } y_i = y_j = \ell, \\ \frac{1}{n}, & \text{if } y_i \neq y_j. \end{cases} \quad (20)$$

It is necessary to weight the values of sample pair in the identical category based on the affinity $A_{i,j}$ so that the samples which are far apart in the identical category have less impact on $\tilde{S}^{(w)}$ and $\tilde{S}^{(b)}$.

2.3. Transfer Component Analysis (TCA). TCA [42] is a typical feature-based transfer learning method, which aims to reduce the difference between the marginal distributions of the different datasets. Given a domain D , it consists of a D -dimensional feature space X , whose marginal probability distribution is $P(X)$; $X = \{x_1, K, x_n\}$ is the training dataset. $T = \{Y, f(X)\}$ is the learning task, where $Y = \{y_1, K, y_n\}$ is the D -dimensional label space and $f(X) = Q(Y|X)$ is the predictive function which represents the conditional probability distribution given D_S (source domain), D_T (target domain), and corresponding learning tasks T_S and T_T . TCA aims to facilitate the predictive function $f_T(X)$ in D_T by learning the d data from D_S and D_T , but in real diagnostic procedure, $D_S \neq D_T$ or $T_S \neq T_T$. A nonlinear mapping function ϕ in a reproducing kernel Hilbert space H exists that $P_S(\phi(X_S)) \approx P_T(\phi(X_T))$ and $P_S(Y_S|\phi(X_S)) \approx P_T(Y_T|\phi(X_T))$. The optimization objective of TCA is that the variance of the feature data can be preserved in a latent space and the marginal distributions between the D_S and D_T datasets can be minimized as much as possible. In TCA, the empirical maximum mean discrepancy (MMD) represents the distance of two marginal distributions $P_S(X)$ and $P_T(X)$, and the definition of MMD is represented as follows [7, 42, 43]:

$$\text{MMD}(X_S, X_T) = \left\| \frac{1}{n_S} \sum_{x_i \in X_S} \phi(x_i) - \frac{1}{n_T} \sum_{x_j \in X_T} \phi(x_j) \right\|_H^2 = \text{tr}(\text{KL}), \quad (21)$$

where n_S and n_T are the sample number of D_S and D_T , respectively. K represents a kernel matrix, as follows:

$$K = \begin{bmatrix} K_{S,S} & K_{S,T} \\ K_{T,S} & K_{T,T} \end{bmatrix}, \quad (22)$$

where $K_{S,S}$, $K_{S,T}$, and $K_{T,T}$ are the kernel matrices in the D_S , cross-domain, and D_T , respectively. The expression of L is shown as follows:

$$L = \begin{cases} \frac{1}{n_S^2}, & (x_i, x_j \in X_S), \\ \frac{1}{n_T^2}, & (x_i, x_j \notin X_S), \\ \frac{1}{n_S n_T}, & (\text{others}). \end{cases} \quad (23)$$

Empirical kernel mapping can reduce the matrix dimension, and it can transform high-dimensional data to low-dimensional data by using matrix $\tilde{W} \in R^{(n_S+n_T)^*m}$ which is embedded into K . The resultant kernel matrix is as follows [7]:

$$\tilde{K} = (K K^{-1/2} \tilde{W}) (\tilde{W} K t K^{-1/2}) = K W W^T K, \quad (24)$$

where $W = K^{-1/2} \tilde{W}$, and equation (20) can be transformed as

$$\text{MMD}(X_S, X_T) = \text{tr}((K W W^T K) L) = \text{tr}(W^T K L K W). \quad (25)$$

Therefore, the kernel learning problem (the objective function of TCA) can be replaced as follows:

$$\begin{aligned} \min_W \text{tr}(W^T K L K W) + \mu \text{tr}(W^T W), \\ \text{s.t. } W^T K H K W = I, \end{aligned} \quad (26)$$

where $\text{tr}(W^T W)$ is used to control the complexity of W and at the same time to avoid the rank deficiency of the denominator. In equation (26), μ is a trade-off parameter. $I \in R^{m \times m}$ and H are identity matrices, respectively. Finally, based on the trace optimization problem, equation (26) can be efficiently solved.

3. Proposed Method and System Framework

3.1. Preferred Feature Selection by Fault Sensitivity and Feature Correlation (PSFFC). In order to select features which are more beneficial to fault pattern recognitions and reduce the redundancy of feature set, in this paper, we suggest that fault sensitivity of statistical feature and correlation between

features should be considered to select preferred features. Therefore, there are two aspects:

- (1) The K -means algorithm (KA) [44] and the SMD of feature data are applied to indicate the fault sensitivity. Each type of statistical feature is processed by KA, and KA can get an index, that is, adjusted rand index (ARI). For each kind of feature, the MD of data samples in each condition can be calculated, and the SMD in all bearing conditions can be further obtained. The ARI and SMD indicate the class discriminative degree and the cohesiveness of the feature data, respectively. The ratio of the ARI and SMD is used to evaluate the fault sensitivity of feature, and the higher the ratio, the greater the fault sensitivity will be.
- (2) PCC [45] is used to evaluate the correlation between features. The higher the PCC, the higher correlation between features will be.

On the basis of the above two aspects, a new feature evaluation index is proposed, feature priority selection degree (FPSD), that is used to select preferred features for fault pattern recognition. The introduction of PSFFC is summarized as follows.

Step 1: given a raw vibration signal dataset, there are M fault types, and each fault type has N vibration signal samples. K types of statistical features can be obtained by the vibration signals processing and original feature extraction. These features can constitute the raw statistical feature set $[RFS^1, RFS^2, K, RFS^K]$, where the expression of RFS^k is as follows:

$$RFS^k = \begin{bmatrix} FS_{11}^k & FS_{12}^k & \cdots & FS_{1N}^k \\ FS_{21}^k & FS_{22}^k & \cdots & FS_{2N}^k \\ \vdots & \vdots & \ddots & \vdots \\ FS_{M1}^k & FS_{M2}^k & \cdots & FS_{MN}^k \end{bmatrix}, \quad (27)$$

where FS_{ij}^k is the k -th feature of the j -th sample in the i -th fault type. Then, by using KA, the ARI of the clustering partitions can be indicated by the cohesiveness of the feature data [46, 47]. The definition of ARI is described as follows:

Given a set of n objects $X = \{x_1, x_2, K, x_n\}$, $U = \{u_1, u_2, K, u_R\}$, and $V = \{v_1, v_2, K, v_C\}$ are supposed to represent two different partitions of the objects in X , the ARI is then defined as follows [46, 47]:

$$\frac{(a - (((a+c)(a+b))/d))}{(((a+c) + (a+b))/2) - (((a+c)(a+b))/d)}, \quad (28)$$

where a is the number of that objects in a pair belonging to the same classes in U and V ; d is the number of that objects in a pair belonging to the different classes in U and V ; b represents the number of that objects belonging to the same classes in U and the different

classes in V ; c represents the number of that objects belonging to the different classes in U and the same classes in V .

The maximum of ARI is 1, which indicates that the correct classification between classes is achieved by KA [17, 47]. Therefore, the value of ARI can be used to indicate the clustering performance, which can reflect the feature's discriminant power [47]. When the feature sets, $[RFS^1, RFS^2, K, RFS^K]$, are performed for clustering analysis by K -means algorithm, the corresponding $ARI = \{ARI(1), ARI(1), K, ARI(K)\}$ can be obtained. When the value of ARI is higher, the class discriminative degree of feature is greater.

Step 2: for each type bearing condition (fault pattern), for a type of statistical feature, the MD of each feature data (the elements of row of RFS^k) is calculated. Thus, the corresponding MD set can be obtained, that is $[MD_1^k, MD_2^k, K, MD_M^k]$. The expression of MD_i^k is as follows:

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

where

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

Then, for the k -th statistical feature of M fault types, the SMD of feature samples is calculated to obtain $SMD(k)$. The expression of $SMD(k)$ is as follows:

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

Therefore, K types of statistical features have a mean deviation sequence $(SMD(1), SMD(2), K, SMD(K))$. We suppose that the MD can be used to indicate the cohesion of feature data. When the value of $SMD(k)$ is smaller, the class cohesion of the feature is greater.

Step 3: evaluation index of fault sensitivity, FSD (fault sensitivity degree), can be obtained by calculating the ratio of ARI and SMD. For K types of features, there is a FSD sequence $FSD = \{FSD(1), FSD(1), K, FSD(K)\}$, where the $FSD(k)$ is defined as follows:

$$FSD(k) = \frac{ARI(k)}{SMD(k)}. \quad (32)$$

When the value of $FSD(k)$ is higher, the fault sensitivity of feature is better.

Step 4: by calculating the PCC between features, for the raw feature set which contains K types of statistical features, the PCC between each feature and the

remaining $K - 1$ features should be calculated, and thus, each feature has $K - 1$ PCCs. Then, the SPCC (sum of the $K - 1$ PCCs) can be obtained. Given two samples $X = \{x_1, x_2, x_3, \dots, x_n\}$ and $Y = \{y_1, y_2, y_3, \dots, y_n\}$, the PCC is defined as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (33)$$

where μ_X and μ_Y are the mean of samples, σ_X and σ_Y are, respectively, the standard deviation of samples X and Y . Next, there is a SPCC sequence $\text{SPCC} = \{\text{SPCC}(1), \text{SPCC}(2), \dots, \text{SPCC}(K)\}$, and the $\text{SPCC}(k)$ is defined as follows:

$$\text{SPCC}(k) = \sum_{i=1}^K \text{PCC}_{ki} (i \neq k), \quad (34)$$

where PCC_{ki} represents the PCC between the k -th type feature and i -th type feature. In this paper, we suppose that the higher the SPCC of a feature, the higher the redundancy degree of raw feature set caused by the feature will be.

Step 5: a new feature evaluation index, FPSD, can be obtained by combining the FSD and SPCC. The expression of FPSD is presented as follows:

$$\text{FPSD}(k) = \mu \text{FSD}(k) + (1 - \mu) \frac{1}{\text{SPCC}(k)}, \quad (35)$$

where $\mu \in [0, 1]$ is a balance factor. When μ is 0, FPSD just takes feature correlation into account, and on the contrary, FPSD just takes fault sensitivity into account when μ is 1. In this paper, it is presumed that the selection priority of feature is better when the value of FPSD is higher. Therefore, in the descending mode, the sorted FPSD sequence can be obtained by sorting the FPSD of features. The sorted FPSD sequence can be used to select features for the implementation of the subsequent fault diagnosis process.

3.2. Transfer Component Analysis with Preserving Local Manifold Structure (TCAPLMS). TCA can keep the variance of the data to the greatest extent and minimize the marginal distribution differences between datasets in different domains as much as possible [41]. However, TCA does not take the label information and the local manifold structure of feature data into consideration. Aiming at fault pattern recognition, the label information of the training feature dataset is beneficial to improve the discriminant performance of feature data and increase the classification accuracy [9, 17]. Furthermore, preserving the local manifold structure of data is beneficial to pattern recognition and classification of multimode feature data [2, 17]. Therefore, TCAPLMS, a novel feature-based transfer learning method,

is proposed in this section. TCAPLMS naturally inherits the merits of TCA and LFDA, that is, the optimization goal of improved LFDA can be integrated into TCA, where the label information of feature data is considered and the local manifold structure of data is preserved.

Based on the introduction of TCA and LFDA in Section 2, the optimization goal of the TCAPLMS can be defined by integrating the optimization goal of TCA and improved LFDA. The goal function of TCAPLMS is presented as follows:

$$\min_W \text{tr}(W^T K L K W) + a \text{tr} \left(\frac{W^T K \tilde{S}_{SM}^{(w)} K W}{W^T K \tilde{S}_{SM}^{(b)} K W} \right) + \mu \text{tr}(W^T W), \quad (36)$$

s.t. $W^T K H K W = I,$

where $\tilde{S}_{SM}^{(b)}$ and $\tilde{S}_{SM}^{(w)}$ are obtained by modifying $\tilde{S}^{(w)}$ and $\tilde{S}^{(b)}$ (equations (17) and (18)). The expressions of $\tilde{S}_{SM}^{(b)}$ and $\tilde{S}_{SM}^{(w)}$ are shown as follows:

$$\tilde{S}_{SM}^{(w)} \equiv \frac{1}{2} \sum_{i,j=1}^n \tilde{W}_{i,j}^{(sw)} (x_i - x_j)(x_i - x_j)^T, \quad (37)$$

$$\tilde{S}_{SM}^{(b)} \equiv \frac{1}{2} \sum_{i,j=1}^n \tilde{W}_{i,j}^{(sb)} (x_i - x_j)(x_i - x_j)^T,$$

where $\tilde{W}_{i,j}^{(sw)}$ and $\tilde{W}_{i,j}^{(sb)}$ are presented as follows:

$$\tilde{W}_{i,j}^{(sw)} \equiv \begin{cases} \frac{A_{i,j}}{n_\ell}, & \text{if } y_i = y_j = \ell, \\ 0, & \text{if } y_i \neq y_j, \end{cases} \quad (38)$$

$$\tilde{W}_{i,j}^{(sb)} \equiv \begin{cases} A_{i,j} \left(\frac{1}{n} - \frac{1}{n_\ell} \right), & \text{if } y_i = y_j = \ell, \\ \frac{n_j}{n}, & \text{if } y_i \neq y_j (j \in \text{Nst}(i)), \\ \frac{1}{n}, & \text{else.} \end{cases} \quad (39)$$

In equation (39), $y_i \neq y_j (j \in \text{Nst}(i))$ represents that j is the nearest neighbor of i . The reason for the above modification is a shortcoming of LFDA, that is, the neighbor relationships between samples in the same classes are taken into account. However, the neighbor relationships between samples in different classes are not taken into consideration. Aiming at this problem, the between-class scatter matrix $\tilde{S}^{(b)}$ can be modified to $\tilde{S}_{SM}^{(b)}$.

The solution of equation (36) can be transformed to solve the trace optimization problem. The Lagrange multiplier is contained in diagonal matrix that is employed to equation (36), and it is presented as follows:

$$L = \text{tr}\left(W^T K L K^T W + W^T \left(a \left(\tilde{S}_{SM}^{(w)} - \tilde{S}_{SM}^{(b)}\right) + \mu I\right) W\right) + \text{tr}\left((I - W^T K H K^T W) \Phi\right). \quad (40)$$

Then, the matrix W can be solved out by solving a generalized eigenvalue problem:

$$\left(K L K + a \left(\tilde{S}_{SM}^{(w)} - \tilde{S}_{SM}^{(b)}\right) + \mu I\right) W = K H K^T W \Phi. \quad (41)$$

Finally, eigenvalues and the corresponding eigenvectors can be obtained by solving the above problem; the first d ($d < D$, D is a higher dimension of the inputs of TCAPLMS) eigenvectors which are corresponding to the first d smallest nonnegative eigenvalues can be selected to compose the transformation matrix W .

With the use of the proposed TCAPLMS, the low-dimensional representation of the training and testing datasets can be obtained with a smaller difference of marginal distributions between them, and they have greater discriminant performance and less redundant information.

3.3. System Framework. The structure block diagram of the proposed system framework for variable-condition bearing fault diagnosis is shown in Figure 1. According to the system framework, the entire procedure has four steps, namely, signal processing, features extraction, feature transfer learning, and fault pattern recognition. There are two phases, training and testing phases. First of all, the vibration signals collected for training and testing are, respectively, decomposed into different packet nodes by MODWPT. Then, the original statistical feature generation is performed. For the training phase, the original feature set is processed by the proposed PSFFC to obtain sorted FPSD sequence. The most preferred features can be selected for the training fault diagnosis model. In the testing phase, the sorted FPSD obtained from the training phase will be directly applied to select preferred features to construct feature subset. Next, labeled feature data from the training phase and unlabeled feature data from the testing phase are chosen as source domain and target domain, respectively. The proposed TCAPLMS is employed to process source domain and target domain, which can obtain the low-dimensional feature dataset. Finally, the low-dimensional feature dataset in the training phase is employed to train the pattern recognition classifier. Finally, the trained pattern recognition classifier is used to test the low-dimensional testing feature dataset and output the fault diagnosis accuracy.

4. Experiments and Analysis Results

In this paper, bearing vibration datasets obtained from two experimental test platforms are used to validate the effectiveness of the proposed fault diagnosis framework towards real industrial scenes. The introduction of two experimental test rigs is as follows:

- (1) Test rig 1, as shown in Figure 2, is from Case Western Reserve University (CWRU) [48–50], and this test rig supports a motor load of 0–3 horsepower (hp). There are three accelerometers that are placed on the

fan-end and drive-end bearings at the 12 o'clock position.

- (2) Test rig 2 is SQI-MFS test rig, as shown in Figure 3 [9, 17]; there are different fault conditions which are presented in Figure 4. SQI-MFS supports the motor speeds of 1200–1800 rpm. Two accelerometers are placed on the fan-end and drive-end bearings to collect vibration signals. A high-speed AD collector is used to collect the vibration data under different working conditions.

4.1. Experiments Based on Test Platform 1 (CWRU)

4.1.1. Experimental Setup and Cases. Aiming at the verification of the effectiveness of the proposed diagnosis framework under different working conditions, the vibration signals collected from test platform 1 under two motor loads (2 hp and 3 hp) are used for experiments. We collect vibration data of four fault types, including normal condition, ball fault condition, inner race fault condition, and outer race fault condition. Therefore, there are 12 bearing conditions, for each bearing condition, we choose 40 random samples as the testing samples and 20 random samples as the training samples, and a sample has 2000 continuous data points. For the experimental cases, two cases (cases 1 and 2) are used in experiments. Cases 1 and 2 have the same training samples, that is, the samples of 2 hp of motor load are chosen as the training samples. However, the testing samples of cases 1 and 2 are different, the samples of 2 hp and 3 hp of motor loads are chosen as the testing samples in Case 1 and Case 2, respectively. The detailed introduction of experimental data is shown in Table 2.

4.1.2. Experimental Analysis. First of all, vibration signals are processed by MODWPT, and some wavelet packet nodes can be obtained. In this section, the “dmey” is selected as mother wavelet and the decomposition level is 4. One normal sample, one ball fault sample, one inner race fault sample, and one outer race fault sample from the training set of 3 hp are presented in Figures 5–8, respectively. 16 terminal nodes are obtained by signal processing, and 16 HES of reconstruction signals of 16 terminal nodes can be calculated. Therefore, 192 statistical features, the composition of the raw feature set (RFS), can be generated by calculating the 6 statistical parameters of 16 reconstruction signals and 16 HES. Table 3 presents the 6 statistical parameters.

For the RFS, the proposed feature selection method PSFFC is employed to evaluate the feature priority selection degree of each statistical feature of the training data. The ARI, SSMD, FSD, SPCC, and FPSD of 192 features of the training samples are, respectively, shown in Figures 9–13. The horizontal axis of Figures 9–13 is the number of features. After the procedure of PSFFC, a sorted FPSD sequence can

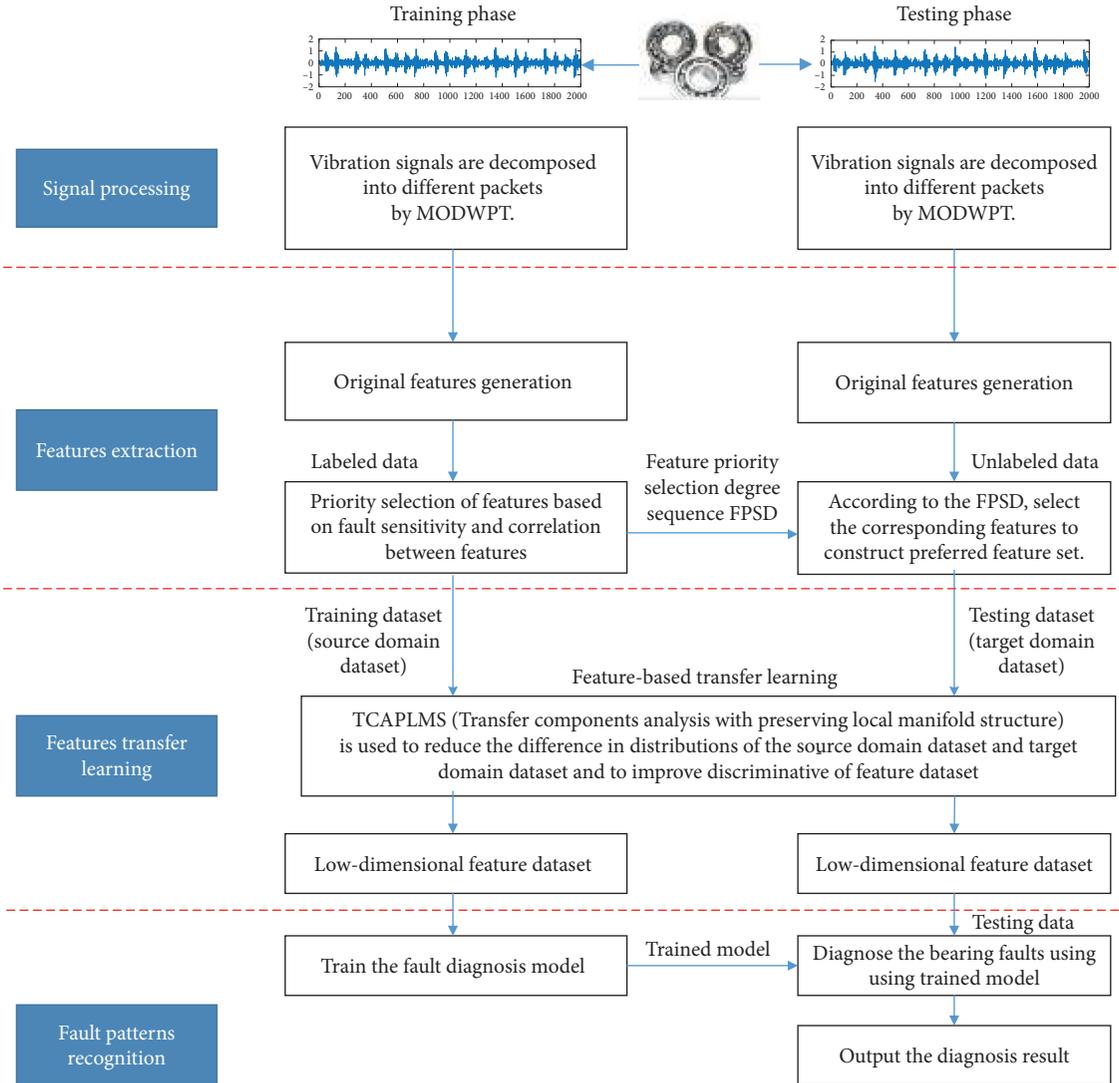


FIGURE 1: The frame structure block diagram of the proposed variable-condition bearing fault diagnosis system.

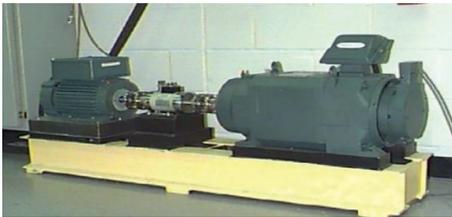


FIGURE 2: Experimental platform 1 (CWRU).

be obtained, and preferred features subset can be formed. Then, the proposed feature-based transfer learning method TCAPLMS is further performed to reduce the marginal distribution differences between the training and testing feature subsets and obtain a low-dimensional feature set with desirable discriminant performance. In this paper, the parameters a and μ in TCAPLMS are 0.5 and 0.3, respectively. Finally, the low-dimensional feature set is employed for the training fault diagnosis model.

For the verification of the effectiveness of the proposed PSFFC and TCAPLMS, two-group comparative experiments are performed. The training and testing datasets are employed for training and testing fault diagnosis model, respectively. In experiments, SVM and KNN are applied for constructing fault diagnosis model, and a series of fault diagnosis models are presented in Table 4. For example, RFS-SVM is a diagnosis model based on SVM, in which the RFS is used as the input of the SVM. TCA is embedded in the RFS-SVM model, which is RFS-TCA-SVM model. PSFFC is embedded in RFS-TCA-SVM model, which is RFS-PSFFC-TCA-SVM model. In this paper, the average diagnostic accuracy of 12 bearing conditions are shown in the experimental analysis, and the detailed description is presented as follows.

In the first group of experiments, PSFFC is not performed. The experimental results of diagnosis models listed in Table 4 are shown in Tables 5–8. For the verification of the superiority of MODWPT, WPT is also used to construct



FIGURE 3: Experimental platform 2 (SQI-MFS).



FIGURE 4: Bearings used in the SQI-MFS test platform.

TABLE 2: The introduction of experimental data.

Fault condition	Fault size (mm)	Number of training samples		Number of testing samples		Class label
		2 hp	3 hp	2 hp (case 1)	3 hp (case 2)	
Normal	0	20	40	40	40	1
	0.007	20	40	40	40	2
	0.014	20	40	40	40	3
	0.021	20	40	40	40	4
	0.028	20	40	40	40	5
Inner race fault	0.007	20	40	40	40	6
	0.014	20	40	40	40	7
	0.021	20	40	40	40	8
	0.028	20	40	40	40	9
Outer race fault	0.007	20	40	40	40	10
	0.014	20	40	40	40	11
	0.021	20	40	40	40	12
Number of samples		240	480	480	480	

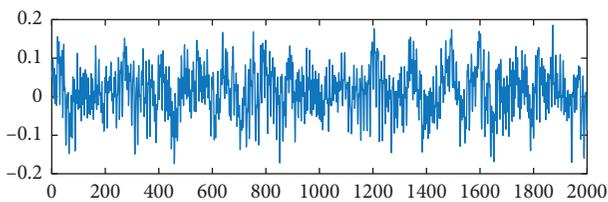


FIGURE 5: A normal vibration signal sample.

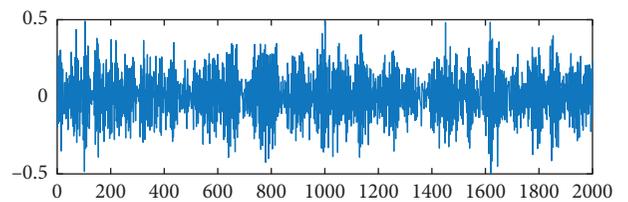


FIGURE 6: A vibration signal sample of ball fault.

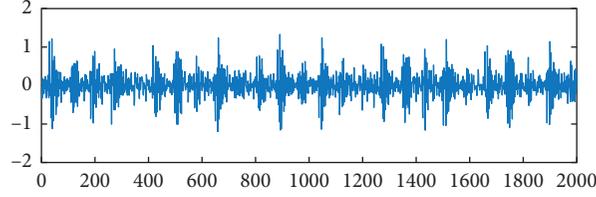


FIGURE 7: A vibration signal sample of inner race fault.

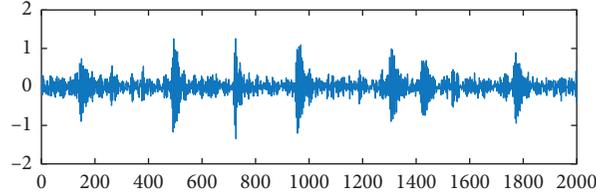


FIGURE 8: A vibration signal sample of outer race fault.

TABLE 3: 6 statistical parameters.

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

TABLE 4: Fault diagnosis models.

Number	The first group experimental models		The second group experimental models	
	The PSFFC is not applied		The PSFFC is applied	
	SVM-based	KNN-based	SVM-based	KNN-based
1	RFS-SVM	RFS-KNN	RFS-PSFFC-SVM	RFS-PSFFC-KNN
2	RFS-LFDA-SVM	RFS-LFDA-KNN	RFS-PSFFC-LFDA-SVM	RFS-PSFFC-LFDA-KNN
3	RFS-TCA-SVM	RFS-TCA-KNN	RFS-PSFFC-TCA-SVM	RFS-PSFFC-TCA-KNN
4	RFS-TCAPLMS-SVM	RFS-TCAPLMS-KNN	RFS-PSFFC-TCAPLMS-SVM	RFS-PSFFC-TCAPLMS-KNN

TABLE 5: Testing accuracies of RFS-SVM and RFS-KNN.

RFS-SVM		RFS-KNN			
Testing accuracy (%)		Testing accuracy (%)			
MODWPT	WPT	MODWPT	WPT	MODWPT	WPT
Case 1	Case 2	Case 1	Case 2	Case 1	Case 2
98.54	83.54	95.21	76.25	98.54	94.63
				95.00	86.67

diagnosis model, and the experimental results of the RFS-SVM and RFS-KNN models using MODWPT and WPT are, respectively, presented in Table 5. According to the diagnosis accuracy in Table 5, it is evident that the diagnosis result of model using MODWPT is better than the model using WPT. Therefore, the experimental results and analysis of all models using MODWPT are introduced below. For the testing set of case 1, all models can obtain desirable diagnosis accuracy. The maximum accuracy of RFS-SVM, RFS-LFDA-SVM, RFS-TCA-SVM, and RFS-TCAPLMS-SVM can attain

98.54%, 99.79%, 91.67%, and 99.38%, respectively. The main reason is that the training and testing samples are from the same working condition and the distribution between them is almost the same. For the testing set of case 2, the highest accuracy of RFS-SVM, RFS-LFDA-SVM, and RFS-TCA-SVM can only attain 83.54%, 87.08%, and 87.29%. However, the highest accuracy of the RFS-TCAPLMS-SVM model can attain 97.50%, which is obviously higher than that of other models. The experimental results of KNN-based models are similar to that of SVM-based models, and the highest accuracy of the RFS-TCAPLMS-KNN model can attain 97.08%, which is obviously higher than that of other models. A problem can be found from the first group experimental results, that is, for the conventional fault diagnosis model, it is not easy to guarantee a preferable diagnosis performance when the distribution between the testing and training set is different. The use of TCAPLMS can help the diagnosis model attain desirable diagnosis performance.

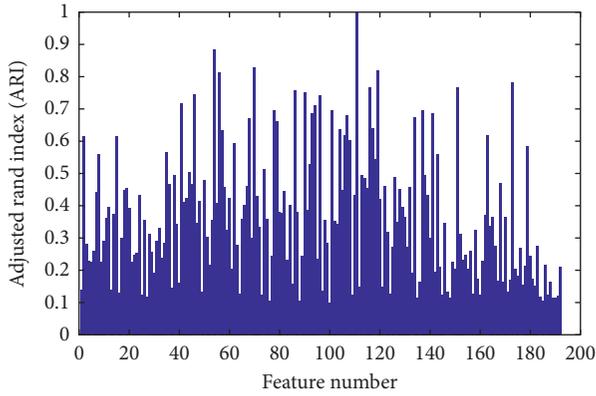


FIGURE 9: The adjust rand index.

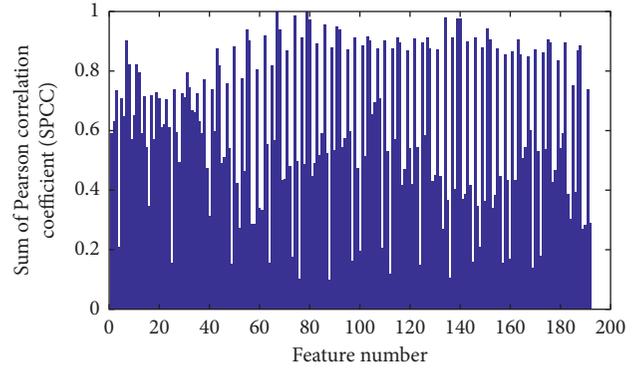


FIGURE 12: The sum of the Pearson correlation coefficient.

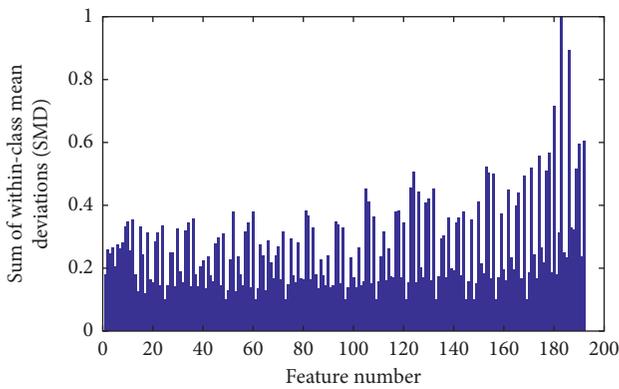


FIGURE 10: The sum of within-class mean deviations.

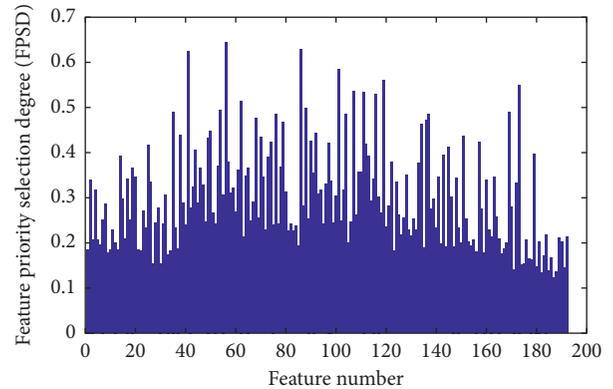


FIGURE 13: The feature priority selection degree of 192 statistical features ($\mu = 0.6$).

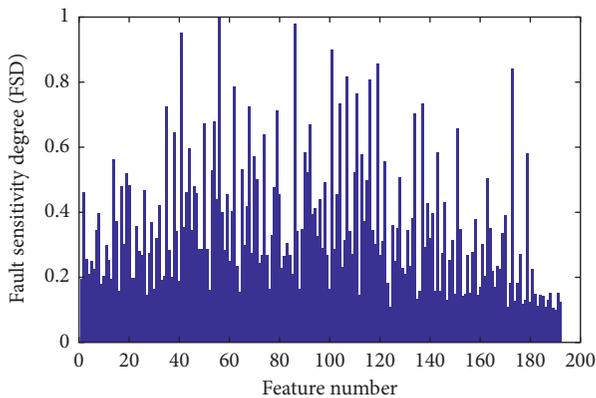


FIGURE 11: The fault sensitivity degree features of the training samples.

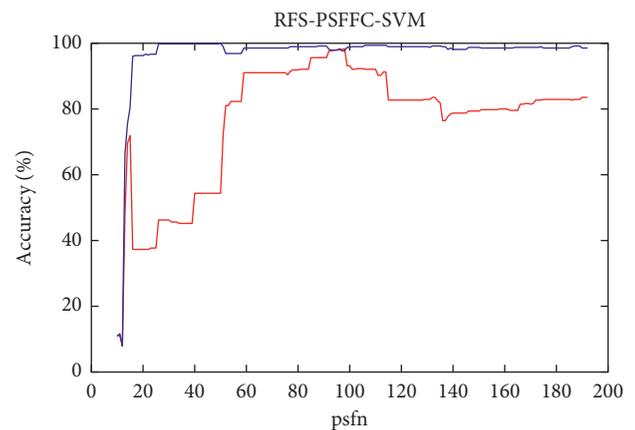


FIGURE 14: The testing accuracies of RFS-PSFFC-SVM with different psfn values.

There is an experimental analysis about the second group of experiments; PSFFC is performed before the steps of features transfer learning and patterns classification. The experimental results of fault diagnosis models listed in Table 4 are shown in Tables 9–12 and Figures 14–22. The dimension represents the number of dimension sizes. All diagnosis models can achieve desirable results when the testing data are from case 1. The highest accuracy can attain over 99.5%, in that the highest

accuracies of RFS-PSFFC-SVM, RFS-PSFFC-LFDA-SVM, and RFS-PSFFC-TCAPLMS-SVM can attain 100%. For the testing set of case 2, compared with RFS-SVM, RFS-PSFFC-SVM, RFS-PSFFC-TCA-SVM, and RFS-PSFFC-TCAPLMS-SVM can

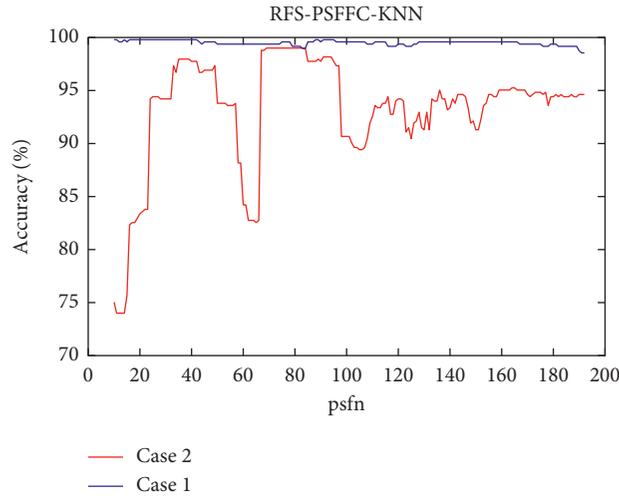


FIGURE 15: The testing accuracies of RFS-PSFFC-KNN with different psfn values.

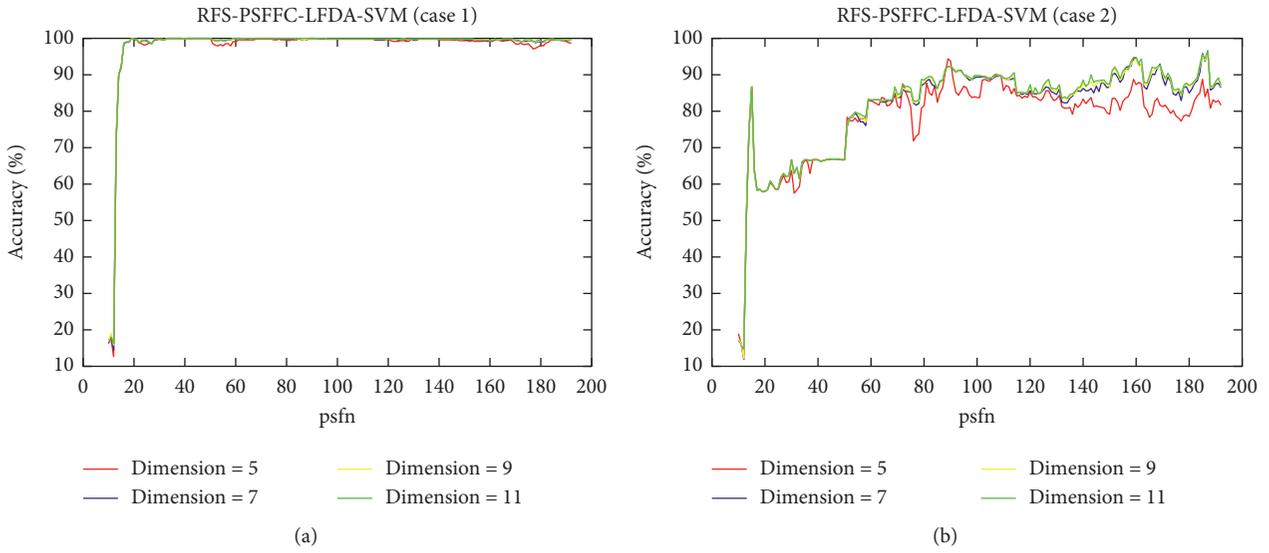


FIGURE 16: The testing accuracies of RFS-PSFFC-LFDA-SVM.

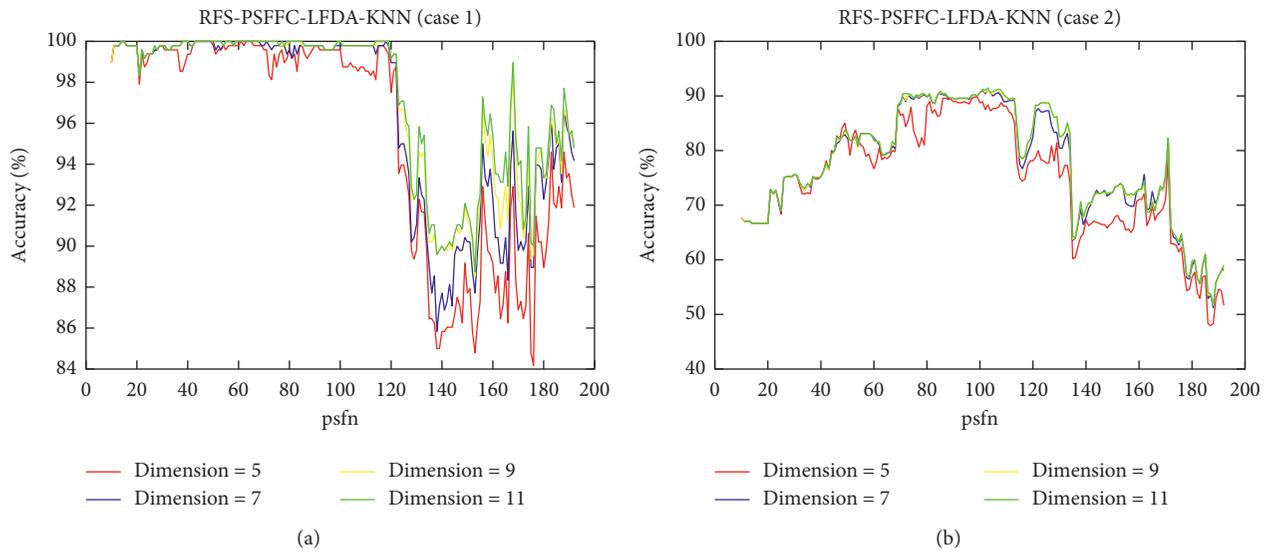


FIGURE 17: The testing accuracies of RFS-PSFFC-LFDA-KNN.

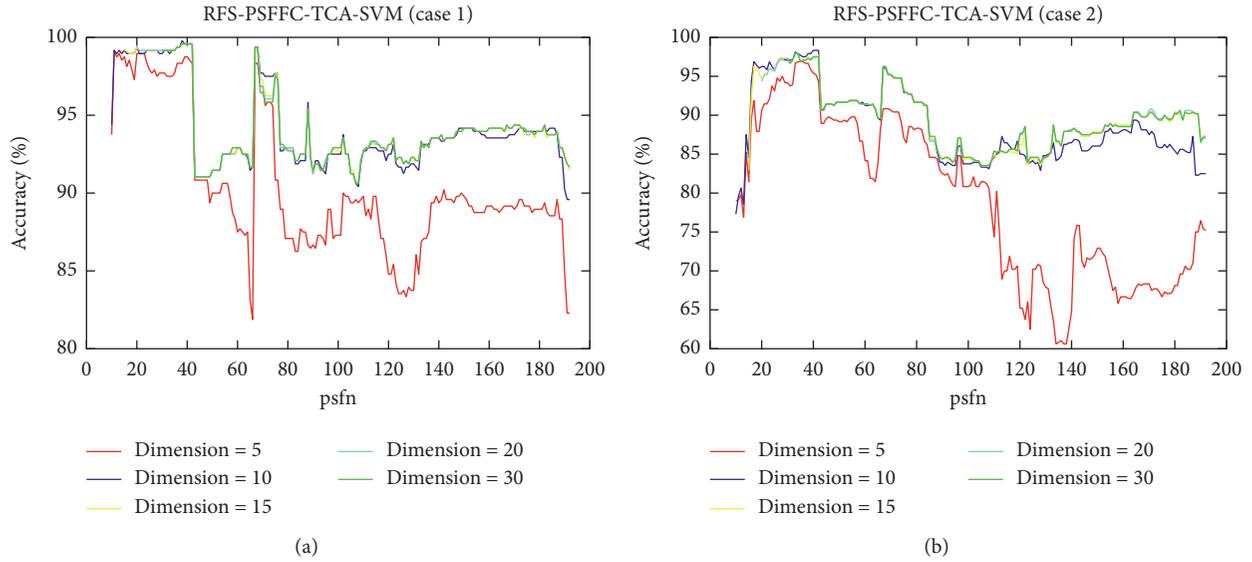


FIGURE 18: The testing accuracies of RFS-PSFFC-TCA-SVM with different dimension sizes for TCA.

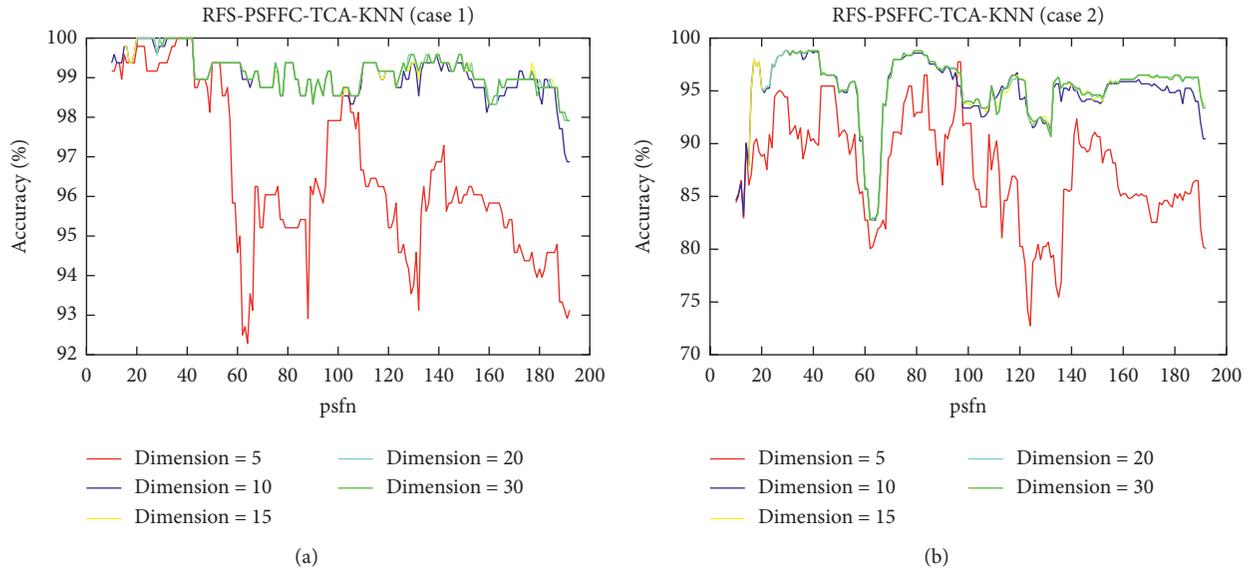


FIGURE 19: The testing accuracies of RFS-PSFFC-TCA-KNN with different dimension sizes for TCA.

achieve desirable performance which is obviously better than RFS-SVM, the maximum diagnosis accuracy of them can attain 95.63% (psfn = 90), 97.50% (psfn = 40), and 100% (psfn = 40, 70, 80), respectively. The experimental results of KNN-based models are similar to that of SVM-based models, and for the testing set of case 2, the highest diagnosis accuracy of RFS-PSFFC-TCAPLMS-KNN can attain 100%. According to the second group experimental results, under different working conditions, it is obvious that the proposed PSFFC can help diagnosis models to improve the diagnosis performance, and the combination of PSFFC and TCAPLMS can attain ideal fault diagnosis accuracy when a good parameter psfn is used; for the testing set of case 2, RFS-PSFFC-TCAPLMS-SVM model can attain 100% accuracy when psfn is 40, 70, or 80. Therefore, the effectiveness and adaptability of the PSFFC and TCAPLMS are verified.

4.2. Experiments Based on Test Platform 2 (SQI-MFS)

4.2.1. Experimental Setup and Cases. The vibration signals collected from the SQI-MFS test platform which is working under different motor speeds are used for experiments. The vibration signal samples of 1200 rpm and 1800 rpm are chosen for a series of experiments; there are 10 bearing conditions, including four (normal, ball fault, inner race fault, and outer race fault) bearing fault types. For each bearing condition, we choose 40 random samples as the testing samples and 20 random samples as the training samples, and a sample has 5000 continuous data points. For the experimental cases, two cases (cases 1 and 2) are used in experiments. Cases 1 and 2 have the same training samples, that is, the samples of 1800 rpm of motor speed are chosen as

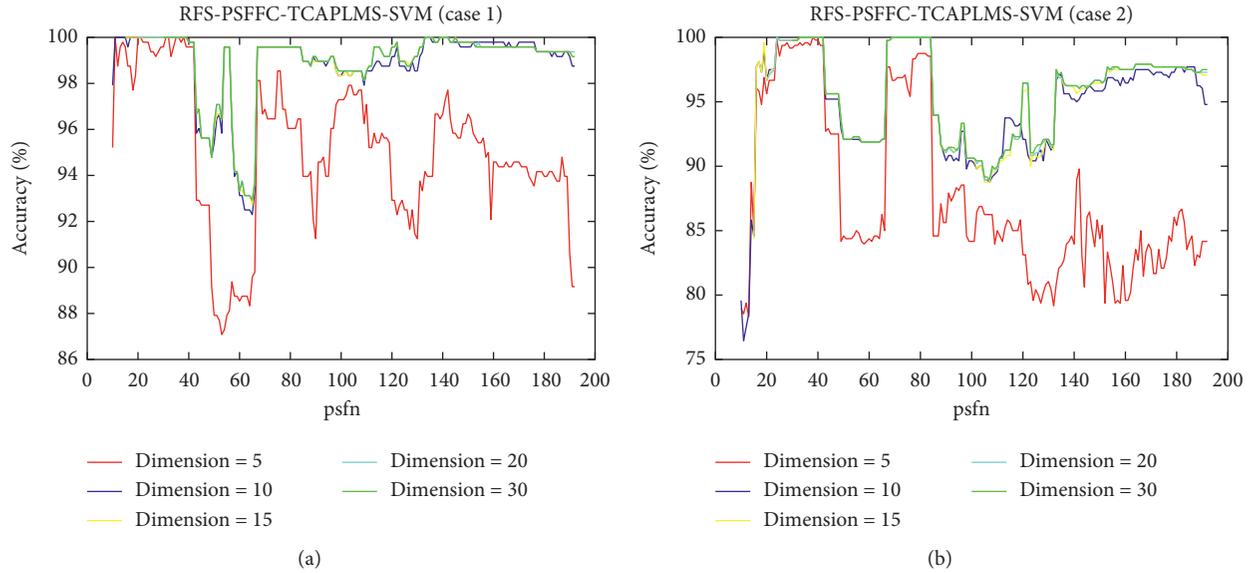


FIGURE 20: The testing accuracies of RFS-PSFFC-TCAPLMS-SVM with different dimension sizes for TCAPLMS.

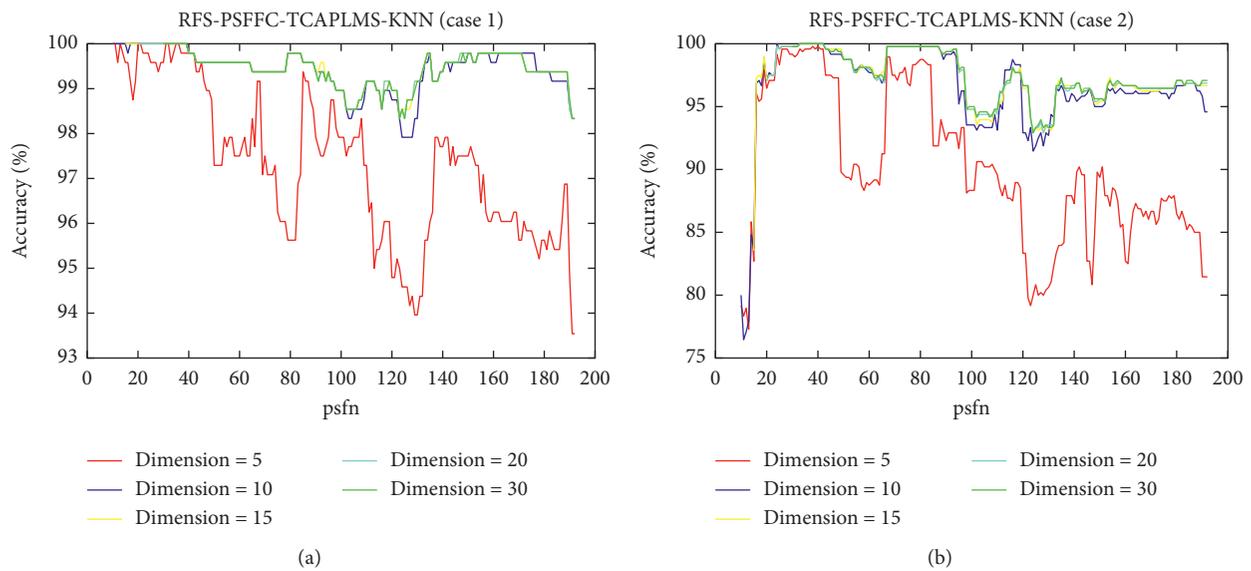


FIGURE 21: The testing accuracies of RFS-PSFFC-TCAPLMS-KNN with different dimension sizes for TCAPLMS.

the training samples. However, the testing samples of cases 1 and 2 are different, the samples of 1800 rpm and 1200 rpm of motor speeds are, respectively, chosen as the testing samples in case 1 and case 2. The detailed introduction of the experimental dataset is shown in Table 13.

4.2.2. Experimental Analysis. The experimental procedure is the same as that of test rig 1; first of all, vibration signals are processed by MODWPT, and different wavelet packet nodes can be obtained. One normal, one ball fault, one inner race fault, and one outer race fault vibration signal samples from

the training set of 1800 rpm are presented in Figures 23–26, respectively. Then, according to the decomposition of a vibration signal, it is similar to Section 4.1.2, 192 statistical features that compose the RFS can be obtained.

For the RFS, the proposed feature selection method PSFFC is employed to evaluate the feature priority selection degrees of 192 statistical features. The ARI, SSMD, FSD, SPCC, and FPSD of 192 statistical features of the training samples are, respectively, shown in Figures 27–31. After the procedure of PSFFC, a sorted FPSD sequence can be obtained, and preferred features subset can be formed. Then, the proposed feature-based transfer learning method

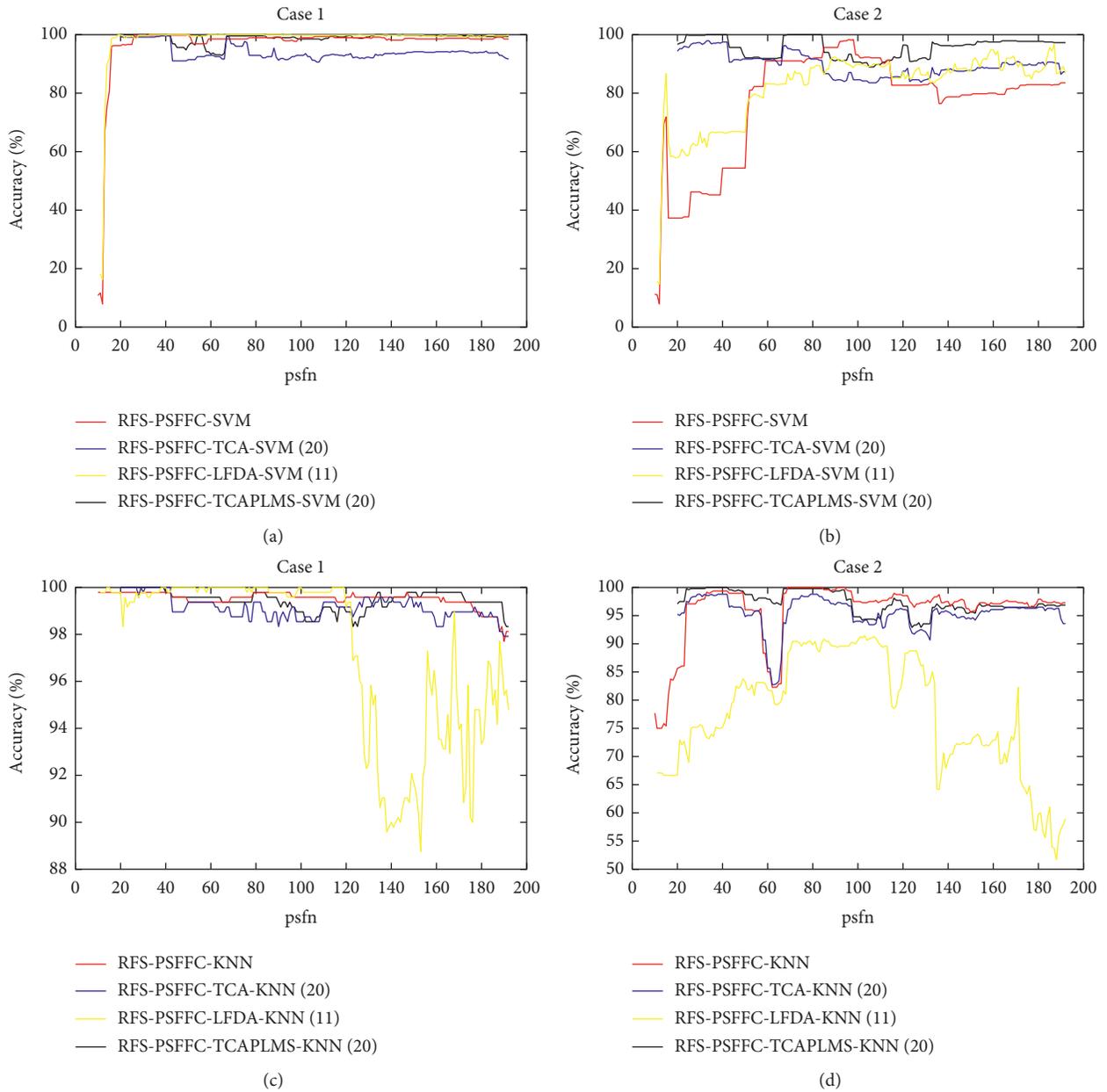


FIGURE 22: The testing accuracies of models with the use of PSFFC, TCA, LFDA, and TCAPLMS, respectively.

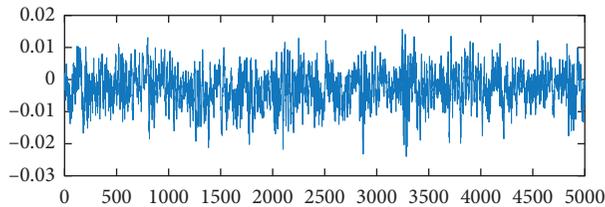


FIGURE 23: A normal vibration signal sample.

TCAPLMS is further performed to process the training and testing feature subsets, which can help to obtain a desirable discriminant performance. Finally, the low-dimensional feature set is employed for the training fault diagnosis model.

The description of the verification of the effectiveness of PSFFC and TCAPLMS is presented as follows. RFS-SVM, RFS-KNN, RFS-TCAPLMS-SVM, RFS-TCAPLMS-KNN, RFS-PSFFC-TCAPLMS-SVM, and RFS-PSFFC-TCAPLMS-KNN are applied for experiments. In Table 14, the

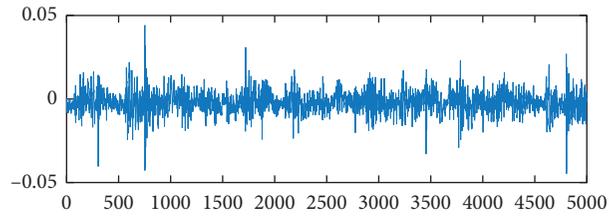


FIGURE 24: A ball fault vibration signal sample.

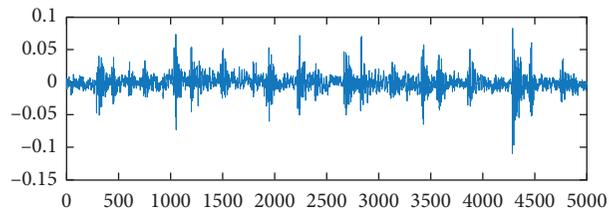


FIGURE 25: An inner race fault signal sample.

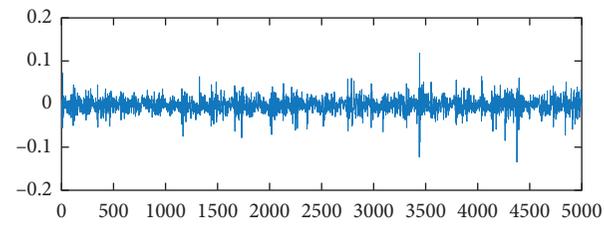


FIGURE 26: An outer race fault signal sample.

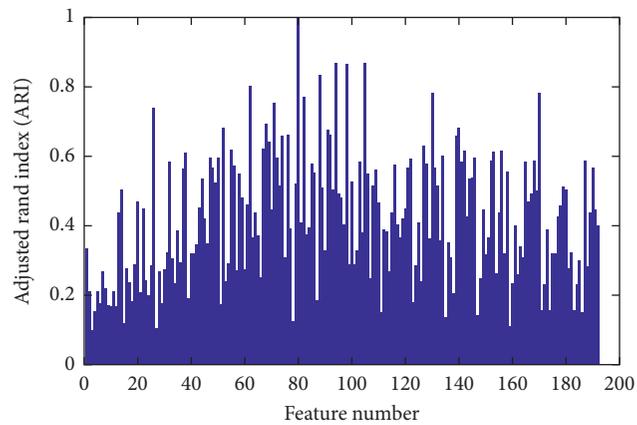


FIGURE 27: The adjusted rand index.

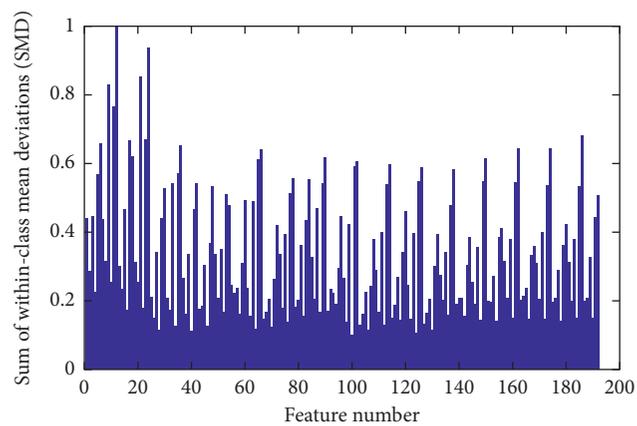


FIGURE 28: The sum of within-class mean deviations.

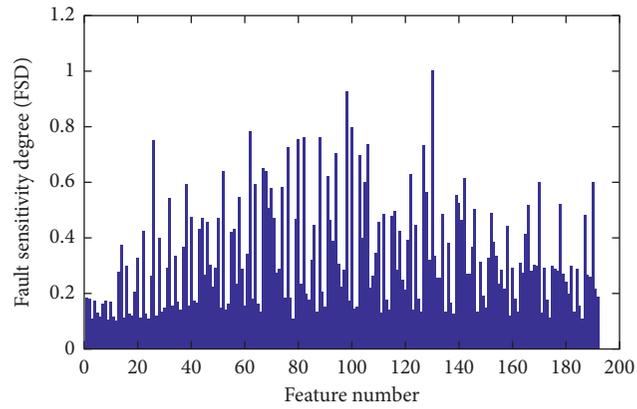


FIGURE 29: The fault sensitivity degree.

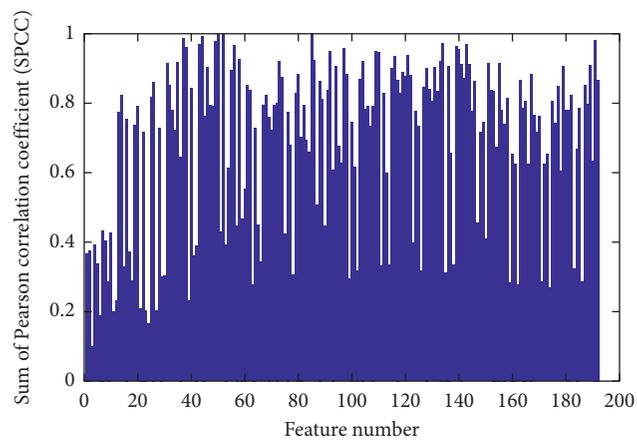
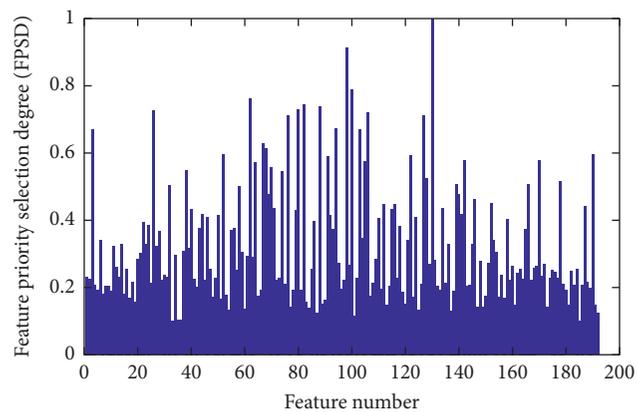


FIGURE 30: The sum of the Pearson correlation coefficient.

FIGURE 31: The feature priority selection degree of 192 statistical features ($\mu = 0.6$).

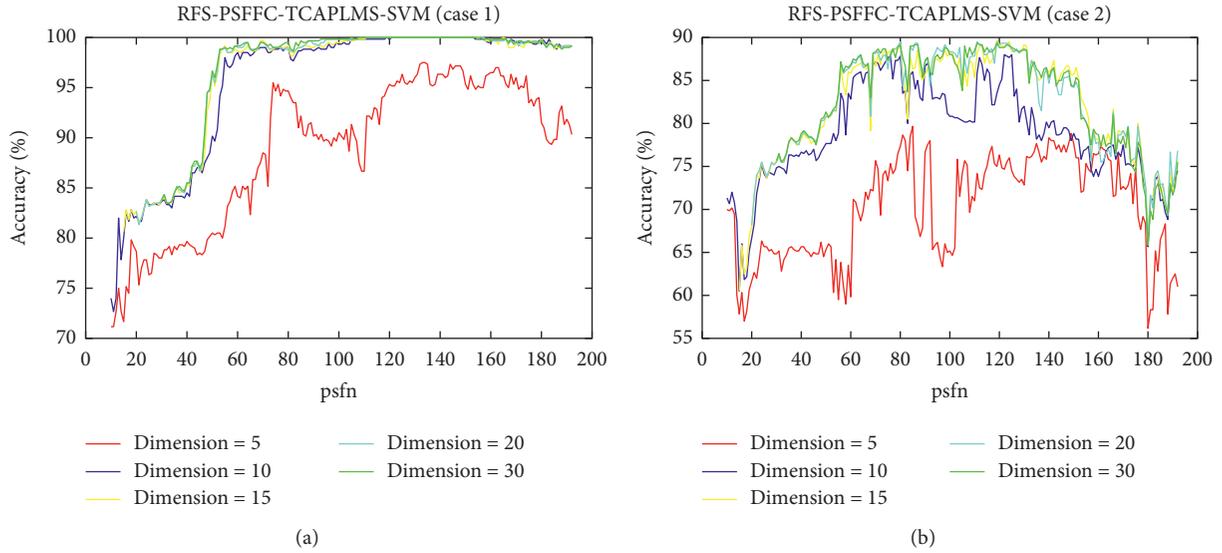


FIGURE 32: The testing accuracies of RFS-PSFFC-TCAPLMS-SVM with different dimension sizes for TCAPLMS.

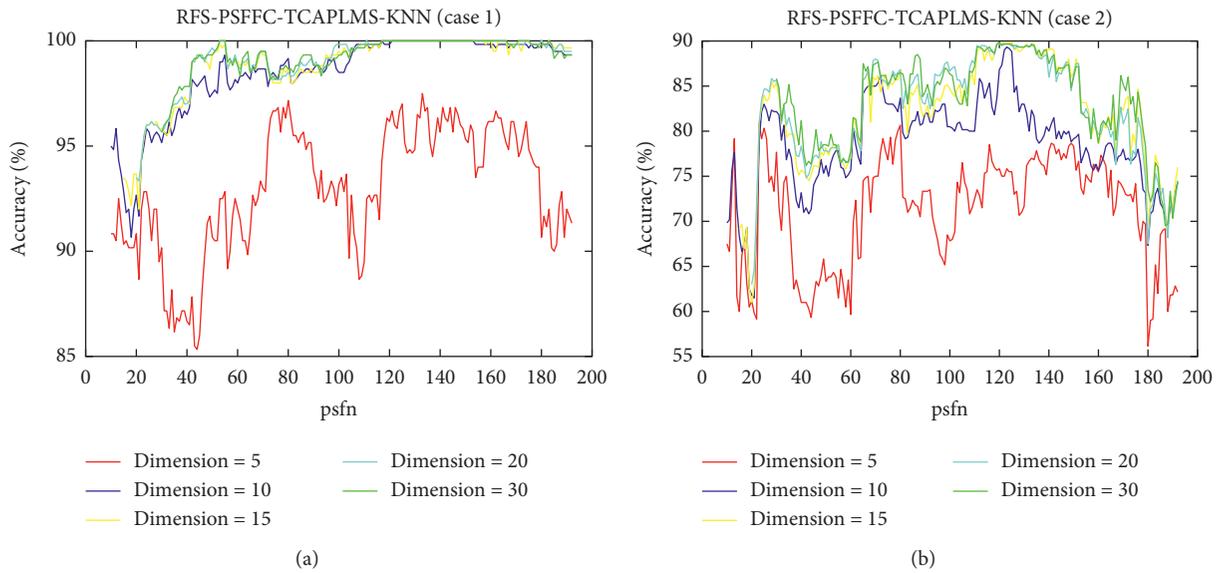


FIGURE 33: The testing accuracies of RFS-PSFFC-TCAPLMS-KNN with different dimension sizes for TCAPLMS.

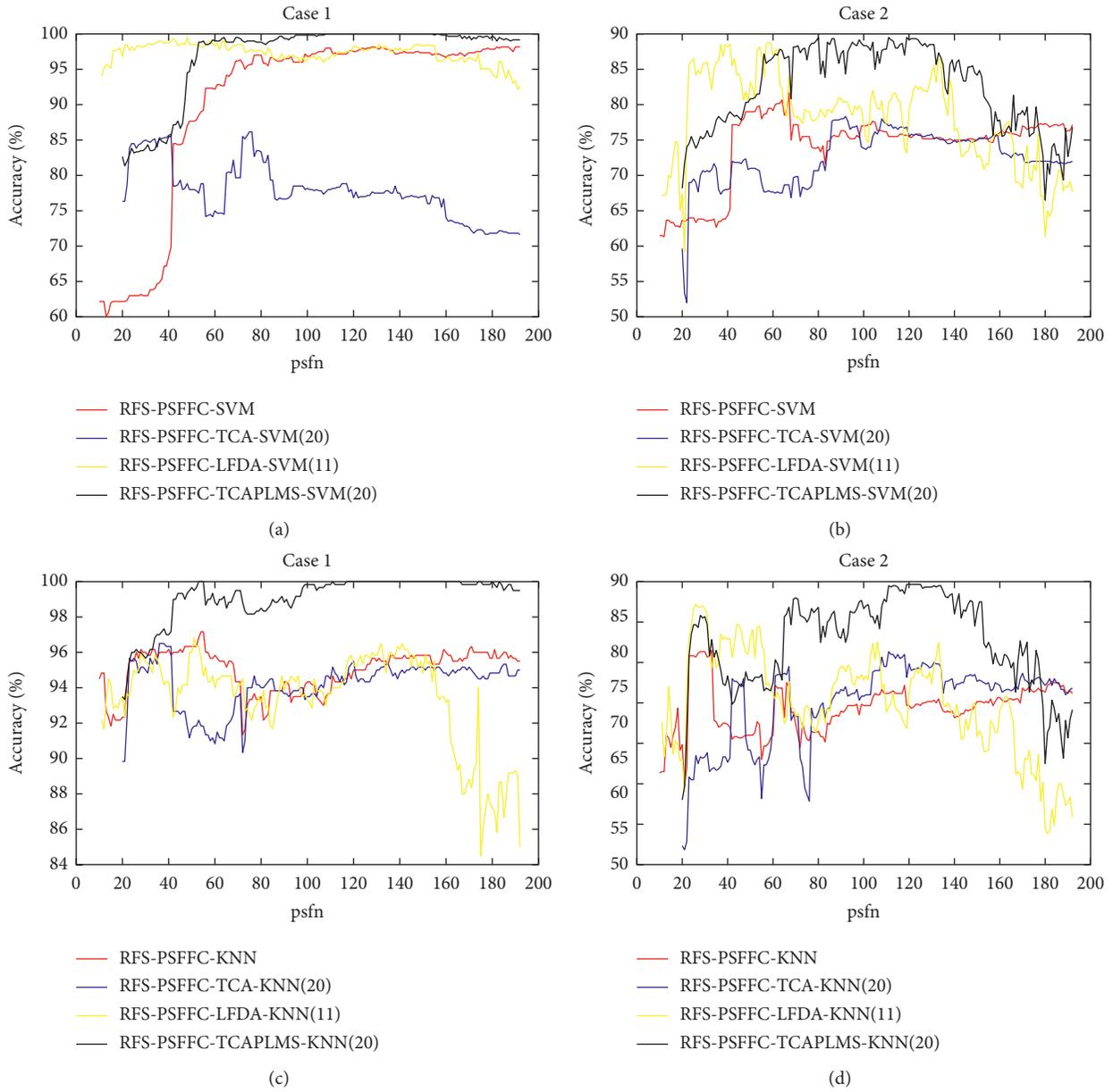


FIGURE 34: The testing accuracies of models with the use of PSFFC, TCA, LFDA, and TCAPLMS.

TABLE 6: Testing accuracies of RFS-LFDA-SVM and RFS-LFDA-KNN.

Dimension size	RFS-LFDA-SVM		RFS-LFDA-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
5	98.75	81.67	91.88	51.67
7	99.58	86.46	94.17	58.33
9	99.58	86.67	94.79	59.17
11	99.79	87.08	94.79	58.96

TABLE 7: Testing accuracies of RFS-TCA-SVM and RFS-TCA-KNN.

Dimension size	RFS-TCA-SVM		RFS-TCA-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
5	82.29	75.21	93.13	80.04
10	89.58	82.50	96.88	90.46
15	91.46	87.29	97.92	93.79
20	91.67	87.29	97.92	93.58
25	91.67	87.08	97.92	93.58
30	91.67	87.08	97.92	93.38

TABLE 8: Testing accuracies of RFS-TCAPLMS-SVM and RFS-TCAPLMS-KNN.

Dimension size	RFS-TCAPLMS-SVM		RFS-TCAPLMS-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
5	89.17	84.17	93.54	81.46
10	98.75	94.79	98.33	94.58
15	99.17	97.08	98.33	96.67
20	99.38	97.29	98.33	96.88
25	99.38	97.50	98.33	96.88
30	99.17	97.50	98.33	97.08

TABLE 9: Testing accuracies obtained by RFS-PSFFC-SVM and RFS-PSFFC-KNN.

sfn	RFS-PSFFC-SVM		RFS-PSFFC-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
10	10.83	11.25	99.79	75.04
20	96.25	37.29	99.79	83.38
30	99.79	46.25	99.79	94.21
40	99.79	54.38	99.79	97.75
50	100.00	54.38	99.38	93.79
60	98.54	91.04	99.38	84.21
70	98.54	91.04	99.38	99.00
80	98.96	91.88	99.17	99.00
90	99.17	95.63	99.58	97.75
100	98.75	93.13	99.58	90.67
110	99.38	92.08	99.38	92.54
120	98.96	82.71	99.38	94.21
130	98.96	82.92	99.58	91.29
140	98.13	78.75	99.58	93.38
150	98.75	79.38	99.58	91.29
160	98.54	80.00	99.58	95.04
170	98.75	81.46	99.38	94.63
180	98.54	82.92	99.38	94.42
190	98.54	83.54	98.75	94.63

TABLE 10: The testing accuracies of RFS-PSFFC-LFDA-SVM and RFS-PSFFC-LFDA-KNN (the dimension size is 11).

sfn	RFS-PSFFC-LFDA-SVM		RFS-PSFFC-LFDA-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
11	18.13	15.63	99.79	67.08
20	99.79	57.92	99.79	66.67
30	99.79	66.67	99.79	75.63
40	99.79	66.67	100.00	75.21
50	100.00	66.67	100.00	83.33
60	100.00	83.13	100.00	81.88
70	100.00	84.58	100.00	88.96
80	100.00	88.54	100.00	89.79
90	100.00	92.29	99.79	89.58
100	99.79	89.79	100.00	90.63
110	99.79	89.17	99.79	89.79
120	99.79	87.29	99.17	84.38
130	99.79	86.04	92.50	82.50
140	99.79	88.54	90.00	69.38
150	99.79	87.50	91.67	72.29
160	99.38	94.79	95.42	72.92
170	99.58	90.63	93.96	75.63
180	99.17	87.08	93.33	59.79
190	99.58	88.13	95.42	57.08

experimental results of RFS-SVM and RFS-KNN models are presented. The experimental results of RFS-TCAPLMS-SVM and RFS-TCAPLMS-KNN are shown in Table 15. The

diagnosis results of RFS-PSFFC-TCAPLMS-SVM and RFS-PSFFC-TCAPLMS-KNN are presented in Table 16. When the testing dataset is from case 1, all models can attain

TABLE 11: The testing accuracies of RFS-PSFFC-TCA-SVM and RFS-PSFFC-TCA-KNN (the dimension size is 20).

sfn	RFS-PSFFC-TCA-SVM		RFS-PSFFC-TCA-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
20	99.38	94.38	100.00	95.25
30	99.17	97.29	99.79	98.79
40	99.58	97.50	100.00	98.79
50	91.46	91.67	99.38	94.83
60	92.92	91.46	99.38	85.67
70	96.88	95.21	98.75	96.29
80	92.92	91.67	99.38	98.79
90	91.25	84.38	98.33	96.92
100	92.92	84.58	98.54	94.00
110	92.71	85.42	99.38	94.63
120	93.13	87.50	99.17	96.08
130	92.29	84.58	99.58	92.13
140	93.54	87.92	99.58	95.67
150	94.17	87.71	99.17	94.42
160	94.17	88.54	98.33	96.08
170	94.38	90.63	98.96	96.50
180	93.96	90.42	98.96	96.50
190	92.29	86.46	97.92	94.42

TABLE 12: The testing accuracies of RFS-PSFFC-TCAPLMS-SVM and RFS-PSFFC-TCAPLMS-KNN (the dimension size is 20).

sfn	RFS-PSFFC-TCAPLMS-SVM		RFS-PSFFC-TCAPLMS-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
20	100.00	96.67	100.00	97.08
30	100.00	99.79	100.00	99.79
40	99.79	100.00	99.79	100.00
50	96.25	92.08	99.58	98.75
60	93.33	91.88	99.58	97.92
70	99.58	100.00	99.38	99.79
80	99.58	100.00	99.79	99.79
90	99.17	91.04	99.38	99.38
100	98.54	90.63	98.96	94.79
110	98.54	89.79	99.17	94.38
120	99.58	96.46	98.96	96.67
130	99.17	92.08	99.17	93.54
140	100.00	96.25	99.38	96.67
150	99.79	96.67	99.79	95.63
160	99.58	97.50	99.79	96.67
170	99.58	97.92	99.79	96.46
180	99.38	97.71	99.38	97.08
190	99.38	97.29	98.54	96.88

TABLE 13: The introduction of experimental data.

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

TABLE 14: The testing accuracies of RFS-SVM and RFS-KNN.

RFS-SVM		RFS-KNN	
Testing accuracy (%)		Testing accuracy (%)	
Case 1	Case 2	Case 1	Case 2
98.17	77.17	95.50	76.17

TABLE 15: The testing accuracies of the RFS-TCAPLMS-SVM and RFS-TCAPLMS-KNN.

Dimension size	RFS-TCAPLMS-SVM		RFS-TCAPLMS-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
5	90.33	61.00	91.33	62.17
10	99.17	74.50	99.33	74.33
15	99.17	75.83	99.67	76.00
20	99.17	76.83	99.50	74.17
30	99.17	75.50	99.33	74.50

TABLE 16: The testing accuracies of RFS-PSFFC-TCAPLMS-SVM and RFS-PSFFC-TCAPLMS-KNN (the dimension size is 20).

sfn	RFS-PSFFC-TCAPLMS-SVM		RFS-PSFFC-TCAPLMS-KNN	
	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)	Case 1 testing accuracy (%)	Case 2 testing accuracy (%)
20	82.67	68.17	93.50	63.00
30	83.83	75.50	95.67	85.67
40	85.50	78.50	97.00	78.33
50	96.33	80.83	99.67	77.83
60	99.00	86.67	98.83	77.50
70	99.50	88.17	99.50	88.00
80	98.83	89.50	98.50	86.83
90	99.33	87.17	98.83	83.00
100	99.83	88.33	99.83	86.67
110	100.00	87.83	99.83	87.50
120	100.00	89.17	100.00	89.67
130	100.00	88.50	100.00	89.33
140	100.00	84.67	100.00	86.00
150	100.00	84.17	100.00	86.33
160	99.83	77.83	100.00	79.67
170	99.67	79.17	99.83	80.83
180	99.67	66.50	100.00	67.50
190	99.17	72.67	99.50	70.67

desirable diagnosis accuracy. The highest diagnosis accuracy of RFS-PSFFC-TCAPLMS-SVM (psfn is 110–150) and RFS-PSFFC-TCAPLMS-KNN (psfn is 120–150) can attain 100%. When a good parameter psfn can be used, the fault diagnosis model for the testing set of cases 2 can attain desirable diagnosis results, for example, the highest testing accuracy of RFS-PSFFC-TCAPLMS-SVM model can attain 89.50% when psfn is 80, which is obviously higher than that of RFS-TCAPLMS-SVM. The curve representation of testing accuracies of RFS-PSFFC-TCAPLMS-SVM and RFS-PSFFC-TCAPLMS-KNN is shown in Figures 32 and 33. The testing

accuracies of models with the use of PSFFC, TCA, LFDA, and TCAPLMS are presented in Figure 34. Therefore, the effectiveness and adaptability of the PSFFC and TCAPLMS can be further validated (Figures 32–34).

5. Conclusions

In the face of real industrial scenarios, the complex working conditions can lead to that data-driven diagnosis methods using conventional machine learning techniques often highlight a limitation that it is difficult to achieve the

desirable fault diagnosis performance, due to that the feature distributions of training and testing data are assumed to be the same. Aiming at this problem, a novel intelligent bearing fault diagnosis framework is proposed towards real industrial scenarios. In this framework, an improved domain adaptation method, transfer component analysis with preserving local manifold structure (TCAPLMS), is proposed to reduce the marginal distributions differences between different domain datasets, and at the same time, take the label information of feature dataset and the local manifold structure of feature data into consideration. Furthermore, preferred feature selection by fault sensitivity and feature correlation (PSFFC) is embedded into this framework for selecting features which are more beneficial to fault pattern recognitions and reduce the redundancy of feature set. Finally, vibration signal datasets collected from two experimental test platforms are used for experiments.

It is obvious that the proposed PSFFC and TCAPLMS have a great potential to be beneficial in actual bearing fault diagnosis applications. In experiments, two cases are selected as comparative cases, and for the experimental test rig 1, cases 1 and 2 have the same training samples, but the testing samples are different. The experimental results show that the diagnosis model using PSFFC and TCAPLMS can attain desirable performance and improve the generalization ability of models, and when a good parameter psfn is used, cases 1 and 2 can both attain 100% diagnosis accuracies. In summary, the experimental results from test rig 2 further demonstrate the effectiveness, adaptability, and great potential of the diagnosis model using PSFFC and TCAPLMS under variable working conditions.

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 supported by the National Key R&D Program of China (nos. 2017YFC0804400 and 2017YFC0804401), "Smart mine" key technology R&D open fund of China University of Mining and Technology and Zibo Mining Group Co., Ltd, and Special translation on innovation ability of scientific research base of China University of Mining and Technology (NO. 2018CXNL02).

References

- [1] Z. Huo, Y. Zhang, P. Francq et al., "Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures," *IEEE Access*, vol. 1, 2017.
- [2] Z. Xiang, Y. Li, P. Wei et al., "Multi-Layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol. 1, 2019.
- [3] H. Zheng, R. Wang, Y. Yang, Y. Li, and M. Xu, "intelligent fault identification based on multisource domain generalization towards actual diagnosis scenario," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 2, pp. 1293–1304, 2020.
- [4] 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, 2019.
- [5] 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.
- [6] H. Zheng, R. Wang, Y. Yang et al., "A diagnosis framework based on domain adaptation for bearing fault diagnosis across diverse domains," *ISA Transactions*, vol. 99, pp. 465–478, 2020.
- [7] S. Liu, X. X. Yu, F. Qian, and F. Dong, "Rolling bearing fault diagnosis based on sensitive feature transfer learning and local maximum margin criterion under variable working condition," *Shock and Vibration*, vol. 2020, pp. 1–34, 2020.
- [8] J. Xingxing, S. Changqing, S. Juanjuan et al., "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, 2018.
- [9] D. Fei, Y. Xiao, D. Enjie et al., "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, pp. 1–29, 2018.
- [10] X. Yu, X. Ren, H. Wan et al., "Rolling bearing fault feature extraction and diagnosis method based on MODWPT and DBN," in *Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, IEEE, New York, NY, USA, 2019.
- [11] H. Shao, J. Lin, L. Zhang et al., "Compound fault diagnosis for a rolling bearing using adaptive DTCWPT with higher order spectra," *Quality Engineering*, vol. 2020, pp. 1–12, 2020.
- [12] Z. Zhang, J. Wu, J. Ma, X. Wang, and C. Zhou, "Fault diagnosis for rolling bearing based on lifting wavelet and morphological fractal dimension," in *Proceedings of the 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 6351–6354, Qingdao, China, 2015.
- [13] W. Fu, J. Zhou, and Y. Zhang, "fault diagnosis for rolling element bearings with VMD time-frequency analysis and SVM," in *Proceedings of the 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*, pp. 69–72, Qinhuangdao, UK, 2015.
- [14] C. Zhang, B. Chen, F. Wan, and B. Song, "Multi-Faults diagnosis of rotating bearings using flexible time-frequency analysis technique," in *Proceedings of the 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, pp. 347–352, Xi'an, China, 2018.
- [15] D. Zhong, W. Guo, and D. He, "An intelligent fault diagnosis method based on stft and convolutional neural network for bearings under variable working conditions," in *Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, pp. 1–6, Qingdao, China, 2019.
- [16] B. Merainani, A. A. Bouzid, A. Ratni, and D. Benazzouz, "Detection of bearing fault using empirical wavelet transform and S transform methods," in *Proceedings of the 20 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, pp. 446–453, Algeria, UK, 2020.

- [17] 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.
- [18] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors Journal*, vol. 17, no. 17, pp. 5581–5588, 2017.
- [19] L. Shuangli and L. Zengli, "Application of improved wavelet packet energy entropy and GA-SVM in rolling bearing fault diagnosis," in *Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–5, Qingdao, China, 2018.
- [20] S. Jing, J. Yuan, X. Li, and J. Leng, "Weak fault feature identification for rolling bearing based on EMD and spectral kurtosis method," in *Proceedings of the 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 235–239, Changchun, China, 2018.
- [21] T. Xiaodong, "Fault prognosis feature extraction and selection for bearings based on statistical indicator optimization," in *Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, pp. 1–6, Qingdao, China, 2019.
- [22] Z. Qing-feng, G. Shao-wei, C. Zong-xiang, and G. Lu-sheng, "Motor bearing faults diagnosis using modified empirical mode decomposition and bi-spectrum," in *Proceedings of the 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 2029–2032, Wuhan, China, 2018.
- [23] O. R. Seryasat, M. Aliyari shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing using FFT, wavelet energy entropy mean and root mean square (RMS)," in *Proceedings of the 2010 IEEE International Conference on Systems, Man and Cybernetics*, pp. 4295–4299, Istanbul, China, 2010.
- [24] B. Chouri, M. E. Aroussi, M. Tabaa, A. Jarrou, M. Fabrice, and A. Dandache, "Bearing fault diagnosis based on Alpha-stable distribution feature extraction and wSVM classifier," in *Proceedings of the 2015 27th International Conference on Microelectronics (ICM)*, pp. 277–280, Casablanca, UK, 2015.
- [25] P. Liu, H. Li, and P. Ye, "A method for rolling bearing fault diagnosis based on sensitive feature selection and nonlinear feature fusion," in *Proceedings of the 2015 8th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, pp. 30–35, Nanchang, China, 2015.
- [26] C. Liu, X. Wu, and T. Liu, "Sparse feature extraction based on sparse representation and dictionary learning for rolling bearing fault diagnosis," in *Proceedings of the 2017 International Conference on Applied System Innovation (ICASI)*, pp. 1733–1735, Sapporo, UK, 2017.
- [27] C. Chen, Z. Li, J. Yang, and B. Liang, "A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis," in *Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC)*, pp. 5622–5626, Chongqing, China, 2017.
- [28] Y. Sun, H. Gao, H. Song, X. Hong, and Q. Liu, "Bearing feature extraction and fault diagnosis algorithm based on convolutional neural networks," in *Proceedings of the 2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pp. 780–784, Chongqing, China, 2018.
- [29] N. BR and P. Geethanjali, "Feature extraction for bearing fault diagnosis in noisy environment. A study," in *Proceedings of the 2019 Innovations in Power and Advanced Computing Technologies (I-PACT)*, pp. 1–5, Vellore, India, 2019.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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. 10, no. 9, pp. 2249–2252, 2017.
- [34] D. Chen, S. Yang, and F. Zhou, "Incipient fault diagnosis based on DNN with transfer learning," in *Proceedings of the 2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 303–308, Hangzhou, China, 2018.
- [35] W. Qian, S. Li, and J. Wang, "A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions," *IEEE Access*, vol. 6, pp. 69907–69917, 2018.
- [36] Y. Zeng, X. Wu, and J. Chen, "bearing fault diagnosis with denoising autoencoders in few labeled sample case," in *Proceedings of the 2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*, pp. 349–353, Xiamen, China, 2020.
- [37] J. Zhou, L.-Y. Zheng, Y. Wang, and C. Gogu, "A multistage deep transfer learning method for machinery fault diagnostics across diverse working conditions and devices," *IEEE Access*, vol. 8, pp. 80879–80898, 2020.
- [38] J. Shao, Z. Huang, and J. Zhu, "Transfer learning method based on adversarial domain adaption for bearing fault diagnosis," *IEEE Access*, vol. 8, pp. 119421–119430, 2020.
- [39] D. K. Alves, F. B. Costa, R. L. D. A. Ribeiro et al., "Real-time power measurement using the maximal overlap discrete wavelet-packet transform," *IEEE Transactions on Industrial Electronics*, vol. 1, no. 99, 2016.
- [40] S. Wan and B. Peng, "The FERgram: a rolling bearing compound fault diagnosis based on maximal overlap discrete wavelet packet transform and fault energy ratio," *Journal of Mechanical Science and Technology*, vol. 33, no. 1, pp. 157–172, 2019.
- [41] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," 2007.
- [42] 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.
- [43] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pp. 677–682, Chicago, IL, USA, July 2008.
- [44] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [45] X. Han, Q. Liu, J. Xu et al., "Face recognition using pearson correlation and hog with single training image per person," in *Proceedings of the 2018 Chinese Automation Congress (CAC)*, Chicago, IL, USA, 2018.
- [46] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

- [47] R. J. G. B. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 833–841, 2007.
- [48] C.W.R. University, 2014, <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [49] 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.
- [50] C. Shen, Y. Qi, J. Wang, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, 2018.