

Research Article

An Enhancement Deep Feature Extraction Method for Bearing Fault Diagnosis Based on Kernel Function and Autoencoder

Fengtao Wang , Bosen Dun , Xiaofei Liu , Yuhang Xue ,
Hongkun Li , and Qingkai Han 

School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Fengtao Wang; wangft@dlut.edu.cn

Received 22 November 2017; Accepted 17 January 2018; Published 27 February 2018

Academic Editor: Murat Inalpolat

Copyright © 2018 Fengtao Wang 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.

Rotating machinery vibration signals are nonstationary and nonlinear under complicated operating conditions. It is meaningful to extract optimal features from raw signal and provide accurate fault diagnosis results. In order to resolve the nonlinear problem, an enhancement deep feature extraction method based on Gaussian radial basis kernel function and autoencoder (AE) is proposed. Firstly, kernel function is employed to enhance the feature learning capability, and a new AE is designed termed kernel AE (KAE). Subsequently, a deep neural network is constructed with one KAE and multiple AEs to extract inherent features layer by layer. Finally, softmax is adopted as the classifier to accurately identify different bearing faults, and error backpropagation algorithm is used to fine-tune the model parameters. Aircraft engine intershaft bearing vibration data are used to verify the method. The results confirm that the proposed method has a better feature extraction capability, requires fewer iterations, and has a higher accuracy than standard methods using a stacked AE.

1. Introduction

Effective health diagnosis of rolling bearing is a significant initiative in today's industry. The bearing of rotating machinery will inevitably experience various faults under harsh working conditions such as large loads, strong impacts, and high speed [1]. The faults may lead to serious casualties if it has not been seasonably detected. Therefore, it is crucial to accurately and automatically diagnose the different faults before they cause serious damage.

The method based on vibration signal has been widely studied and applied in virtue of vibration signals usually carrying rich information [2], and the intelligent diagnosis method is especially considered in recent years. Intelligent fault diagnosis of rotating machinery is a type of pattern recognition problem consisting of three steps, including data preprocessing, feature extraction and selection, and fault classification [3, 4]. First, the raw data collected by the sensor are preprocessed. Then the time-domain, frequency-domain, and time-frequency-domain features are extracted

and selected manually. Finally, a classifier is applied using these features to provide a fault diagnosis. The rapid development of artificial intelligence has resulted in an increased use of machine learning methods for bearing fault diagnosis and examples include artificial neural network (ANN) and support vector machine (SVM) methods. Bin et al. [5] utilized wavelet packets-empirical mode decomposition to decompose the original signal and extracted statistical features as multilayer perceptron network for fault classification. Zhang et al. [6] extracted nineteen statistical features from the measured vibration signals as inputs for SVM to recognize the roller bearing operation conditions. In [7], the signal is firstly filtered by morphological filter and then decomposed by the empirical mode decomposition (EMD) method; the extract features are mapped into the LTSA to extract the character features used as an input to an SVM for diagnosis. However, the traditional neural network method requires manual feature selection, which requires considerable theoretical knowledge and practical experience. The widely used ANN and SVM methods represent supervised learning models

with a shallow structure that lacks sufficient representation of the fault features and a large number of labeled data are required [8].

Deep learning represents a novel pattern recognition approach and was proposed by Hinton and Salakhutdinov in 2006 [9]; the method has developed rapidly in recent years. In addition to being used for image recognition and speech recognition, this method has resulted in breakthroughs in the field of bearing fault diagnosis. Due to the multilayered structure of deep learning, it can derive fault information of the bearings from historical data and provide an accurate assessment. Jia et al. [10] used a denoising AE to construct a deep learning model, selected the Fourier coefficient of the original signal as an input, and achieved good results according to the fault diagnosis test of rolling bearings and gears. Shao et al. [11] combined a contractive autoencoder with the denoising autoencoder, and adopted the locality preserving projection algorithm to carry out the feature fusion and achieved a good performance. Chen and Li [12] used a sparse AE to fuse the features of multiple sensors and developed a deep belief network (DBN) to conduct a fault diagnosis of rolling bearings. AEs are a widely used model in deep learning and are capable of extracting deep features from unlabeled data using a multilayer coding process. However, for many applications, the model training is difficult and the performance is affected by the structure of the hidden layer and the number of iterations. Moreover, manual signal processing or feature selection is still required for a bearing fault diagnosis [13, 14], which reduces the applicability of the deep learning method. In order to further improve the diagnostic performance of the deep learning network and improve the model applicability, a kernel function method is applied to the AE.

The kernel function is an effective method utilized in machine learning to solve the nonlinear problems; examples include SVM [15] and radial basis function (RBF) neural network [16]. The input space is mapped to a high-dimensional feature space using a nonlinear transform and the calculation in the high-dimensional features space is performed using a kernel function; this approach reduces the computational complexity. Some conventional methods have been improved by various approaches of the kernel function method, including the kernel principal component analysis (KPCA) [17], the kernel independent component analysis (KICA) [18], and the kernel discriminant analysis (KDA) [19]. Hence, a novel KAE network based on a kernel function combined with an AE is proposed.

In this study, an enhancement deep feature extraction method is developed with one KAE and N AEs; the input data is mapped to a high-dimensional space and a coding network is used for coding the calculations of the high-dimensional data by KAE; then the N AEs are carried out to extract deep feature layer by layer. Three experiments were applied to validate the proposed method on an aircraft engine intershaft bearing test rig. The results show that the proposed method is capable of extracting the features of the bearing faults with fewer iterations and with a higher accuracy compared to standard methods.

The remainder of this paper is organized as follows. Section 2 presents the fundamental theory of autoencoder

and stacked autoencoder network. Section 3 presents the proposed method and the procedure. Section 4 discusses the fault diagnosis result in three experiments and features extraction capability which are visualized by principal component analysis method. The conclusions are given in Section 5.

2. The Fundamental Theory

2.1. Autoencoder. A common AE is a three-layer network consisting of an encoder network and a decoded network. The encoder network connects the input layer and the hidden layer, which can obtain the features of the original data. The hidden layer and the output layer are connected by the decoder network that reconstructs the output, which is equal to the input based on the low-dimensional coding data.

The encoder network is defined as an encoding function denoted by f_θ [20]. For the training sample $\mathbf{x} = [x_1, x_2, \dots, x_N]^T \in \mathfrak{R}^N$, the encoder takes the input vector \mathbf{x} nonlinear mapping to a hidden representation $\mathbf{h} = [h_1, h_2, \dots, h_M]^T \in \mathfrak{R}^M$ through f_θ :

$$\mathbf{h} = f_\theta(\mathbf{x}) = s_f(\mathbf{W}\mathbf{x} + \mathbf{b})$$

$$s_f(t) = \frac{1}{(1 + e^{-t})}, \quad (1)$$

where $s_f(t)$ is the activation function of encoder, and the parameter set of encoder is $\theta = \{\mathbf{W}, \mathbf{b}\}$, where \mathbf{W} is the weight matrix and \mathbf{b} is bias vector.

The decoder network is defined as a reconstruction function denoted by $g_{\theta'}$. It maps \mathbf{h} transform back into a reconstruction vector $\mathbf{z} = [z_1, z_2, \dots, z_N]^T \in \mathfrak{R}^N$:

$$\mathbf{z} = g_{\theta'}(\mathbf{h}) = s_f(\mathbf{W}'\mathbf{h} + \mathbf{b}'), \quad (2)$$

where $\theta' = \{\mathbf{W}', \mathbf{b}'\}$ are the parameter set of encoder, weight matrix $\mathbf{W}' = \mathbf{W}^T$, and \mathbf{b}' is bias vector.

The parameter set $\theta = \{\theta, \theta'\} = \{\mathbf{W}, \mathbf{b}, \mathbf{W}', \mathbf{b}'\}$ of the autoencoder is optimized to minimize the reconstruction error through the training process

$$\begin{aligned} J_{\text{AE}}(\theta) &= L(\mathbf{x}, \mathbf{z}) \\ &= -\sum_{i=1}^N [x_i \log(z_i) + (1 - x_i) \log(1 - z_i)], \end{aligned} \quad (3)$$

where $L(\mathbf{x}, \mathbf{z})$ is the loss function that means the discrepancy between \mathbf{x} and \mathbf{z} .

When the loss function is sufficiently small, it can be assumed that the coding vector is capable of reconstructing the original input vector; that is, most of the information contained in the original data is included in the encoding vector. The automatic coding network is also a nonlinear reduction method, in which the dimensions are lower for the encoding vector than the input vector.

2.2. Stacked Autoencoder Network. An AE is an unsupervised three-layer learning network but its information extraction

ability is limited and it lacks sufficient structure to represent the deep characteristics of the signal. A stacked autoencoder (SAE) uses multiple AE layers to develop more hidden layers; each AE layer performs a nonlinear transformation of the input samples from the preceding layer to the following one. During the training process, the hidden layers of the AE layers represent the inputs to the succeeding AE layer and the network uses an unsupervised learning algorithm layer by layer to extract the features from the input data.

The input layer and the first hidden layer of the SAE are regarded as the encoder network of the first AE. After the first AE is trained through minimizing the reconstruction error in (3), the first encode vector \mathbf{h}_1 of the \mathbf{x} is calculated as follows:

$$\mathbf{h}_1 = f_{\theta_1}(\mathbf{x}), \quad (4)$$

where θ_1 is the parameter set of the first AE.

Then the encode vector \mathbf{h}_1 is treated as input data; the first hidden layer and the second hidden layer of the SAE are regarded as the encoder network of the second AE. The process is conducted in the sequence until the N th AE is trained for initializing the final hidden layer of the SAE. And the N th encode vector \mathbf{h}_N is calculated as

$$\mathbf{h}_N = f_{\theta_N}(\mathbf{h}_{N-1}), \quad (5)$$

where θ_N is the parameter set of the N th AE.

Subsequently, a backpropagation (BP) algorithm is used to fine-tune the network parameters using a supervised approach. The SAE is a type of deep neural network (DNN) that combines a supervised and an unsupervised approach [21, 22].

3. The Proposed Method

3.1. Autoencoder Based on a Kernel Function. The encoding process of the AE is a nonlinear calculation but for low-dimensional raw data an accurate classification requires a large number of iterations and a long calculation time; this approach is also prone to misclassifications. In order to solve this problem, the kernel function method is combined with the AE.

A kernel function is defined as a nonlinear mapping ϕ from the input space R to the characteristic space H ; for all $x, y \in R$, function $K(x, y) : K(x, y) = K(\phi(x), \phi(y))$. Because the calculation of the hidden layer in the feature space is provided by the kernel function, a specific mapping relationship does not have to be defined, which greatly reduces the computational complexity of the problem.

Based on the above theory, an improved method KAE, which combines the kernel function and the AE, is proposed. First, the Gram matrix of the kernel function is calculated and its input is the new automatic encoder; the coding process changes to

$$\mathbf{h} = s_f [\mathbf{W}^{(1)} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{b}^{(1)}], \quad (6)$$

where $\mathbf{x}_i, \mathbf{x}_j$ are any two samples of the training data.

Correspondingly, the decoding function is changed to

$$\mathbf{K}(\mathbf{z}_i, \mathbf{z}_j) = s_g (\mathbf{W}^{(2)} \mathbf{h} + \mathbf{b}^{(2)}). \quad (7)$$

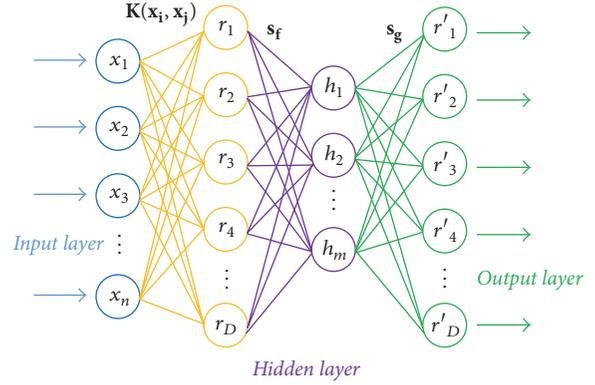


FIGURE 1: The structure of KAE.

The improved AE network firstly maps the data to a high-dimensional space; then the high-dimensional data are coded and calculated and the nonlinear low-dimensional features are obtained. By adding the kernel functions, the original signal components are mapped to a high-dimensional space, which speeds up the coding process and improves the efficiency of extracting the signal characteristics and the classification accuracy. The algorithm structure diagram is shown in Figure 1.

The proposed method can be summarized as follows: choose the KAE as the first layer of the deep network and use the hidden layer of the KAE as the input of the next layer of the AE. Connect multiple AE layers to form a deep network and use the BP algorithm to fine-tune the parameters and obtain the diagnosis model.

3.2. Selection of the Kernel Function. Based on the Mercer theorem, any semidefinite function can be used as a kernel function. Common kernel functions include the linear kernel function, polynomial kernel function, and radial basis kernel function. The radial basis function is a real-valued function that only depends on distance. The Gaussian radial basis kernel function uses the Euclidean distance as the distance function; the transformation matrix is qualitatively good and only has one undetermined parameter; as a result, the complexity of the model is low. Therefore, the Gaussian radial basis function is used in this study; its mathematical expression is as follows:

$$K(\mathbf{x}, \mathbf{y}) = \exp \left[-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{(2\sigma^2)} \right], \quad (8)$$

where σ is an independent variable; it indicates the width of the nucleus.

The Gaussian kernel function has only one pending parameter but its performance is directly dependent on the choice of the nuclear parameters [23] because the kernel function and its parameters directly determine the corresponding feature space by nonlinear mapping. When an inappropriate kernel function or nuclear parameters are selected, the accuracy of the results decreases compared with the results of the original space [24]. Therefore, choosing proper nuclear parameters is of key importance when using Gaussian kernel

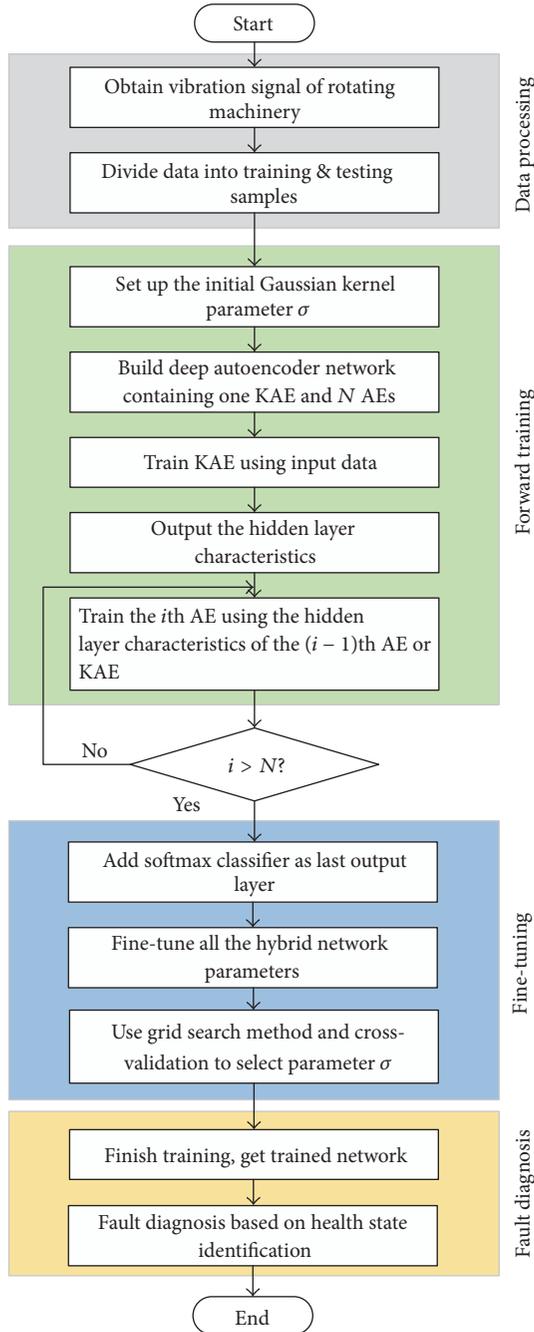


FIGURE 2: The flowchart of the proposed method.

functions. In this study, we use grid-searching and cross-validation to search for the optimal parameters in the search optimization space and to determine the parameters of the kernel function.

3.3. Procedure of the Proposed Method. In this study, an enhanced deep learning method is developed for the fault diagnosis of rotating machinery. A flowchart of the proposed method is shown in Figure 2.

The fault diagnosis process takes place in a stepwise manner.

Step 1. The original vibration signal of the bearing is obtained and the selected signal is divided into samples.

Step 2. The Gaussian kernel parameters σ are initialized, and the number of AE layers are determined.

Step 3. The input data are used to train the KAE and the output data are used to train the next AE.

Step 4. N AEs are trained in the unsupervised method and the hidden layers of the AE are the input for the subsequent AE layer until the training is completed.

Step 5. The output layer is classified and the reverse error propagation algorithm is used to fine-tune the network parameters.

Step 6. The kernel parameters σ are optimized by grid-searching and cross-validation until the network training is completed.

4. Experimental Verification

4.1. Test Rig for the Aircraft Engine Intershaft Bearing. The intershaft bearing is one of the key components of an aircraft engine. In order to verify the proposed method for the bearing fault diagnosis, a test rig for aircraft engine intershaft bearing based on a double rotor is used to simulate the different fault types of the bearing. Subsequently, the vibration signal data are analyzed. The test rig is shown in Figure 3 and consists of two motors, a rotor, three mass disks, and four accelerometers.

The intershaft bearing is fixed at the joint of the low-voltage axis and the high-voltage axis and is connected to two motors. The bearing's outer ring is connected to the high-voltage end of the shaft and the inner ring is connected to the low-voltage end of the shaft. Four acceleration sensors are installed to collect the vibration signal of the intershaft bearing on the support bearing pedestal of high and low-voltage axis. The hardware acquisition system uses an NI acquisition card to collect the data and the sampling frequency is 25.6 K.

In this case, ten operating conditions are considered, including the inner race fault, outer race fault, and roller fault. The faults are introduced to the intershaft bearing under the running conditions of high-voltage-motor single rotation (HR), low-voltage-motor single rotation (LR), and high-voltage-motor/low-voltage-motor relative rotation (HLR), respectively. In addition, a normal condition of a two-motor relative rotation is tested. The rotation speed of the motors is 20 Hz. The artificial axial crack fault sizes of inner race, outer race, and roller are all 2.0 mm in width and 0.8 mm in depth. The fault grooves in the bearing are machined by an electric spark, as shown in Figure 4. The outer ring fault is displayed by taking one of roller elements because the outer race and the holder cannot be removed.

Due to the distance of the sensors from the bearing, the bearing signals will attenuate during transmission and contain noise. In this study, we chose the closest vertical acceleration sensor installed on the high-pressure axis bracket seat of the intershaft bearing for recoding the vibration of

TABLE 1: Description of the intershaft operation conditions.

	Bearing operating condition	Motor speed (Hz) HR/LR	Size of training/testing samples	Label
ILR	Inner fault (LR)	0/20	150/50	1
IHR	Inner fault (HR)	20/0	150/50	2
IHLR	Inner fault (HLR)	20/20	150/50	3
OLR	Outer fault (LR)	0/20	150/50	4
OHR	Outer fault (HR)	20/0	150/50	5
OHLR	Outer fault (HLR)	20/20	150/50	6
RLR	Roller fault (LR)	0/20	150/50	7
RHR	Roller fault (HR)	20/0	150/50	8
RHLR	Roller fault (HLR)	20/20	150/50	9
NHLR	Normal (HLR)	20/20	150/50	10

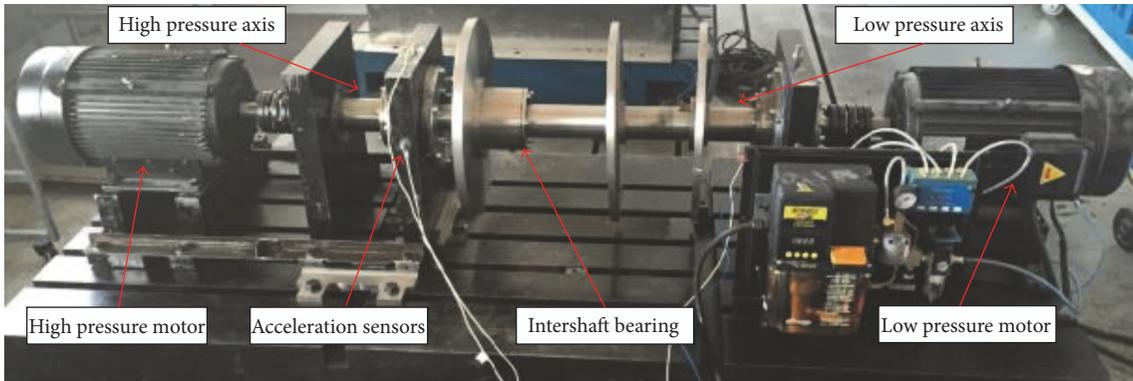


FIGURE 3: The test rig of aeroengine intershaft bearing.



FIGURE 4: Faults in the intershaft bearing.

the signal data. The experiment was performed four times under the same conditions and each condition consisted of 10 seconds of data. The data was divided into 200 samples, in which each sample is a measured vibration signal consisting of 1200 sampling data points. Random three-quarters of the data were randomly selected to serve as the training set and the remaining 1/4 was used as the test set, the details of the samples are shown in Table 1.

In this case study, three experiments are taken into account. In order to verify the diagnosis result, the standard SAE and standard DBN method are compared. A polynomial kernel function (PK) and a power exponent kernel function (PEK) are also used for comparison.

Experiment 1. The raw vibration data were used as the input without performing any signal preprocessing or manual feature extraction.

Experiment 2. The fast Fourier transformation was implemented on each signal to get the 1200 Fourier coefficients. Then the Fourier coefficients are used as input to feed into different methods for fault classification.

Experiment 3. The eighteen statistical features, same as [12], are manually extracted from each signal. Then the extracted features are used as input to feed into different methods for fault classification.

4.2. Diagnosis Results and Analysis. Four trials are carried out for the diagnosis and the results of the different methods are shown in Figure 5. The multiclass confusion matrix of the proposed method in Experiment 1 for the second trials is shown in Figure 6. The multiclass confusion matrix demonstrates the details of the classification results for all conditions and shows the classification accuracy and the misclassification errors. The ordinate axis of the confusion matrix refers to the actual classification labels and the horizontal axis shows the predict classification labels. The color bar on the right illustrates the correspondence between the colors and the numbers from 0 and 1 [25]. The average testing accuracies and standard deviations of four experiments are shown in Table 3.

In Experiment 1, the main parameters of different methods are listed in Table 2. The structure and network

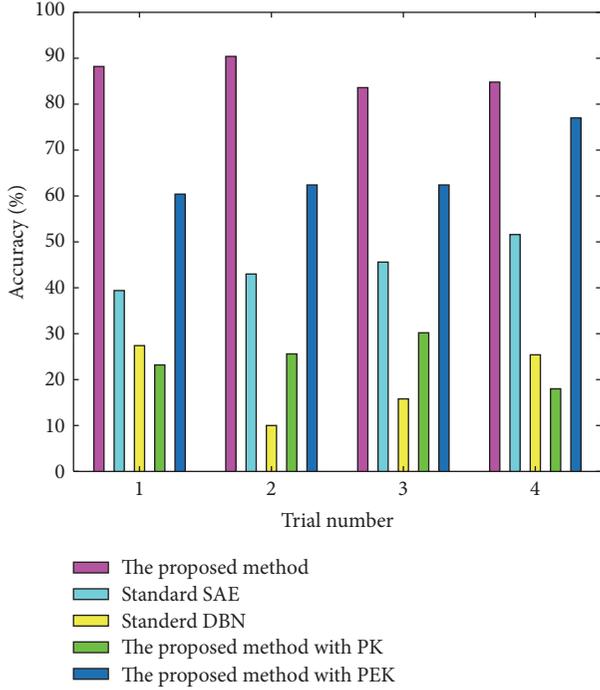


FIGURE 5: Detailed diagnosis results of the 4 trials in Experiment 1.

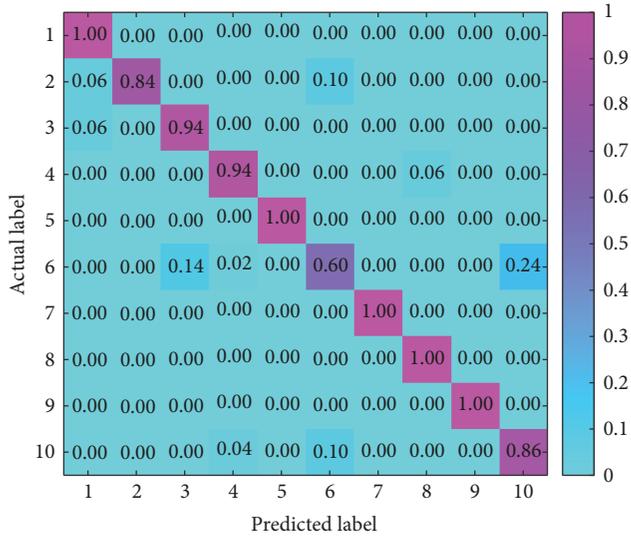


FIGURE 6: Multiclass confusion matrix of the proposed method for the second trial.

parameters of deep learning model are a great challenge, and there is not a mature method in theory to select the optimal structure of deep learning models at present [26]. In this study, we developed the network consistent with [10] and the architecture of the network and the parameter set is determined after several tests.

It can be seen from Figure 5 that the diagnostic accuracy of the proposed method in the four trials ranges from 83.6% to 90.4%, indicating that the proposed method accurately identifies the 10 healthy states from the raw vibration data

TABLE 2: Parameter description of the five methods in Experiment 1.

Methods	Parameter description
The proposed method	The network structure parameters are 1200-800-100-20-10, learning rate is 0.3, momentum is 0.5, training iteration number is 10, fine-tuning iterations are 50, the Gaussian kernel parameter is 26.40.
Standard SAE	The network structure parameters are 1200-800-100-20-10, learning rate is 0.3, momentum is 0.5, training iteration number is 10, fine-tuning iterations are 50.
Standard DBN	The network structure parameters are 1200-800-100-20-10, learning rate is 0.3, momentum is 0.5, training iteration number is 10, fine-tuning iterations are 50.
The proposed method with PK	The network structure parameters are 1200-800-100-20-10, learning rate is 0.3, momentum is 0.5, training iteration number is 10, fine-tuning iterations are 50, the PK parameters are $b = 0$, $d = 1.20$.
The proposed method with PEK	The network structure parameters are 1200-800-100-20-10, learning rate is 0.3, momentum is 0.5, training iteration number is 10, fine-tuning iterations are 50, the PEK parameter is 30.00.

under three different conditions and four different fault locations. However, the average classification accuracy of the SAE is below 50% and the accuracy of the DBN network with the same structure is in the range of 10–27.4%. The standard deviation of the proposed method in the four trials is 2.70%, 4.45% for the SAE method and 7.09% for the DBN method (Table 3). It can be seen from Figure 5 and Table 3 that the proposed method has a better stability compared with the two traditional deep learning methods when using raw data as input. The main reason is that the complex raw data have better separability in a high-dimensional space and are more reliable for learning robust fault characteristics from the measured vibration signals.

The classification accuracy of the PK is in the range of 18–30.2%, which is lower than the value for the standard SAE method. The classification accuracy of the PEK for the three trials' accuracy is about 60% and 77% for the fourth test. This is higher than the accuracy of the traditional SAE, but lower than the accuracy of the Gaussian kernel function method; it also has greater fluctuations.

Figures 7 and 8 are the diagnosis results of the 4 trials in Experiments 2 and 3. From Table 2 we can find that the average testing accuracy of the proposed method is 99.50% and 98.00% in Experiments 2 and 3 which is higher than other methods, while standard SAE and standard DBN are greatly influenced by input strategy. When we use Fourier coefficients as input in Experiment 2, the standard SAE and

TABLE 3: The mean and standard deviation of the test results.

Method	Average of accuracy (%)			Standard deviation of accuracy (%)		
	Ex 1	Ex 2	Ex 3	Ex 1	Ex 2	Ex 3
The proposed method	86.75	99.50	98.00	2.70	0.75	2.01
Standard SAE	44.90	96.35	43.45	4.45	1.40	5.25
Standard DBN	19.65	92.55	12.50	7.09	2.46	4.33
The proposed method with PK	24.25	87.30	86.15	4.40	0.87	6.15
The proposed method with PEK	65.55	98.80	97.15	6.66	0.93	2.23

Ex 1: Experiment 1; Ex 2: Experiment 2; Ex 3: Experiment 3.

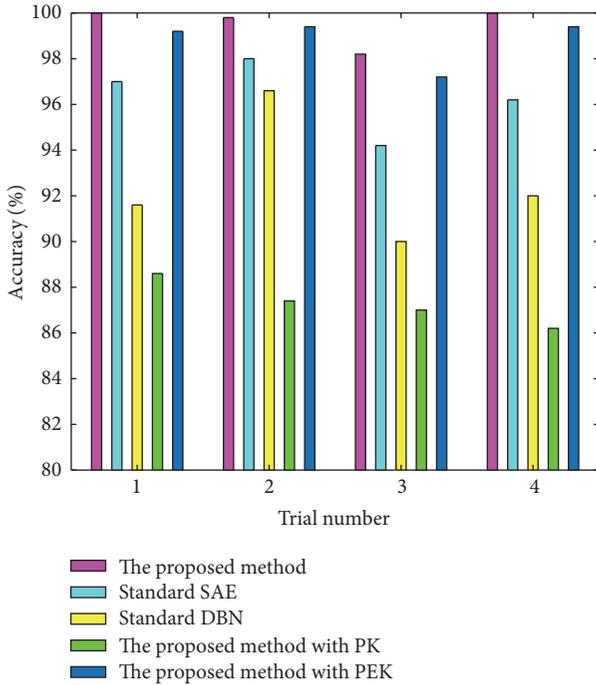


FIGURE 7: Detailed diagnosis results of the 4 trials in Experiment 2.

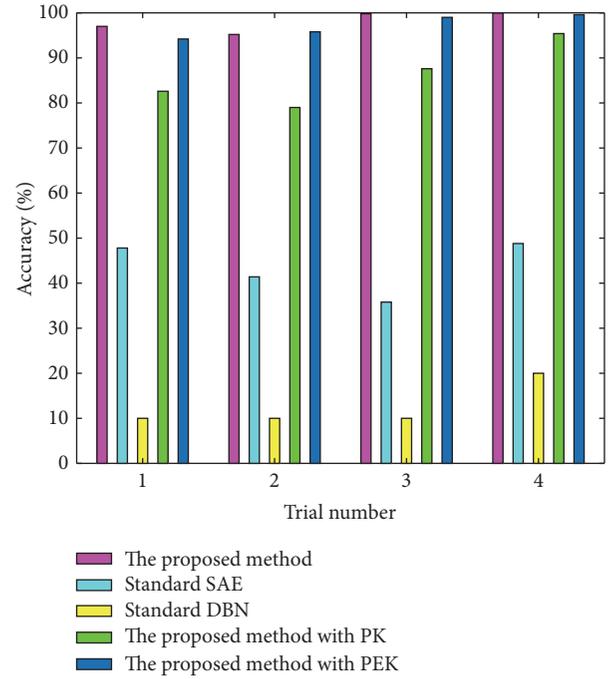


FIGURE 8: Detailed diagnosis results of the 4 trials in Experiment 3.

DBN diagnosis accuracy are 96.35% and 92.55% but 43.45% and 12.50% in Experiment 3. We can conclude that the proposed method has greater applicability in different input strategy than standard SAE and DBN when the signal has much noise.

The average accuracy of the PK is 87.30% in Experiment 2 which is lowest of all methods and 86.15% in Experiment 3. The classification accuracy of the PEK in Experiments 2 and 3 is 98.80% and 97.15% which is slightly lower than the proposed method but higher than the standard SAE and DBN. Above all, the results show that the choice of the kernel function influences the diagnosis; therefore, choosing an appropriate kernel function improves the diagnostic accuracy. In addition, Gaussian kernel function can be used to obtain the best results.

In Experiment 2, the proposed method network structure parameters are 1200-800-100-20-10, the learning rate is 0.3, the momentum is 0.5, the number of training iterations is 10,

the number of fine-tuning iterations is 50, and the Gaussian kernel parameter is 23.35. In comparison, the architectures of the standard SAE, standard DBN, and the proposed method with PK and PEK are all the same as the proposed method. The PK parameters are $b = 0$, $d = 1.10$ and the PEK parameter is 27.52.

In Experiment 3, the proposed method network structure parameters are 1200-800-100-20-10, the learning rate is 0.3, the momentum is 0.5, the number of training iterations is 10, the number of fine-tuning iterations is 50, and the Gaussian kernel parameter is 1.15. In comparison, the architectures of the proposed method with PK and PEK are all the same as the proposed method. The PK parameters are $b = 0$, $d = 1.20$; and the PEK parameter is 0.54. The architecture of the network and the parameter set are determined after several tests. The structure parameters are 18-30-20-10-10 of standard SAE and DBN, the network parameters are the same as the proposed method.

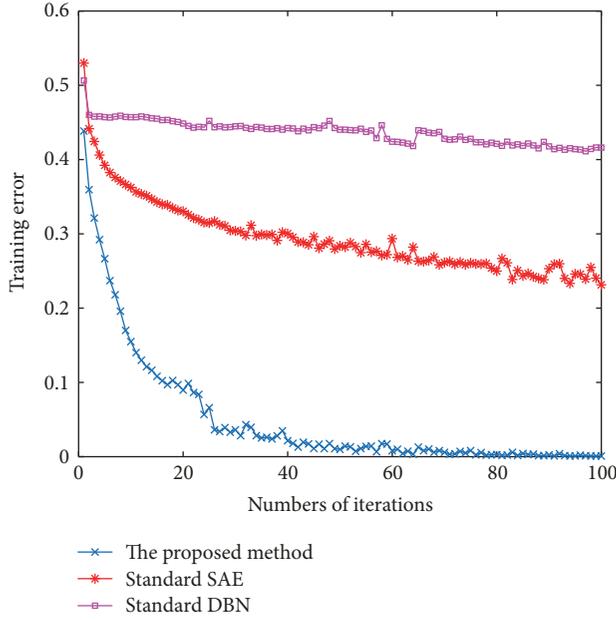


FIGURE 9: Curves of the training error of the proposed method, standard SAE method, and standard DBN method.

To further investigate the reasons behind the higher accuracy of the proposed method in Experiment 1, the relationship between the training errors and the number of iterations is analyzed for the traditional SAE, the DBN, and the proposed method. As shown in Figure 9, for 41 iterations, the training error of the proposed method is below 0.02. Although there are certain fluctuations in the training error as the number of iterations increases, the error eventually converges to zero. However, for the traditional SAE method, the error decreases but does not reach a steady state, even with 100 iterations. For 30–100 iterations, the error decreases by only 0.0726; this indicates that the traditional SAE method is not sensitive to the original signal characteristics and that a large number of iterations are needed to reduce the training error. During the iterative process of the DBN method, the error basically remains stable, indicating that the method achieves a local optimal state during the early stage and is unable to classify the original signal. The results show that the proposed method achieves a lower training error after a few iterations and that the results are stable and better than those of the traditional SAE method.

The accuracy rates of the proposed method, standard SAE, and DBN are all very high in Experiment 2, so the relationship between accuracy and times of fine-tuning iteration is analyzed. In Figure 10, the proposed method can get 86.40% accuracy rate in 5 iterations, but 43.60% and 61.00% for SAE and DBN. Meanwhile, when the iterations are more than 30, the accuracy of the proposed method is stable at 100%, while the standard SAE and DBN have small fluctuations with the iterations growth; this indicates that the proposed method needs fewer iterations to reach a steady diagnosis model than standard SAE and DBN.

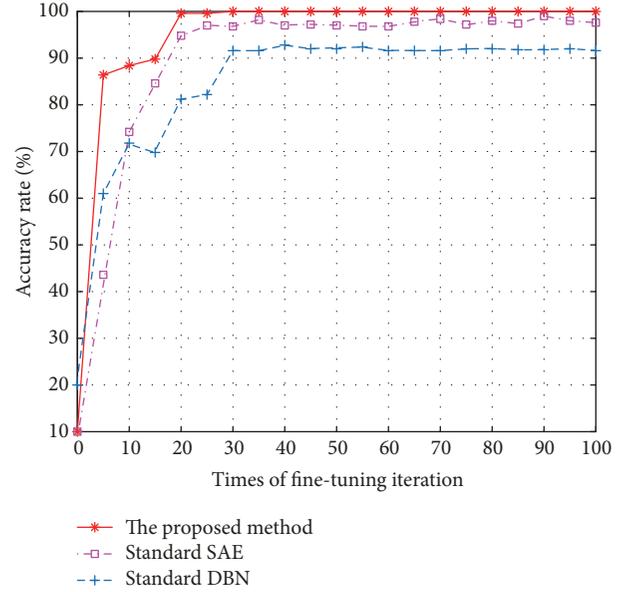


FIGURE 10: Curves of the accuracy and fine-tuning iteration of the proposed method, standard SAE, and standard DBN.

4.3. Feature Extraction Capability Evaluation. In order to verify the feature extraction ability of the proposed method, the principal components (PCs) of the last AE layer are extracted using the principal component analysis method (PCA).

As shown in Figure 12, the features for the same state are well gathered, the clustering center is clear, and the state of the different categories can be effectively separated; so the accuracy is close to 90% in Experiment 1. In contrast, Figure 11 shows that, for the same state, the feature distribution range of the SAE is large, there is no regular pattern of aggregation, and the characteristics of the different categories are intermixed. The traditional SAE method is not suitable for feature extraction for the raw data and does not produce effective results. The results show that the feature extraction ability is poor for the traditional SAE method when the original signal is used as the input and when a limited number of iterations are used because the raw data contain too much noise and interference. Because the kernel function can be regarded as a measure of similarity in the eigenspace, the undivided information in the raw data can result in a better representation in a high eigenspace, which allows for extracting the state characteristics of the raw data. The result shows that the feature extraction ability is better for the proposed method than for the traditional deep learning methods.

In Experiment 2, the features can be well extracted and the state of the different categories can be effectively separated in both standard SAE and the proposed method as in Figures 13 and 14; this is the reason why the diagnosis accuracy is close to 100%. However there are also some differences between the standard SAE and the proposed method. The features of the same state in Figure 14 are gathered better than in Figure 13, the clustering center is clearer, and the states of the different

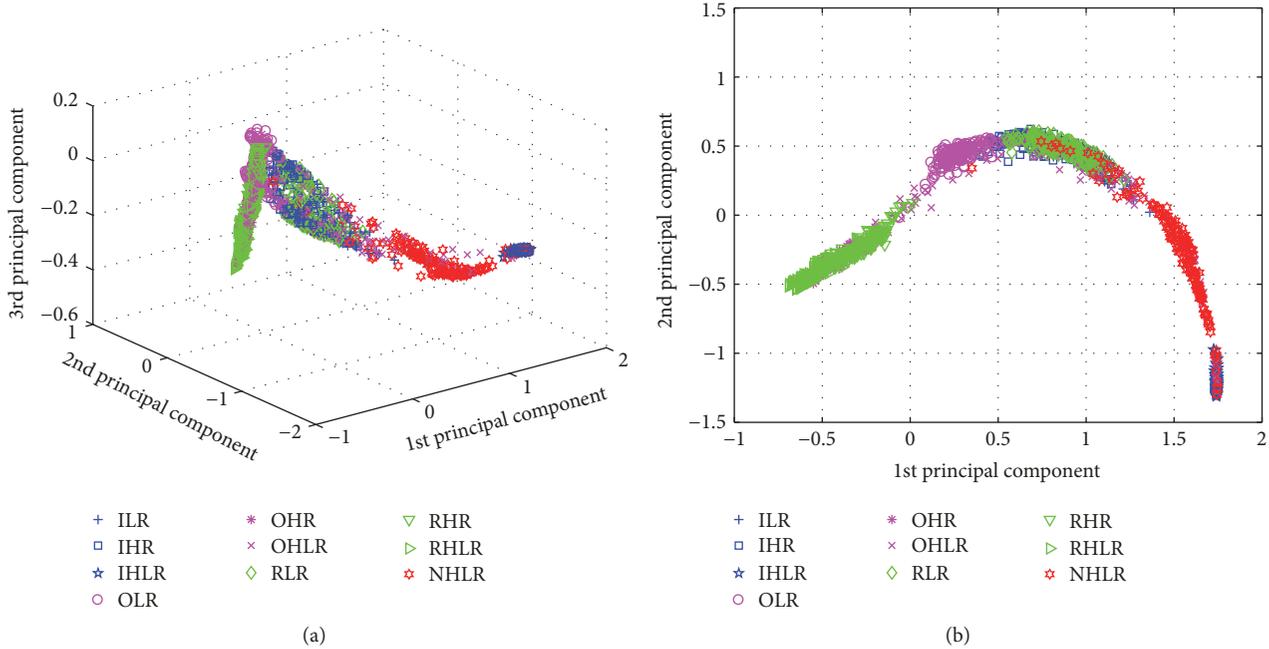


FIGURE 11: Feature visualization maps of standard SAE in Experiment 1. (a) The first three PCs; (b) the first two PCs.

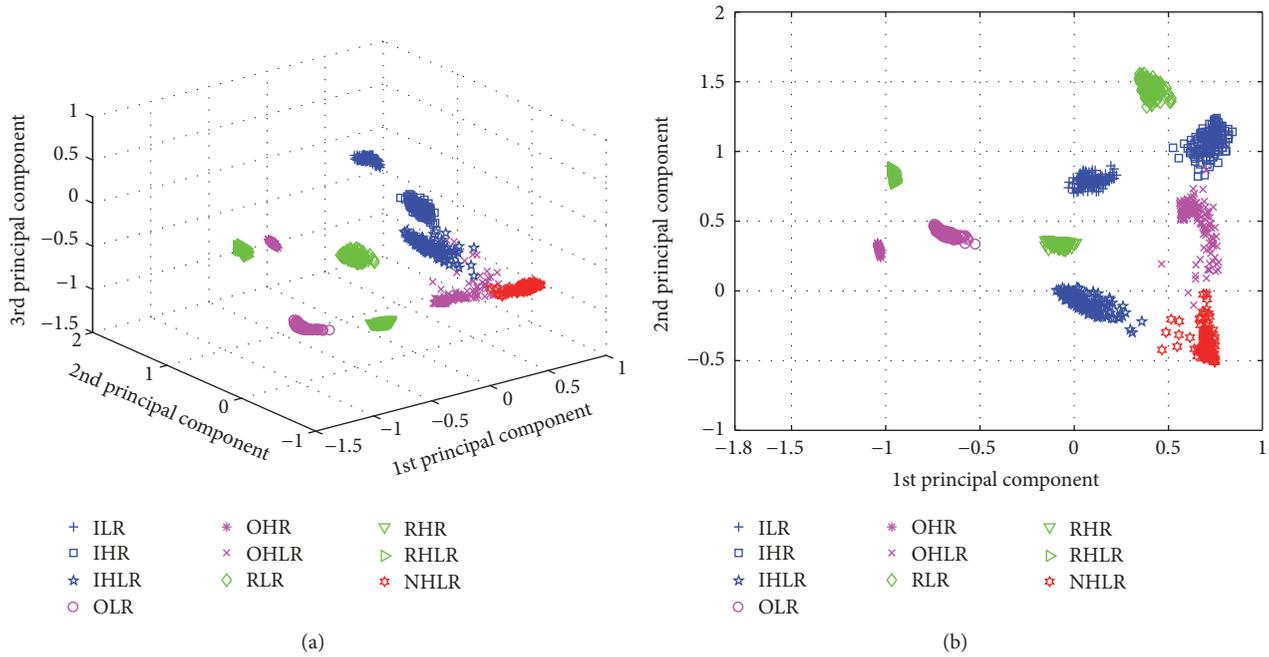


FIGURE 12: Feature visualization maps of the proposed method in Experiment 1. (a) The first three PCs; (b) the first two PCs.

categories are farther in three-dimensional space. The results show that standard SAE can extract effective features from Fourier coefficients and make good diagnosis result just as [10]. But the proposed method can extract better feature than SAE due to the fact that the kernel function maps the input data to a high eigenspace, which helps to get precise state features.

In Experiment 3, the features of standard SAE are intermixed and form a linear shape and the distribution range is large which makes it difficult to distinguish between different states in Figure 15. This may result from the data containing much noise and interference signal; the different state time-domain and frequency-domain features cannot be effectively separate. In Figure 16, the different state features

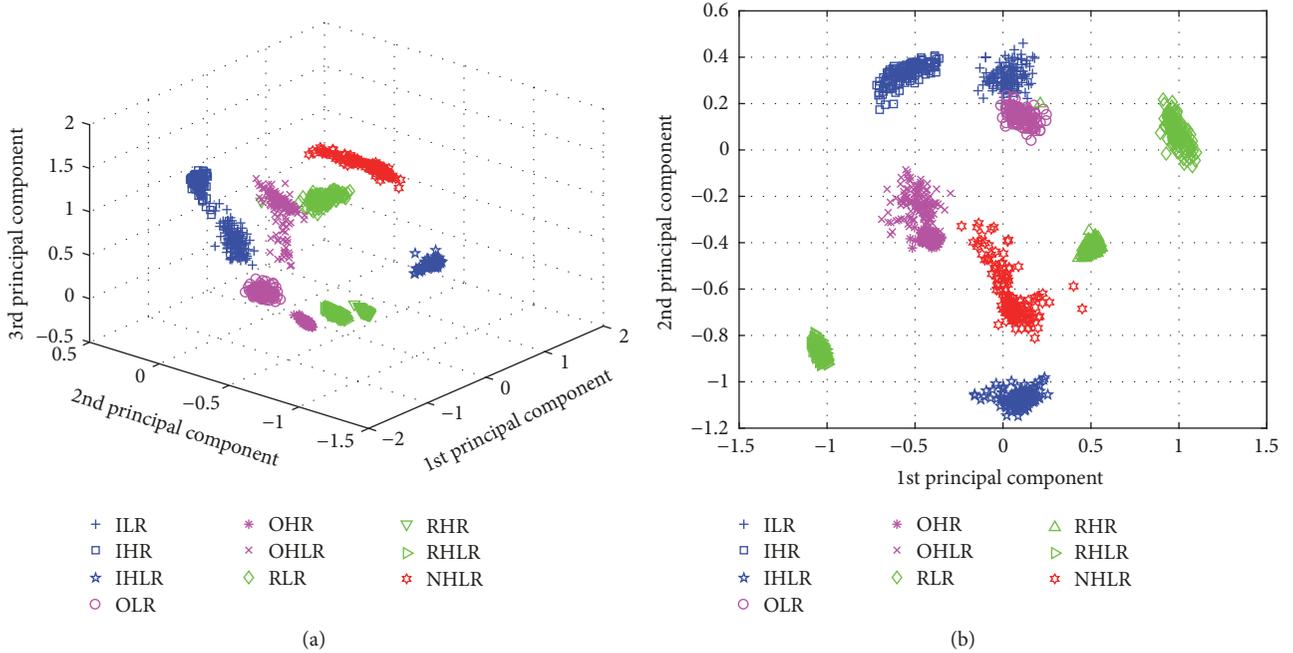


FIGURE 13: Feature visualization maps of standard SAE in Experiment 2. (a) The first three PCs; (b) the first two PCs.

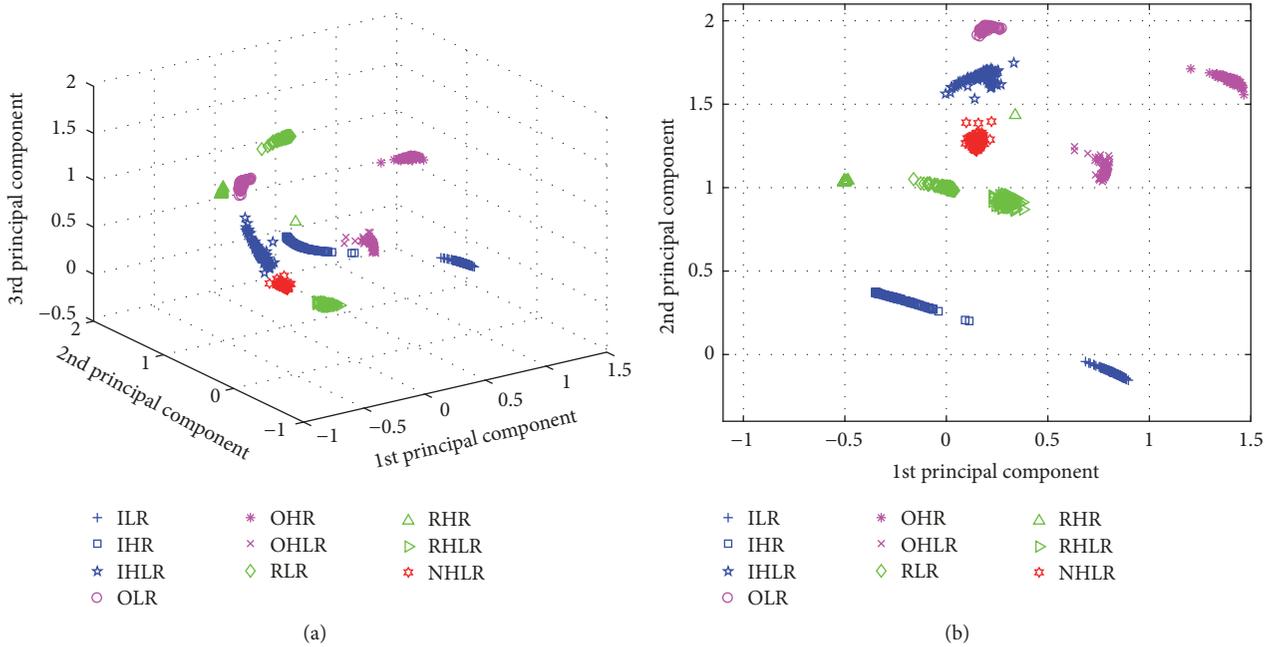


FIGURE 14: Feature visualization maps of the proposed method in Experiment 2. (a) The first three PCs; (b) the first two PCs.

of the proposed method can be clearly distinguished, and the same sort features gathered in a range. The feature extraction capability of the proposed method is promoted significantly because of the kernel function compared to SAE.

In conclusion, the combination of the Gaussian kernel function and the deep AE network expands the applications of the traditional SAE method, achieves a higher accuracy with improved extraction capability, a more obvious

clustering center, and a better diagnosis effect, and requires fewer iterations.

5. Conclusion

In this paper, a novel method which combines a Gaussian kernel function with the deep AE network is proposed for bearing fault diagnosis. The proposed method can be divided

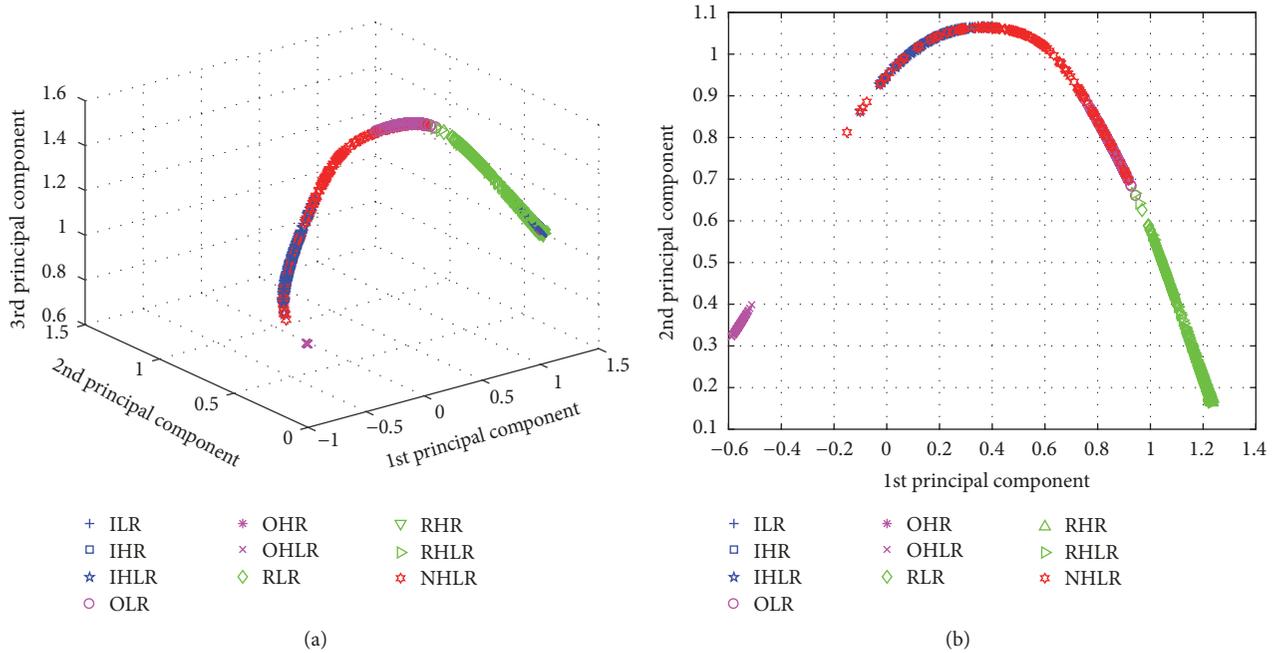


FIGURE 15: Feature visualization maps of standard SAE in Experiment 3. (a) The first three PCs; (b) the first two PCs.

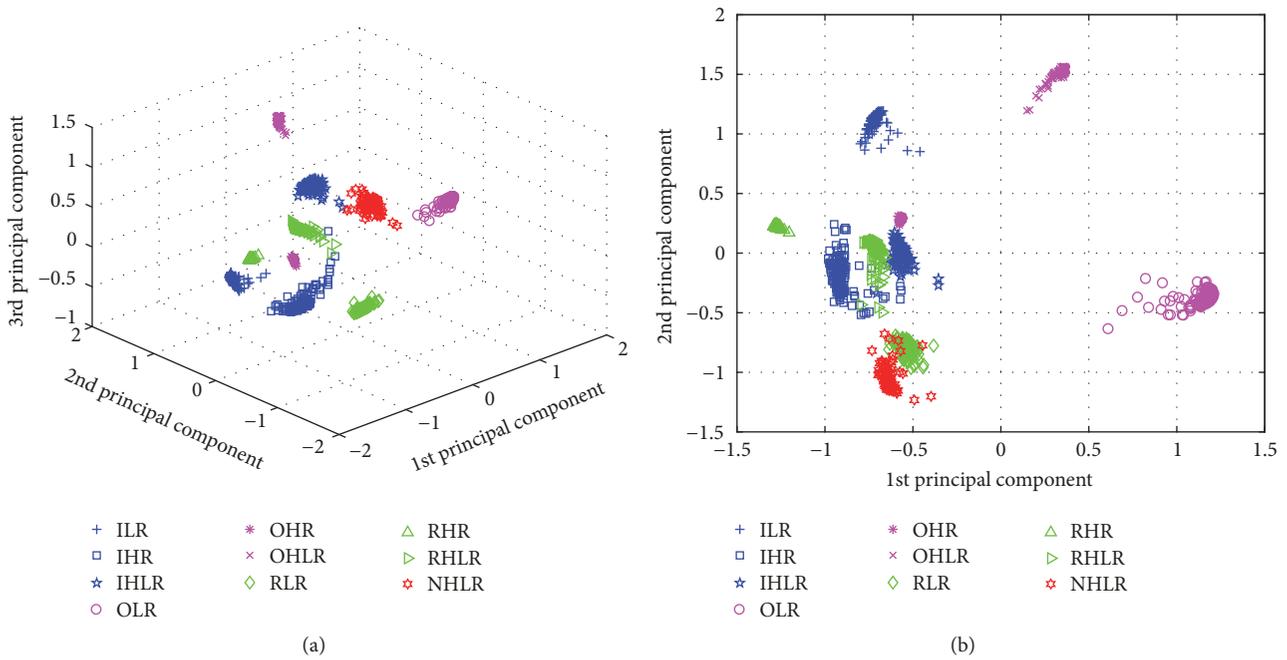


FIGURE 16: Feature visualization maps of the proposed method in Experiment 3. (a) The first three PCs; (b) the first two PCs.

into three major steps. Firstly, kernel function is employed to enhance the feature learning capability, and a new AE is designed termed kernel AE (KAE). Subsequently, a deep neural network is constructed with one KAE and multiple AEs to extract inherent features layer by layer. Finally, softmax is adopted as the classifier to accurately identify different bearing faults, and error backpropagation algorithm is used to fine-tune the model parameters.

Compared to conventional deep learning methods, the proposed method has a better feature extraction ability with better clustering effect, and results in a better diagnosis effect, higher accuracy, and wider range of application. The main contributions of this paper are (i) to introduce the kernel function to AE and form a new KAE network for better processing the nonlinear component; (ii) to propose a new deep feature learning method constructed with one KAE and

multi -AEs to automatically and effectively learn the fault features from the raw vibration signals; and (iii) to conduct three experiments on the aircraft engine intershaft bearing test rig, which verify that the proposed method has better feature extraction ability and better fault diagnosis results than standard SAE.

In addition, the investigation indicates that the proposed fault diagnosis method has great potential to be an effective tool for fault diagnosis of rolling bearings and the authors will continue to investigate this topic in the future.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is financially supported by the National Natural Science Foundation of China (no. 51375067).

References

- [1] H. Jiang, C. Li, and H. Li, "An improved EEMD with multi-wavelet packet for rotating machinery multi-fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 225–239, 2013.
- [2] Y. Lei, M. J. Zuo, Z. He, and Y. Zi, "A multidimensional hybrid intelligent method for gear fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1419–1430, 2010.
- [3] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3398–3407, 2013.
- [4] 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.
- [5] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, no. 1, pp. 696–711, 2012.
- [6] X. Zhang, B. Wang, and X. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowledge-Based Systems*, vol. 89, pp. 56–85, 2015.
- [7] S. J. Dong, L. L. Chen, B. P. Tang, X. Y. Xu, Z. Y. Gao, and J. Liu, "Rotating machine fault diagnosis based on optimal morphological filter and local tangent space alignment," *Shock and Vibration*, vol. 2015, Article ID 893504, 9 pages, 2015.
- [8] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *American Association for the Advancement of Science: Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [10] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72–73, pp. 303–315, 2016.
- [11] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowledge-Based Systems*, vol. 119, pp. 200–220, 2017.
- [12] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017.
- [13] L. Guo, H. Gao, H. Huang, X. He, and S. Li, "Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring," *Shock and Vibration*, vol. 2016, Article ID 4632562, 10 pages, 2016.
- [14] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectronics Reliability*, vol. 75, pp. 327–333, 2017.
- [15] M. Ring and B. M. Eskofier, "An approximation of the Gaussian RBF kernel for efficient classification with SVMs," *Pattern Recognition Letters*, vol. 84, pp. 107–103, 2016.
- [16] X. Qi, Z. Yuan, and X. Han, "Diagnosis of misalignment faults by tachless order tracking analysis and RBF networks," *Neurocomputing*, vol. 169, pp. 439–448, 2015.
- [17] F. Wang, X. Chen, B. Dun, B. Wang, D. Yan, and H. Zhu, "Rolling bearing reliability assessment via kernel principal component analysis and weibull proportional hazard model," *Shock and Vibration*, vol. 2017, Article ID 6184190, 11 pages, 2017.
- [18] X. Rui, Z. Sanyuan, and X. Lei, "Nonlinear process monitoring based on improved kernel ICA," in *Proceedings of the 2006 International Conference on Computational Intelligence and Security, ICCIAS 2006*, pp. 1742–1746, China, October 2006.
- [19] W.-A. Yang, M. Xiao, W. Zhou, Y. Guo, W. Liao, and G. Shen, "Trace ratio criterion-based kernel discriminant analysis for fault diagnosis of rolling element bearings using binary immune genetic algorithm," *Shock and Vibration*, vol. 2016, Article ID 8631639, 15 pages, 2016.
- [20] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [21] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proceedings of the 20th Annual Conference on Neural Information Processing Systems (NIPS '06)*, pp. 153–160, Cambridge, Mass, USA, December 2006.
- [22] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, "Why does unsupervised pre-training help deep learning?" *Journal of Machine Learning Research*, vol. 11, pp. 625–660, 2010.
- [23] Z. Xu, M. Dai, and D. Meng, "Fast and efficient strategies for model selection of Gaussian support vector machine," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 5, pp. 1292–1307, 2009.
- [24] H. Xiong, M. N. S. Swamy, and M. O. Ahmad, "Optimizing the kernel in the empirical feature space," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 16, no. 2, pp. 460–474, 2005.
- [25] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 95, pp. 187–204, 2017.
- [26] S. Kim, Y. Choi, and M. Lee, "Deep learning with support vector data description," *Neurocomputing*, vol. 165, pp. 111–117, 2015.



Hindawi

Submit your manuscripts at
www.hindawi.com

