

Research Article

KPCA and AE Based Local-Global Feature Extraction Method for Vibration Signals of Rotating Machinery

Xiao Hu,¹ Zhihui Xiao ,² Dong Liu,³ Yongjun Tang,⁴ O. P. Malik ,⁵ and Xiangchen Xia⁶

¹School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China

²Key Laboratory of Hydraulic Machinery Transients, Ministry of Education, Wuhan University, Wuhan 430072, China

³State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

⁴Technology Center of State Grid Xinyuan Co., Ltd., Beijing 100000, China

⁵Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

⁶School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China

Correspondence should be addressed to Zhihui Xiao; xiaozhihui@126.com

Received 2 March 2020; Revised 30 May 2020; Accepted 5 June 2020; Published 29 June 2020

Academic Editor: Guido Ala

Copyright © 2020 Xiao Hu 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.

Feature extraction plays a key role in fault diagnosis of rotating machinery. Many methods reported in the literature are based on masses of labeled data and need much prior knowledge to select the most discriminating features or establish a complex deep-learning model. To solve the dilemma, a novel feature extraction method based on kernel principal component analysis (KPCA) and an autoencoder (AE), namely, SFS-KPCA-AE, is presented in this paper to automatically extract the most discriminative features from the frequency spectrum of vibration signals. First, fast Fourier transform is calculated on the entire vibration signal to get the frequency spectrum. Next, the spectrum is divided into several segments. Then, local-global feature extraction is performed by applying KPCA to these segments. Finally, an AE is employed to obtain the low-dimensional representations of the high-dimensional global feature. The proposed feature extraction method combined with a classifier achieves fault diagnosis for rotating machinery. A rotor dataset and a bearing dataset are utilized to validate the performance of the proposed method. Experimental results demonstrate that the proposed method achieved satisfactory performance in feature extraction when training samples or motor load changed. By comparing with other methods, the superiority of the proposed SFS-KPCA-AE is verified.

1. Introduction

Rotating machinery is widely used in the modern industry. However, after long operating time under harsh working conditions, it will inevitably break down and even result in catastrophic accidents and huge economic losses [1, 2]. Therefore, developing accurate fault diagnosis technologies is of great significance to avoid potential losses [3, 4].

In many studies, fault diagnosis for rotating machinery is converted to classification of the features extracted from the vibration signals [5–9]. Hence, high accuracy of fault diagnosis always benefits by fault features extracted from original signals. Fault features in time domain, frequency domain, and time-frequency domain extracted by signal

analysis technology have been widely investigated [10, 11]. Wavelet transform was adopted in [12] to extract 2-D time-frequency features from raw 1-D vibration. Complementary ensemble empirical mode decomposition was used in [13] to remove undesirable noise and extract fault signature in the vibration signals of rolling bearings. Optimized variational mode decomposition and envelope demodulation analysis were utilized in [14] to extract features from vibration signals of bearings. However, most of the aforementioned feature extraction methods require comprehensive experience of domain experts on the specific problems to select the most discriminative features, which makes the choice of features sensitive to different cases. Instead of selecting features by a human operator, deep-learning techniques extract useful

features in a data-driven way and have become a research hotspot. A weighted sparse representation classification based on dictionary learning was proposed in [15] to extract the fault features and identify the fault status. This method considers the locality of the sample and solves the time-shift deviation problem of vibration signals. An improved convolutional neural network based on LeNet-5 was constructed in [16] to learn and enhance the features from the samples of multisensor data fusion and achieved good results. Although fault diagnosis based on deep-learning techniques, such as convolutional neural networks [17–19], can automatically learn useful information from original signals, they are supervised feature extraction processes that require large amounts of labeled data. In a practical situation, it is difficult to acquire self-contained negative samples from the online monitoring of critical plants [20]. Therefore, a majority of unlabeled data bring about difficulties for supervised methods.

To solve the above dilemma, unsupervised learning which discovers useful representations from unlabeled data has caught the attention of researchers. The commonly used methods of unsupervised feature learning include kernel principal component analysis (KPCA) and an autoencoder (AE). In the field of fault diagnosis, these two unsupervised feature learning methods are often used to extract the time-domain features of signals, and they also have great potential in the frequency-domain feature extraction. In practice, analysis of the spectrum obtained by the fast Fourier transform (FFT) is a commonly used fault diagnosis technique. On one hand, the change of the main frequency components in the spectrum contains a wealth of mechanical motion state information. On the other hand, the frequency components with significant amplitude changes or the frequency components which have a fixed ratio relationship with the rotation frequency are often used for fault detection and regarded as key objects of fault type identification. Consequently, there is a lot of information in the frequency spectrum of vibration signal that is very critical for fault diagnosis, and the two unsupervised feature learning methods mentioned above can be used to extract the important information. The redundant information in the frequency spectrum of vibration signals, such as those frequency components with the same amplitude in different samples, can be removed by using KPCA as a data preprocessing method. Then, the most important information in the frequency spectrum can be extracted by the autoencoder in an unsupervised way.

In this paper, in order to take advantage of the aforementioned unsupervised feature learning techniques and extract key information that can distinguish the state of the device from the frequency spectrum of vibration signals of rotating machinery, an unsupervised feature learning method based on the KPCA and AE is proposed. KPCA is a technique for nonlinear principal component analysis. Experiments comparing the utility of kernel principal component analysis features for pattern recognition using a linear classifier have shown that nonlinear principal components afforded better recognition rates than corresponding numbers of linear principal components [21]. In

the proposed method, KPCA reduces useless information in the frequency spectrum and roughly extracts useful features in the frequency spectrum. Next, an autoencoder is used to learn the effective encoding of the useful features, so that the final extracted features can retain the most important information in the frequency domain of the signal with a very low dimension. Generally, it is difficult to predict which frequency components will change abnormally and can be used as a fault feature when a device fails, so attention should be paid to both local and global features of the frequency spectrum to ensure that important failure information is not missed. As KPCA is a popularly used method for global structure analysis of datasets and it omits the local information [22–25], the method of segmenting the frequency spectrum is adopted in this study. First, FFT is utilized to obtain the frequency spectrum of vibration signals. Next, the frequency spectrum is divided into several segments. Then, KPCA is employed to extract the local features from these segments, and these local features are connected in sequence to obtain one global feature vector. After that, the global feature vector is sent to the AE to get the most discriminative features with low dimension. The proposed feature extraction method combined with a classifier can realize fault diagnosis for rotating machinery. A rotor dataset and a rolling bearing dataset participated in verifying the effectiveness and generalization of the proposed method. Several experiments were conducted to confirm the robustness of the proposed method under the condition of different numbers of randomly selected training samples and the condition of load fluctuation. Besides, the superiority of the proposed method was validated by designing comparative experiments.

2. Background Knowledge and the Proposed Method

2.1. Principal Component Analysis. Principal component analysis (PCA) is an unsupervised learning method that can be used in feature extraction, noise reduction, and pattern recognition. PCA is a basis transformation to diagonalize an estimate of the covariance matrix of the data \mathbf{x}_k , $k = 1, \dots, l$, $\mathbf{x}_k \in \mathbb{R}^N$, $\sum_{k=1}^l \mathbf{x}_k = 0$, defined as follows [21]:

$$C = \frac{1}{l} \sum_{j=1}^l \mathbf{x}_j \mathbf{x}_j^T. \quad (1)$$

The new coordinates in the eigenvector basis, i.e., the orthogonal projections onto the eigenvectors, are called principal components.

2.2. Kernel Principal Component Analysis. Kernel principal component analysis performs a nonlinear form of principal component analysis by using integral operator kernel functions [21]. Assuming that the data \mathbf{x}_k , $k = 1, \dots, l$, $\mathbf{x}_k \in \mathbb{R}^N$ is mapped into the feature space F by

$$\begin{aligned} \Phi: \mathbb{R}^N &\longrightarrow F, \\ \mathbf{x}_k &\longrightarrow \Phi(\mathbf{x}_k), \end{aligned} \quad (2)$$

and $\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_l)$ is centered, i.e., $\sum_{k=1}^l \Phi(\mathbf{x}_k) = 0$. To do PCA for the covariance matrix,

$$\bar{\mathbf{C}} = \frac{1}{l} \sum_{j=1}^l \Phi(\mathbf{x}_j) \Phi(\mathbf{x}_j)^T, \quad (3)$$

is used to find eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{V} \in F \setminus \{0\}$ satisfying

$$\lambda \mathbf{V} = \bar{\mathbf{C}} \mathbf{V}. \quad (4)$$

Considering the equivalent system of (4),

$$\lambda (\Phi(\mathbf{x}_k) \cdot \mathbf{V}) = (\Phi(\mathbf{x}_k) \cdot \bar{\mathbf{C}} \mathbf{V}), \quad \text{for all } k = 1, \dots, l, \quad (5)$$

and there exist coefficients $\alpha_1, \dots, \alpha_l$ such that

$$\mathbf{V} = \sum_{i=1}^l \alpha_i \Phi(\mathbf{x}_i). \quad (6)$$

Substituting (3) and (6) into (5) and defining an $l \times l$ matrix \mathbf{K} by

$$\mathbf{K}_{ij} := (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)), \quad (7)$$

we arrive at

$$l \lambda \mathbf{K} \boldsymbol{\alpha} = \mathbf{K}^2 \boldsymbol{\alpha}, \quad (8)$$

where $\boldsymbol{\alpha}$ denotes the column vector with entries $\alpha_1, \dots, \alpha_l$. To find solutions of (8) is to solve the eigenvalue problem

$$l \lambda \boldsymbol{\alpha} = \mathbf{K} \boldsymbol{\alpha}, \quad (9)$$

for nonzero eigenvalues. Then, the solutions $\boldsymbol{\alpha}^k$ belonging to nonzero eigenvalues are normalized by normalizing the corresponding vectors in F , i.e., $(\mathbf{V}^k \cdot \mathbf{V}^k) = 1$. By virtue of (6), (7), and (9), this can be converted to

$$1 = \sum_{i,j=1}^l \alpha_i^k \alpha_j^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) = (\boldsymbol{\alpha}^k \cdot \mathbf{K} \boldsymbol{\alpha}^k) = \lambda_k (\boldsymbol{\alpha}^k \cdot \boldsymbol{\alpha}^k). \quad (10)$$

For principal component extraction, projections of the image of a test point $\Phi(\mathbf{x})$ onto the eigenvectors \mathbf{V}^k in F can be computed according to

$$(\mathbf{V}^k \cdot \Phi(\mathbf{x})) = \sum_{i=1}^l \alpha_i^k (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})). \quad (11)$$

Note that neither (7) nor (11) requires the $\Phi(\mathbf{x}_i)$ in explicit form, which means they are only needed in dot products. Therefore, kernel functions can be used for computing these dot products without actually performing the map Φ . The kernel functions commonly used in KPCA are shown in Table 1.

Assuming that the dimension of the data \mathbf{x}_k , $k = 1, \dots, l$, $\mathbf{x}_k \in \mathbb{R}^N$ is to be reduced to J . The KPCA algorithm is described in the following steps [26]:

- (1) Substitute kernel functions for all occurrences of $(\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j))$.
- (2) Compute the dot product matrix $\mathbf{K}_{ij} = (\kappa(\mathbf{x}_i, \mathbf{x}_j))_{ij}$.

- (3) Diagonalize \mathbf{K} according to

$$\tilde{\mathbf{K}}_{ij} = \mathbf{K} - \mathbf{I}_l \mathbf{K} - \mathbf{K} \mathbf{I}_l + \mathbf{I}_l \mathbf{K} \mathbf{I}_l, \quad (12)$$

where $\tilde{\mathbf{K}}_{ij}$ is the centralized dataset and \mathbf{I}_l is an $l \times l$ matrix whose every element is 1. Normalize the eigenvector expansion coefficients α^k by requiring equation (10).

- (4) Solve (9).

- (5) Extract principal components of a test point \mathbf{x} by computing projections onto eigenvectors according to equation (11). Take the principal components corresponding to the largest J eigenvalues to be the output.

2.3. Autoencoder. An autoencoder is one of the best-known unsupervised machine learning algorithms which is generally used for dimension reduction and feature extraction [27–30]. Assuming that there is a D -dimensional dataset, $\mathbf{x}^{(n)} \in \mathbb{R}^D$, $1 \leq n \leq N$, the encoder maps each sample in this dataset to the feature space to get the encoding, $\mathbf{z}^{(n)} \in \mathbb{R}^M$, $1 \leq n \leq N$, of each sample. The goal is to use the encoding of each sample to reconstruct the original sample. The original AE is a three-layer feedforward neural network, consisting of an input layer, a hidden layer, and an output layer. The input layer and the hidden layer constitute the encoder:

$$f: \mathbb{R}^D \longrightarrow \mathbb{R}^M, \quad (13)$$

which transforms the input data into low-dimensional features. The hidden layer and the output layer constitute the decoder:

$$g: \mathbb{R}^M \longrightarrow \mathbb{R}^D, \quad (14)$$

which reconstructs the input data from the corresponding features. The learning goal of the autoencoder is to minimize reconstruction error:

$$\mathcal{L} = \sum_{n=1}^N \|\mathbf{x}^{(n)} - g[f(\mathbf{x}^{(n)})]\|^2. \quad (15)$$

If $M < D$, the autoencoder is equivalent to a dimension reduction or feature extraction method. After completing the training of the AE, the decoder is generally removed. Only the encoder is retained, and the output of the encoder can be used as a feature of each sample.

2.4. Proposed Feature Extraction Method. The time-domain statistical characteristics of rotating machinery vibration signals, such as mean square value, standard deviation, and kurtosis, can only reflect whether the state of the equipment is normal. To know the location and the type of the failure requires further analysis. In practical situations, spectrum analysis is an important analysis method, and fast Fourier transform is the most commonly used tool for spectrum analysis. In the frequency spectrum of the vibration signal of a rotating machine, the frequency components which have a

TABLE 1: Common kernel functions.

Kernel functions	Expression	Parameters
Polynomial kernels	$\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d$	$d \geq 1$ is the degree of polynomial
Radial basis functions	$\kappa(\mathbf{x}, \mathbf{y}) = \exp(-(\ \mathbf{x} - \mathbf{y}\)/2\sigma^2)$	$\sigma > 0$ is the width of rbf kernel
Sigmoid kernels	$\kappa(\mathbf{x}, \mathbf{y}) = \tanh(\kappa(\mathbf{x} \cdot \mathbf{y}) + \theta)$	$\theta < 0$

fixed relationship with the rotation frequency are the main objects of concern. However, neglect of the other frequency components may cause a misjudgment when a device fails. On one hand, it is difficult to predict which frequency component's amplitude will change when the state of the device is abnormal. This means it is difficult to determine which information should become the key feature of the frequency spectrum. On the other hand, abnormal changes in some frequency components of some samples may correspond to new types of faults, and these are unlabeled samples. These unlabeled samples can bring difficulties to the supervised fault diagnosis method. To sum up, more frequency components need to be paid attention to, so that the local feature and the global feature of the frequency spectrum can be extracted and used to judge the state of the rotating machine.

In this study, in order to extract the key information that can distinguish the state of the device in the frequency spectrum of vibration signals of rotating machinery, an unsupervised feature extraction method combining KPCA and AE is proposed. In addition, to overcome the shortcomings of KPCA in ignoring the local feature of the data, the frequency spectrum segmentation method is used to extract the features of each frequency spectrum segment. Then, merge the features of these frequency spectrum segments to obtain the global feature of the frequency spectrum, which not only avoids missing local features of the frequency spectrum but also reduces the influence of noise components on the result of frequency spectrum feature extraction.

The feature extraction for vibration signals using the proposed method is implemented using the following steps:

- (1) Calculate the FFT on the entire signal to obtain its frequency spectrum. For a given signal data \mathbf{T} with N points, $[t_1, t_2, \dots, t_N]$, let \mathbf{R} with L spectrum lines, $[r_1, r_2, \dots, r_L]$ be its frequency spectrum, and \mathbf{R} is normalized into $[-1, 1]$.
- (2) Divide \mathbf{R} into N_s segments, and each segment contains L/N_s points:

$$\mathbf{R} = \{r^1, r^2, \dots, r^{N_s}\}. \quad (16)$$

- (3) Apply KPCA to each item in $\{r^1, r^2, \dots, r^{N_s}\}$ to extract the principal components, $\mathbf{C}_{N_s} = \{c_{N_s}^1, c_{N_s}^2, \dots, c_{N_s}^{V_s}\}$, where V_s is the number of principal components. The principal components \mathbf{C}_{N_s} of each segment are regarded as local features of the frequency spectrum and are connected in sequence to obtain a global feature vector \mathbf{x} with a dimension of $N_s \times V_s$ as follows:

$$\mathbf{x} = \{c_1^1, c_1^2, \dots, c_1^{V_s}, c_2^1, c_2^2, \dots, c_2^{V_s}, \dots, c_{N_s}^1, c_{N_s}^2, \dots, c_{N_s}^{V_s}\}. \quad (17)$$

- (4) All samples in the dataset are processed as the form of \mathbf{x} according to the previous three steps, and then \mathbf{x} is normalized into $[-1, 1]$.
- (5) Select some processed samples as training samples and send them to the constructed AE for training, as detailed in Figure 1.
- (6) After training, the AE has the capability of reconstructing the input \mathbf{x} : $\mathbf{x} \rightarrow \mathbf{h} \rightarrow \mathbf{x}$, where \mathbf{h} is a three-dimensional vector output from the trained encoder. \mathbf{h} is also the latent feature of frequency spectrum desired to be obtained.

Fault diagnosis integrating the proposed method and a classifier is depicted in Figure 2. As shown in Figure 2, fault diagnosis can be divided into three parts: data preprocessing, feature extraction, and fault recognition [1]. Therefore, the effectiveness of feature extraction can be verified by ensuring the faults are correctly identified. The accuracy of fault identification is always converted into the classification accuracy of the classifier. The classification model establishes the relationship between fault features and fault types during the training phase, and then such a relationship is employed to make decisions for testing samples [31]. After that, the classification accuracy of the testing dataset is calculated and used as an index to evaluate the performance of feature extraction methods. However, except for feature extraction methods, there are other factors that affect classification accuracy, such as the number of training samples [32, 33] and how they are selected (fixed or randomly) and the type and the parameter configuration of the classifier. On the whole, an outstanding feature extraction method should have strong stability, which means that it can lead to a high classification accuracy in the case of random selection of training samples, changing the training sample number or changing classifier type.

3. Case Study and Discussion

3.1. Case Study 1: Feature Extraction of Rotor Signals

3.1.1. Data Description. The data were acquired from the experimental rotating machinery system displayed in Figure 3. The rotor laboratory bench is driven by a DC motor controlled by the DH5600 speed controller. The rotor with a diameter of 10 mm and a length of 850 mm is supported by the bearing brackets located at the two sides and contains two single shafts coupled together by a coupling. Two rub

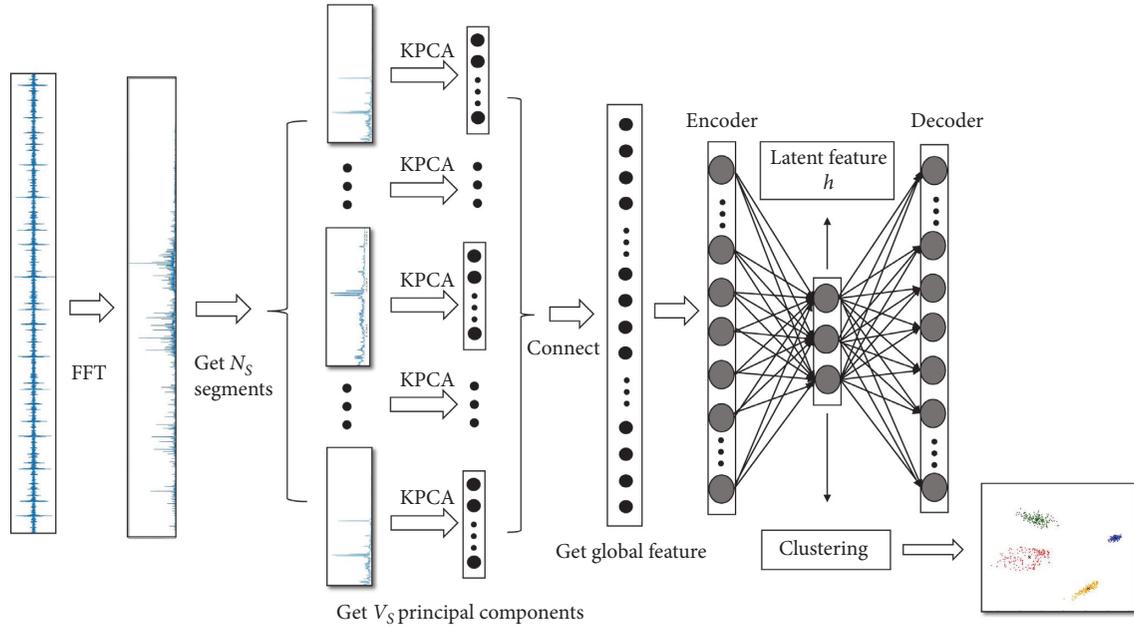


FIGURE 1: Flowchart of the proposed method.

screw housings are installed on the rack of the system, and two mass disks with a diameter of 75 mm are fixed on the rotor. The sensors for signal acquisition contain a photoelectric sensor for speed measurement, a piezoelectric accelerometer for vibration measurement, and two eddy current sensors for displacement measurement. The signals collected by eddy current sensors and the piezoelectric accelerometer are sent to a front-end processor for filtering and amplification and then transmitted to the computer for storage and analysis.

The system can simulate many types of rotating machinery faults. Here, three kinds of faults, including unbalance, misalignment, and contact-rubbing, are simulated. The unbalance condition is simulated by screwing 2 g mass block into the threaded hole near the edge of the mass disk. The misalignment condition is simulated by changing the relative position of the coupling shafts, while the contact-rubbing is simulated by screwing the rub screw into the rub screw housing. The speed of the system is assigned to 1200 rpm, the sampling frequency is 2048 Hz, and the sample length is 1 s during the experiments [1]. Under each of the three fault conditions and normal condition, 45 vibration signals are acquired, respectively, for a total of 180 samples. The time-domain diagram of the vibration signals in the four states is illustrated in Figure 4.

3.1.2. Parameter Selection. Following the steps described in Section 2.4, FFT is calculated on the entire rotor signal to get the frequency spectrum. The frequency spectrum of one group containing four different rotor health states is illustrated in Figure 5.

It can be observed in Figure 4 that the signals are buried in noise, which causes difficulties to distinguish different states of rotors in time domain, while it can be seen from

Figure 5 that the frequency spectrum of the four rotor states has the following differences: (1) the amplitude of the first maximum value (the rotor's rotation frequency, 20 Hz) is significantly different; (2) the six spectrum lines with the largest amplitude values in each spectrum are circled in red. These spectrum lines can be regarded as the dominant frequency components in the frequency spectrum. It can be seen that the dominant frequency components are different under different device states. Meanwhile, in some frequency bands, the amplitude of the same frequency component is obviously different, such as the frequency bands 150 Hz ~ 200 Hz, 400 Hz ~ 500 Hz, and 600 Hz ~ 700 Hz. Therefore, aiming to extract the features of each frequency band in the frequency spectrum and avoid missing the local information of the frequency spectrum, a strategy of spectrum segmentation is adopted in this study.

In the proposed method, there are two parameters that should be selected before using KPCA, i.e., N_s and V_s . As described in Section 2.4, N_s is the number of segments that the frequency spectrum is to be divided into, while V_s is the output dimension of KPCA. These two parameters not only determine the input dimension of the AE but also have influence on the structure complexity and training difficulty of the AE. Therefore, they should be determined carefully.

The parameter N_s , which determines the frequency range included in each frequency spectrum segment should be determined first. In this experiment, the sampling frequency of the rotor's vibration signal is 2048 Hz and the sampling time is 1 s, so the frequency resolution in the frequency spectrum is about 1 Hz, and the rotation frequency of the rotor corresponds to the 21st spectrum line in the frequency spectrum. According to the comparison result of Figure 5, it is inferred that the performance of feature extraction will be better when the first spectrum line is included in the first frequency spectrum segment, which means N_s should be less

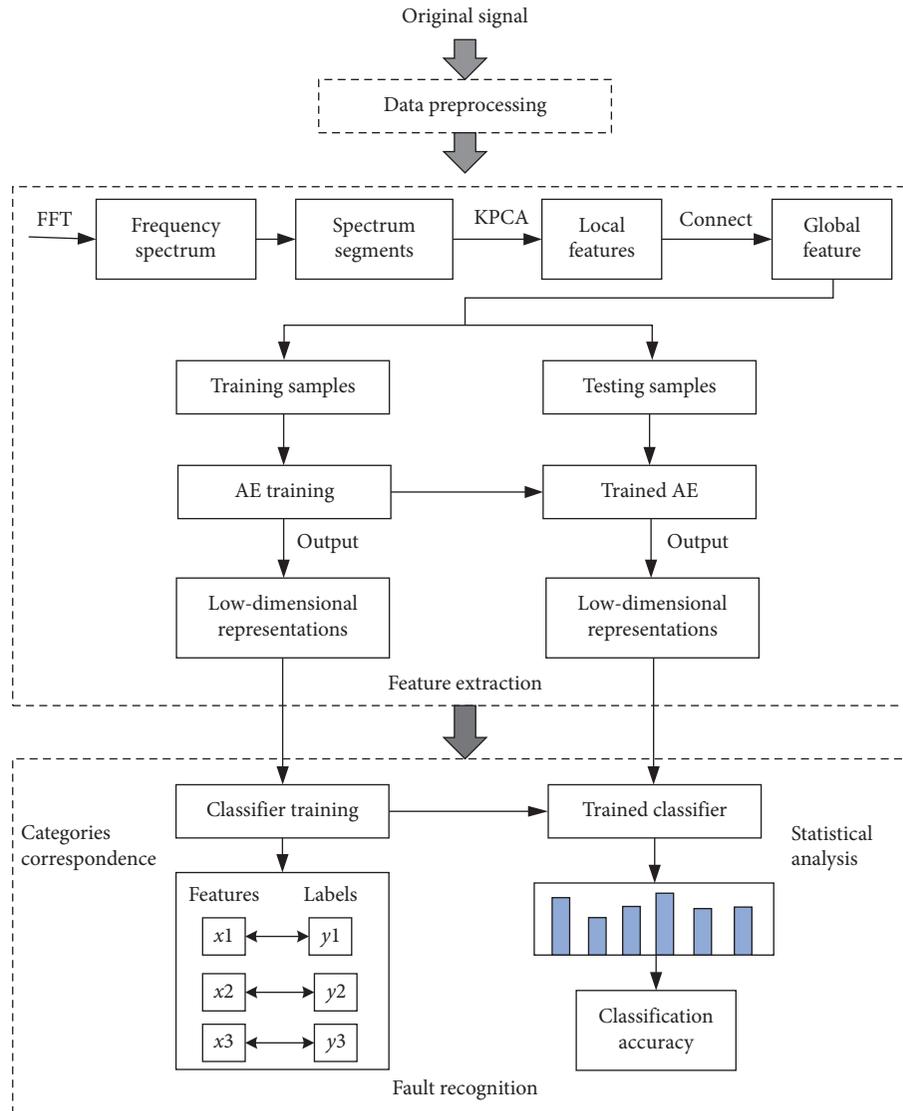


FIGURE 2: Flowchart of fault diagnosis.

than 48. Meanwhile, as the proposed method divides the entire frequency spectrum equally, the set of values of N_s is [2, 4, 8, 16, 32]. And comparative experiments were conducted to verify this inference.

Here, local features obtained by applying KPCA to each segment of the frequency spectrum are connected in sequence to obtain a global feature and then classified by the k -nearest neighbor (KNN) classifier. Classification accuracy is thereafter calculated to help select the appropriate parameters. The classification accuracies of the KNN classifier configured with different parameters are shown in Figure 6. The parameter configurations with mean classification accuracy higher than 95% are shown in Table 2. The effect of randomness is reduced by conducting 100 trials of experiments, and the mean accuracy was calculated. It is noted that when $N_s = 1$, it means that KPCA is applied to the frequency spectrum without division. Unfortunately, there has been no standard method for choosing the best kernel for a given problem until now [21]. So, "Gaussian" which is commonly

used with only one hyperparameter (γ) as the kernel function is chosen, and the γ is set as $\gamma = 1/V_s$.

Comparing the mean classification accuracies in Figure 6 and Table 2, the following conclusions can be drawn: (1) the highest accuracy is 98.89%, when $N_s = 32$ and $V_s = 2$, which means each frequency spectrum contains 32 frequency spectrum lines, and the first frequency segment only includes one frequency component that has an integer proportional relationship with the rotor rotation frequency, that is, 1 times rotation frequency; (2) when $N_s = 1$, that is, the principal components are extracted without frequency spectrum segmentation, no matter how many principal components are extracted, the classification accuracy is lower than 95%, indicating that the segmentation of the frequency spectrum is necessary to help improve the performance of KPCA; (3) the more segments there are, the fewer principal components need to be kept in each segment. For example, when $N_s = 1$, the classification accuracy of V_s equal to 5, 6, or 7 is the highest; when $N_s = 128$, the

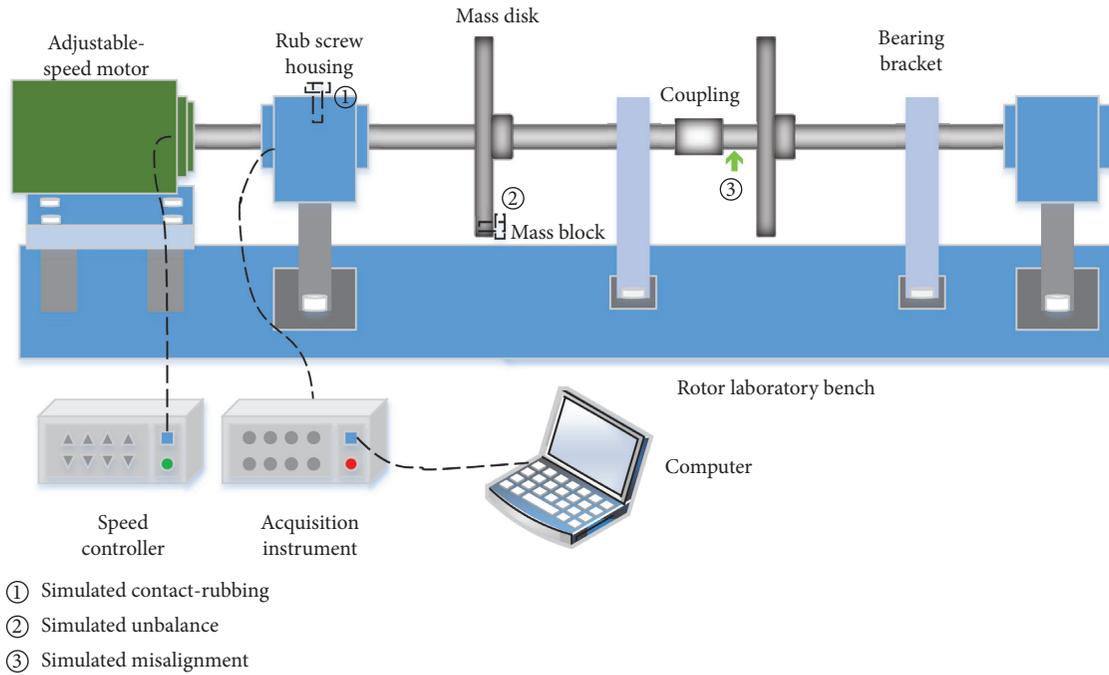


FIGURE 3: Rotor laboratory bench.

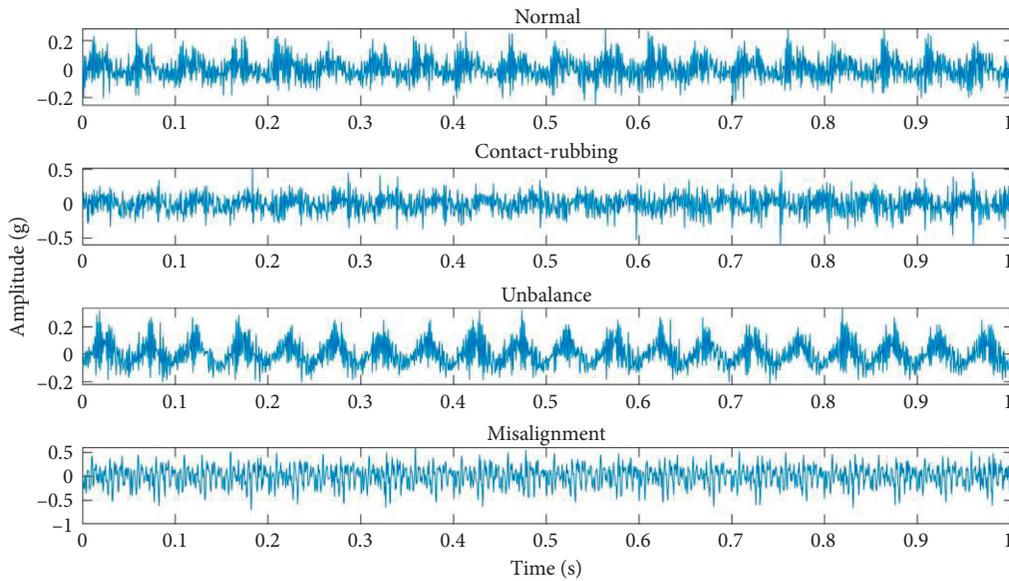


FIGURE 4: Vibration signal in different conditions.

classification accuracy of $V_s = 1$ is the highest. When N_s takes the value of 2, 4, 8, 16, 32, or 64, $V_s = 2$ has the highest accuracy. Finally, $N_s = 32$, $V_s = 2$, and $\gamma = 0.5$ is regarded as the best parameter configuration.

Based on the above experimental results, the experience of selecting N_s and V_s is summarized. When the proposed method is applied to other projects, the following methods can be used to initially determine the optimal values of N_s and V_s : (1) when selecting N_s , the upper limit of the frequency range of the first spectrum segment should be more

than 1 times but less than 2 times the rotation frequency. (2) The number of V_s is less than or equal to 3.

According to the above analysis, the proposed method is applied to the rotor signals as follows. First, each frequency spectrum with 1024 spectrum lines is divided into 32 segments. Next, two principal components considered as local features of the spectrum are extracted from each spectrum segment using KPCA. Then, these local features are connected in sequence to obtain the global feature vector. Finally, the global feature vectors are sent to the constructed

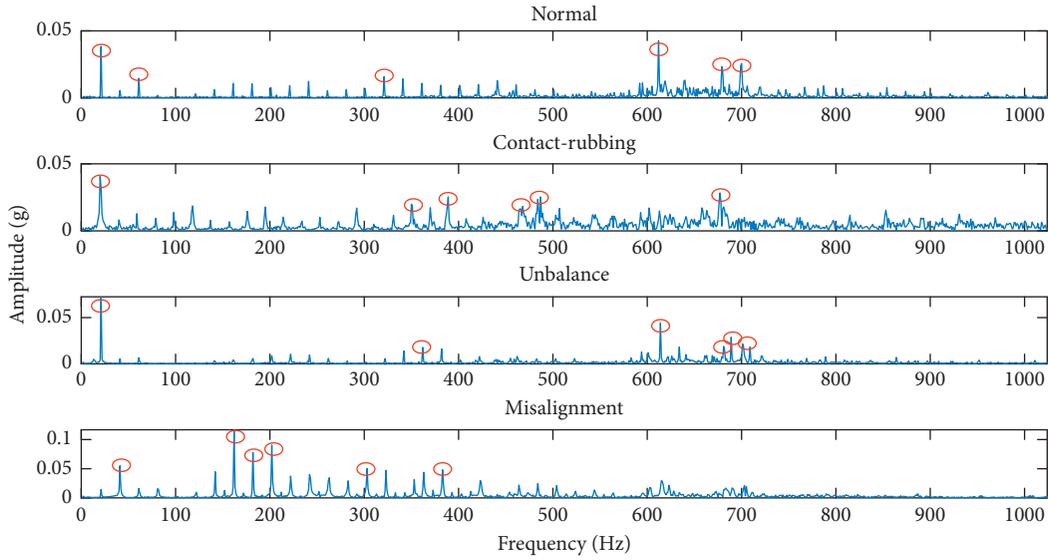


FIGURE 5: Frequency spectrum of different rotor health states.

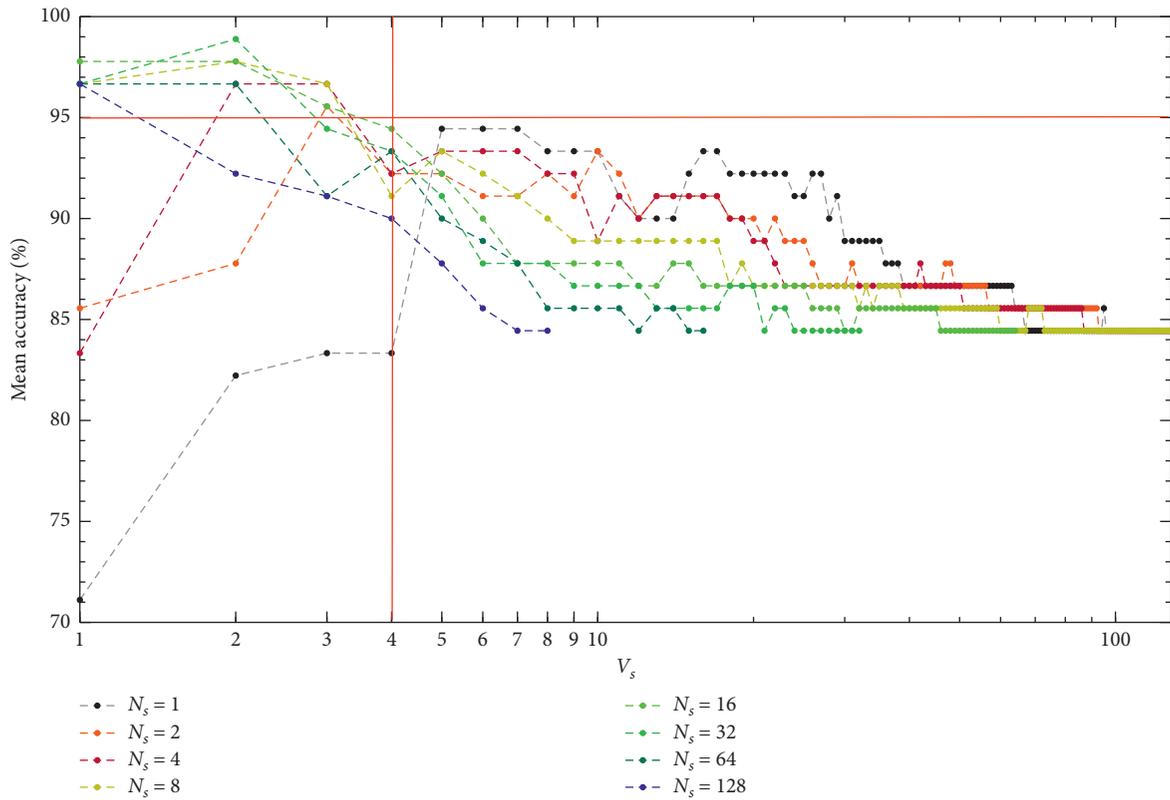


FIGURE 6: Classification accuracies of the KNN classifier configured with different parameters.

AE to extract the most critical and lower-dimensional features of the frequency spectrum.

In this study, the modules in AE are parameterized through neural networks. Both the encoder and decoder have only one layer, containing 100 neurons, and the activation function is Rectified Linear Units (ReLUs), given by the following:

$$f(x) = \max(0, x). \tag{18}$$

There are only three neurons in the output layer of the encoder, and the activation function is linear. Activation of the last layer of the decoder is hyperbolic tangent ($\tan h$) to match the input which is normalized into $[-1, 1]$.

TABLE 2: Parameter configuration with mean classification accuracy higher than 95%.

N_s	Number of spectrum line	V_s					
		1		2		3	
		Dimension of global feature	Mean accuracy (%)	Dimension of global feature	Mean accuracy (%)	Dimension of global feature	Mean accuracy (%)
2	512	2	—	4	—	6	95.56
4	256	4	—	8	96.67	12	96.67
8	128	8	97.78	16	96.67	24	96.67
16	64	16	97.78	32	97.78	48	95.56
32	32	32	96.67	64	98.89	96	—
64	16	64	96.67	128	96.67	192	—
128	8	128	96.67	256	—	384	—

Full-batch training was used during the experiments because of the limited size of samples. Then, the loss of the validation dataset was investigated to choose an appropriate learning rate. Training samples, validation samples, and test samples account for 60%, 20%, and 20%, respectively. All samples are randomly selected during the experiments.

In most cases, the learning rate (lr) can be fixed, decreasing, momentum-based, or adaptive and is set in the range $[1e-1, 1-4]$. In this study, two learning rate setting strategies are studied: one is fixed learning rate which means the learning rate does not change with epochs; the other is decreasing learning rate which means the learning rate decreases with the increase of epochs. The learning rate setting schemes are displayed in Table 3, where patience = 10 ($p = 10$) means that if the loss of the validation dataset no longer decreases after 10 epochs, the learning rate will reduce in the form of $lr_{n+1} = lr_n * 0.1$. The validation loss varies with the increase of training epochs, as shown in Figure 7.

It can be seen from Figure 7 that (1) when the fixed learning rate is $1e-4$ and $1e-2$, the validation loss converges at the slowest and fastest speed, respectively; (2) when the learning rate is set as $1e-1$, the validation loss suddenly increases around the third epoch; (3) when the learning rate is set as $1e-1$ and $1e-2$, the validation loss first decreases and then increases, indicating that the model is over-fitted. The above information suggests that the model shows bad performance when the learning rate is set as $1e-1$, $1e-2$, and $1e-4$. It is also observed from Figure 7 that the validation loss converges around 800 epochs under different learning rate settings, and the lowest validation loss is 0.0092 in scheme 7. Ultimately, scheme 7 is adopted and training epoch = 1000 is set during the following experiments.

3.1.3. Comparative Experiments. The effectiveness of the proposed method was verified by combining with the KNN classifier. Comparative experiments were designed to validate the superiority of the proposed method. In Method 1, the SFS-KPCA-AE model was set up as described above. In Method 2, the steps were the same as those in Method 1, except that the frequency spectrum in Method 1 was replaced by the time series. In Method 3, the first four steps were the same as the proposed method, while at Step 5, AE was replaced by KPCA. In Method 4, the first two steps were the same as the proposed method, while at Step 3, KPCA was

replaced by AE. It is noted that in order to reduce the complexity of Method 4, only two autoencoders were trained, one for extracting local features, namely, AE-1, and one for extracting global features, namely, AE-2. During model training, 90 samples were randomly selected in each round of experiments. Because each sample contained 32 spectrum segments, a total of 2880 spectrum segments were obtained and constituted the training dataset. First, AE-1 was trained to get the local features of each frequency spectrum. Second, these local features were connected in sequence to obtain a global feature. Finally, AE-2 was trained with the global features to get a three-dimensional feature.

In Method 5, the same features (F20), as in [34], were extracted from time domain and frequency domain to comprehensively describe the healthy status through FFT and statistical methods. For a given signal data \mathbf{T} with N points, $[t_1, t_2, \dots, t_N]$, let \mathbf{R} with L spectrum lines, $[r_1, r_2, \dots, r_L]$ be its frequency spectrum, and f_l is the frequency of the l -th spectrum line. Ten time-domain feature parameters, $(TF_1 - TF_{10})$, and ten frequency features, $(FF_1 - FF_{10})$, can be calculated as shown in Tables 4 and 5, respectively. Let $\mathbf{F} = [TF_1 \dots TF_{10} FF_1 \dots FF_{10}]$ denote the extracted features from each sample. Normalize \mathbf{F} into $[-1, 1]$ to reduce the huge discrepancy of different feature parameters. Finally, the dimension of the normalized mixed-domain feature is reduced to 3 by using KPCA. In Method 6, the same twenty features were regarded as the input of the AE whose hyperparameter configuration was the same as Method 1. Then, these 20-dimensional features were converted into 3-dimensional features as the output of the trained encoder.

In Method 7, Ensemble Empirical Mode Decomposition and Curve Code (EEMD-CC) was used for the rotor dataset. Details of the method can be found in reference [32].

Each of the experiments was conducted in 20 trials. Training samples (90) and testing samples (90) were selected randomly to reduce the influence of randomness. The classification accuracies of each method in each trial are detailed in Figure 8. The mean accuracy with corresponding standard deviation and cost time of each method are calculated and listed in Table 6.

As illustrated by Table 6, Method 3 takes 1.04 s which is the minimum, Method 7 requires 28.31 s which is the maximum, and the remaining methods spend no more than 4 s. Because the autoencoder needs to be trained in

TABLE 3: Learning rate setting schemes.

	No.	Initial learning rate	Learning rate	Patience	Epoch
Fixed learning rate	1	—	$1e-1$	—	1000
	2	—	$1e-2$	—	1000
	3	—	$1e-3$	—	1000
	4	—	$1e-4$	—	1000
Decreasing learning rate	5	$1e-2$	—	10	1000
	6	$1e-2$	—	30	1000
	7	$1e-2$	—	50	1000
	8	$1e-2$	—	100	1000
	9	$1e-2$	—	200	1000
	10	$1e-2$	—	300	1000

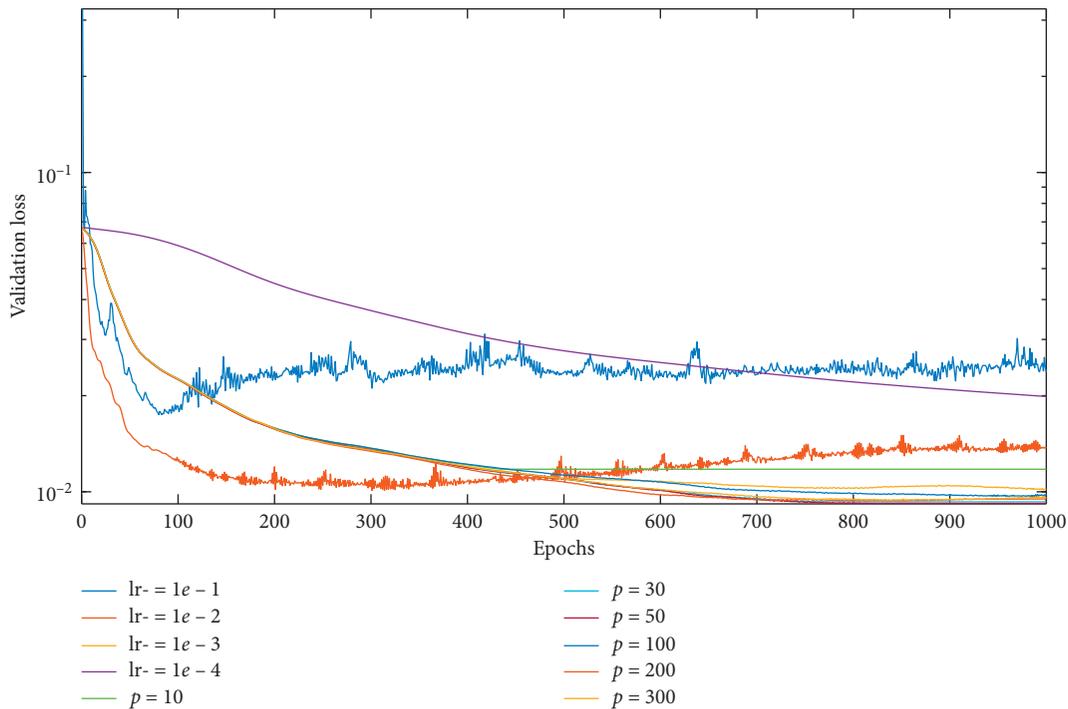


FIGURE 7: Validation loss with different learning rates.

1,000 epochs, the methods using the autoencoder cost more time than the method without the autoencoder. As shown in Figure 8, Method 1 achieves high classification accuracy with strong robustness. Moreover, as can be seen from Table 6, the mean testing accuracy of Method 1 is higher than that of other methods. As for Method 2, the testing accuracy of 47.00% is the lowest in Table 6. It indicates that the features extracted from time series are less discriminative than features extracted from the frequency spectrum. Thus, it leads to lower classification accuracy than that of Method 1. The mean testing accuracies of Method 3 and Method 4 are 86.67% and 84.55%, both are lower than that of Method 1, which indicates that both the KPCA algorithm and the autoencoder have great influence on the performance of feature extraction and combining KPCA and AE can extract the most discriminative features. Moreover, the mean testing accuracies of Method 4 and Method 6 are lower than that of Method 3 and Method 5,

which suggests that KPCA has a better performance than AE in feature extraction on the frequency spectrum and the mixed-domain features. It is noted that the dataset used here is the same as in [31], but the data were not denoised, which implies why the EEMD-CC method did not perform as well as in the literature.

For further comparison, the proposed method was applied to the denoised vibration signals as in reference [31], and a test accuracy of 97.92% and a standard deviation value of 1.15% were achieved. This accuracy is lower than that of Method 5. However, comparison of the computing time between Method 1 and Method 7 in Table 6 shows that the latter takes significantly more time than the former. Hence, Method 1 is more suitable for rapid fault diagnosis than Method 7. Most importantly, this comparative experiment indicates that the proposed method has excellent robustness and can reduce the interference caused by noise on feature extraction.

TABLE 4: Feature parameters in time domain.

No.	Feature expression
1	$TF_1 = (1/N) \sum_{n=1}^N t_n$
2	$TF_2 = \sqrt{(1/N) \sum_{n=1}^N [t_n - TF_1]^2}$
3	$TF_3 = \sqrt{(1/N) \sum_{n=1}^N (t_n)^2}$
4	$TF_4 = \max t_n $
5	$TF_5 = (N/(N-1)(N-2)) \sum_{n=1}^N ((t_n - TF_1)/TF_2)^3$
6	$TF_6 = \{(N(N \pm 1))/((N-1)(N-2)(N-3))\} \sum_{n=1}^N ((t_n - TF_1)/TF_2) - (3(N-1)^2)/((N-2)(N-3))$
7	$TF_7 = TF_4/TF_3$
8	$TF_8 = TF_4/((1/N) \sum_{n=1}^N t_n)^2$
9	$TF_9 = TF_3/((1/N) \sum_{n=1}^N t_n)$
10	$TF_{10} = TF_4/((1/N) \sum_{n=1}^N t_n)$

TABLE 5: Feature parameters in frequency domain.

No.	Feature expression
1	$FF_1 = 1/L \sum_{l=1}^L r_l$
2	$FF_2 = \sum_{l=1}^L f_l r_l / \sum_{l=1}^L r_l$
3	$FF_3 = \sqrt{\sum_{l=1}^L f_l^2 r_l / \sum_{l=1}^L r_l}$
4	$FF_4 = \sqrt{\sum_{l=1}^L (f_l - FF_2)^2 r_l / L}$
5	$FF_5 = \sqrt{\sum_{l=1}^L f_l^4 r_l / \sum_{l=1}^L f_l^2 r_l}$
6	$FF_6 = \sqrt{\sum_{l=1}^L f_l^2 r_l / \sum_{l=1}^L r_l \sum_{l=1}^L f_l^4 r_l}$
7	$FF_7 = FF_4/FF_1$
8	$FF_8 = \sum_{l=1}^L (f_l - FF_1)^3 r_l / LFF_4^3$
9	$FF_9 = \sum_{l=1}^L (f_l - FF_5)^4 r_l / LFF_6^4$
10	$FF_{10} = \sum_{l=1}^L r_l \sqrt{f_l - FF_5} / L \sqrt{FF_6}$

3.2. Case Study 2: Feature Extraction of Rolling Bearing Signals

3.2.1. Data Description. To study the generalization of the proposed method, the classic rolling bearing datasets provided by the Case Western Reserve University Bearing Data Center Website [35] were used for the test. The datasets include acceleration measured in the vertical direction on the housing of the drive-end (DE) bearing and acceleration measured in the vertical direction on the fan-end (FE) bearing housing and on the motor supporting base plate (BA). Vibration signals were collected under four different loads: 0 hp, 1 hp, 2 hp, and 3 hp, and each load condition contained four health states of bearing: normal, fault in inner race (IR), fault in ball (BA), fault in outer race (OR). Bearing damage is a single-point damage caused by electro discharge machining, and each fault case contains three different fault widths: 0.007 in, 0.014 in, and 0.021 in. The sampling frequency is 12,000 Hz, and the sampling time for each sample in the original dataset is approximately 10 s.

Drive-end bearing faults were adopted in this case study. In order to obtain enough samples for training and testing, random sampling from the original dataset was utilized to get samples.

For instance, the data of inner race faults with 0.007 in under 1 hp, of which ID is 106, contain a total of 121991 points. One hundred nonrepeating points were selected randomly from these data points as the start point for each sample [11]. Then, 12000 points suitable for FFT were collected from those start points to obtain a high frequency resolution. As detailed in Figure 9, 100 samples each including 12000 points could be acquired from the original sample through the described approach. Four datasets, labeled as set 0, set 1, set 2, and set 3, respectively, were constructed for the experiments. A total of 4000 samples participated in the experiments. The details of each dataset are listed in Table 7.

3.2.2. Validation of the Proposed Method

(1) Fault Diagnosis with Different Training Sample Numbers. According to the previous parameter selection knowledge detailed in Section 3.1.2, firstly, the value of N_s should be determined, and the upper limit of the frequency range of the first spectrum segment should be more than 1 times but less than 2 times the rotation frequency. As the rotation frequency of the motor is about 30 Hz and the frequency resolution is about 1 Hz, the frequency spectrum of each sample in each dataset is divided into 100 segments ($N_s = 100$). Secondly, two local features ($V_s = 2$) are extracted by applying the KPCA algorithm to each segment, and these local features are connected in sequence to obtain a global feature vector. Thirdly, the global feature vector is regarded as input of the AE whose hyperparameters are the same as those of the rotor dataset.

Different training sample numbers were adopted to verify the stability of the proposed method. Each of the experiments was conducted in 50 trials. In addition, training samples and testing samples were selected randomly to reduce the influence of randomness. The KNN classifier was again used for classification during the experiments. The testing accuracies of SFS-KPCA-AE+KNN for bearing dataset are shown in Figure 10. As illustrated by Figure 10, testing accuracies of each dataset with different training sample numbers are not less than 99%, and the average

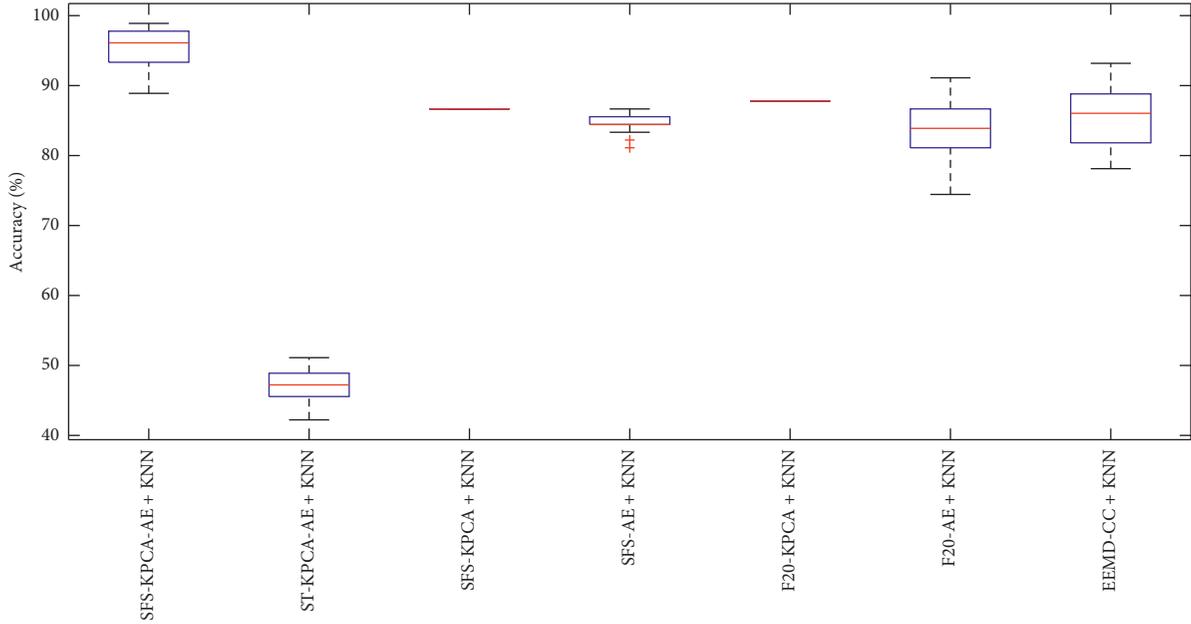


FIGURE 8: Classification accuracies of different methods for the rotor dataset.

TABLE 6: Classification comparison of the rotor dataset.

Methods	Description	Dimension of feature vector	Testing accuracy (mean accuracy \pm standard deviation)	Cost time (s)
1 (proposed)	SFS-KPCA-AE + KNN	3	95.17% \pm 2.87%	3.63
2	ST-KPCA-AE + KNN	3	47.00% \pm 2.53%	3.48
3	SFS-KPCA + KNN	3	86.67%	1.04
4	SFS-AE + KNN	3	84.55% \pm 1.34%	19.99
5	F20-KPCA + KNN	3	87.78%	1.33
6	F20-AE + KNN	3	83.67% \pm 4.40%	3.87
7	EEMD-CC + KNN	8	85.49% \pm 4.38%	28.31

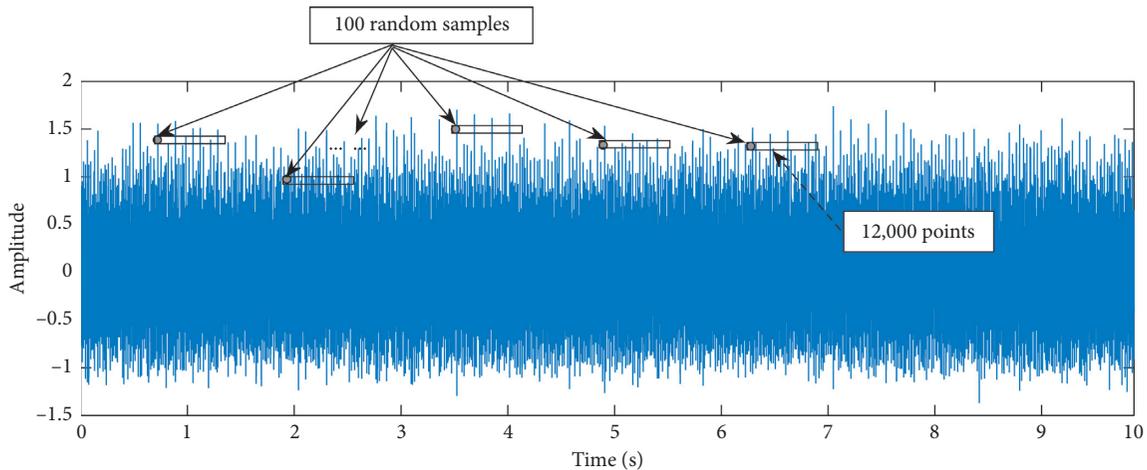


FIGURE 9: Diagram of data sampling.

testing accuracy of each dataset is not less than 99.5%. It is observed from Figure 10 that for samples in set 2 and set 3, testing accuracies with different training sample numbers all reach 100%. The results indicate that the feature extraction capability of the proposed method is stable when the number of training samples changes.

(2) *Fault Diagnosis under Load Fluctuation.* Experimental results under the same load show that the proposed method has a good feature extraction ability for the rolling bearing vibration signals. However, in practice, bearings usually have to work under unstable load. Rotation speed variation appears with the load fluctuations, which brings about

TABLE 7: Detail of datasets (the coding format is motor load-fault location-fault severity, for example, 1-IR-1 means the load is 1 hp, fault in inner race, and the fault width is 0.007 in).

Fault location	Fault width (in)	Motor load (hp)/motor rotation speed (r/min)			
		0/1797	1/1772	2/1750	3/1730
Inner race	0.007	0-IR-1	1-IR-1	2-IR-1	3-IR-1
	0.014	0-IR-2	1-IR-2	2-IR-2	3-IR-2
	0.021	0-IR-3	1-IR-3	2-IR-3	3-IR-3
Ball	0.007	0-BA-1	1-BA-1	2-BA-1	3-BA-1
	0.014	0-BA-2	1-BA-2	2-BA-2	3-BA-2
	0.021	0-BA-3	1-BA-3	2-BA-3	3-BA-3
Outer race	0.007	0-OR-1	1-OR-1	2-OR-1	3-OR-1
	0.014	0-OR-2	1-OR-2	2-OR-2	3-OR-2
	0.021	0-OR-3	1-OR-3	2-OR-3	3-OR-3
Normal	None	0-normal	1-normal	2-normal	3-normal

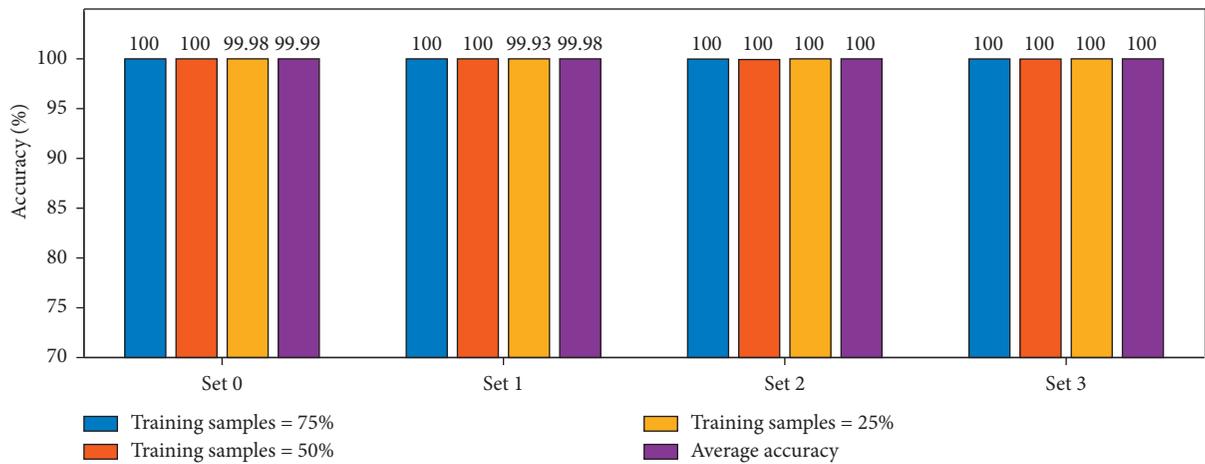


FIGURE 10: Classification accuracies using different training sample numbers.

difficulty for feature extraction [11]. Hence, another experiment was conducted to research the stability of the proposed method under different motor loads. In this experiment, twelve conditions were tested. Condition 0_1 means that the autoencoder in the proposed method is trained on 50% of the samples randomly selected in set 0 and tested on all samples in set 1. After that, the KNN classifier is used to classify the samples in set 1. Fifty percent of the samples are randomly selected for training, and the rest for testing.

Results obtained by repeated experiments are given in Figure 11. It can be observed from Figure 11 that the testing accuracy keeps high under different conditions of motor load fluctuation, and none of them is less than 94%. These results suggest that the proposed method has good stability under load fluctuation.

3.2.3. Comparative Experiments. As discussed previously, a good feature extraction could be the key step of fault diagnosis. Yet, the performance of the classifier used also has an impact on the result of fault diagnosis. It is difficult to judge whether the high accuracy of a trained classifier is due to the classifier itself or to the strong discriminating ability of the extracted features. A good feature extraction method

should be able to extract the most discriminative features to help different classifiers obtain a high classification accuracy.

Here, 25% of samples of each dataset were used to train the AE in the proposed method and the rest were used to test. The latent coding spaces of the AE are displayed in Figure 12, where the coordinate axes are the three features extracted by the proposed SFS-KPCA-AE. As can be seen from Figure 12, the feature space successfully divides the 10 healthy states of the bearing, which indicating that the extracted three features are discriminative, verifying the strong feature extraction ability of the proposed method.

Furthermore, the features extracted by the proposed method were sent to the KNN classifier and SVM classifier, respectively, for classification. The average testing accuracies of SFS-KPCA-AE + KNN and SFS-KPCA-AE + SVM for four datasets are shown in Table 8. It can be seen that the testing accuracy of SFS-KPCA-AE + KNN is 99.93%, and that of SFS-KPCA-AE + SVM is 99.97%. Then, the proposed method was compared with other methods published in the literature using the same dataset. The results of comparison are listed in Table 8.

In reference [36], the experimental samples included different fault locations: inner race, ball, outer race, and different fault widths: 0.007 in and 0.021 in with the same load of 0, 1, 2, or 3 hp. The method was based on ensemble

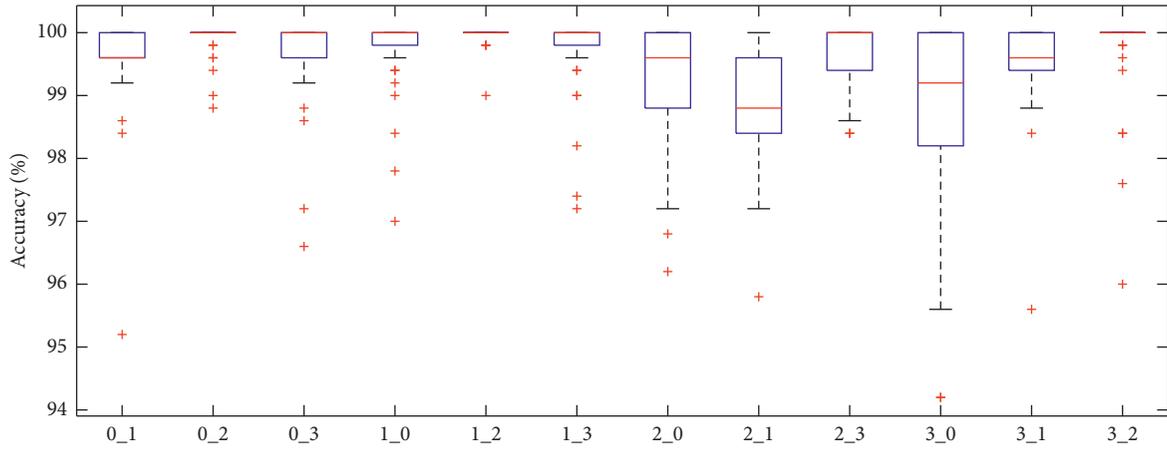


FIGURE 11: Testing accuracy under load fluctuation.

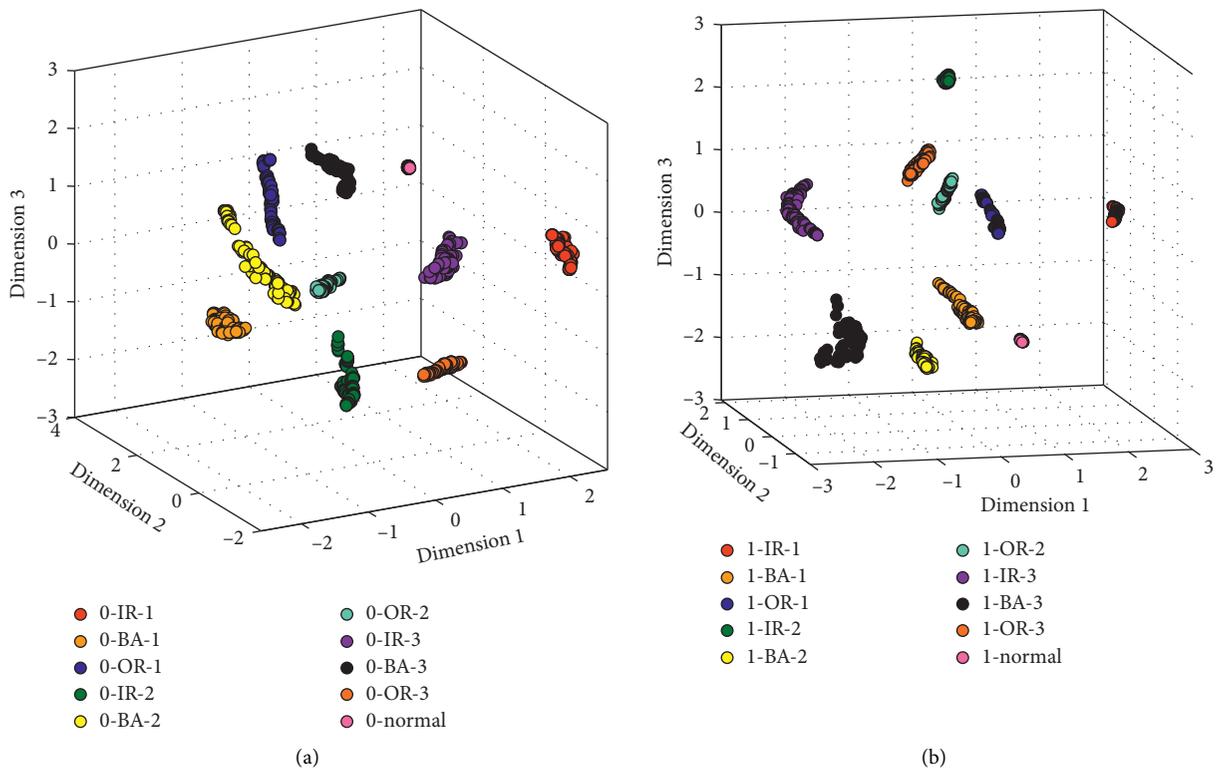


FIGURE 12: Continued.

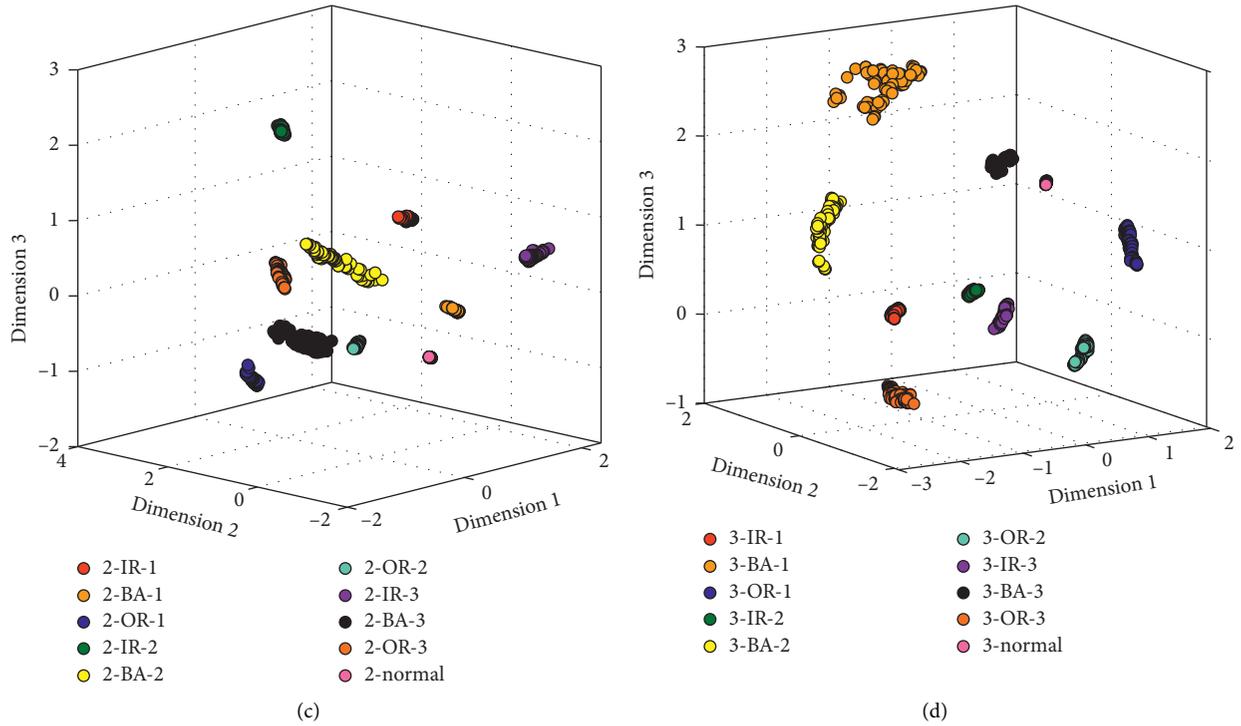


FIGURE 12: Latent coding spaces of the AE (training samples = 25%). (a) Set 0. (b) Set 1. (c) Set 2. (d) Set 3.

TABLE 8: Classification comparison of the bearing dataset.

Methods	Description	No. of classes	Training samples (%)	Testing accuracy (%)
Proposed	SFS-KPCA-AE + KNN	10	25	99.93
Proposed	SFS-KPCA-AE + SVM	10	25	99.97
[36]	EEMD-ICD + SVM	6	25	98.22
[37]	EMD-WKLFDA + SVM	10	40	98.80
[38]	Bispectrum features + SVM	4	50	96.98
[39]	SDA-DNN	10	25	92.35
[40]	GRU-NP-DAE	10	75	99.22
[41]	NSAE-LCN	10	25	99.92

empirical mode decomposition (EEMD) and optimized support vector machines. The average testing accuracy under different load was 98.22%. In reference [37], empirical mode decomposition (EMD), wavelet kernel local fisher discriminant analysis (WKLFDA), and SVM were integrated to distinguish ten health conditions in the bearing dataset and the average testing accuracy reached 98.80%. In reference [38], bispectrum features were used to obtain features. SVM was used to classify four bearing health conditions, and the classification accuracy was 96.98%. In reference [39], the stacked denoising AE-based deep neural network (SDA-DNN) was used to realize the bearing health state identification and the average classification accuracy of 92.35% was obtained. In reference [40], Gated Recurrent Unit-based nonlinear predictive denoising AEs (GRU-NP-DAEs) were constructed to classify different fault types of the bearings, and an average accuracy of 99.22% was achieved. A local connection network (LCN) constructed by the normalized sparse AE (NSAE) was proposed in [41] to learn meaningful

and dissimilar features from raw vibration signals and recognize mechanical health conditions. This method achieved an accuracy of 99.92%.

It is noted that the methods in references [36–38] extract features by using different signal processing technologies, such as EEMD and wavelet analysis, while in references [39–41], features are extracted by using deep-learning techniques, such as DNN and DAE. It is evident that all these methods are thoroughly developed, yet they get lower diagnosis accuracies compared with SFS-KPCA-AE + KNN and SFS-KPCA-AE + SVM. In addition, compared with the proposed method, these methods may rely too much on prior experience to select the most discriminative features or require a lot of time in the training complex model, while the proposed feature extraction method extracts features based on unsupervised learning and is easy to train. The results of comparative experiments demonstrate the robustness and superiority of the proposed SFS-KPCA-AE.

4. Conclusions

To extract the most discriminating fault signature from the frequency spectrum and ease the burden of designing features for fault diagnosis, a feature extraction method, namely, SFS-KPCA-AE, is proposed in this paper. In this method, FFT is first used to acquire the frequency spectrum of vibration signal, later the frequency spectrum is divided into several segments. Next, several principal components considered as local features of the spectrum are extracted from each spectrum segment using KPCA. Then, these local features are connected in sequence to obtain the global feature vector. Finally, the global feature vectors are sent to the constructed AE to extract the most critical and low-dimensional features of the frequency spectrum. A rotor dataset and a bearing dataset are used in the experiments to verify the effectiveness of the proposed SFS-KPCA-AE approach. By discussing the experimental results, three main conclusions are obtained:

- (1) The proposed method integrating the segmented frequency spectrum, KPCA and AE, is based on unsupervised learning techniques, which means that it is independent of labeled data and a lot of human labor to design the most discriminative features.
- (2) The strategy of extracting features from the segments of the frequency spectrum and connecting local features to get the global feature vector can not only obtain both local and global characteristics of the frequency spectrum but also reduce the impact of noise frequency components on fault diagnosis. Plenty of experimental results suggest that the proposed method has good generalization and strong robustness under different training samples, different load conditions, and fluctuating motor loads.
- (3) Features extracted using the proposed method have both a very low dimension and an excellent distinction among different device states. Moreover, the proposed method in combination with different classifiers realizes accurate fault diagnosis for rotating machinery. Through comparison with other methods, the superiority of the proposed SFS-KPCA-AE is demonstrated.

However, since the proposed method is used for feature extraction of the frequency spectrum, the sensitivity to the sampling length of the signal remains to be studied further. In order to adapt to more types of machine vibration signals, it is necessary to further study how to segment the spectrum (evenly or unevenly).

Data Availability

Some data used during the study are available in <https://data.mendeley.com/datasets/p9bsmj4xwg/1> and the Bearing Data Center: <http://csegroups.case.edu/bearingdatacenter/home>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 51979204).

References

- [1] N. Lu, Z. Xiao, and O. P. Malik, "Feature extraction using adaptive multiwavelets and synthetic detection index for rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 393-415, 2015.
- [2] J. Liang, Y. Zhang, J.-H. Zhong, and H. Yang, "A novel multi-segment feature fusion based fault classification approach for rotating machinery," *Mechanical Systems and Signal Processing*, vol. 122, pp. 19-41, 2019.
- [3] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *Ieee Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137-3147, 2016.
- [4] D. Abboud, M. Elbadaoui, W. A. Smith, and R. B. Randall, "Advanced bearing diagnostics: a comparative study of two powerful approaches," *Mechanical Systems and Signal Processing*, vol. 114, pp. 604-627, 2019.
- [5] X. Zhang, Y. Liang, J. Zhou, and Y. zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164-179, 2015.
- [6] J. Y. Chen, D. Zhou, C. Lyu et al., "An integrated method based on CEEMD-SampEn and the correlation analysis algorithm for the fault diagnosis of a gearbox under different working conditions," *Mechanical Systems and Signal Processing*, vol. 113, pp. 102-111, 2018.
- [7] H. Ao, J. S. Cheng, K. L. Li et al., "A roller bearing fault diagnosis method Based on LCD energy entropy and ACROA-SVM," *Shock and Vibration*, vol. 2014, Article ID 825825, 12 pages, 2014.
- [8] X. Yan and M. Jia, "Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 56-86, 2019.
- [9] Q. Jiang, Y. Shen, H. Li, and F. Xu, "New fault recognition method for rotary machinery based on information entropy and a probabilistic neural network," *Sensors*, vol. 18, no. 2, pp. 337-349, 2018.
- [10] C. Malla and I. Panigrahi, "Review of condition monitoring of rolling element bearing using vibration analysis and other techniques," *Journal of Vibration Engineering & Technologies*, vol. 7, no. 4, pp. 407-414, 2019.
- [11] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412-424, 2018.
- [12] P. F. Liang, C. Deng, J. Wu et al., "Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform," *Computers In Industry*, vol. 113, pp. 1-10, 2019.
- [13] Y. Lu, R. Xie, and S. Y. Liang, "CEEMD-assisted kernel support vector machines for bearing diagnosis," *The International Journal of Advanced Manufacturing Technology*, vol. 106, no. 7-8, pp. 3063-3070, 2020.

- [14] H. Li, T. Liu, X. Wu et al., "Application of optimized variational mode decomposition based on kurtosis and resonance frequency in bearing fault feature extraction," *Transactions of the Institute of Measurement And Control*, vol. 42, no. 3, pp. 518–527, 2019.
- [15] H. Wang, B. Ren, L. Song, and L. Cui, "A novel weighted sparse representation classification strategy based on dictionary learning for rotating machinery," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 712–720, 2020.
- [16] H. Wang, S. Li, L. Song et al., "An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2648–2657, 2019.
- [17] M. Sohaib, C.-H. Kim, and J.-M. Kim, "A hybrid feature model and deep-learning-based bearing fault diagnosis," *Sensors*, vol. 17, no. 12, pp. 2876–2891, 2017.
- [18] W. Zhang, G. L. Peng, C. H. Li et al., "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 21, 2017.
- [19] Y. Zhang, K. S. Xing, R. X. Bai et al., "An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image," *Measurement*, vol. 157, p. 12, 2020.
- [20] J. Yuan and X. Liu, "Semi-supervised learning and condition fusion for fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 615–627, 2013.
- [21] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [22] B. Zhou and X. Gu, "Multi-block statistics local kernel principal component analysis algorithm and its application in nonlinear process fault detection," *Neurocomputing*, vol. 376, pp. 222–231, 2020.
- [23] X. Deng, X. Tian, and S. Chen, "Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis," *Chemo-metrics and Intelligent Laboratory Systems*, vol. 127, pp. 195–209, 2013.
- [24] P. Cui, C. Zhan, and Y. Yang, "Improved nonlinear process monitoring based on ensemble KPCA with local structure analysis," *Chemical Engineering Research and Design*, vol. 142, pp. 355–368, 2019.
- [25] J. Yu, "Local and global principal component analysis for process monitoring," *Journal of Process Control*, vol. 22, no. 7, pp. 1358–1373, 2012.
- [26] J.-M. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 59, no. 1, pp. 223–234, 2004.
- [27] R. Katuwal and P. N. Suganthan, "Stacked AE based deep random vector functional link neural network for classification," *Applied Soft Computing*, vol. 85, pp. 1–12, 2019.
- [28] P. Park, P. D. Marco, H. Shin, and J. Bang, "fault detection and diagnosis using combined autoencoder and long short-term memory network," *Sensors*, vol. 19, no. 21, pp. 4612–4628, 2019.
- [29] E. Principi, D. Rossetti, S. Squartini, and F. Piazza, "Unsupervised electric motor fault detection by using deep autoencoders," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 2, pp. 441–451, 2019.
- [30] A. P. Valentine and J. Trampert, "Data space reduction, quality assessment and searching of seismograms: autoencoder networks for waveform data," *Geophysical Journal International*, vol. 189, no. 2, pp. 1183–1202, 2012.
- [31] D. Liu, Z. Xiao, X. Hu, C. Zhang, and O. P. Malik, "Feature extraction of rotor fault based on EEMD and curve code," *Measurement*, vol. 135, pp. 712–724, 2019.
- [32] L. M. Bachmann, M. A. Puhon, G. T. Riet, and P. M. Bossuyt, "Sample sizes of studies on diagnostic accuracy: literature survey," *Bmj*, vol. 332, no. 7550, pp. 1127–1129, 2006.
- [33] S. Y. Kim, "Effects of sample size on robustness and prediction accuracy of a prognostic gene signature," *BMC Bioinformatics*, vol. 10, pp. 1–10, 2009.
- [34] X. Xue and J. Zhou, "A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery," *ISA Transactions*, vol. 66, pp. 284–295, 2017.
- [35] Bearing Data Center, <http://csegroups.case.edu/bearingdatacenter/home>.
- [36] X. Zhang and J. Zhou, "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines," *Mechanical Systems and Signal Processing*, vol. 41, no. 1–2, pp. 127–140, 2013.
- [37] M. Van and K. Hee-Jun, "Bearing defect classification based on individual wavelet local Fisher discriminant analysis with particle swarm optimization," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 124–135, 2016.
- [38] L. Saidi, J. Ben Ali, and F. Fnaiech, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Transactions*, vol. 54, pp. 193–206, 2015.
- [39] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol. 130, pp. 377–388, 2017.
- [40] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Transactions*, vol. 77, pp. 167–178, 2018.
- [41] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, 2018.