

Research Article

Fault Diagnosis Method for Rolling Bearing Based on Sparse Principal Subspace Discriminant Analysis

Hongdi Zhou,^{1,2} Lin Zhu,^{1,2} and Xixing Li ^{1,2}

¹School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China

²Key Laboratory of Modern Manufacturing Quality Engineering in Hubei Province, Wuhan 430068, China

Correspondence should be addressed to Xixing Li; li_xi_xing@126.com

Received 25 December 2021; Accepted 23 March 2022; Published 13 April 2022

Academic Editor: Wahyu Caesarendra

Copyright © 2022 Hongdi Zhou 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.

Rolling bearings are omnipresent parts in industrial fields. To comprehensively reflect the status of rolling bearing and improve the classification accuracy, fusion information is widely used in various studies, which may result in high dimensionality, redundancy information of dataset, and time consumption. Thus, it is of crucial significance in extracting optimal features from high-dimensional and redundant feature space for classification. In this study, a fault diagnosis of rolling bearings model based on sparse principal subspace discriminant analysis is proposed. It extracts sparse discrimination information, meanwhile preserving the main energy of original dataset, and the sparse regularization term and sparse error term constrained by $l_{2,1}$ -norm are introduced to improve the performance of feature extraction and the robustness to noise and outliers. The multi-domain feature space involved a time domain, frequency domain, and time-frequency domain is first derived from the original vibration signals. Then, the intrinsic geometric features extracted by sparse principal subspace discriminant analysis are fed into a support vector machine classifier to recognize different operating conditions of bearings. The experimental results demonstrated that the feasibility and effectiveness of the proposed fault diagnosis model based on a sparse principal subspace discriminant analysis algorithm can achieve higher recognition accuracy than fisher discriminant analysis and its extensions, and it is relatively insensitive to the impact of noise and outliers owing to the sparse property.

1. Introduction

Rolling element bearings play an indispensable role in rotating machinery, which are inclined to damage for they are often working under awful conditions, such as high temperature, high torque, and high rotating speed, and almost 45–55% of rotating machinery failures are bearing faults [1–3]. Thus, accurate and efficient diagnosis of incipient bearing faults is of great significance to ensure the safety of the mechanical system, since unexpected failures may cause huge economic losses and even lead to catastrophic casualties [2, 4]. Hitherto, various vibration signal-based methodologies for bearing faults diagnosis are emerging because vibration signals can be easily obtained and offer more useful information than other types of signals [5–7]. Generally, these methods are composed of three stages, namely, signal collection and preprocessing, feature extraction, and condition recognition, while the crucial

procedure is feature extraction, which determines the accuracy of fault identification, and how to extract reliable and sensitive features from vibration signals is a research focus [8].

To comprehensively reflect the machine status, the fusion information including time domain, frequency domain, and time-frequency domain is utilized to recognize the fault type [9–11]. Although information fusion is reported to be beneficial to improve the classification accuracy of fault diagnosis, it may result in a curse of dimensionality and information redundancy that would significantly increase the computational cost of subsequent classifiers or even decrease the performance [12]. Fortunately, dimension reduction is an effective feature extraction method, and various dimension reduction methods for fault diagnosis have been developed in the past few decades. The most well known is the principal component analysis (PCA) [13], which aims to preserve the main energy of the original

feature set and maximize the second-order statistics of original features [3]. For its superior capability of maintaining global data structure, the PCA is widely used as a data preprocessing and feature extraction technique for face recognition, computer vision, and fault detection [14–17]. Despite some advantages, PCA is not suitable for manifold-structure data because it is a linear method that only considers the global Euclidean structure of samples. Thus, some manifold-based dimension reduction methods such as locality preserving projection (LPP) [18], neighborhood preserving embedding (NPE) [19], and sparsity preserving projections (SPP) [20] are presented to deal with dimension reduction related to manifold-structure samples, with a view to preserve geometric structures of original dataset in the subspace. Though these methods are effective in feature extraction, they are unsupervised that may be unsuitable for classification issues since the features are extracted without discrimination information [21].

Fisher discriminant analysis (FDA), as a classical supervised subspace discriminant analysis (SDA) method, is widely used in fault diagnosis, which extracts discriminant features by finding an optimal projection direction with the purpose of simultaneously maximizing the between-class scatter and minimizing the within-class scatter [22, 23]. However, the classification results of FDA can be easily affected by outliers and noise since it minimizes the sum of squared errors. Moreover, FDA tends to obtain undesired results for multi-classification tasks, where an overlap of latent features of different attributes may be obtained [23, 24]. To address these issues, various extensions of FDA are investigated to enhance the performance and efficiency. Sugiyama et al. proposed a local FDA (LFDA) to deal with multimodal labeled data by integrating LPP and FDA [25]. Chen and Hao exploited PCA-LDA (short for PSDA) for radio frequency fingerprint feature dimension reduction, and simulation results showed the superiority of PSDA in a high signal-noise ratio [26]. Jin et al. introduced the trace ratio linear discriminant analysis (TRLDA) for motor bearing fault diagnosis [10]. Jiang et al. designed a rolling bearing fault diagnosis model based on marginal fisher analysis (MFA), which is an extension of FDA [27]. Gao et al. presented a novel feature extraction and dimension reduction method called joint global and local structure discriminant analysis (JGLDA) that integrated the local intrinsic structure into FDA [28]. Feng et al. devised a kernel joint Fisher discriminant analysis (KJFDA) method for fault diagnosis, in which both the local and global discriminant information are extracted [29]. Van and Kang put forward a bearing fault diagnosis method based on wavelet kernel local Fisher discriminant analysis (WKLFDA), and particle swarm optimization (PSO) was applied to optimize the parameters of WKLFDA [23]. Zhong et al. use the sparse kernel local Fisher discriminant analysis (SLFDA) for fault diagnosis of diesel engine working process [30]. Although the FDA-based methods mentioned above all can improve the diagnostic performance to overcome the unimodal weakness of classical FDA, they are sensitive to outliers and noise since the scatter matrixes are calculated by l_2 -norm, which is likely to magnify the impact of outliers as distance

criterion of the objective function [24]. Meanwhile, the l_2 -norm regularization depends on Gaussian noise, which may complicate the precise estimation of two scattering matrices. Compared to l_2 -norm regularization, the l_1 -norm regularization can reduce the influence of outliers since it accumulates the absolute value of elements, which would decrease the deviation degree of dataset with large errors. Recently, various techniques related to l_1 -norm have been reported to be more robust than l_2 -norm-related methods [31–33]. Wang et al. reported the FDA with l_1 -norm method for image recognition, and the experimental results proved its robustness to outliers [31]. Zhang et al. presented a novel feature extraction method called l_1 -norm-based global optimal locality preserving LDA (GLDA_L1), which utilized the FDA and LPP to integrate both global and local structure information via a unified l_1 -norm optimization framework [32]. Wang et al. devised the sparse LFDA for facial expression recognition based on LFDA and l_1 -norm [34]. However, the classification performance of these methods related on the l_1 -norm may decline since each projection vector requires to be iteratively solved. Meanwhile, the essence of l_1 -norm is a derivation of absolute value that may result in some defects in robust feature selection, because they cannot comprehensively cover the intuitive difference across features [24]. To address this issue, some subspace learning techniques are proposed, such as sparse discriminant analysis (SDA) [35], sparsity regularization discriminant projection (SRDP) [36], and robust sparse linear discriminant analysis (RSLDA) [21], in which the sparse constraint is adopted to select sensitive features and remove the redundancy of dataset. Here, the $l_{2,1}$ -norm regularization based on the sparse technique is adopted to constraint the error function and discriminant matrix, which can adaptively select the optimal mapping direction. The $l_{2,1}$ -norm-related algorithm can improve the recognition performance in the presence of outliers, because the influence on the residual of the objective function constrained by the $l_{2,1}$ -norm is less than the squared residual constrained by the l_2 -norm.

Inspired by the subspace discriminant analysis and the $l_{2,1}$ -norm-related regularization methods, this study presented a rolling bearing fault diagnosis model based on sparse principal subspace discriminant analysis (SPSDA) by integrating PCA, LDA, and sparse constraint. The proposed method can simultaneously extract discrimination information and preserve the main energy of original dataset with respect to the number of projection directions. The projection matrix of PSDA is constrained by $l_{2,1}$ -norm, which can improve the performance of feature extraction owing to the row-sparsity property. A sparse error term adapted noise during feature extraction is introduced to improve the robustness to noise and outliers. Experimental investigations are carried out to demonstrate the feasibility and effectiveness of the proposed method for rolling bearing fault diagnosis.

The rest of the article is organized as follows: in Section 2, the principle of FDA and PSDA is introduced, and an improved PSDA method based on sparse technique is discussed in detail. Then, the fault diagnosis model based on the

SPSDA algorithm is presented in Section 3. After that, the practical cases are studied to validate the superior performance of the proposed model in Section 4. Finally, some concluding remarks are summarized in Section 5.

2. Principle of the Proposed Method

2.1. Brief Review of FDA. Given a set of C -class training sample patterns $\mathbf{x}_j^i \in R^N$ ($i = 1, 2, \dots, C; j = 1, 2, \dots, n_i$), where n_i is the number of samples with i th class, $N = \sum_{i=1}^C n_i$ is the total number of all samples, and \mathbf{x}_j^i is the j th training sample of i th class. FDA tries to seek an optimal mapping matrix to project high-dimensional original feature space into the low-dimensional feature space, with the purpose of simultaneously separating the different attribute samples and gathering the same attribute samples. FDA acquires the optimal projection matrix by the following equation based on Fisher's discriminant problem [19]:

$$\mathbf{W} = \arg \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}}, \quad (1)$$

where \mathbf{S}_b and \mathbf{S}_w are the between-class scattering matrix and within-class scattering matrix, respectively, which are defined as follows:

$$\begin{aligned} \mathbf{S}_b &= \frac{1}{N} \sum_{i=1}^C n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \\ \mathbf{S}_w &= \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{x}_j^i - \boldsymbol{\mu}_i)(\mathbf{x}_j^i - \boldsymbol{\mu}_i)^T, \end{aligned} \quad (2)$$

where $\boldsymbol{\mu}_i = 1/n_i \sum_{j=1}^{n_i} \mathbf{x}_j^i$ is the mean matrix of the i th class and $\boldsymbol{\mu} = 1/N \sum_{i=1}^C \sum_{j=1}^{n_i} \mathbf{x}_j^i$ denotes the mean feature of total samples. Generally speaking, the equation (1) is equal to the following optimal problem [21]:

$$\mathbf{W} = \arg \min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \text{Tr}(\mathbf{W}^T (\mathbf{S}_w - \lambda \mathbf{S}_b) \mathbf{W}), \quad (3)$$

where the global solution to the optimization (1) can be obtained by the eigenvectors corresponding to the first d maximum nonzero eigenvalues of the generalized problem $\mathbf{S}_b \mathbf{W} = \lambda \mathbf{S}_w \mathbf{W}$, where \mathbf{S}_w is nonsingular and λ is a small positive constant. Note that the definition of \mathbf{S}_b is based on l_2 -norm, which may result in the sensitivity of outliers and noise since the error of distance increases in square form.

2.2. Introduction of SPSDA. Despite FDA being a widely used SDA method in classification, its performance may decline when encounters multimode or nonGaussian data that is common in an actual industrial environment. Meanwhile, preremoval of redundancy and coupling information between features can improve the recognition performance. Therefore, PSDA integrated PCA and FDA is presented to overcome the shortcomings [26], which utilizes a linear transformation to preproject the original features into latent features. For a given original matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, the preprojection of latent features can be obtained

by $\mathbf{S} = \mathbf{P}\mathbf{X}$ and $\mathbf{P}^T \mathbf{P} = \mathbf{I}$, where \mathbf{P} and \mathbf{S} are the loading matrix and latent features, respectively. Here, \mathbf{S} is calculated by eigen decomposition to remove redundant information. Subsequently, the between-class distance \mathbf{S}'_b and within-class distance \mathbf{S}'_w of latent feature space can be redefined as follows:

$$\mathbf{S}'_b = \frac{1}{N} \sum_{i=1}^C n_i (\boldsymbol{\mu}'_i - \boldsymbol{\mu}')(\boldsymbol{\mu}'_i - \boldsymbol{\mu}')^T, \quad (4)$$

$$\mathbf{S}'_w = \frac{1}{N} \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{p}_i \mathbf{x}_j^i - \boldsymbol{\mu}'_i)(\mathbf{p}_i \mathbf{x}_j^i - \boldsymbol{\mu}'_i)^T.$$

Here, the $\boldsymbol{\mu}'$ is the mean matrix of latent features and the $\boldsymbol{\mu}'_i$ is the mean matrix of the i th class of latent features. Thus, the objective function of PSDA can be described as follows:

$$\mathbf{W}' = \arg \min_{\mathbf{W}'^T \mathbf{W}' = \mathbf{I}} \text{Tr}(\mathbf{W}'^T (\mathbf{S}'_w - \lambda \mathbf{S}'_b) \mathbf{W}'). \quad (5)$$

However, PSDA acquires optimal discriminant projection by using principal subspace latent features, which may make the centre of original scattering matrix easily deviate from the original model since the measured signals are usually polluted by noise and outliers under the complex industrial environment. To surmount the limitations of PSDA with vast quantities of noise and outlier corruptions, a sparse principal subspace discriminant analysis algorithm (SPSDA) based on $l_{2,1}$ -norm regularization is presented in this section. Compared with l_2 -norm and l_1 -norm regularization, $l_{2,1}$ -norm regularization has a good sparse property, which can effectively suppress the influence of outliers and noise and make the learned features have better interpretability. Thus, a new objective function of SPSDA with $l_{2,1}$ -norm regularization is defined as follows:

$$\mathbf{W} = \arg \min_{\mathbf{W}} \text{Tr}(\mathbf{W}^T (\mathbf{S}'_w - \lambda \mathbf{S}'_b) \mathbf{W}) + \beta \|\mathbf{W}\|_{2,1}, \quad (6)$$

where \mathbf{W} is the principal discriminant projection matrix, and \mathbf{S}'_b and \mathbf{S}'_w are the between-class matrices and within-class matrices, respectively. λ is a small positive constant that is utilized to balance the importance of \mathbf{S}'_b and \mathbf{S}'_w , and β is a trade-off parameter determining the importance of the corresponding term. $\|\cdot\|_{2,1}$ is the $l_{2,1}$ -norm that is defined as $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^m w_{i,j}^2}$.

Considering the reconstruction relationship between extracted features and original features, the objective function of SPSDA can be rewritten as follows:

$$\mathbf{W} = \arg \min_{\mathbf{W}, \mathbf{Q}, \mathbf{E}} \text{Tr}(\mathbf{W}^T (\mathbf{S}'_w - \lambda \mathbf{S}'_b) \mathbf{W}) + \beta_1 \|\mathbf{W}\|_{2,1} + \beta_2 \|\mathbf{E}\|_{2,1},$$

$$\text{s.t. } \mathbf{S} = \mathbf{Q}\mathbf{W}^T \mathbf{S} + \mathbf{E} = \mathbf{Q}\mathbf{W}^T \mathbf{S} + \mathbf{E}, \mathbf{Q}^T \mathbf{Q} = \mathbf{I}, \quad (7)$$

where \mathbf{E} denotes the sparse outliers and random noise and β_1 and β_2 are the trade-off parameter determining the importance of the corresponding term. Finally, the alternating direction method of multipliers (ADMM) [37] is adopted to solve the optimization problem (7).

3. Process of Fault Diagnosis by SPSDA

High-dimensional heterogeneous feature set should be constructed before the implementation of feature extraction by using the proposed SPSDA. Once faults such as rubbing and loosening appear in equipment, both the amplitude and distribution of the original vibration signals differ from those of the state of being healthy, and the frequency spectrum and distribution vary accordingly. Time-domain characteristics are sensitive to the incipient failures, and frequency-domain features can reveal the instantaneous cyclical component of the high frequency, while time-frequency features can reflect the frequency components and the time-varying characteristic of nonstationary signals. Thus, 11 statistical characteristics related to time-domain signals and 13 statistical characteristics associated with frequency-domain spectrum are selected [38], which is listed in Table 1. Among them, $x(n)$ is the original signal series, N is the total number of the data points in a single signal sample, and $s(n)$ and $f(n)$ are the normalized power spectral density (PSD) of $x(n)$ and the corresponding frequency of $s(n)$, respectively.

In addition, the time-frequency domain features associated with wavelet packet decomposition (WPD) and empirical model decomposition (EMD) are considered to adequately represent the different types of bearing faults. Thus, 32 relative energy and energy spectrum entropy features related to 4-level db2 WPD coefficients are also employed to discriminate different types of faulty signals since WPD can meticulously analyse nonstationary signals, which can adequately describe the energy distribution in time-frequency domain [39]. Similarly, 6 EMD relative energy features and 6 EMD energy entropy features corresponding to the first 6 intrinsic mode functions (IMF) are also adopted. The relative energy and energy spectrum entropy in time-frequency domain are defined as follows:

$$\text{RE}(i) = \frac{\sum_{j=1}^K (C_i^j)^2}{\sum_{m=1}^N \sum_{j=1}^K (C_m^j)^2}, \quad (8)$$

$$\text{En}(i) = -\sum_{j=1}^K p_i^j \log_2(p_i^j),$$

where $p_i^j = (C_i^j)^2 / \sum_{j=1}^K (C_i^j)^2$, C_i^j is the j th coefficients of the i th wavelet packet node or IMF, N is the total number of wavelet packet nodes or IMFs, and K is the total number of coefficients in each wavelet packet node or IMF. $\text{RE}(i)$ and $\text{En}(i)$ denote the relative energy and energy spectrum entropy features of the i th level wavelet decomposition or IMF. Therefore, a high-dimensional feature set containing 68 features is constructed to describe the state of rolling bearing.

After the construction of high-dimensional heterogeneous feature set, feature extraction should be conducted to remove redundant information and improve diagnosis accuracy and diagnosis efficiency. Since the dimension reduction methods based on subspace discriminant analysis can effectively extract the representative features, a rolling

bearing fault diagnosis model based on the SPSDA by integrating PCA, LDA, and sparse constraint is proposed in this study, and the flowchart of proposed fault diagnosis method is shown in Figure 1. The high-dimensional heterogeneous features of training samples and testing samples are first constructed from original vibration signals. The SPSDA is subsequently employed in the feature extraction from the heterogeneous feature set of vibration signals under the circumstances of massive samples. Finally, the obtained low-dimensional features are fed into the support vector machine (SVM) to recognize the running state of rolling bearing. Thus, we can put forward the corresponding decisions or control measures to determine the type of failures by the classification results.

4. Experimental and Results

4.1. Experimental Configuration and Vibration Signals Collection. In order to evaluate the effectiveness of the proposed algorithm, an experimental study on fault diagnosis of rolling bearings was carried out. The vibration measuring system is mainly composed of mechanical system and hardware of electrical system, and the hardware structure chart of electrical system and the actual vibration measurement system are shown in Figures 2(a) and 2(b), respectively. The 63/28-2RZ series deep groove ball bearings were utilized as tested bearings, which were delivered by the automatic machinery system that was composed of the preset mechanism, the measuring mechanism, the sorting mechanism, and the feeding mechanism [39]. The geometric parameters of tested bearings were listed in Table 2. The piezoelectric acceleration sensor (YD-1) was mounted on the top of tested bearings to collect the single-point radial vibration signals, which would be amplified by a charge amplifier (DHF-2). The charge sensitivity and frequency response of acceleration sensor were $6\text{--}10 \text{ pC/ms}^{-2}$ and $1\text{--}10 \text{ kHz} \pm 1 \text{ dB}$, respectively, and the frequency range of the amplifier was $0.3 \text{ Hz--}100 \text{ kHz}$. Subsequently, the vibration signals were transformed into voltage signals by an A/D converter (PCI-9114) and imported into a computer for further processing. The rotational speed of the driving motor and sampling frequency were 1500 rpm and 25 kHz, respectively. A radial load of 0 kN and axial load of 1.0 kN were loaded onto the shaft and bearings by the cylinder.

Four different operating conditions including inner race fault, outer race fault, ball fault, and normal condition were introduced. The scratch defects of the bearings were machined by electric engraving pen, and the width of the scratch defects of the inner race, outer race, and ball was $65 \pm 22 \mu\text{m}$, $70 \pm 20 \mu\text{m}$, and $70 \pm 20 \mu\text{m}$, respectively, and the depths of the scratch defects were $0.2 \pm 0.05 \text{ mm}$. Therefore, the characteristic bearing defect frequencies can be calculated according to the kinematic parameters and the rotational speed, and the characteristic bearing defect frequencies of the inner race, outer race, and ball are 121.75 Hz, 78.25 Hz, and 55 Hz, respectively. Figure 3 shows the normalized time-series vibration signals in conjunction with frequency spectrums of those four working conditions. Obviously, there are certain differences among those signals,

TABLE 1: The time-domain and frequency-domain feature parameters.

Time-domain features	Frequency-domain features
$t_1 = \sum_{n=1}^N x(n)/N$	$f_1 = \sum_{n=1}^N s(n)/N$
$t_2 = \sqrt{\sum_{n=1}^N (x(n) - t_1)^2 / N - 1}$	$f_2 = \sum_{n=1}^N (s(n) - f_1)^2 / N - 1$
$t_3 = (\sqrt{\sum_{n=1}^N (x(n))^2} / N)$	$f_3 = \sum_{n=1}^N (s(n) - f_1)^3 / N (\sqrt{f_2})^3$
$t_4 = (\sum_{n=1}^N \sqrt{ x(n) } / N)^2$	$f_4 = \sum_{n=1}^N (s(n) - f_1)^4 / N (f_2)^2$
$t_5 = \max(x(n)) - \min(x(n)) / 2$	$f_5 = \sum_{n=1}^N s(n)f(n) / \sum_{n=1}^N s(n)$
$t_6 = \sum_{n=1}^N (x(n) - t_1)^3 / (N - 1)t_2^3$	$f_6 = \sqrt{\sum_{n=1}^N (f(n) - f_5)^2 f(n) / N}$
$t_7 = \sum_{n=1}^N (x(n) - t_1)^4 / (N - 1)t_2^4$	$f_7 = \sqrt{\sum_{n=1}^N f(n)^2 s(n) / \sum_{n=1}^N s(n)}$
$t_8 = \max x(n) / t_3$	$f_8 = \sqrt{\sum_{n=1}^N f(n)^4 s(n) / \sum_{n=1}^N f(n)^2 s(n)}$
$t_9 = \max x(n) / t_4$	$f_9 = \sqrt{\sum_{n=1}^N f(n)^2 s(n) / \sum_{n=1}^N s(n) \sum_{n=1}^N f(n)^4 s(n)}$
$t_{10} = Nt_3 / \sum_{n=1}^N x(n) $	$f_{10} = f_6 / f_5$
$t_{11} = N \max x(n) / \sum_{n=1}^N x(n) $	$f_{11} = \sum_{n=1}^N (f(n) - f_5)^3 s(n) / N f_6^3$
	$f_{12} = \sum_{n=1}^N (f(n) - f_5)^4 s(n) / N f_6^4$
	$f_{13} = \sum_{n=1}^N \sqrt{(f(n) - f_5)s(n)} / N \sqrt{f_6}$

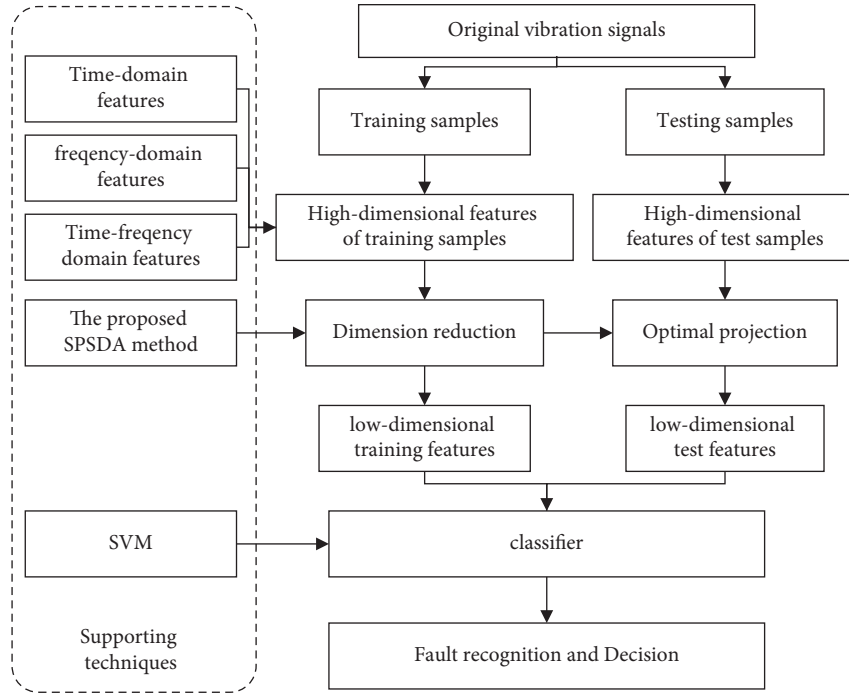


FIGURE 1: Implementation process of the proposed fault diagnosis method.

such as the signals of normal status are almost white noise, the signals of inner race fault and outer race fault are characterized by period impulse, and the signals of ball fault are submerged into white noise. It is difficult to distinguish different faults only from the vibration signals and the frequency spectrums because of the influences of the noise. About 100 bearings of each status were tested, and the radial vibration signals under those four operating conditions are collected as samples. Therefore, 400 datasets can be acquired, and each sample is collected for one second. Then, 50 samples for each status are randomly selected as the training

dataset and the remaining samples as the test dataset to evaluate the classification accuracy.

4.2. Dimensionality Reduction Performance of the Proposed Method. In order to intuitively authenticate the performance of the proposed SPSDA method, the following comparative experiments with other existing dimension reduction methods including PCA, FDA, PSDA, and MFA on the measured vibration signals were conducted by the above bearing failure simulation test rig. The compared dimension reduction

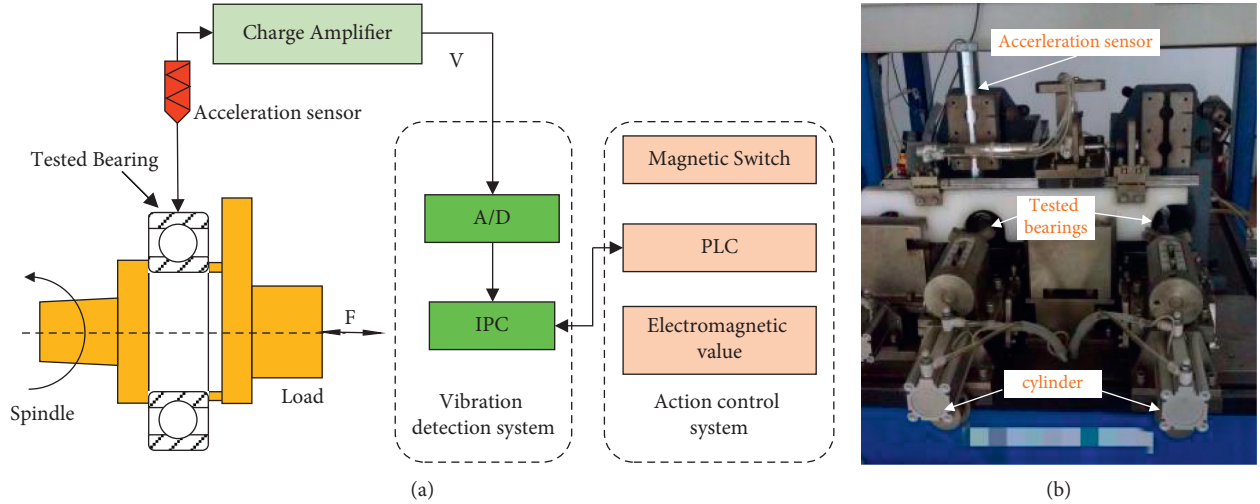


FIGURE 2: Test rig. (a) The hardware structure chart of electrical system and (b) the actual experimental system.

TABLE 2: The geometric parameters of 63/28-2RZ.

Items	Inner ring diameter D_i (mm)	Outer ring diameter D_o (mm)	Ball diameter D_b (mm)	Pitch diameter D_m (mm)	Number of balls Z	Contact angle α (rad)
Value	28	68	11.509	48.5	8	0.274

methods include an unsupervised method and three supervised methods. Specifically, PCA is adopted as the compared unsupervised method, which extracts latent features without considering class information. The supervised methods consist of FDA, PSDA, and MFA, which can exploit the labeled information into the objective functions related to l_2 -norm that projects original feature space into optimization low-dimensional subspace. After the high-dimensional feature dataset is constructed, it would be projected into the embedded subspace by the corresponding projection matrix of different dimension reduction methods, which can effectively eliminate redundant information and extract low-dimensional features. The target dimensionality of the embedded subspace for each technique is set to a certain number so that the cumulative variance contribution rate is more than 95%. Figures 4–8 display the plots of the first three principal components of their projection results for visualization, where Figures 4(a), 5(a), 6(a), 7(a), and 8(a) represent the training results, Figures 4(b), 5(b), 6(b), 7(b), and 8(b) represent the testing results, and the hollow dots and the solid dots denote the training samples and the testing samples, respectively. Obviously, since PCA is a linear and unsupervised method, the PCA has certain limitations in the extraction of low-dimensional sensitive features of rolling bearings, and the extracted features of PCA are mostly overlapped as shown in Figure 4. Compared to the unsupervised method, FDA and its extensions MFA and PSDA can extract satisfactory discriminant features in the embedded subspace. As seen in Figures 5–7, most of the low-dimensional features of FDA and its variants can be distinguished, and only a small number of extracted features are confused. From Figure 5, most of the low-dimensional features dataset of FDA can be distinguished and partial samples are overlapped. As seen in Figure 6, the inner race fault in the low-dimensional features dataset of

PSDA can be recognized, while the normal condition, ball fault, and outer race fault samples are overlapped. From the clustering results of MFA as shown in Figure 7, most of the low-dimensional features of MFA can be identified and only several data points are overlapped. From the clustering results of the proposed SPSDA as shown in Figure 8, the low-dimensional feature dataset of SPSDA separates from each other in 3-D space and only a few data points are mixed. The experimental results indicated that the low-dimensional reduced features obtained by SPSDA in the embedded subspace have obvious discriminant characteristics than other methods.

Within the fault diagnosis related to pattern recognition in conjunction with feature extraction techniques that find low-dimensional sensitive features for original signals, classifiers should be adopted to recognize the type of bearing faults. SVM was exploited for classification in this study for its well-developed statistical learning theory. Then, 50 data for each status including inner race fault, outer race fault, ball fault, and normal condition were randomly selected for SVM training, and the remaining samples were employed for testing. To evaluate the effectiveness of the proposed SPSDA-SVM method, the failure detection rate of the proposed method was compared with the classification results of SVM, PCA-SVM, FDA-SVM, PSDA-SVM, and MFA-SVM. The quantitative evaluation procedure for different methods was repeated for 10 times, and the recognition results are listed in Table 3. The classification accuracies of original features, PCA, FDA, PSDA, MFA, and SPSDA are 78.4%, 83.7%, 94.5%, 95.6%, 93.55%, and 97.65%, respectively. All accuracies of low-dimensional features are higher than original features, which implies that the recognition accuracies can be improved and the redundancy information can be removed by means of dimension

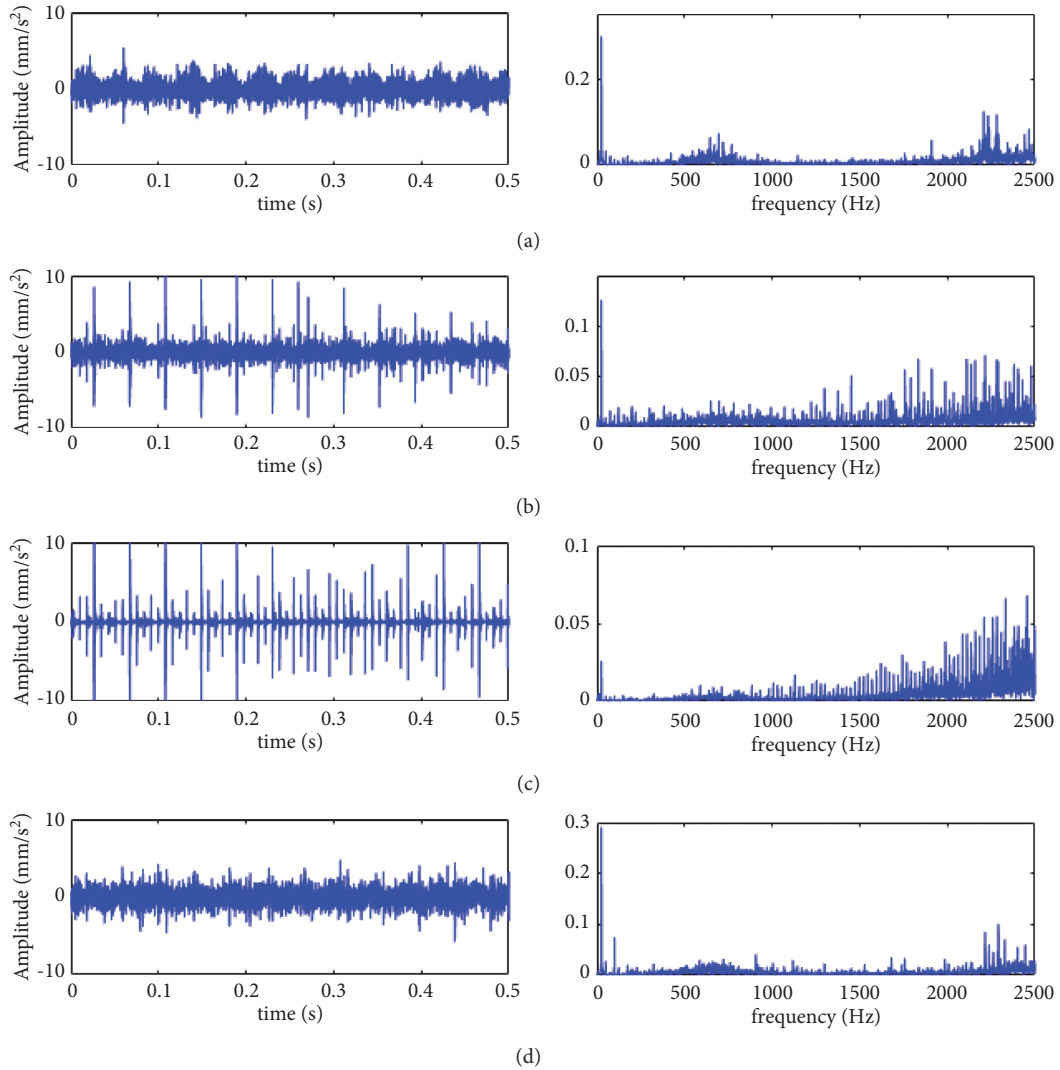


FIGURE 3: The time domain and frequency domain figures of vibration signals for the four bearing conditions. (a) Normal condition, (b) inner race fault, (c) outer race fault, and (d) ball fault.

reduction methods. The classification performance of supervised method is better than that of unsupervised method, since the supervised method can take full advantage of discriminant information. SPSDA performs better than other dimensionality reduction techniques in terms of extracting discriminative features, which can achieve high recognition results. This is primarily due to the fact that the proposed SPSDA method can simultaneously extract discrimination information and preserve the main energy of original dataset with respect to the number of projection directions and considers the row-sparsity property by introducing the $l_{2,1}$ -norm regularization terms, which can make the low-dimensional features more beneficial to the discrimination among different classes.

4.3. Robust Performance Comparison on Noisy Signals. To authenticate the effectiveness and robustness of the proposed SPSDA method on noisy bearing fault datasets, a

series of experiments on the noisy bearing fault datasets were performed. To simulate the real industrial environment, the white noise is added into the original vibration signals to yield noisy data with SNR of 10 dB, 5 dB, and 2 dB, respectively. For each status, each class has 100 instances containing 25000 sampling points, where 50% samples are randomly selected for training and the remaining instances are adopted for testing. The dimensionality of the embedded subspace is set to a certain number, so that the cumulative variance contribution rate is more than 95%. SVM was employed as the base classification in this case, and the quantitative evaluation procedure for different methods was repeated for 10 times to avoid randomness. The average classification results are listed in Table 4. As seen, the classification accuracies of all dimensionality reduction methods decrease with the decrease of SNR, which means that the presence of noises can weaken the performance of these dimension reduction methods in feature extraction. The classification results of

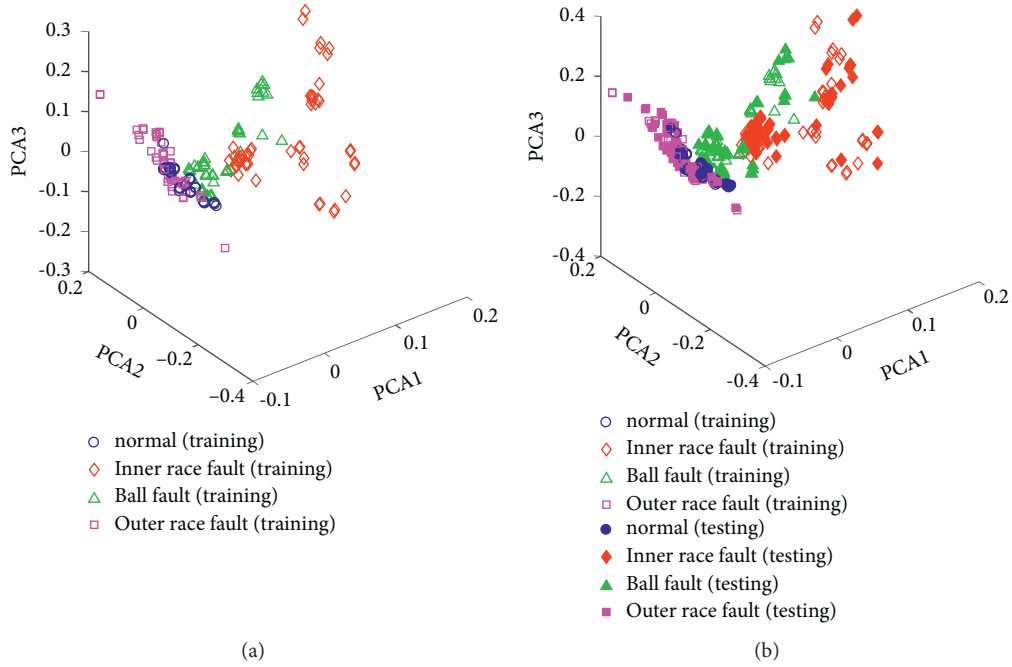


FIGURE 4: Clustering results with PCA. (a) Training results and (b) testing results.

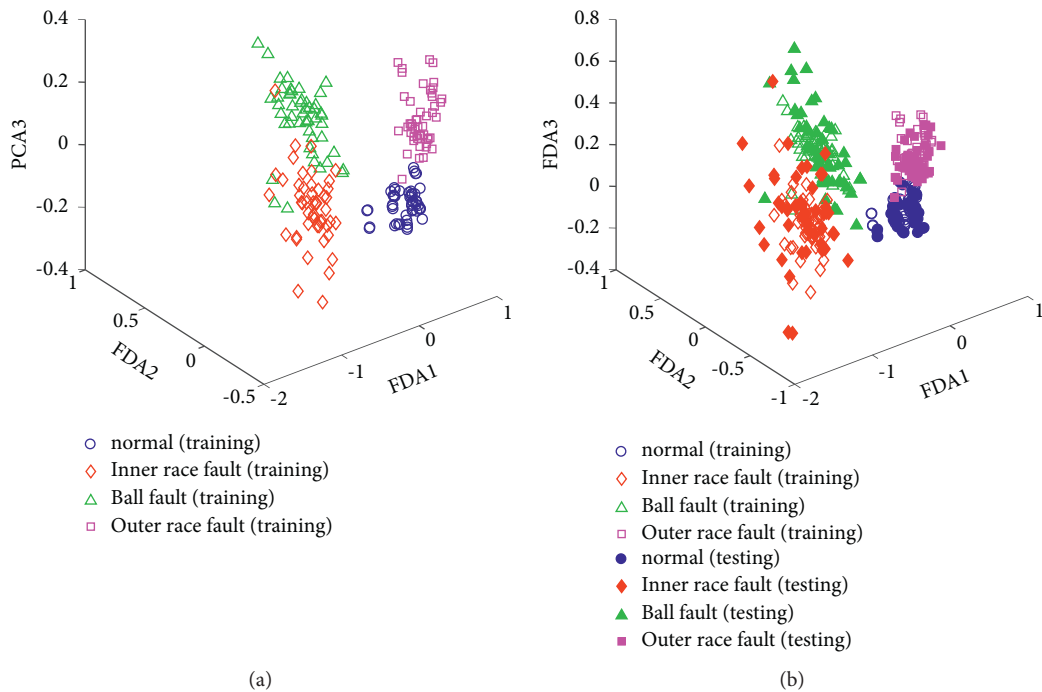


FIGURE 5: Clustering results with FDA (a) Training results and (b) testing results.

original features are lowest for all cases, which implies that effective dimensionality reduction techniques can remove redundancy information, alleviate the impact of noises, and improve identification performance. The proposed SPSDA dimension reduction method can effectively relieve the influence of noises by incorporating sparse regularization terms, and the recognition accuracy of SPSDA features is obviously higher than other peer

algorithms. The sparse regularization term and sparse error term constrained by the $l_{2,1}$ -norm can adapt noise during feature extraction and improve the robustness to noise and outliers. Above all, the proposed SPSDA method has better robustness compared with other peer methods, and SPSDA can effectively extract low-dimensional sensitive features from noisy signals, which can improve the classification performance of rolling bearings.

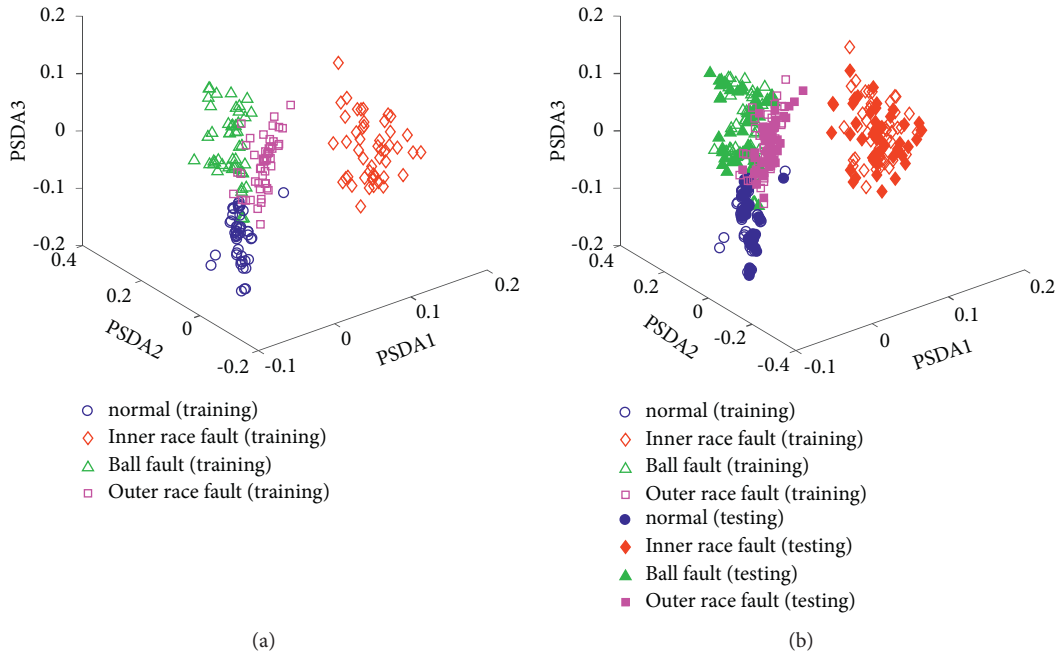


FIGURE 6: Clustering results with PSDA. (a) Training results and (b) testing results.

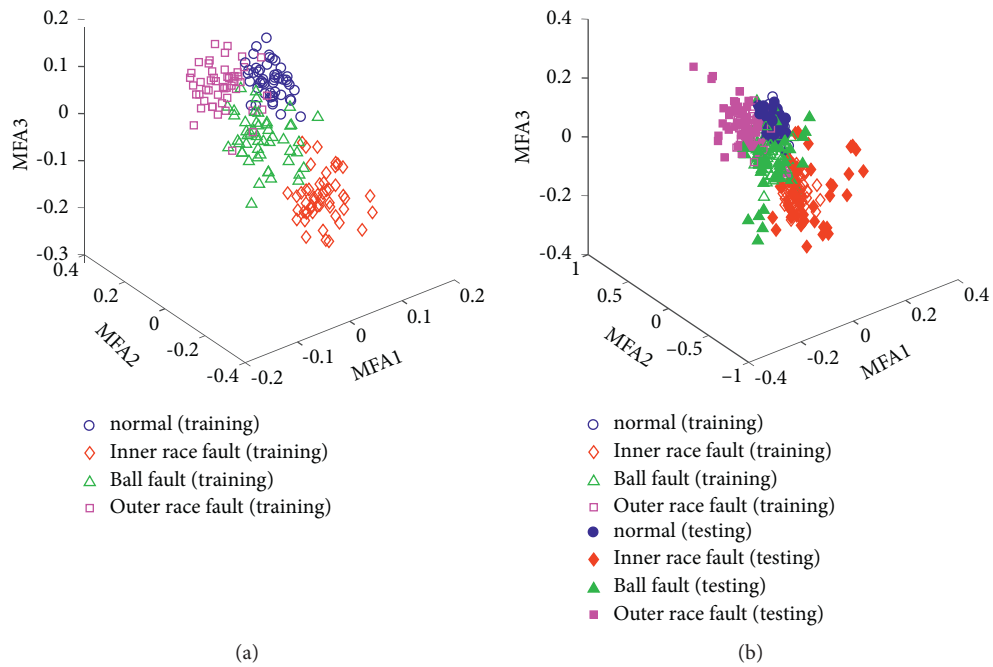


FIGURE 7: Clustering results with MFA. (a) Training results and (b) testing results.

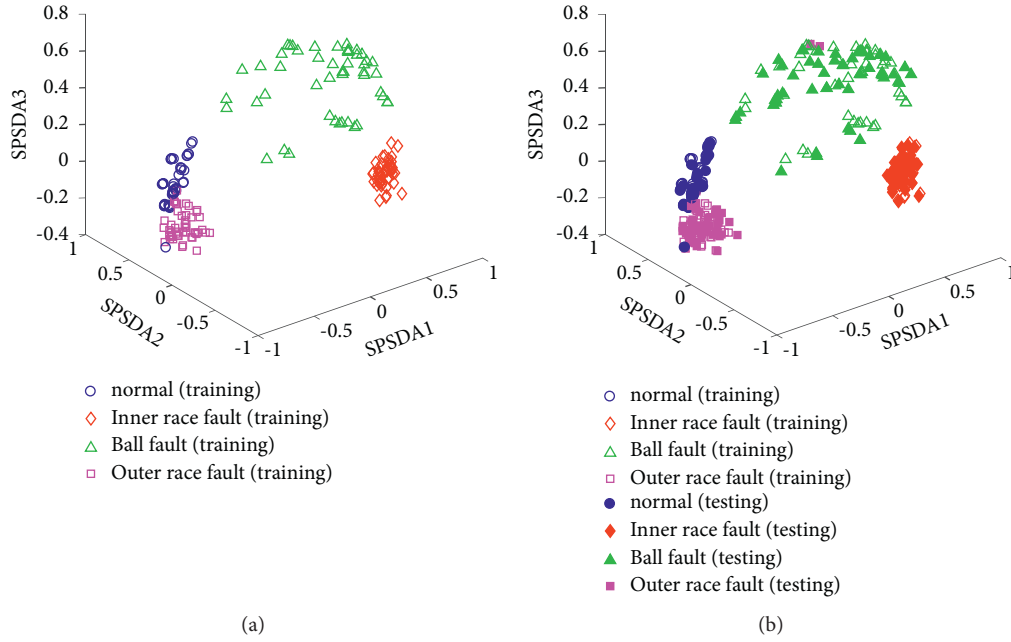


FIGURE 8: Clustering results with the proposed SPSDA. (a) Training results and (b) testing results.

TABLE 3: The recognition accuracies of different methods to the bearing sets with the SVM classifier.

Methods	Normal (%)	Outer race fault (%)	Inner race fault (%)	Ball fault (%)	Average accuracy
Original	77.8	76.8	82.4	76.6	78.4
PCA	81.6	78.6	90.4	84.2	83.7
FDA	94.2	94	95.4	94.4	94.5
PSDA	94.4	94.2	100	93.8	95.6
MFA	88.4	91.8	98.2	95.8	93.55
SPSDA	96.4	94.2	100	100	97.65

TABLE 4: The classification performance of different dimensionality reduction methods on four noisy bearing fault datasets.

Methods	Data	Data (10 dB)	Data (5 dB)	Data (2 dB)	Average
Original	78.4	77.6	74.4	73.4	75.95
PCA	83.7	83.2	82.2	81.6	82.68
FDA	94.5	94.2	93.6	92.8	93.78
PSDA	95.6	94.6	94.2	93.2	94.40
MFA	93.55	92.8	91.5	90.6	92.11
SPSDA	97.65	97.2	96.2	95.6	96.66

5. Conclusions

In order to address the problems of high-dimensionality, strong relevance, and redundancy generated by information fusion in the bearing fault diagnosis domain, a rolling bearing fault diagnosis model based on sparse principal subspace discriminant analysis (SPSDA) is proposed in this study. The proposed method can extract discrimination information, meanwhile preserving the main energy of original dataset and introducing sparse regularization term and sparse error term constrained by $l_{2,1}$ -norm to improve the performance of feature extraction and the robustness to noise and outliers. Firstly, the high-dimensional

heterogeneous feature set involved the time domain, frequency domain, and time-frequency domain is constructed. Subsequently, SPSDA is applied to extract low-dimensional features to remove redundancy information and improve the diagnosis accuracy and diagnosis efficiency under the circumstances of massive samples, because dimension reduction methods based on subspace discriminant analysis can effectively extract the representative features. Finally, the obtained low-dimensional features are fed in SVM to recognize the running state of rolling bearing. Experimental investigations are conducted to validate the superiority and effectiveness of the proposed SPSDA method. Compared with other peer methods, the SPSDA outperformed the PCA, FDA, PSDA, and MFA by yielding improvements of 13.95%, 3.15%, 2.05%, 4.1%, and SPSDA has a better fault diagnosis effect on noised vibration signals, which implies that the proposed method possesses more stable performance than other compared ones and is relatively insensitive to the impact of noises and outliers owing to the sparse property. Next, we are trying to extend our algorithm to recognize different fault severities. The challenge is that the optimal projections of proposed method are obtained only considering global structure features, which may decline the recognition performance. Therefore, locally joint sparse strategies are deserved further investigation.

Data Availability

The data that support the finding of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was sponsored by the National Natural Science Foundation of China (grant no. 52005168), the Natural Science Foundation of Hubei Province (grant no. 2019CFB326), the Scientific Research Foundation for High-level Talents of Hubei University of Technology (grant no. GCRC2020009), and the Technology Innovation Special Project of Hubei Province (grant no. 2019AEE014).

References

- [1] A. Moshrefzadeh, "Condition monitoring and intelligent diagnosis of rolling element bearings under constant/variable load and speed conditions," *Mechanical Systems and Signal Processing*, vol. 149, Article ID 107153, 2021.
- [2] K. Zhao, H. Jiang, Z. Wu, and T. Lu, "A novel transfer learning fault diagnosis method based on Manifold Embedded Distribution Alignment with a little labeled data," *Journal of Intelligent Manufacturing*, vol. 33, no. 1, pp. 151–165, 2020.
- [3] H. D. Zhou, F. Zhong, T. L. Shi, W. Lai, J. Duan, and Y. Zhang, "Class-information-incorporated kernel entropy component analysis with application to bearing fault diagnosis," *Journal of Vibration and Control*, vol. 27, no. 5-6, pp. 543–555, 2021.
- [4] A. Glowacz, R. Tadeusiewicz, S. Legutko et al., "Fault diagnosis of angle grinders and electric impact drills using acoustic signals," *Applied Acoustics*, vol. 179, Article ID 108070, 2021.
- [5] Z. Xu, C. Li, and Y. Yang, "fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks," *Applied Soft Computing*, vol. 95, Article ID 106515, 2020.
- [6] S. Wang, I. Selesnick, G. Cai, Y. Feng, X. Sui, and X. Chen, "Nonconvex sparse regularization and convex optimization for bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 9, pp. 7332–7342, 2018.
- [7] A. Kumar, G. Vashishtha, C. P. Gandhi, Y. Zhou, A. Glowacz, and J. Xi, "Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, Article ID 3510710, 2021.
- [8] S. Xing, Y. Lei, S. Wang, and F. Jia, "Distribution-Invariant deep belief network for intelligent fault diagnosis of machines under new working conditions," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2617–2625, 2021.
- [9] M. Gan, C. Wang, and C. A. Zhu, "Multiple-domain manifold for feature extraction in machinery fault diagnosis," *Measurement*, vol. 75, pp. 76–91, 2015.
- [10] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2441–2451, 2014.
- [11] W. Wang, F. Yuan, and Z. Liu, "Sparsity discriminant preserving projection for machinery fault diagnosis," *Measurement*, vol. 173, Article ID 108488, 2021.
- [12] X. Tao, C. Ren, Q. Li et al., "Bearing defect diagnosis based on semi-supervised kernel Local Fisher Discriminant Analysis using pseudo labels," *ISA Transactions*, vol. 110, pp. 394–412, 2021.
- [13] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.
- [14] Q. Q. Wang, Q. X. Gao, and X. B. Gao, "Norm based PCA for image recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1336–1346, 2018.
- [15] R. Sharma and M. S. Patterh, "A new pose invariant face recognition system using PCA and ANFIS," *Optik*, vol. 126, no. 23, pp. 3483–3487, 2015.
- [16] C. Z. Liu and M. Kavakli, "Extensions of principle component analysis with applications on vision based computing," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10113–10151, 2016.
- [17] H. X. Chen, W. J. Huang, J. M. Huang, C. Cao, L. Yang, and L. Zeng, "Multi-fault condition monitoring of slurry pump with principle component analysis and sequential hypothesis test," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. No. 7, p. 2059019, 2020.
- [18] T. Long, Y. Sun, J. Gao, and B. Yin, "Locality preserving projection based on Euler representation," *Journal of Visual Communication and Image Representation*, vol. 70, Article ID 102796, 2020.
- [19] J. Gui, Z. sun, W. Jia, R. Hu, Y. Lei, and S. Ji, "Discriminant sparse neighborhood preserving embedding for face recognition," *Pattern Recognition*, vol. 45, no. 8, pp. 2884–2893, 2012.
- [20] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.
- [21] J. Wen, X. Fang, J. Cui et al., "Robust sparse linear discriminant analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 390–403, 2019.
- [22] Y. Zhou, S. Yan, Y. Ren, and S. Liu, "Rolling bearing fault diagnosis using transient-extracting transform and linear discriminant analysis," *Measurement*, vol. 178, Article ID 109298, 2021.
- [23] M. Van and H.-J. Kang, "Wavelet kernel local Fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 12, pp. 3588–3600, 2015.
- [24] M.-Q. Zhang and X.-L. Luo, "Novel dynamic enhanced robust principal subspace discriminant analysis for high-dimensional process fault diagnosis with industrial applications," *ISA Transactions*, vol. 114, pp. 1–14, 2021.
- [25] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.
- [26] X. Chen and X. Hao, "Feature reduction method for cognition and classification of IoT devices based on artificial intelligence," *IEEE Access*, vol. 7, pp. 103291–103298, 2019.
- [27] L. Jiang, T. Shi, and J. Xuan, "Fault diagnosis of rolling bearings based on Marginal Fisher analysis," *Journal of Vibration and Control*, vol. 20, no. 3, pp. 470–480, 2014.
- [28] Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint global and local structure discriminant analysis," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 626–635, 2013.
- [29] J. Feng, J. Wang, H. Zhang, and Z. Han, "fault diagnosis method of joint Fisher discriminant analysis based on the local and global manifold learning and its kernel version,"

- IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 122–133, 2016.
- [30] K. Zhong, M. Han, T. Qiu, and B. Han, “fault diagnosis of complex processes using sparse kernel local Fisher discriminant analysis,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 5, pp. 1581–1591, 2020.
 - [31] H. X. Haixian Wang, X. S. Xuesong Lu, and Z. L. Wenming Zheng, “Fisher discriminant analysis with L1-norm,” *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 828–842, 2014.
 - [32] D. Zhang, X. Li, J. He, and M. Du, “A new linear discriminant analysis algorithm based on L1-norm maximization and locality preserving projection,” *Pattern Analysis & Applications*, vol. 21, no. 3, pp. 685–701, 2018.
 - [33] C.-N. Li, Y.-H. Shao, W. Yin, and M.-Z. Liu, “Robust and sparse linear discriminant analysis via an alternating direction method of multipliers,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 3, pp. 915–926, 2020.
 - [34] Z. Wang, Q. Ruan, and G. An, “Facial expression recognition using sparse local Fisher discriminant analysis,” *Neurocomputing*, vol. 174, pp. 756–766, 2016.
 - [35] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, “Sparse discriminant analysis,” *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.
 - [36] S. Yuan, X. Mao, and L. Chen, “Sparsity regularization discriminant projection for feature extraction,” *Neural Processing Letters*, vol. 49, no. 2, pp. 539–553, 2019.
 - [37] Y. Kanno and S. Kitayama, “Alternating direction method of multipliers as a simple effective heuristic for mixed-integer nonlinear optimization,” *Structural and Multidisciplinary Optimization*, vol. 58, no. 3, pp. 1291–1295, 2018.
 - [38] Y. Lei, Z. He, and Y. Zi, “A new approach to intelligent fault diagnosis of rotating machinery,” *Expert Systems with Applications*, vol. 35, no. 4, pp. 1593–1600, 2008.
 - [39] H. Zhou, T. Shi, G. Liao et al., “Weighted kernel entropy component analysis for fault diagnosis of rolling bearings,” *Sensors*, vol. 17, no. 3, p. 625, 2017.