

Research Article

Combining DBN and FCM for Fault Diagnosis of Roller Element Bearings without Using Data Labels

Fan Xu ^{1,2}, Yan jun Fang ^{2,3}, Dong Wang ¹, Jia qi Liang ^{2,3} and Kwok Leung Tsui¹

¹School of Data Science, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong

²Department of Automation, School of Power and Mechanical Engineering, Wuhan University, Wuhan, Hubei, China

³Department of Automation, School of Electrical Engineering and Automation, Wuhan University, Wuhan, Hubei, China

Correspondence should be addressed to Yan jun Fang; 2014102080001@whu.edu.cn

Received 28 June 2018; Revised 7 November 2018; Accepted 12 November 2018; Published 4 December 2018

Academic Editor: Marc Thomas

Copyright © 2018 Fan Xu 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.

Because deep belief networks (DBNs) in deep learning have a powerful ability to extract useful information from the raw data without prior knowledge, DBNs are used to extract the useful feature from the roller bearings vibration signals. Unlike classification methods, the clustering method can classify the different fault types without data label. Therefore, a method based on deep belief networks (DBNs) in deep learning (DL) and fuzzy C-means (FCM) clustering algorithm for roller bearings fault diagnosis without a data label is presented in this paper. Firstly, the roller bearings vibration signals are extracted by using DBN, and then principal component analysis (PCA) is used to reduce the dimension of the vibration signal features. Secondly, the first two principal components (PCs) are selected as the input of fuzzy C-means (FCM) for roller bearings fault identification. Finally, the experimental results show that the fault diagnosis of the method presented is better than that of other combination models, such as variation mode decomposition- (VMD-) singular value decomposition- (SVD-) FCM, and ensemble empirical mode decomposition- (EEMD-) fuzzy entropy- (FE-) PCA-FCM.

1. Introduction

With the development of science and technology, aerospace equipment, industrial equipment, and other fields of mechanical and electrical equipment have become increasingly complex, intelligent, and integrated, so that the operating conditions and the working environment are becoming more complex and changeable. Therefore, accurate and effective fault diagnosis in complex equipment systems becomes an effective way to improve the reliability and safety of the systems and to reduce the maintenance cost [1]. Roller bearings as one of the most common components in mechanical systems and their operating conditions will directly affect the performance of the entire mechanical equipment [2–5]. Using vibration signals for roller bearings, fault diagnosis has become one of the commonly used ways in recent years. Analyzing the roller bearing vibration signals and extracting their characteristics effectively are very important and of practical significance because the vibration signals can reflect the state of the roller bearings and the

quality of the feature extraction, which determines the accuracy of the fault diagnosis.

For signal feature extraction, many different traditional methods for vibration signal feature extraction have been presented, such as statistical analysis, wavelet transform (WT), and various mode decomposition models. In reference [6], different statistical indexes, such as mean value, kurtosis, and clearance factor, are employed to calculate the vibration signals, and they are regarded as the eigenvectors for assessing the degradation of slurry pumps by using vibration signals. Wang et al. proposed a method based on WT for gear fault diagnosis [7]. The vibration signals are decomposed into continuous statistical features on different scales by using WT. Because the dimension of the features is high, principal component analysis (PCA) is used to reduce the dimension of the eigenvectors. However, WT needs to select the wavelet function and the number of decomposition layers. As the vibration signals have nonlinear and nonstationary features, this is not a self-adaptive method. To overcome this drawback, empirical mode decomposition

(EMD) [8] can decompose the vibration signals into series of intrinsic mode functions (IMFs) and a residual self-adaptivity. However, EMD has a mode-mixing problem. Ensemble empirical mode decomposition (EEMD) [9] can solve the mode-mixing problem self-adaptively by introducing Gaussian white noise and decomposing a complicated signal into IMFs. Many scholars use the EEMD and entropy combination models to extract the vibration signal features. Zhang and Zhou employed the EEMD to decompose the roller bearings' vibration signals into some IMFs, and then fuzzy entropy (EE) is used to calculate the IMF entropy values; the extracted features are selected as the input of the support vector machine (SVM) for roller bearing fault diagnosis [10]. In reference [11], a method based on EEMD, sample entropy (SE), and SVM for fault diagnosis is developed, and the main purpose of this paper is similar to that of reference [11]; the only difference is that the fact SE replaces fuzzy entropy (FE).

EEMD cannot separate the vibration signals correctly for the closely located frequencies, but variation mode decomposition (VMD) [12] decomposes the signal into variation and nonrecursive modes. Because its essence is a number of adaptive wiener filter groups, VMD can separate two pure harmonic signals with similar frequencies. In [13], the roller bearings vibration signals are decomposed into some band-limited intrinsic mode functions (BLIMFs), and then singular value decomposition (SVD) is used to compute the eigenvalue of each BLIMF.

However, for some complex systems, the traditional feature extraction methods, regardless of self-adaptivity or not, are not enough to extract the sensitive features of all fault types due to the interaction of the external environment and the internal structure. Sometimes, several fault feature extraction methods need to achieve a certain effect for fault.

DBN, PCA, and FCM models are reviewed. Section 3 describes the experimental data sources, evaluation of the clustering effect, and the fault diagnosis methodology. The validation of the experiments is given in Section 4. Finally, the conclusion is given in Section 5.

2. The Theoretical Framework of DBN, PCA, and FCM Models

2.1. Theoretical Framework of DBN. DBN was proposed by Hilton and Salakhutdinov [14]. It is widely applied in object and speech identification and image classification. DBN contains input layer X , hidden layer (multilayer unsupervised restricted Boltzmann machine (RBM)), and an output layer. The network structure of DBN is shown in Figure 1.

The RBM is a classic energy-based model, which includes a visible and a hidden layer. The structure of RBM is shown in Figure 2, where vector v and h denote the visible and hidden layers, respectively. W denotes the connection weight values between the visible and hidden layers. For these layers, the connection is complete between the intercellular nodes, and there is no separate connection in each layer.

The invisible and hidden layers' neuron values are binary variables, and the neurons' numbers in the visible and hidden layers are I and J , respectively. v_i and h_j represent the status between the i th visible neuron and the j th hidden layer neuron. For a group of a specific combination (v, h) , RBM as a system with energy is listed as follows:

$$E(v, h)_\theta = - \sum_{i=1}^I a_i v_i - \sum_{j=1}^J b_j h_j - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j, \quad (1)$$

where $\theta = (w_{ij}, a_i, b_j)$ is the parameter matrix, w_{ij} denotes the connection weight values between the visible layer v_i and hidden layer h_j , and a_i and b_j denote the bias values of the visible layer v_i and hidden layer h_j . The joint probability distribution based on the energy function is obtained by

$$p(v, h)_\theta = \frac{-e^{E(v, h)_\theta}}{Z(\theta)}, \quad (2)$$

where $Z(\theta) = \sum_v \sum_h e^{-E(v, h)_\theta}$ is called the partition function and the distribution function (the likelihood function) $p(v)_\theta$ is the edge distribution of joint probability $p(v, h)_\theta$. For a given visible layer, each neuron in a hidden layer is independent. Therefore, the active probability is

$$p(h_j = 1, v)_\theta = \delta \left(b_j + \sum_{i=1}^I v_i w_{ji} \right), \quad (3)$$

where $\delta(x) = 1/(1 + e^{-x})$ is the sigmoid active function. For a given hidden layer, the active probability of the i th neuron node in the invisible layer is

$$p(v_j = 1, h)_\theta = \delta \left(a_j + \sum_{i=1}^I h_i w_{ji} \right). \quad (4)$$

RBM is the trained iterative way, and the purpose of the training is to learn the value of the parameter $\theta = (w_{ij}, a_i, b_j)$ to the fitting when the training data is used, where the parameter θ can be obtained by finding the maximal log-likelihood function by using the training set (where N is the number of samples):

$$\theta^* = \arg_\theta \max L(\theta) = \arg_\theta \max \sum_{t=1}^N \ln p(v^{(t)})_\theta. \quad (5)$$

Update the parameters w_{ij} , a_i , and b_j according to the following equations using the contrastive divergence (CD) algorithm:

$$\begin{aligned} \Delta w_{ij} &= \varepsilon (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}), \\ \Delta a_i &= \varepsilon (\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}}), \\ \Delta b_j &= \varepsilon (\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}}), \end{aligned} \quad (6)$$

where ε is the learning rate in the pretraining phase and $\langle \cdot \rangle_{\text{data}}$ and $\langle \cdot \rangle_{\text{recon}}$ denote the mathematical expectation.

2.2. Theoretical Framework of PCA. The essence of PCA is to retain the coordinates of the main components as

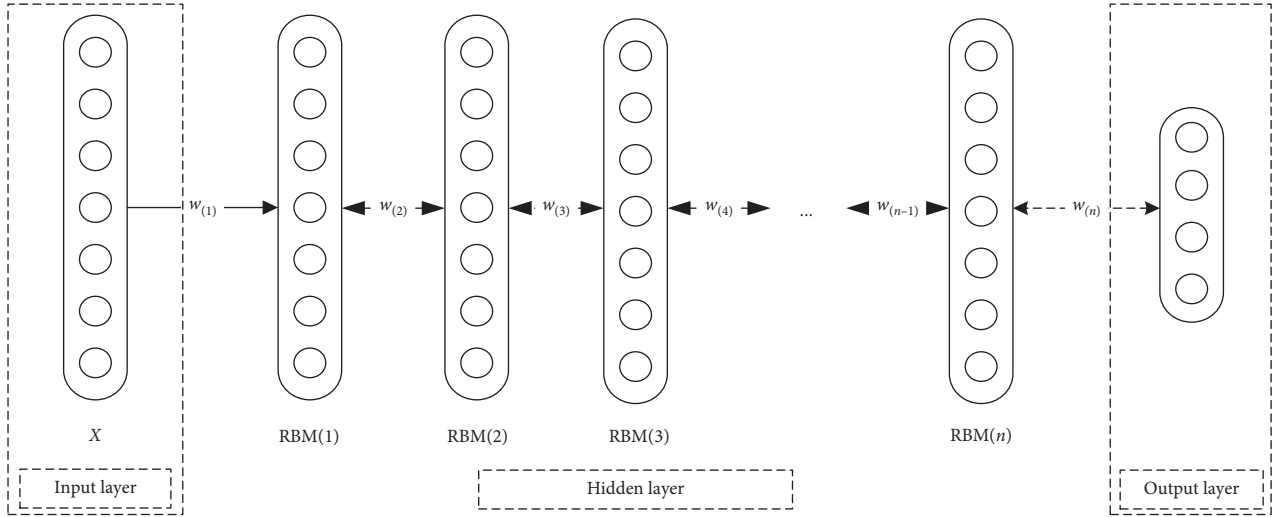


FIGURE 1: The structure of DBN.

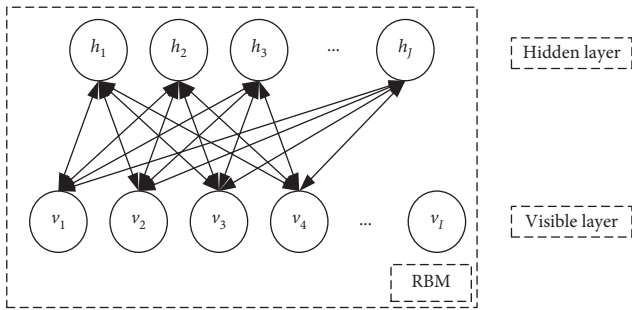


FIGURE 2: The structure of RBM.

the new data space direction to achieve the goal of dimension reduction. Assuming $X_N^k \in R$, ($k = 1, 2, \dots, K$), the K dimensional matrix with N samples, x_i^k ($i = 1, 2, \dots, N$) is the i th sample, where R is the covariance matrix. R is defined by

$$R = P \begin{bmatrix} \lambda_1 & 0 & \dots & \dots \\ \dots & 0 & \dots & \dots \\ \dots & \dots & \dots & \lambda_k \end{bmatrix} P^T = \sum_{i=1}^k P_i \lambda_i P_i^T, \quad (7)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$, P_i is the unit orthogonal eigenvector of the λ_i , and λ_i is the eigenvalue. Therefore, the matrix $X_N^k \in R$ is decomposed as follows:

$$\begin{aligned} X &= [p_1, p_2, \dots, p_k] [p_1, p_2, \dots, p_k]^T x_i^T \\ &= [p_1, p_2, \dots, p_i] [p_1, p_2, \dots, p_i]^T x_i^T \\ &\quad + [p_{i+1}, p_{i+2}, \dots, p_k] [p_{i+1}, p_{i+2}, \dots, p_k]^T x_i^T \quad (8) \\ &= C x_i^T + (I - C) x_i^T \\ &= \overline{x}_i^T + e_i^T, \end{aligned}$$

where $[p_1, p_2, \dots, p_i]$ denotes the first to i th PCs, $[p_{i+1}, p_{i+2}, \dots, p_k]$ is the residual space (RS), and hence, e_i is the projection of x_i in RS. C is the projection matrix.

The process of obtaining the projection matrix C by means of the covariance matrix R is called the modeling process. The number of PCs directly affects the merits of

the model and the final fault detection and diagnostics. This paper uses the main component contribution rate method to select the number of PCs as follows:

$$\eta \geq \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^K \lambda_i} * 100\%, \quad (9)$$

where η is the percentage of the total variance explained by the first $k \leq K$ PCs.

2.3. Theoretical Framework of FCM. FCM is one of the most common clustering algorithms, based on the objective function to minimize the Euclidean distance between each sample and all clustering centers. Correct cluster centers and classification matrices should be used to meet the termination criteria condition constantly, and hence, the data samples with similar characteristics are clustered into a class.

For a given vector $X = [x_1, x_2, \dots, x_n]$, the corresponding fuzzy classification matrix is $A = [u_{ij}]_{c \times n}$ and c is the number of clusters. The clustering centers are $C = [c_1, c_2, c_3, \dots, c_c]$, as mentioned above, and FCM is described as follows:

$$\left\{ \begin{array}{l} \min \left\{ J(A, c_1, \dots, c_c) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \right\} \\ \text{s.t. } \sum_{i=1}^c u_{ij} = 1, j = 1, 2, \dots, n \end{array} \right\}, \quad (10)$$

where n , c , and m represent the number of samples, clusters, and weighted index, respectively. u_{ij}^m is used to determine the degree of membership of each sample for each cluster. The greater the value u_{ij}^m is, the greater the likelihood of the i th cluster also is. d_{ij}^2 is the Euclidean distance between the j th point and i th clustering center point.

FCM converted the extreme value problems with constraints to unconstrained issues by introducing the operator $\lambda = [\lambda_1, \lambda_1, \dots, \lambda_n]$:

$$\begin{aligned} \bar{J}(A, C^T, \lambda) &= J(A, c_1, \dots, c_c) + \sum_{j=1}^n \lambda \left(\sum_{i=1}^c u_{ij} - 1 \right) \\ &= \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c u_{ij} - 1 \right). \end{aligned} \quad (11)$$

Equation (11) is the objective function, and the necessary conditions for it to reach the minimum value under the following conditions are as follows:

$$c_{ij} = \frac{\left(\sum_{j=1}^n u_{ij}^m x_j \right)}{\left(\sum_{j=1}^n u_{ij} \right)}, \quad (12)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{ij}/d_{kj})^{2/(m-1)}}. \quad (13)$$

The purpose of the FCM is to find the classification matrix and the clustering centers, which will minimize the value of the objective function to smallest. The procedure of FCM is as follows:

Step 1. Initialize the cluster center point number c , weighted index m , classification matrix $A = [u_{ij}]_{c \times n}$ and iteration number $l = 0$.

Step 2. Calculate the cluster centers C according to Equation (12).

Step 3. Update the classification matrix A according to Equation (13).

Step 4. If $\|A^{l+1} - A^l\| < \varepsilon$, stop the loop, or else set $l = l + 1$, and return to step 2.

3. Data Source and Clustering Effect Evaluation

3.1. Data Source. The experimental data came from Case Western Reserve University Bearing Data [15]. Three faults (inner Race Fault (IRF), Outer Race Fault (ORF), and Ball Fault (BF)) with fault diameters of 0.18 mm (1 hp), 0.36 mm (2 hp), and 0.54 mm (3 hp) were employed in this paper. The sampling frequency was 12000 Hz.

Table 1 shows the working conditions which are under consideration in this study. In Table 1, "NR" represents the bearings with no faults. The fault diameters are 0.18 mm (1 hp), 0.36 mm (2 hp), and 0.54 mm (3 hp). A, B, and C represent the three datasets. Each subset contains ten types of roller bearings faults under different conditions. Each type of the fault dataset has 30 samples with 2048 points, and hence, different datasets A, B, and C have a total of 300 samples.

3.2. Clustering Effect Evaluation. The two indicators partition coefficient (PC) and classification entropy (CE) are applied to evaluate the quality of the clustering results [16]. Partition coefficient is defined as

TABLE 1: The roller bearings experimental data under different conditions.

Datasets	Fault diameters (hp)	Fault type	The number of samples
A/B/C	1/2/3	NR	30/30/30
		BF1	30/30/30
		IRF1	30/30/30
		ORF1	30/30/30
		BF2	30/30/30
		IRF2	30/30/30
		ORF2	30/30/30
		BF3	30/30/30
		IRF3	30/30/30
		ORF3	30/30/30

$$PC = \alpha = \frac{1}{N} \sum_{i=1}^c \sum_{q=1}^N (\mu_{iq})^2, \quad (14)$$

where μ_{iq} denotes the membership value of the q th point in the i th cluster. The disadvantage of a PC is the lack of a direct connection to certain attributes of the data itself.

- (1) Classification entropy measures the fuzziness of the cluster partition only:

$$CE = \beta = -\frac{1}{N} \sum_{i=1}^c \sum_{q=1}^N \mu_{iq} \log(\mu_{iq}). \quad (15)$$

When the PC value is close to 1, it means that the effect of clustering is good; when the CE value is close to 0, it indicates that the effect of the clustering is better [16].

3.3. The Procedures of the Method Presented. The roller bearings vibration signal features were extracted by DBN, and then PCA was used to reduce the dimension of the eigenvectors. The first two PCs were regarded as the input of FCM for fault diagnosis. The procedures of the method presented are listed below:

- (1) Because the frequency spectra of rotating machinery can reflect how their important components are distributed with discrete frequencies, they can potentially provide useful information about the health and working conditions of the machine [17]. Therefore, fast Fourier transformation (FFT) is used to resolve the original vibration signal into a coefficient symmetrical matrix. As a result, the half coefficient symmetrical matrix is selected as the input vector for training DBN. Before DBN training, the input data is normalized to [0, 1].
- (2) Several hidden layers are used to extract the features of the vibration signals.
- (3) Reduce the dimensions of the features of the vibration signals in step 2 by using PCA. The first two PCs are regarded as the input of FCM for fault diagnosis
- (4) PC, CE, and classification accuracy were employed to evaluate the clustering performance of the different combination models, such as EEMD-FE-PCA-FCM, VMD-SVD-FCM, and DBN-PCA-FCM.

The detailed flow chart is shown in Figure 3.

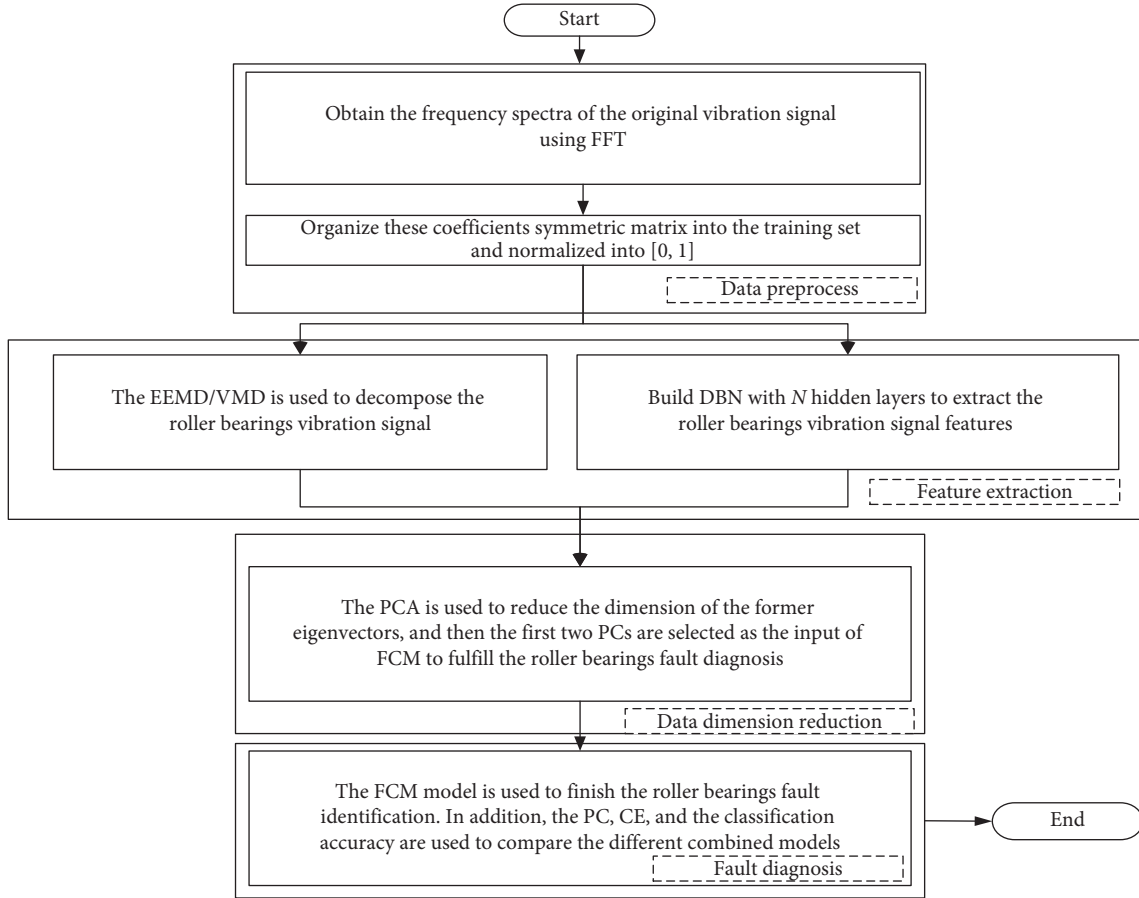


FIGURE 3: The flow chart of the method presented.

4. Feature Extraction and Fault Diagnosis

In this section, various vibration signals in Table 1 are first preprocessed by FFT, and then DBN is used to extract the useful feature information through several hidden layers. The time-domain figure of the various original vibration signals is shown in Figure 4.

The ten kinds of vibration signals are difficult to distinguish. There are no obvious vibration patterns in the NR and IRF signals. Unlike NR and IRF signals, the BF and ORF signals have obvious vibration patterns because the bearing and outer race components experience a certain impact when the roller bearings are working. Compared with NR and BF signals, IRF and ORF vibration signals have fixed vibration periods in some unique frequency bands, and the self-similarity is high. Especially when the inner ring is fixed, the outer ring rotates with the roller bearings; hence, the vibration regularity in ORF signals is more obvious. IRF and ORF vibration signals have strong periodic regularity, but it is difficult to distinguish these vibration signals under different conditions. To mine the signal features, the DBN, VMD, and EEMD models are used to decompose the vibration signals and PCA is used to reduce the dimension of the extracted features.

4.1. VMD Decomposition. To decompose vibration signals effectively, the number of mode m in VMD should first be

predetermined. When the value of m is too small, the decomposition of the mode cannot fully reflect the original signal with the time-frequency information, and therefore, the VMD decomposition cannot be achieved. A larger m produces a similar frequency for each BLIMF component, which may result in overdecomposition. Therefore, in order to select the appropriate m , we observe the center frequency of the signal to determine the m according to references [13]. The results of the center frequency under different modes m are shown in Table 2.

Here, it is shown that when dataset A is used, the center frequency in the IRF2 signals ranges from 0.0507 kHz to 0.3469 kHz especially when $m = 5$, for example, 0.2279 kHz in BLIMF3, 0.2980 kHz in BLIMF4, and 0.3469 kHz in BLIMF5. The center frequencies of these three modes are very close to one another. This indicates that the vibration signals are not decomposed effectively. The same also happens when $m = 3$ (like 0.0535 kHz in BLIMF1, 0.2145 kHz in BLIMF2, and 0.2979 kHz in BLIMF3). However, the decomposition results ($m = 4$) contain four frequency components which are separated well. Hence, the parameter m in VMD is selected as 4. The penalty factor ∂ is often set at 2000 [13]. The VMD composition results and its envelope spectrum for each BLIMF are shown in Figure 5.

As shown in Figure 5(a), the IRF2 vibration signals are decomposed into four BLIMFs components. The range of amplitude of each BLIMF is gradually increased; each BLIMF frequency is also increased. To further verify the

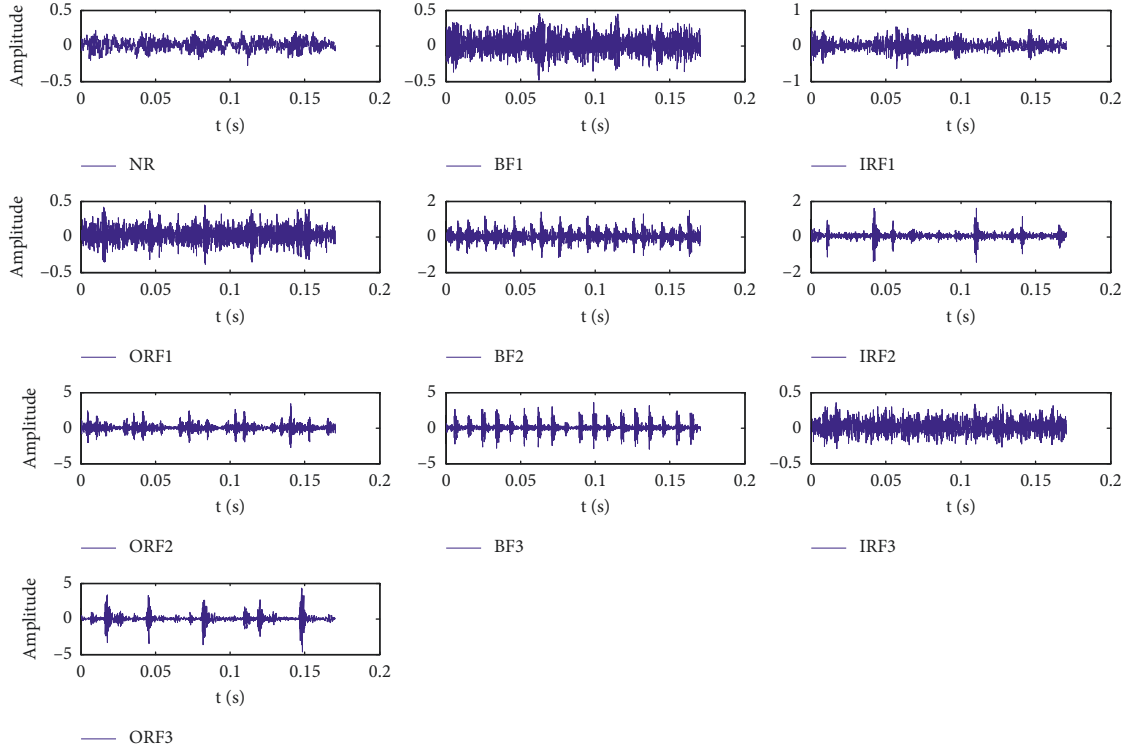


FIGURE 4: The time-domain waveforms for each working condition.

TABLE 2: Center frequency corresponding to different m in VMD.

m	Center frequency (kHz)					
	BLIMF1	BLIMF2	BLIMF3	BLIMF4	BLIMF5	BLIMF6
2	0.0616	0.2294	—	—	—	—
3	0.0535	0.2145	0.2979	—	—	—
4	0.0508	0.1088	0.2288	0.3004	—	—
5	0.0511	0.1101	0.2279	0.2980	0.3469	—
6	0.0507	0.1102	0.2106	0.2356	0.2980	0.3443

effect of VMD decomposition, Figure 5(b) shows that the envelope spectrum of each BLIMF, and it can be seen from Figure 5(b) that the decomposition results contain IRF2 double fault frequency (164 Hz) and fault frequency (58 Hz) (58 Hz is the IRF2 signal fault frequency).

4.2. Feature Extraction Using DBN. Firstly, The FFT is used to transform the time-domain vibration signal to the frequency domain; here, we take an IRF2 signal for example. The result of the spectrum envelope analysis in the frequency domain is shown in Figure 6.

As shown in Figure 6(b), the IRF2 signal working frequencies mainly focus on 0–1000 Hz. Because the working frequency for the IRF2 signal is 58 Hz, the frequency is mainly focused on 58 Hz and the double frequency (164 Hz). This indicates that the frequency-domain signal contains useful feature information. Therefore, we use the FFT to preprocess the different vibration signals in the first step. Then, we use the DBN in this section to extract the feature. The number of input layer nodes is set at 1024 because each

sample contains 2048 points and only half of the coefficient matrix is used after FFT decomposition before the DBN training procedure. The numbers of neural nodes for the second to fourth layers are set at 512, 256, and 128, respectively. The learning rate is 0.15, the momentum value is 0.65, and the number of epochs is 1200. After the vibration signals' features have been extracted by each layer in DBN, reduce the dimensions of the feature vectors by using PCA. The results of the first two PCs for each hidden layer are shown in Figure 7.

Obviously, data, of the same fault type, are discrete in the first three layers, while there may be overlapping between data of different fault types in all the datasets. As the number of hidden layers increases, these scattered data points are more concentrated at one point and these data of different fault types are more separated from one another. As can be seen in Figure 7, from all datasets, the data points of the same shapes are more concentrated at one point and there is a clear separation between different fault data types in the final hidden layer as compared with that of the first hidden layer. For example, all NR signal data, which have a triangular shape, are concentrated (overlapping with each other), in dataset B in the final hidden layer, and they are, however, discrete in the first hidden layer.

The results of PC1 and PC2 through the final hidden layer when PCA is used are shown in Table 3, where “total” denotes the sum of all eigenvalues λ in Equation (9) and η is the cumulative contribution rate calculated by the first two PCs. The two largest eigenvalues (λ_1 - λ_2) in Equation (9), when PCA is used, are the first two PCs; the greater the λ is, the more useful the information contained in the corresponding

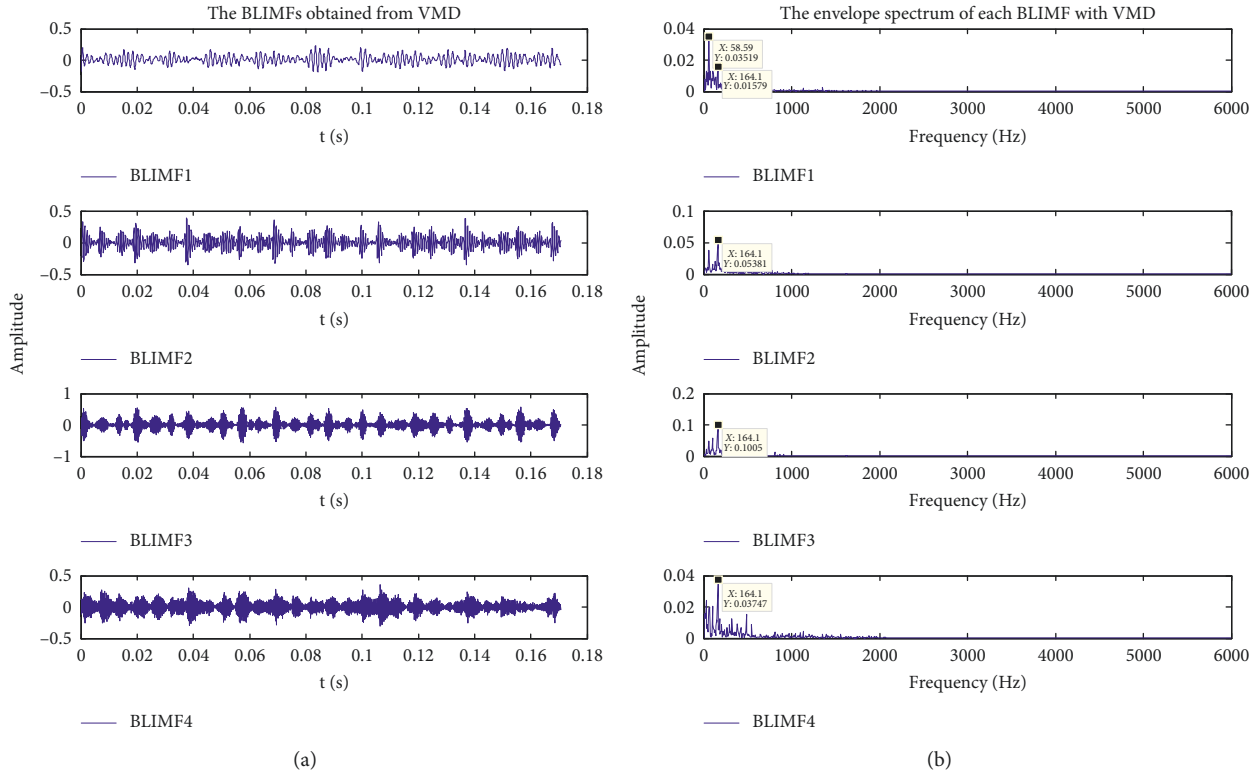


FIGURE 5: Signal decomposition diagram (VMD): (a) the BLIMF components; (b) the envelope spectrum of each BLIMF.

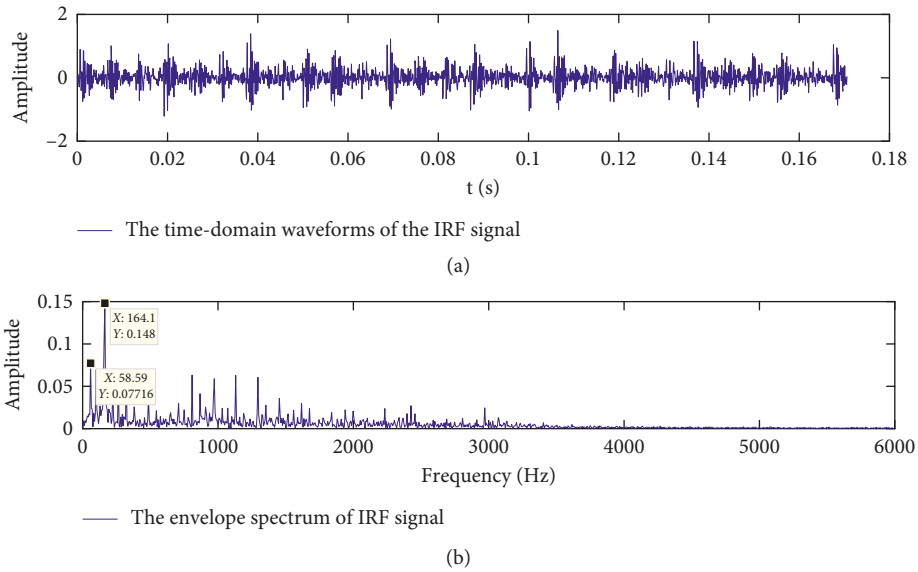


FIGURE 6: The time domain waveforms and the envelope spectrum of the IRF2 signal: (a) time domain; (b) envelope spectrum.

PC is. The available number of PCs is often selected as 2 when the cumulative contribution rate η is more than 80% [16]. In Table 3, it is up to more than 85% with different datasets, for example, 87.81% in dataset A. Moreover, with the increment of the number of PCs, the eigenvalues are decreased, and the first two PCs are often selected as the input of FCM for roller bearing fault diagnosis (as space is limited, only the first six PCs are shown in Table 3).

4.3. Fault Diagnosis and a Comparison Analysis. Before roller bearing fault diagnosis, the EEMD is also used to decompose the vibration signal into some IMFs. Some parameters in EEMD should be preset, such as Gaussian white noise amplitude mm , and the number of inserted white noise nn in EEMD, embedded dimension m , and similarity tolerance r in FE should generally be set before calculation for parameter nn in EEMD. If the additional noise is standard

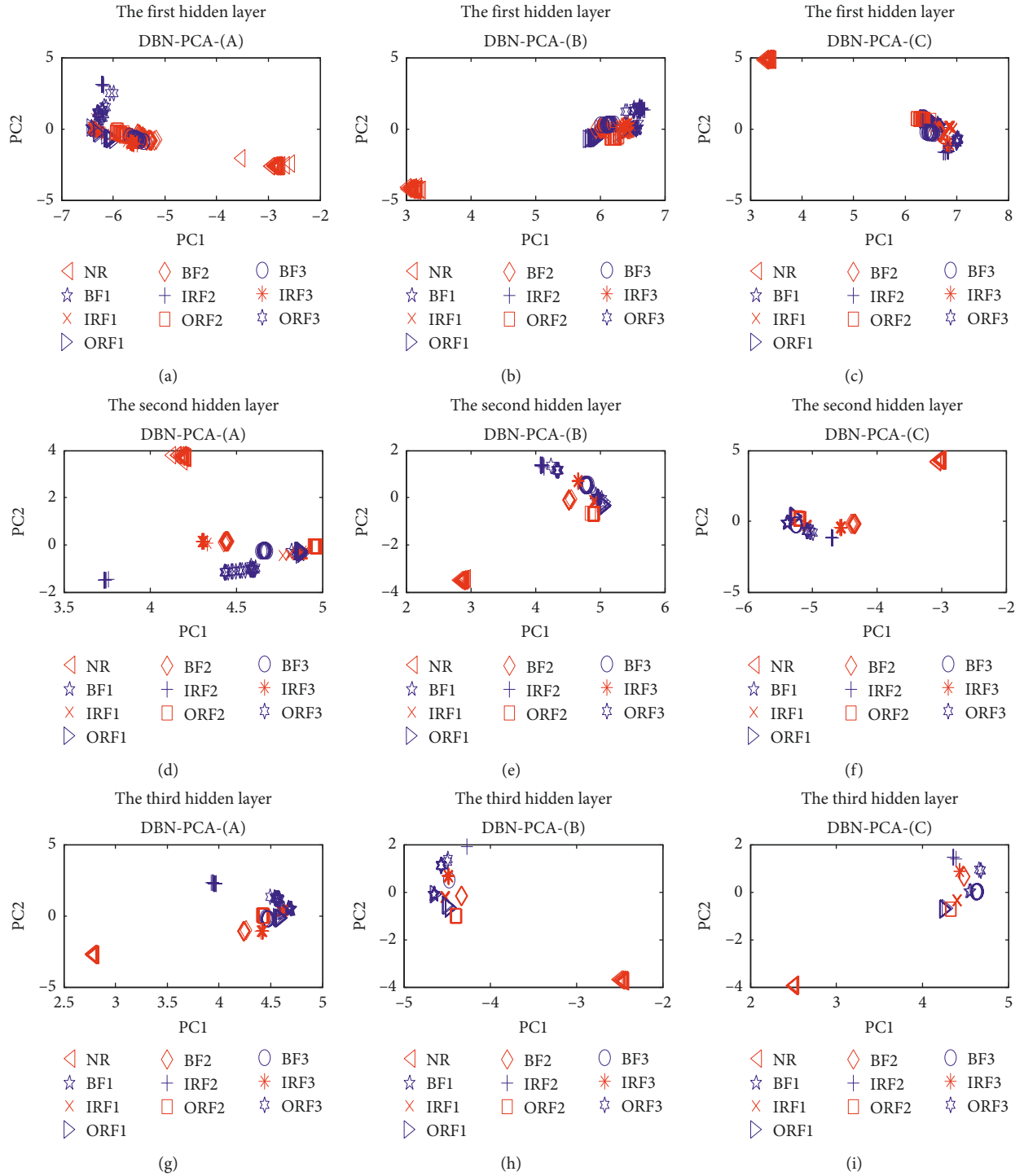


FIGURE 7: The PC1 and PC2 obtained by each hidden layer in DBN: (a, d, g) dataset A; (b, e, h) dataset B; (c, f, i) dataset C.

deviation and this is only a small part of the standard deviation of the input signal, then the remaining noise will result in less than 1% error. The authors suggest that the value of added white noise mm is usually fixed at about 20% of the standard deviation of the input signal [18–24]. The parameter mm is set at 100.

In FE, the greater embedded dimension m allows more detailed reconstruction of the dynamic process. But too great an m value is unsuitable due to the need for too great

$N = 10^m - 30^m$, which is difficult to meet the general requirements and will bring about loss of information. m is often fixed at 2 [16]. Here, similarity tolerance r is often fixed at $0.1 - 0.25 * SD$. SD denotes the standard deviation of the original vibration signals [16].

For the FCM model, the parameter $c = 4$ is set, where c is the number of clusters. Meanwhile, the value of the termination tolerance $\varepsilon = 1e - 6$. FCM is used to identify the different roller bearing faults, and the results of two-dimensional

TABLE 3: The results of PC1-PC2 and the value of η .

Model	Dataset	PC1 (λ_1)	PC2 (λ_2)	PC3 (λ_3)	PC4 (λ_4)	PC5 (λ_6)	PC6 (λ_6)	Total	η (%)
DBN-PCA	A	18.5424	1.6845	0.8434	0.7880	0.3610	0.2697	23.0347	87.81
	B	18.5430	2.1037	1.0733	0.7564	0.6150	0.2566	23.7500	86.93
	C	18.5427	2.0779	0.9061	0.7110	0.4287	0.3223	23.4881	87.79

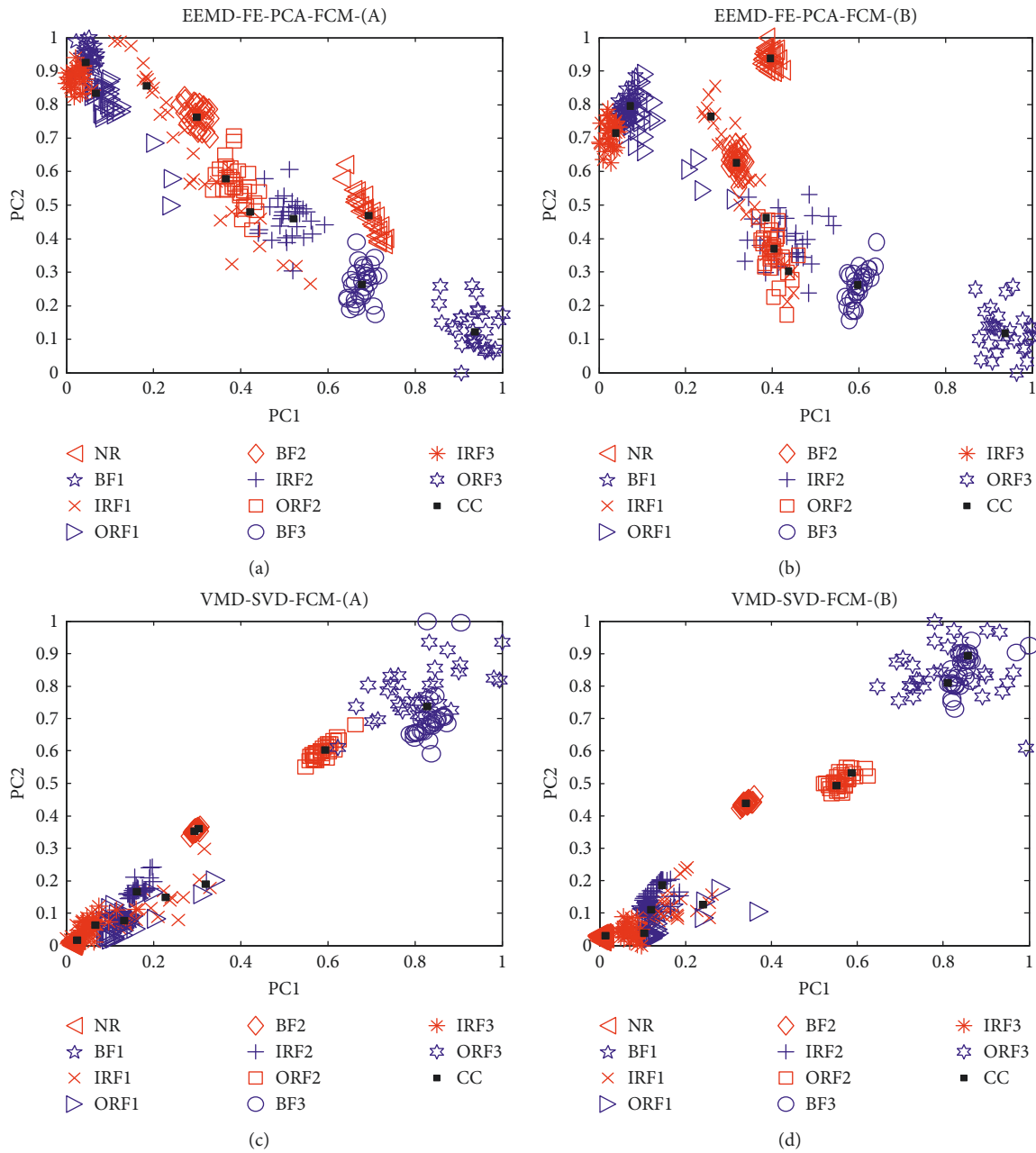


FIGURE 8: Continued.

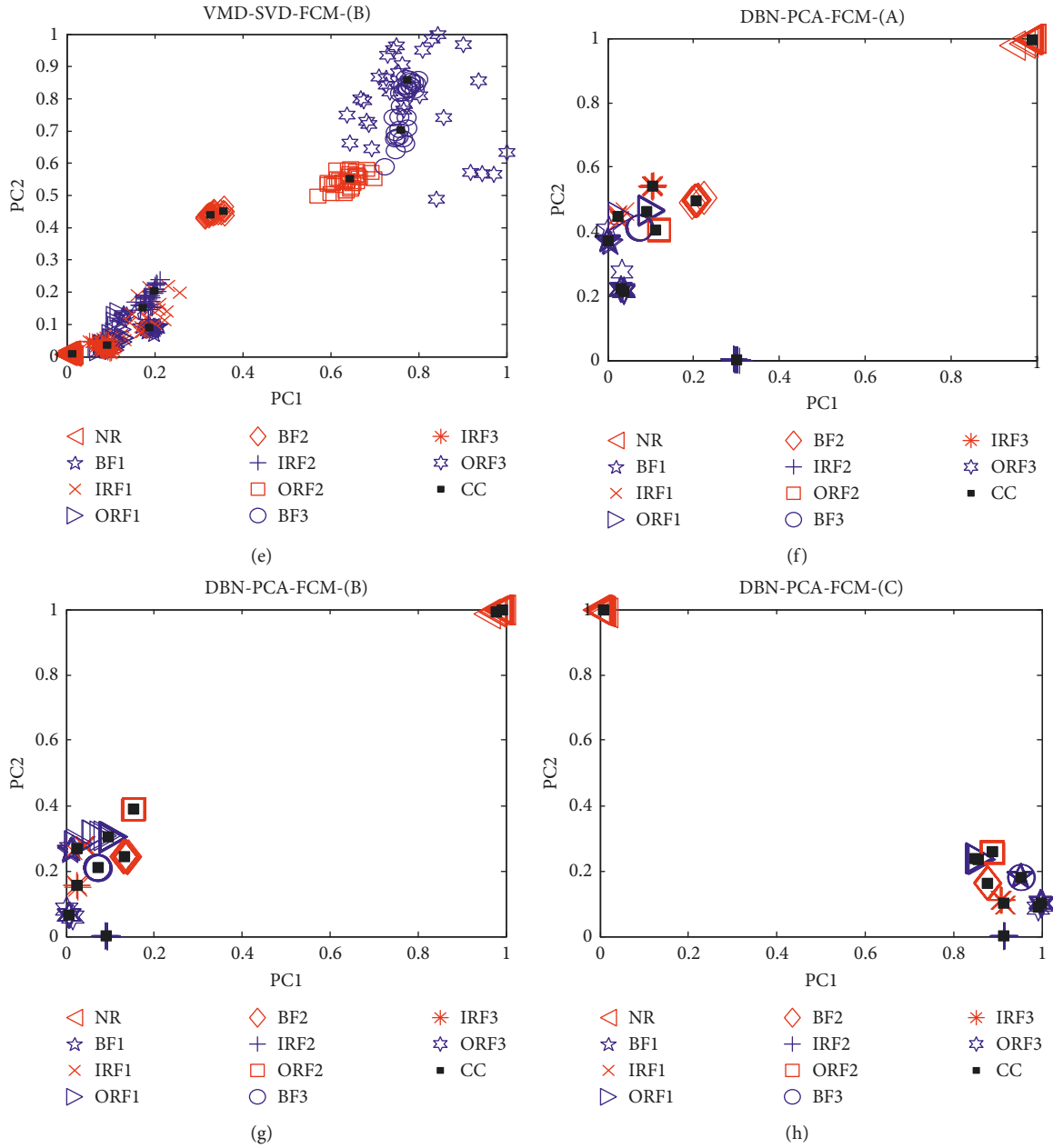


FIGURE 8: 2-dimensional clustering with DBN-PCA-FCM: (a, c, f) dataset A; (b, d, e, g) dataset B; (h) dataset C.

clustering with different datasets are shown in Figure 8. The symbol cc denotes each clustering center. PC1 and PC2 are horizontal and vertical coordinates [25-30]

(1) In Table 4, the greatest PC value is 0.9959 with dataset C when DBNs are used, and the smallest PC value is 0.6580 with dataset A in EEMD. The smallest CE value is 0.0117 in DBN, but the greatest CE value is 0.7735 in EEMD.

(2) As shown in Table 4, the results of PC in DBN are overall greater than those of VMD and EEMD, and the results of CE in DBN are smaller than those of

TABLE 4: The results of α (PC) and β (CE).

Model	Dataset	α (PC)	β (CE)
EEMD-FE-PCA-FCM	A	0.6580	0.7735
	B	0.6894	0.6962
	C	0.7123	0.6440
VMD-SVD-FCM	A	0.7308	0.5990
	B	0.7119	0.6050
	C	0.7582	0.5308
DBN-PCA-FCM	A	0.9777	0.0623
	B	0.9892	0.0287
	C	0.9959	0.0117

TABLE 5: The results of classification accuracy.

Mode	Dataset	The accuracy of each cluster (%)										Total (%)
		NR	BF1	IRF1	ORF2	BF2	IRF2	ORF2	BF3	IRF3	ORF3	
EEMD-FE-PCA-FCM ($k = 2$)	A	93.3	100	60	100	100	100	93.3	100	100	26.7	87.33
	B	100	100	100	63.3	100	73.3	100	100	100	30	86.69
	C	100	100	56.7	43.3	96.77	96.7	66.7	100	100	100	76.34
MD-SVD-FCM ($k = 2$)	A	100	100	100	70	100	100	33.3%	100	23.3	13.3	77
	B	90	100	100	56.7	100	86.7	100	100	30	43.3	80.67
	C	76.7	100	100	76.7	23.67	100	23.3	100	60	100	76.037
DBN-PCA-FCM ($k = 2$)	A	100	100	96.7	100	100	100	100	100	90	100	98.67
	B	100	100	100	100	100	100	100	93.3	100	100	99.33
	C	100	100	100	100	100	100	100	100	100	100	100

VMD and EEMD. These results indicate that the clustering performance of the method presented is superior to that of VMD-SVD-FCM and EEMD-FE-PCA-FCM.

The BF3 and ORF3 samples in Figures 8(a)–8(e) using the EEMD-PCA-FCM and VMD-SVD-FCM models are scattered randomly, but in Figures 8(f)–8(h), these scattered data points are more concentrated at one point, and the data of different fault types are more separated. This demonstrates that the DBN has a good feature extraction ability.

To verify the clustering effect, the three indicators PC, CE, and classification accuracy are used to estimate and compare the method presented, namely, the EEMD-FE-PCA-FCM, and VMD-SVD-FCM models. The results of α (PC) and β (CE) are shown in Table 4. The values of PC and CE are calculated by u_{ij}^m , and in Equation (14), the greater the PC value close to 1, the better the effect of FCM. However, the smaller the CE value is close 0, the better the effect of FCM.

To demonstrate that DBN can extract the signals effectively, classification accuracy is used to compare the DBN-PCA-FCM, VMD-SVD-FCM, and EEMD-FE-PCA-FCM models. The corresponding clustering accuracy is shown in Table 5:

- (1) The greatest classification accuracy is up to 100% with DBN, and the lowest classification accuracy is 76.037%.
- (2) The overall classification accuracy of DBN is greater than that of the VMD and EEMD models, about 10%–20%.
- (3) For different vibration signals, particularly in dataset C, the accuracy is up to 100% in DBN. But it is slightly lower in VMD, for example, 23.3% and 60%.

The method presented can extract the vibration signals and diagnose faults effectively, and its clustering is superior to that of the EEMD-FE-PCA-FCM and VMD-SVD-FCM models.

5. Conclusions

A method based on DBN and FCM for roller bearing fault diagnosis is presented in this paper. Unlike many traditional feature extraction methods, the different roller bearing

vibration signals are extracted by using DBN. To visualize the data, PCA is used to reduce the dimension of the eigenvectors. Then, the first two PCs are selected as the input of FCM for fault diagnosis, and the experimental results show that the feature extraction is better than that of the other models, such as the VMD-SVD/EEMD-FE-PCA combination model. The classification accuracy shows that the FCM clustering model can identify the roller bearing faults well under various conditions without data labels.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project no. T32-101/15-R).

References

- [1] S. A. Mcinerny and Y. Dai, "Basic vibration signal processing for bearing fault detection," *IEEE Transactions on Education*, vol. 46, no. 1, pp. 149–156, 2003.
- [2] Y. Huang, B. X. Wu, and J. Q. Wang, "Test for active control of boom vibration of a concrete pump truck," *Journal of Vibration and Shock*, vol. 31, no. 2, pp. 91–94, 2012.
- [3] F. Resta, F. Ripamonti, and G. Cazzluani, "Independent modal control for nonlinear flexible structures: an experimental test rig," *Journal of Sound and Vibration*, vol. 329, no. 8, pp. 961–972, 2011.
- [4] G. Bagordo, G. Cazzluani, F. Resta, and F. Ripamonti, "A modal disturbance estimator for vibration suppression in nonlinear flexible structures," *Journal of Sound and Vibration*, vol. 330, no. 25, pp. 6061–6069, 2011.
- [5] X. B. Wang and S. G. Tong, "Nonlinear dynamical behavior analysis on rigid flexible coupling mechanical arm of hydraulic excavator," *Journal of Vibration and Shock*, vol. 33, no. 1, pp. 63–70, 2014.

- [6] P. W. Tse and D. Wang, "Enhancing the abilities in assessing slurry pumps' performance degradation and estimating their remaining useful lives by using captured vibration signals," *Journal of Vibration and Control*, vol. 23, no. 12, pp. 1925–1937, 2017.
- [7] D. Wang, K. L. Tsui, P. W. Tse, and M. J. Zuo, "Principal components of superhigh-dimensional statistical features and support vector machine for improving identification accuracies of different gear crack levels under different working conditions," *Shock and Vibration*, vol. 2015, Article ID 420168, 14 pages, 2015.
- [8] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [9] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [10] X. Y. Zhang and J. Z. 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.
- [11] K. H. Zhu, X. Song, and D. X. Xue, "Fault diagnosis of rolling bearings based on IMF envelope sample entropy and support vector machine," *Journal of Information & Computational Science*, vol. 10, no. 16, pp. 5189–5198, 2013.
- [12] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [13] C. Liu, Y. J. Wu, and C. G. Zhen, "Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering," *Proceedings of the CSEE*, vol. 35, no. 13, pp. 3358–3365, 2015.
- [14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [15] The Case Western Reserve University Bearing Data Center Website, Bearing data center test seeded fault test data [EB/OL]. [2013-05-02], <http://csegroups.case.edu/bearingdatacenter/pages/download-data-file>.
- [16] F. Xu, Y. J. Fang, and R. Zhang, "PCA-GG rolling bearing clustering fault diagnosis based on EEMD fuzzy entropy," *Computer Integrated Manufacturing System*, vol. 22, no. 11, pp. 2631–2642, 2016.
- [17] S. Ng, P. Tse, and K. Tsui, "A one-versus-all class binarization strategy for bearing diagnostics of concurrent defects," *Sensors*, vol. 14, no. 1, pp. 1295–1321, 2014.
- [18] R. Zhang, F. Xu, Y. J. Fang, Z. M. Kong, and R. L. Tang, "A fault diagnosis method combined with ensemble empirical mode decomposition, base-scale entropy and clustering by fast search algorithm for roller bearings," *Journal of Vibroengineering*, vol. 18, no. 7, pp. 4472–4490, 2006.
- [19] 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.
- [20] Y. Huang, Z. Q. Liu, R. Du, and H. Tang, "Simulation and experiments of active vibration control for ultra-long flexible manipulator under impact loads," *Journal of Vibration and Control*, article 107754631879422, 2018.
- [21] R. L. Tang, X. Li, and J. G. Lai, "A novel optimal energy-management strategy for a maritime hybrid energy system based on large-scale global optimization," *Applied Energy*, vol. 228, pp. 254–264, 2018.
- [22] R. L. Tang, Z. Wu, and X. Li, "Optimal operation of photovoltaic/battery/diesel/cold-ironing hybrid energy system for maritime application," *Energy*, vol. 162, pp. 697–714, 2018.
- [23] J. Lai, X. Lu, X. Yu, W. Yao, J. Wen, and S. Cheng, "Distributed multi-DER cooperative control for master-slave-organized microgrid networks with limited communication bandwidth," *IEEE Transactions on Industrial Informatics*, 2018.
- [24] X. Lu, J. Lai, X. Yu, Y. Wang, and J. M. Guerrero, "Distributed coordination of islanded microgrid clusters using a two-layer intermittent communication network," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 3956–3969, 2018.
- [25] C. Affonso, R. Debiasio, and L. Andre, "Deep learning for biological image classification," *Expert System with Application*, vol. 85, pp. 114–122, 2017.
- [26] S. N. Sun, B. B. Zhang, L. Xie, and Y. Zhang, "An unsupervised deep domain adaptation approach for robust speech recognition," *Neurocomputing*, vol. 257, pp. 79–87, 2017.
- [27] P. Tamilselvan and P. F. Wang, "Failure diagnosis using deep belief learning based health state classification," *Reliability Engineering & System Safety*, vol. 115, pp. 124–135, 2013.
- [28] V. T. Tran, F. AlThobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014.
- [29] H. Shao, H. K. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, pp. 1–17, 2015.
- [30] S. Q. Zhang, G. X. Sun, and L. Li, "Study on mechanical fault diagnosis method based on LMD approximate entropy and fuzzy C-means clustering," *Chinese Journal of Scientific Instrument*, vol. 34, no. 3, pp. 714–720, 2013.
- [31] Y. J. Wu, C. L. Liu, C. G. Zhen, and Z. yaliang, "Gearbox condition monitoring based on variational mode decomposition," *Journal of Mechanical Transmission*, vol. 164, pp. 161–164, 2016.



Hindawi

Submit your manuscripts at
www.hindawi.com

