IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 1, 2020, accepted June 16, 2020, date of publication June 22, 2020, date of current version July 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003909

A Method of Rolling Bearing Fault Diagnose Based on Double Sparse Dictionary and Deep Belief Network

JUNFENG GUO AND PENGFEI ZHENG
School of Mechanical and Electronic Engineering, Lanzhou University of Technology, Lanzhou 730050, China

Corresponding author: Junfeng Guo (junf_guo@163.com)

This work was supported in part by the National Science Foundation of China under Grant 51465034.

ABSTRACT Feature extraction is the key technology in the data-driven intelligent fault diagnosis methods of rolling bearing. However, the acquired features by the traditional methods, which mainly based on time-frequency domain, sometimes cannot well represent the characteristics of the signal and are difficult to accurately identify because of their complexity and subjectivity. Aiming at this problem, the sparse representation theory is used to the field of fault diagnosis because the different types of rolling bearing fault signals only has the highest matching degree with the dictionary atoms trained by the same type of fault signals. A novel method based on the double sparse dictionary model joint with Deep Belief Network (DBN) is proposed for the fault diagnosis of rolling bearing. Firstly, each type of fault signal is trained according to the double sparse dictionary learning algorithm and the corresponding double sparse subdictionaries is obtained. In order to reduce the feature dimension of sparse representation coefficients, the low contribution atoms of all subdictionaries are removed and the rest are recombined into a comprehensive double sparse dictionary. Then, the Orthogonal Matching Pursuit (OMP) algorithm is adopted to obtain the corresponding sparse feature coefficients of each fault signal on the comprehensive double sparse dictionary. Finally, the coefficient is used as the input of DBN to train and judge the faults of the rolling bearing. The experimental results show that the proposed method has higher diagnosis accuracy and stability compared with the traditional intelligent fault diagnosis methods, and the training and testing time of DBN is greatly reduced.

INDEX TERMS Sparse representation, double sparse dictionary, deep belief network (DBN), bearing fault diagnosis, feature extraction.

I. INTRODUCTION

Rolling bearing is a key component which is widely used in rotating machinery, the overall performance of the whole machine is often directly affected by its running state. In case of failure, shutdown maintenance or even major production accidents will be inevitable. The mechanical vibration signal contains abundant information which can reflect the working state of rolling bearing. Therefore, it is an effective method to reduce the downtime and the maintain cost and ensure the safe operation of the enterprise by monitoring and diagnosing the working state of rolling bearing based on vibration signal.

There are two main directions for the research of rolling bearing fault diagnosis technology: physical model based on fault mechanism and signal processing model based on data-driven [1]. For a long time, the methods of dynamic analysis based on physical models are popular in traditional fault diagnosis technology. The methods mainly focus on the analysis of the changes of stiffness, damping and mass when the defect occurs, and attempt to establish the dynamic models from the displacement, velocity and acceleration to explain the failure mechanism. For example, Niu *et al.* [2] established the dynamic model of angular contact ball bearing with roller defect and analyzed its vibration response. In order to quantitatively diagnose the fault of rolling bearing, Cui *et al.* [3] quantitatively analyzed the rolling bearing vibration response mechanism with the outer ring failure and established a nonlinear vibration model of the rolling bearing failure severity. Yakout *et al.* [4] studied the effect of the internal radial clearance on the damping characteristics,

The associate editor coordinating the review of this manuscript and approving it for publication was Youqing Wang.

natural vibration modes and fatigue life of rolling bearings. Zhao *et al.* [5] established a bearing dynamic model to explore the effect of raceway defects on the nonlinear dynamic behavior of rolling bearing. However, due to the dynamic analysis methods based on physical models have strong dependence on the mechanism of fault generation, especially for some complex mechanical systems, the physical model behind the damage is very difficult to establish and analyze [6]–[8]. Compared with the traditional physical model-based analysis methods, the data-driven methods have less dependency on the fault mechanism and are more efficient in fault diagnosis without prior knowledge, so they are more fit in with the actual needs of the project. In recent years, with the development of sensor technology and signal processing technology, data-driven methods have become the mainstream methods in the field of fault diagnosis [9].

The main process of data-driven fault diagnosis methods include: signal acquisition, feature extraction and pattern recognition. Among them, feature extraction is the crucial step. Many scholars have done a lot of research on feature extraction and achieved fruitful results. For example, in references [10]–[11], the fault features of rolling bearings were extracted by modal decomposition and wavelet analysis and applied to fault diagnosis. Samanta and Nataraj [12] utilized time-domain features to characterize the bearing health conditions and employed artificial neural networks (ANNs) and support vector machines (SVM) to diagnose faults of bearings. In reference [13], the spectrum of bearing fault signal was transformed into two-dimensional spectrum image to complete fault diagnosis. Cai and Xiao [14] put forward a fault feature extraction method based on the combination of generalized S-transform and singular value decomposition for bearing fault diagnosis. Wang *et al.* [15] summarized and studied the related problems of multivariate statistical process monitoring methods based on time domain signals. In references [16], a new improved kernel partial least squares method, which considered the related key performance indicator information in the residual subspace, has been proposed for related key performance indicator process monitoring. In references [17], [18], the time-frequency domain method combined with convolutional neural network (CNN) was used for bearing fault feature extraction. The above methods mainly focus on feature extraction using time-domain, frequency-domain and time-frequency domain technologies, and then use intelligent algorithm to complete the fault diagnosis. In addition, some scholars have improved the pattern recognition algorithm to achieve the ideal classification effect. For example, in references [19], [20], fault diagnosis was carried out by improving convolutional neural network, and in reference [21], fault diagnosis was carried out by improving the automatic encoder. Though these methods did work in intelligent fault diagnosis of rolling bearing, they still have the following deficiencies: the construction of feature extraction method is more difficult and the features are manually extracted depending on much prior knowledge about signal processing techniques and diagnostic expertise. Furthermore, these manual features are extracted according to a specific diagnosis issue and probably unsuitable for other issues.

To solve the problems mentioned above, some researchers proposed some fully intelligent fault diagnosis methods. These methods do not need to extract features manually, but only need to provide vibration data for the intelligent fault diagnosis model to automatically learn features and complete fault diagnosis. For example, in reference [22], Jia *et al.* established a stack of autoencoder network structure, which can adaptively extract the fault characteristics from raw signals of rolling bearing and automatically classify the bearing health conditions into different group. Shao *et al.* [23] proposed an ensemble deep autoencoders (EDAEs) method, which can extract features directly from the original data to complete intelligent fault diagnosis of rolling bearings. Janssens *et al.* [24] proposed a feature learning model based on CNN for condition monitoring, this model can autonomously learn useful features for bearing fault detection from the data itself. Jing *et al.* [25] developed a CNN to learn features directly from frequency data of vibration signals and tested the different performance of feature learning from raw data, frequency spectrum and combined time-frequency data.

Although these methods have achieved better classification results, there are still some problems, such as complex model, long training time and low diagnosis accuracy, etc. In order to prevent the information lose, the vibration data of one cycle or more must be input to the intelligent fault diagnosis model. With the increase of vibration frequency, the amount of vibration data in a cycle becomes larger. As a result, the fault diagnosis model will become quite complex. It is well known that too many model parameters will make the model difficult to train, and lead to a diagnosis process time consuming, which is not conducive to online condition monitoring and fault diagnosis. Therefore, it is urgent to develop a simple or effective fault diagnosis method which is suitable for practical engineering application.

Compared with the entire vibration signal, the percentage of fault information is relatively less in most instances. In other words, most of the information taken by vibration signal is useless, when we encounter fault diagnosis classification and recognition problems. If there is a way to reduce or remove these components that are not helpful for fault classification, and obtain more effective and concise fault information, it will make the process of fault diagnosis more efficient. Sparse representation [26], [27] is a hotspot in the field of signal processing in recent years. The theory states that by constructing a reasonable dictionary model, a linear combination of a small number of dictionary atoms can be used to approximate the signal. To obtain the ideal sparse coefficient, how to choose the appropriate dictionary is the primary task that must be considered. Most researches on dictionary construction can be mainly divided into two basic approaches: the analysis approach and the learning-based approach [28]. All analysis dictionaries are formed by highly structured mathematical models and provide the fast-numerical implementation, but at the same time they are fixed and limited in their ability to adapt to different types of data. Such dictionaries include Wavelet, discrete cosine transformation (DCT), Contourlets etc. Another type of dictionaries obtained from the training of a set of sample signals themselves is produced by Machine Learning algorithms, thus their atoms have a good self-adaptability to sample signals, but due to their unstructured form, it is relatively difficult to solve this problem. These dictionaries mainly include principal component analysis (PCA), method of optimal directions (MOD) and K-singular value decomposition (K-SVD) [29]. In 2010, Rubinstein *et al.* [30] proposed a dictionary learning model named double sparse (DS), it combines the advantages of analysis dictionary and learning dictionary, and greatly improves the efficiency of dictionary learning.

Deep learning is a machine learning method proposed by professor Hinton and Salakhutdinov [31] in 2006. This method efficiently classifies and recognizes data samples by fully extracting the hidden features in a large number of samples. In recent years, it has made great achievements in machine vision, speech recognition, natural language processing, translation online, data mining and other fields. It has also been favored by a large number of scholars in mechanical fault diagnosis. Deep Belief Network (DBN) is a probabilistic artificial neural network model [32]. This model stacks multiple Restricted Boltzmann Machine (RBM) which has a strong data learning capability by unsupervised learning data features.

Therefore, this paper combines the double sparse dictionary model and DBN to achieve rolling bearing fault diagnosis in the sparse domain. Firstly, the double sparse dictionary is obtained from the vibration signal of rolling bearing by iterative training according to the double sparse dictionary learning method. Then, the original signal is sparse decomposed in double sparse dictionary by using the Orthogonal Matching Pursuit (OMP) algorithm [33], [34], and the sparse coefficient is taken as the feature vector. Finally, the sparse representation feature signals corresponding to different types of fault signals are input into DBN to complete fault diagnosis. Compared with the traditional methods of intelligent fault diagnosis using original vibration signal, the method proposed in this paper can express the essential characteristics of the original signal with less data through sparse decomposition, which greatly simplifies the intelligent fault diagnosis model and reduces the training time of the model, and the method has small diagnosis fluctuation, good stability and high diagnosis accuracy.

The remainder of the paper is organized as follows. Section II briefly introduces the sparse representation theory and double sparse dictionary learning algorithm. In Section III, a description of the proposed diagnosis technique is given, and the overall framework of the proposed method is illustrated. Section IV explains the experimental effect of the proposed method through experimental analysis and comparison with other commonly used methods. We summarize and conclude the paper in Section V.

II. DOUBLE SPARSE DICTIONARY

A. SPARSE REPRESENTATION THEORY

The sparse representation theory states that by constructing an appropriate over-complete dictionary D, any discrete signal x can be represented by a linear combination of dictionary atoms.

$$x = D\alpha \tag{1}$$

where $D = \{d_k \mid k = 1, 2, \cdots, K\}$ is an over-complete dictionary, d_k is a column vector of length N, which is called an atom. x is the N-dimensional original signal, α represents the sparse coefficient vector.

The goal of the sparse representation is to make the number of non-zero elements in α as small as possible. In mathematical form, it can be described as:

$$\alpha = argmin \, \|\alpha\|_0^0 \quad s.t. \, \|x - D\alpha\|_2^2 \leq \epsilon \tag{2}$$

where $\|\alpha\|_0^0$ means the number of nonzero elements in α. ϵ means error tolerance term.

When the dictionary is known, the process of solving the sparse coefficient is called sparse decomposition. Because of the redundancy of dictionary atoms, it is a NP-hard problem to solve the above formula by using the l_0 norm, which can be converted into solving the l_1 norm [35]–[37].

$$\alpha = argmin \, \|\alpha\|_1^1 \quad s.t. \|x - D\alpha\|_2^2 \leq \epsilon \tag{3}$$

The commonly used solution methods to the above formula are divided into two categories: one is convex relaxation algorithms, such as base pursuit algorithm (BP) [38], interior point methodology (IPM) [39] etc. The other is greedy algorithm, such as matching pursuit Algorithm (MP) [26], OMP [33] etc. Since OMP algorithm is the most widely used method to solve this kind of optimization problem, it is used to solve the sparse coefficient in this paper.

B. DOUBLE SPARSE DICTIONARY

The double sparse dictionary learning algorithm combines the analysis dictionary and the learning-based dictionary, inherits the respective advantages of the two dictionaries, and has a more efficient sparse representation ability [30], [40]. The basic rule is that the double sparse dictionary Ψ is sparsely represented under the base dictionary B again, and the mathematical form is described as follows:

$$X = \Psi\Gamma = BA\Gamma \tag{4}$$

where X is the N-dimensional original signal matrix, A is the sparse dictionary matrix of dictionary Ψ under base dictionary B, Γ represents the sparse representation coefficient matrix.

Given the signal sample X, the objective function to obtain a double sparse dictionary is described as:

$$\hat{A},\Gamma = argmin\,\|X - BA\Gamma\|_F^2 \quad s.t.\begin{cases}\forall i, & \|\gamma_i\|_0^0 \le t\\ \forall j, & \|\alpha_j\|_0^0 \le p,\ \ \|B\alpha_j\|_2 \le p\end{cases} \tag{5}$$

where t represents the sparsity of the sparse representation coefficient, p represents the sparsity of column vector of sparse dictionary matrix A, γ_i represents the i-th column coefficient vector of the sparse representation coefficient matrix Γ, and α_j represents the j-th column vector of the sparse dictionary matrix A.

In the sparse coding stage, the algorithm framework of sparse dictionary $\hat{A}$ and sparse coefficient matrix Γ is the same as that of K-SVD. But they are different in the dictionary update stage. The specific algorithm flow is shown as follows.

Input: The matrix of signal sample set $X = [x_1, x_2, \cdots, x_N]$, base dictionary B trained based on learning dictionary algorithm, initial dictionary representation A_0, the upper limit p of sparsity of atoms in sparse dictionary, the upper limit t of sparsity of sparse representation coefficient of training samples; number of algorithm training iterations $j = 1, 2, \cdots, J$.

Output: double sparse dictionary Ψ.

Step 1: Sparse coding

Let $A = A_0$, the OMP algorithm is used to solve the sparse coefficient of the sample signal in the current sparse dictionary A, and the solution objective function is described as:

$$\forall i, \quad \Gamma_i = argmin\,\|x_i - BA\gamma\|_2^2 \quad s.t.\|\gamma\|_0^0 \le t \tag{6}$$

Step 2: Dictionary update

The atoms of the sparse dictionary A are updated column by column. Assuming that the j-th column atom α_j of the sparse dictionary A is currently updated, let $\alpha_j = 0$, and matrix I is introduced to store the sample signal index of α_j used in sparse representation. Let the current sparse coefficient $g = \Gamma_{j,I}^T$. In order to meet the constraints requirements of the objective function, g is standardized to make $g = \frac{g}{\|g\|_2}$. At this time, the objective function form of dictionary update is described as:

$$\{\alpha_j, g\} = argmin\,\|E_j - B\alpha_j g\|_2^2 \quad s.t.\,\|\alpha\|_0^0 \le p, \quad \|B\alpha_j\|_2 = 1 \tag{7}$$

where E_j represents the reconstruction error when only the j-th atom is used for sparse representation.

$$E_j = X_I - BA\Gamma_I \tag{8}$$

Multiply g at both ends of the above formula and let:

$$z = E_j g = X_I g - BA\Gamma_I g \tag{9}$$

After the above processing, when the dictionary atom is updated, the objective function is converted into:

$$\alpha_j = argmin\,\|z - B\alpha_j\|_2^2 \quad s.t.\,\|\alpha_j\|_0^0 \le p \tag{10}$$

First standardize $\|B\alpha_j\|_2 = 1$, and $\hat{\alpha}_j = \frac{\alpha_j}{\|B\alpha_j\|_2}$, A_j is replaced with $\hat{\alpha}_j$, at this point, the atomic update of the sparse dictionary is completed.

After updating g according to the formula described below, update the sparse coefficient and obtain the final sparse coefficient matrix. Update sparse coefficient matrix: the sparse coefficient $\Gamma_{j,I}^T$ will be updated after g is modified according to the following formula:

$$g = \left(E_j^I\right)^T B\alpha_j \tag{11}$$

Step 3: After several iterations, all the sparse dictionary atoms are updated and the updated sparse dictionary matrix A is output. Then A is multiplied by the base dictionary B to get the double sparse dictionary Ψ.

III. FAULT DIAGNOSIS METHOD BASED ON DOUBLE SPARSE DICTIONARY JOINT WITH DBN

Aiming at some of the limitations of traditional rolling bearing condition monitoring methods in the context of big data, a rolling bearing fault diagnosis method based on the double sparse dictionary model and DBN is proposed in this paper. The vibration signals are sparsely coded in the double sparse dictionary, and the DBN performs intelligent fault diagnosis.

A. IMPLEMENTATION OF THE PROPOSED METHOD

There are four types of vibration signals: the normal state (N), the inner race fault (IF), the outer race fault (OF), and the rolling element fault (RF) which were downloaded from the bearing data center of the Case Western Reserve University (CWRU) in the United States. Each original health condition data set comprises 100 samples and each sample comprises 1024 sampling points. The data is trained separately to obtain the double sparse subdictionaries Ψ_1, Ψ_2, Ψ_3 and Ψ_4 corresponding to their respective states. The calculation method of the reconstruction accuracy δ when the signal is sparsely represented is defined as follows:

$$\delta = \left(1 - \frac{\left\|\hat{f} - f\right\|_2}{\|f\|_2}\right) \times 100\% \tag{12}$$

where $\hat{f}$ represents the reconstructed signal, and f represents the original signal. The relative reconstruction accuracy of different type of signals respectively sparsely decomposed in the subdictionaries obtained from self-type sample signals and other types is analyzed, the histogram of the average value distribution of the reconstruction accuracy of each type of signal under different state double sparse sub-dictionaries is shown in Fig. 1.

It can be clearly seen that each type of signal has the highest reconstruction accuracy only in the subdictionary trained with the identical fault category signal as itself. Therefore, after the subdictionaries of different signals trained by double sparse dictionary learning algorithm are combined into a comprehensive dictionary and when the vibration signal of a specific

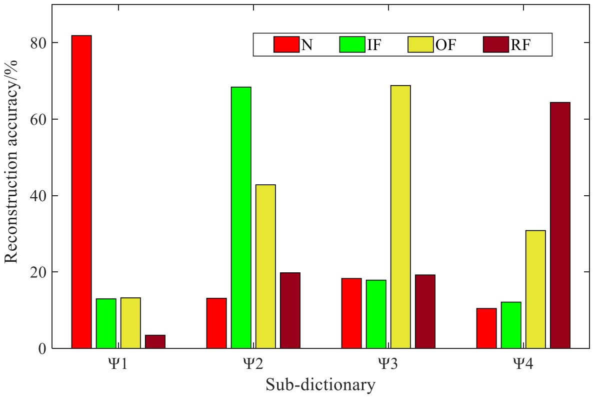


FIGURE 1. Relative reconstruction accuracy of signals sparsely decomposed under different dictionary bases.

health condition is sparse decomposed in the comprehensive dictionary, it is more likely to choose the dictionary atoms obtained from the same fault signal training as the decomposition atoms.

So, there's such an idea, if different double sparse sub-dictionaries obtained from all types of fault states are combined in order, a comprehensive double sparse dictionary is created, and its form can be described as $\hat{\Psi} = \{\Psi_c \mid \Psi_1, \Psi_2, \cdots, \Psi_C\}$. The non-zero elements of the sparse decomposition vector of each fault type's original signal under the comprehensive dictionary show the characteristics of high block concentration, as is described in (13).

$$x_\tau = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1\tau} & \cdots & d_{1\mathrm{T}} \\ d_{21} & d_{22} & \cdots & d_{2\tau} & \cdots & d_{2\mathrm{T}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{m\tau} & \cdots & d_{m\mathrm{T}} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \gamma_\tau \\ \vdots \\ 0 \end{bmatrix} \tag{13}$$

where x_τ is the sample signal of the τ-type failure of the rolling bearing, $\Psi_\tau = [d_{1\tau}, d_{2\tau}, \cdots, d_{m\tau}]^T$ represents a double sparse subdictionary obtained from the training of the τ-th state signal, γ_τ represents the non-zero element set of the sparse decomposition vector of the τ-th fault signal x_τ in the comprehensive double sparse dictionary $\hat{\Psi}$. The sparse decomposition coefficients of this centralized block representation are very distinguishable and can be used as input of neural networks for fault classification and recognition.

DBN is a multi-hidden layer probabilistic generating artificial neural network model constructed by stacking multiple RBMs. Its core component, RBM, is energy-based probability distribution model algorithm. In the method proposed in this paper, the sparse points of different types of fault sparse representation feature signals obtained by sparse decomposition under double sparse comprehensive dictionary have the characteristic of energy distribution concentration, which largely conforms to the characteristic of RBM feature learning. Therefore, the method of fault diagnosis based on sparse representation and DBN has good classification and recognition results theoretically.

RBM is a shallow learning model. Its basic structure consists of a visible layer and a hidden layer. Neurons in the same layer are independent of each other. Neurons in different layers are connected in two directions. Information flows in two directions between neurons during network training and use. Given the state vectors v and h, the energy function of the RBM can be expressed as:

$$E\{v, h \mid \theta\} = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i h_j w_{ij} \tag{14}$$

where v is visible layer, h is hidden layer. $\theta = \{a_i, b_j, w_{ij}\}$ represents the parameter obtained from RBM training. The updating rules of parameters during RBM training are performed according to the following formulas.

$$\begin{aligned} \Delta w_{ij}^n &= \mu \Delta w_{ij}^{n-1} + \varepsilon \left(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{recon}\right) \\ \Delta a_i^n &= \mu \Delta a_i^{n-1} + \varepsilon \left(\langle h_j \rangle_{data} - \langle h_j \rangle_{recon}\right) \\ \Delta b_j^n &= \mu \Delta b_j^{n-1} + \varepsilon \left(\langle v_i \rangle_{data} - \langle v_i \rangle_{recon}\right) \end{aligned} \tag{15}$$

where, μ represents the momentum learning rate, ε represents the learning rate, n represents the number of iterations of RBM training; $\langle \cdot \rangle_{data}$ represents the mathematical expectation of the input data set during training, $\langle \cdot \rangle_{recon}$ represents the model mathematical expectation obtained by one iteration reconstruction using the contrast divergence algorithm.

During RBM pre-training, the sparse representation feature signal set Γ is inputted into the visible layer v_1 of the first RBM, and the output layer h_1 and the weights and bias parameters $\theta = \{a_{1i}, b_{1j}, w_{ij}^1\}$ of the present RBM are calculated. Then, the remaining RBM is trained by the same method and the corresponding parameters are obtained. Finally, the trained RBMs are stacked in series and BP neural network is added at the end to form a complete DBN. The basic structure is shown in Fig. 2.

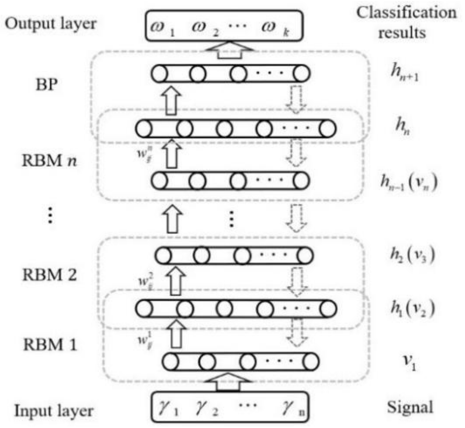


FIGURE 2. The structure of DBN.

By using RBM for feature learning, we can get more profound feature expression. Finally, the back-propagation (BP)

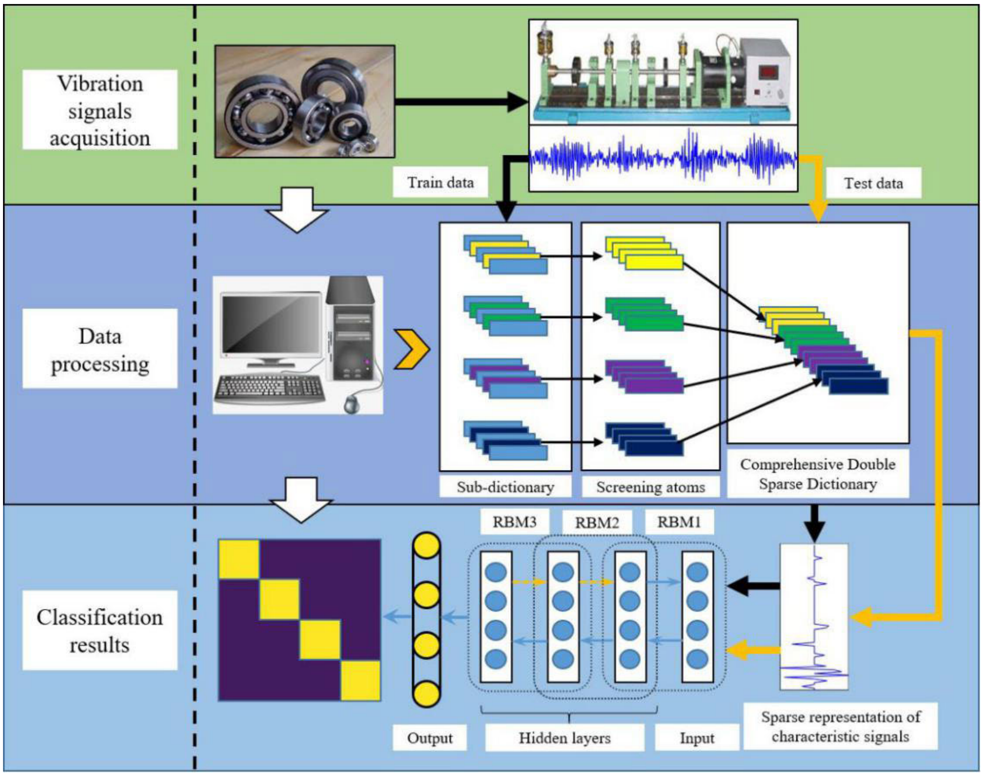


FIGURE 3. **The overall framework of the proposed method.**

algorithm is used to fine tune the parameters of the whole network in order to obtain higher classification accuracy.

B. OVERALL FRAMEWORK OF THE PROPOSED METHOD

The proposed method gives up dependence on the physical model characteristics of the signal and uses the sparse representation coefficients of the original signal in the comprehensive double sparse dictionary as the input of the DBN for fault diagnosis. The overall framework of the proposed method is shown in Fig. 3. When using the proposed method for fault diagnosis, we need to focus on the following two stages: comprehensive dictionary acquisition based on double sparse dictionary model and DBN model training based on sparse representation features.

In the comprehensive dictionary acquisition stage, the original time-domain vibration signals of rolling bearing in various states are firstly used to train the subdictionaries according to the double sparse dictionary learning algorithm flowchart described in section II. Since the K-SVD dictionary has excellent performance in sparse representation of vibration signals and is widely used in dictionary learning, this article uses the K-SVD algorithm to train and obtain the dictionary as the base dictionary D in the double sparse dictionary training. Then OMP algorithm is used to solve the sparse coefficient matrix of the original signals under the base dictionary, and the sparse coefficient matrix is used as the sparse dictionary A. Finally the double sparse dictionary learning algorithm is used to solve the double sparse subdictionary Ψ.

However, the obtained dictionary atoms in this training method are redundant. A large number of dictionary atoms are not used or adopted at a low frequency in sparse decomposition, and the existence of these atoms makes the dimension of sparse representation feature vectors obtained in sparse decomposition increase, which is not conducive to the subsequent training and testing of DBN. Therefore, the dictionary atoms need to be screened to eliminate the atoms with low contribution rate.

How to screen atoms depends on the sparse decomposition algorithm. This article uses the OMP algorithm as the sparse decomposition algorithm. In the OMP algorithm, the priority of the atom to be decomposed is determined by the largest absolute value of the inner product of the signal to be decomposed x_i and the dictionary atom d_k. That is, the dictionary atom with the largest value of $\left|\langle x_i \cdot d_k^T \rangle\right|$ is used first. Therefore, the atoms that have the largest absolute value of the inner product with the training set signal and are used most frequently during sparse decomposition are selected as candidate atoms in the sub-dictionary set. After the candidate atoms are selected in the same way, each type of sub-dictionary atoms are arranged and combined in sequence to form the final double sparse comprehensive dictionary $\hat{\Psi}$. Then the decomposition coefficients of various signals are solved by OMP under the double sparse comprehensive dictionary $\hat{\Psi}$ and used as feature vectors, and fault labels are added for DBN training.

In the DBN training stage, a DBN model with multiple hidden layers needs to be established firstly, and the vibration signals of different health status categories are decomposed under the double sparse comprehensive dictionary $\hat{\Psi}$ to obtain the corresponding sparse feature signal set Γ. Then the set is input into the DBN to train the network. The dimension of the DBN visible layer is consistent with the dimension of the input sparse feature vector. According to Gibbs sampling, the probability of activation $\mathrm{P}(h_1|v_1)$ of neurons in the hidden layer h_1 is calculated, and the output value of the hidden layer $h_1 = \{h_{11}, h_{12}, \ldots, h_{1j}\}$ is obtained, simultaneously the weights of the layer and the biases of the corresponding neurons are calculated. A sample is extracted from the calculated hidden layer by Gibbs sampling again and the probability of the lower edge distribution of the joint distribution is solved to obtain the activation probability $\mathrm{P}(v_1|h_1)$ of v_1 neurons in the visible layer. Then the visible layer $v_1 = \{v_{11}, v_{12}, \ldots, v_{1j}\}$ is reconstructed and the bias of each neuron is solved. The reconstructed visible layer is then re-input into the visible layer and the output is calculated by the Sigmoid activation function to obtain the final output result of the hidden layer. The weights and biases are calculated according to Eq. (15). The above steps are repeated until the mean square error reaches the minimum, and the first RBM training is completed. Then the next RBM is trained according to the same steps, and the corresponding weights and biases of each layer are saved. However, the weights of each RBM and the bias of each neuron only perform optimally when the feature vector of the corresponding layer is mapped, and it is not optimal on the entire network structure, so the BP algorithm is also needed to fine-tune the whole DBN network parameters from top to bottom. The parameters of the trained network model will be explained in detail in Section IV-B.

After the parameters of the whole network model are determined, another set of original vibration signals is taken as the test set, and the sparse representation features of the signals under the double sparse comprehensive dictionary $\hat{\Psi}$ are input into the trained network model for testing, and the classification and recognition results of the proposed method are analyzed and evaluated.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental data provided by CWRU Bearing Data Center [41] are analyzed in this section to verify the effectiveness of the proposed method. The data acquisition test rig is shown in Fig. 4, and the test rig is arranged from left to right with a 2-horsepower motor (left), a torque transducer/encoder (center), a dynamometer (right) and control electronics (not shown). Vibration data are collected using accelerometers, which are attached to the housing with magnetic bases. In the experiment, the bearings used in the drive end are the 6205-2RS JEM SKF deep groove ball bearings, and the number of rolling elements is 8, and single point faults are introduced to the test bearings using electro-discharge machining.

All experiments in this paper are based on a PC platform with Intel (R) Core (TM) i5-4590 CPU @ 3.30GHz, 8.00 GB of memory, and Windows 10 64-bit operating system. MATLAB is the software used in this paper and its version is R2017b. There are 4 types of data used in this experiment, including normal baseline signal data (N), inner ring fault (IF), outer ring fault (OF), and rolling element fault (RF) data. The measurement points are all located at the drive end, and the applied load is 3 horsepower. The reference speed is 1730 r / min, the sampling frequency is 12 kHz, and the degree of failure is 0.18mm. In order to ensure the validity of the experimental data, the length of each sample signal is set to 1024, that is, it contains the sampled signals within two complete vibration cycles. The training set and testing set of the specific data are shown in Table 1.

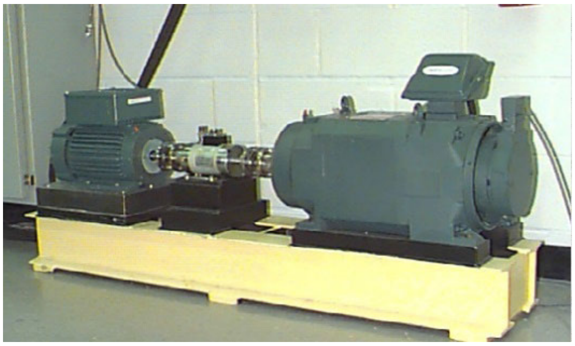


FIGURE 4. The data acquisition test rig of CWRU.

TABLE 1. Bearing data set.

Sample type	Number of samples		Data label
	Training set	Testing set	
N	500	200	1000
IF	500	200	0100
OF	500	200	0010
RF	500	200	0001

According to the proposed method, the corresponding double sparse comprehensive dictionary is first obtained according to different types of signals in the experiment. The impact of the double sparse dictionary training parameters and the subdictionary atom screening parameters on the diagnosis results is evaluated. Then, the DBN model is constructed, and the relevant parameters of the model are trained step by step. Next, the testing set signals are used for the fault classification and recognition of the rolling bearing by using the proposed method in the paper. The average accuracy of the rolling bearing fault recognition can reach 98. 12% in 15 times experiments. Finally, the diagnostic effect between the proposed method and the commonly used intelligent fault diagnosis methods is compared and analyzed.

A. DOUBLE SPARSE DICTIONARY TRAINING AND INFLUENCE ANALYSIS OF MAIN PARAMETERS

When the vibration signal is sparsely decomposed by using double sparse dictionaries, the base dictionary and parameter values in the dictionary model need to be reasonably selected in order to obtain an over-complete dictionary with better performance, which is conducive to fault classification.

There are eight parameters of the double sparse comprehensive dictionary that have an effect on the classification accuracy of the fault signal, which are listed as follows, the atoms number of the base dictionary B is n, and its training iterations number is k, N is the atoms number of the sparse dictionary A, the sparsity of the double sparse comprehensive dictionary is p, the sparsity of the vibration signal is t, K represents the iteration number during the training of the double sparse dictionary model, m is the atoms number of the sub-dictionary to be selected when the comprehensive dictionary is obtained, and s indicates the number of non-zero elements of sparse representation eigenvector when the comprehensive dictionary is obtained. Single factor analysis is used to determine the parameters of double sparse dictionary, that is, the remaining parameters are fixed, and only the influence of a single parameter change on the diagnostic accuracy is considered. When each parameter value is changed, the average value and standard deviation of classification accuracy of 15 experiments are taken to analyze the influence of each parameter on the diagnosis accuracy, as shown in Fig. 5.

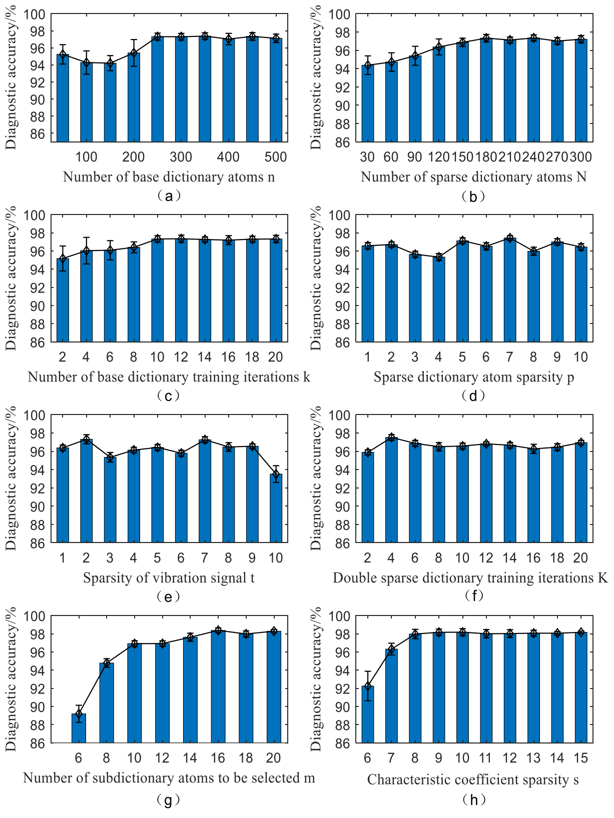


FIGURE 5. The influence of dictionary training parameters on diagnosis accuracy.

Firstly, the influence of the number of atoms n in the base dictionary on the diagnosis accuracy is considered. The initial values of other parameters are tentatively determined as: $N = 300$, $k = 10$, $K = 4$, $t = 2$, $p = 5$, $m = 10$, $s = 10$. The initial number of layers of DBN is set as 2 layers, the number of neurons in the hidden layer is set as 40, the learning rate is 0.01, the momentum is set as 0.5, and the number of RBM training is set as 100. It can be seen from Fig. 5-(a) that as the number of atoms in the base dictionary changes from 50 to 500 in steps of 50, the diagnosis accuracy is the highest and the diagnosis result is relatively stable when the number of atoms in the base dictionary is 250, so the number of atoms in the base dictionary is determined as 250. When the number of atoms in the sparse dictionary is determined, it changes from 30 to 300 in steps of 30. The corresponding diagnosis accuracy is shown in Fig.5-(b), it can be seen that when the number of atoms in the sparse dictionary is small, the diagnosis accuracy is low and the result is unstable. When the number of atoms in the sparse dictionary is 180 or more, the diagnosis accuracy is high and the result is stable. Considering the time of dictionary training, the number of atoms in the sparse dictionary is determined to be 180. Next, the influence of the number of iterations of the base dictionary on the diagnosis accuracy is studied, It can be seen from Fig 5-(c) that when the training times of the base dictionary reach 10 times, the diagnosis accuracy is high and the fluctuation is not obvious, and the diagnosis accuracy is basically unchanged with the increase of the number of iterations, so the number of training iterations of the base dictionary are determined as 10 times. From figure Fig 5-(d), we can see that the value of atom sparsity of sparse dictionary has little effect on the diagnosis accuracy, therefore, the sparse value 7 is selected when the diagnosis accuracy is the highest. Then, the influence of vibration signal sparsity on diagnosis accuracy is considered, we can see from Fig. 5-(e) that its value has a good result when it is taken as 2. Therefore, the sparsity of vibration signal is 2. Then the influence of the number of iterations of double sparse dictionary on the diagnosis accuracy is considered. Considering the influence of dictionary training time, the value with high diagnosis accuracy and low iteration times is selected. According to the results shown in Fig. 5-(f), the number of training iterations of double sparse dictionary is finally determined to be 4. After the double sparse subdictionary of four types of signals are trained, if four subdictionaries are directly used as comprehensive dictionary, the dimension of sparse representation feature signals is too high, Therefore, it is necessary to further select the atoms of the sub-dictionary. The screening method is determined by the maximum absolute value of the inner product of the double sparse dictionary atom and the original signal. The larger the absolute value of the inner product is, the higher the matching degree between the signal and the dictionary atom is. Fig. 5-(g) shows that that when the number of atoms in the selected sub dictionary is 16 or above, the improvement of average accuracy is not obvious, but relatively speaking, when the number of atoms selected is 20, the result of diagnosis accuracy is more stable, so the number of atoms selected is determined as 20. Finally, the sparsity of the sparse representation feature signal in the double sparse comprehensive dictionary is determined, it can be seen from Fig. 5-(h) that when the sparsity of sparse representation feature signal is 9, the diagnosis accuracy is the highest, and the diagnosis accuracy is basically unchanged with the increase of its value. For the convenience of calculation, integer value 10 is selected as the final sparse representation of the sparse degree of the feature signal. To sum up, the final parameters of the double sparse dictionary are determined as: $n = 250$, $N = 180$, $k = 10$, $p = 7$, $t = 2$, $K = 4$, $m = 20$, and $s = 10$.

B. DBN TRAINING AND PARAMETER ANALYSIS

After the sparse representation feature signal is obtained, the fault diagnosis is realized by DBN. When DBN is used for sample classification, the following parameters should be paid more attention: the number of RBM δ, the number of hidden neurons λ, the learning rate ε, momentum μ, and the number of network training ρ. In general, the smaller the learning rate ε is, the better the learning effect is, but it will increase the network training time. Usually, the value of learning rate ε ranges from 0.01 to 0.8. In this paper, the learning rate ε is 0.01. The other parameters are determined by experimental analysis. The influence of various parameters change on the diagnosis accuracy is shown in Fig.6. The average value and corresponding standard deviation of 15 experiments are taken for analysis of each experiment result.

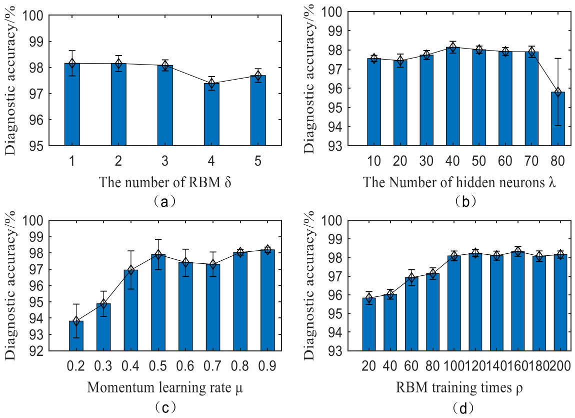


FIGURE 6. The influence of DBN parameters of on diagnosis accuracy.

Fig. 6-(a) shows the influence of RBM number on the diagnosis accuracy. It can be seen that when the number of RBM used in DBN is 1 to 3, the diagnosis accuracy is relatively high. When the number of RBM is 3, the standard deviation of diagnosis accuracy is the smallest, which indicates that the diagnosis result is relatively stable when this value is taken. The diagnosis result is slightly low when the number of RBM continues to increase. Considering the factors of saving training time, therefore, the number of RBM of DBN is determined to be 3. Fig. 6-(b) shows the effect of the number of neurons in the hidden layer of DBN on the diagnosis results. The number of neurons in each hidden layer is the same value. It can be seen that when the DBN structure is 40-40-40, the diagnosis accuracy is the highest and the standard deviation of accuracy is small. Therefore, the hidden layer network structure of DBN is determined as 40-40-40. Fig. 6-(c) shows the influence of momentum on the diagnosis accuracy. The momentum value varies from 0.2 to 0.9. It can be seen that when the momentum value is 0.5, the diagnosis result reaches a high value, but it is not the optimal value. With the further increase of the momentum value, the diagnosis result basically remains unchanged. The average diagnosis accuracy dimension is above 97%. When the momentum values are 0.8 and 0.9, the standard deviation of accuracy is relatively small, but when the momentum value is 0.9, the diagnosis result standard deviation is the smallest and the diagnosis accuracy is relatively higher, so the momentum value is 0.9. Fig. 6-(d) shows the influence of RBM training times on the diagnosis results. When the RBM training times increase from 20 to 200, the diagnosis results reach the highest state after the training times are 100, but relatively speaking, the training times are 120 with higher diagnosis result and the standard deviation of accuracy is lower. Therefore, the training times of RBM are determined as 120.

According to the analysis result shown in Fig. 6, under the consideration of classification accuracy and model training time costs, the main parameters of DBN are finally determined as follows: $\delta = 3$, $\lambda = 40$, $\varepsilon = 0.01$, $\mu = 0.9$, and $\rho = 120$.

C. COMPARISON OF DIAGNOSTIC RESULTS OF THIS METHOD WITH COMMONLY USED METHODS

In the traditional mechanical fault intelligent diagnosis methods, the most commonly used methods are the classification method based on the combination of time domain statistical features and support vector machine (SVM) fault classification and the classification method based on single layer and deep layer back propagation neural network (BPNN). This section compares the proposed method with deep layer BPNN with the same architecture as DBN, a single hidden layer BPNN with shallow structure, as well as the SVM-based method combined SVM with commonly used 14 time domain statistical features. The 14 time domain statistical features are maximum, minimum, mean, peak-to-peak value, rectified mean value, variance, standard deviation, kurtosis, skewness, root-mean-square, waveform factor, peak factor, impulse factor and margin factor respectively. The number of samples of each type of fault signal in the test set is 200, so there are 800 samples for four types of fault in the test set. Four different methods are used to classify and test the four types of fault signals.

In order to demonstrate the ability of the proposed method to automatically learn valuable features, t-SNE method is used to illustrate the results of feature learning of the proposed method, deep layer BPNN and single hidden layer BPNN methods respectively, as shown in Fig. 7. Because the SVM-based method combined SVM with commonly used time domain statistical features does not have the process of feature learning, its feature results are not displayed. It can be seen from Fig. 7 that the propose method can represent the input data in a more precise and identifiable way than other methods do. Among the four types of state signals,

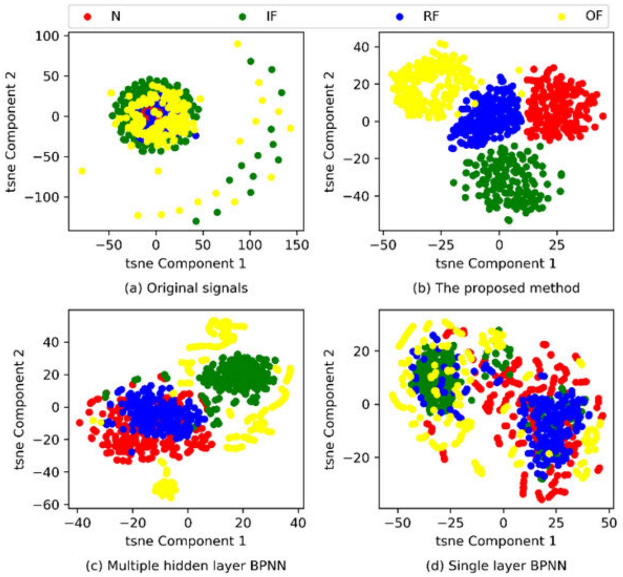


FIGURE 7. Two dimensional projection of feature learning effect of different methods.

the features of the same type of signals extracted by the proposed method have better aggregation, while the different types of signals have better separability. The main reason is that the sparse vectors with lower dimensions obtained by double sparse decomposition contain abundant essential information of the original signal and hidden layers of DBN allow the deep architectures to be more powerful in modeling the complex nonlinear relationship hidden in sparse vectors. The feature diagrams of four types of original state signals obtained by t-SNE method are almost all overlapped. This is because although the original signals contain a lot of information, their essential distinguishing features have not been obtained. The features obtained by the deep layer BPNN method and the single hidden layer BPNN method are also overlapped. The main reason is that these two methods have poor ability to learn complex nonlinear relationship of input data.

The confusion matrices of the four classification results are shown in Fig. 8. The confusion matrix demonstrates the classification results of all the conditions in detail, which contains classification accuracy and misclassification error. The ordinate axis of the confusion matrix refers to actual category, and the horizontal axis represents prediction category. The color bar in right illustrates the correspondence between colors and numbers from 0 and 1. It can be seen from Fig. 8-(a) that the sparse decomposition features of normal condition signals are more distinct from those of other fault category signals, see Fig. 10 for details, so the diagnostic accuracy is ideal to reach 100%. Among the other fault categories, only the inner ring fault signal has the lowest diagnostic accuracy, but the diagnostic accuracy still reaches 96%.

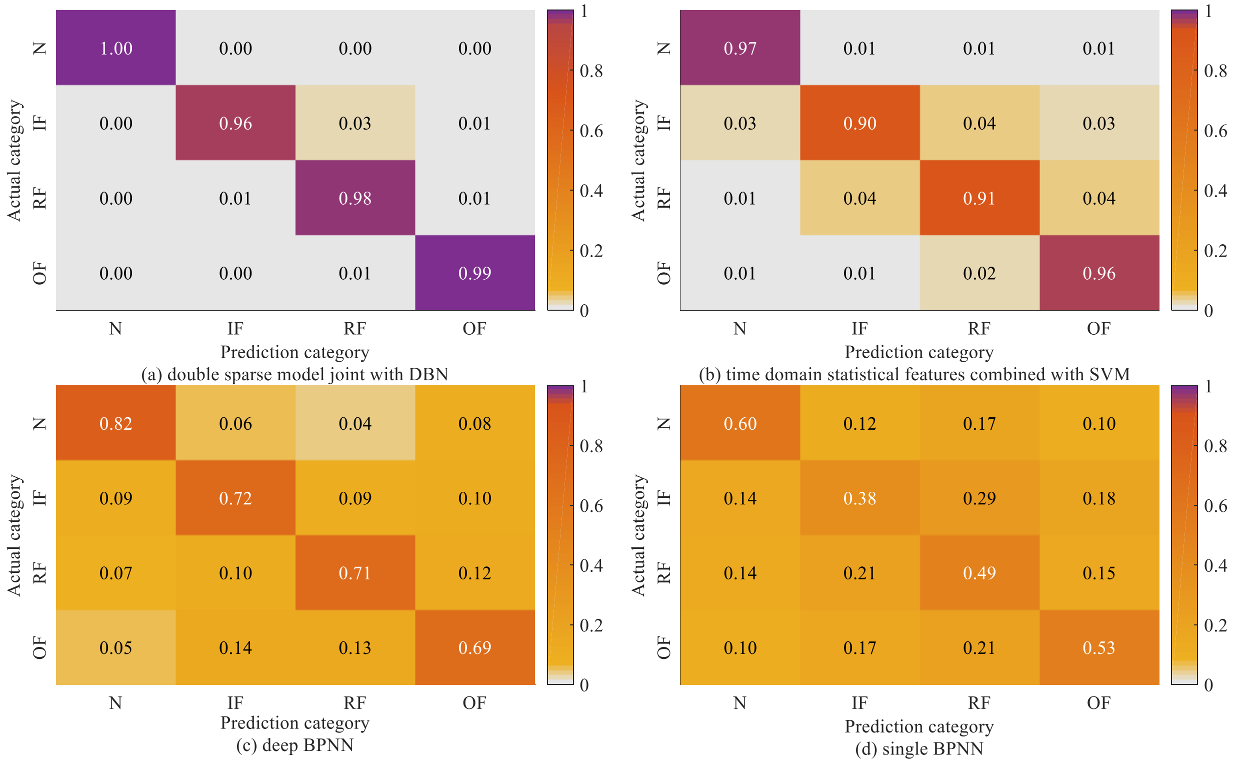


FIGURE 8. The confusion matrix between the proposed method and the common intelligent fault diagnosis method. (a) Classification effect based on double sparse model joint with DBN; (b) Classification effect of time domain statistical features combined with SVM; (c) Classification effect of deep BPNN; (d) Classification effect of single BPNN.

Among 200 samples of inner ring fault, 6 samples are misdiagnosed as rolling element fault and 2 samples are misdiagnosed as outer ring fault. The diagnosis accuracy of rolling element fault is 98%. Among 200 samples of rolling element fault, 2 samples are misdiagnosed as inner ring fault and 2 samples are misdiagnosed as outer ring fault. The diagnosis accuracy of outer ring fault is 99%. Only 2 samples of the 200 outer ring fault samples are misdiagnosed as rolling element fault. The reason for the misclassification is that all of large values of the sparse decomposition features of inner ring fault, outer ring fault and rolling element fault have obvious interval differentiation, but there are also a few amount of small value interval overlap. On the whole, the average classification accuracy is over 98%, and there are less than 14 misclassified samples. Considering that the acquisition time of each signal length is only 8.5ms, even if one sample signal is misjudged, the next sample signal can quickly correct the system fault when it is collected, so the method proposed in this paper has great application potential in the field of engineering practice.

For other commonly used intelligent fault diagnosis methods, only the method of the combination of statistical characteristics and SVM is better, but compared with Fig. 8-(*a*) and Fig. 8-(*b*), the classification effect of this method is slightly worse. The other two methods of BPNN classification, whether deep or single-layer, are not very effective, and because the sample signal dimension is too high, single-layer BPNN can hardly complete the classification, as shown in Fig. 8-(c) and Fig. 8-(d).

The diagnostic results of 15 experiments with four different methods are shown in Fig. 9. It can be seen that the results of 15 experiments of the method proposed in this paper have high diagnostic accuracy, and the diagnostic result curve is approximately a straight line, which shows that the method has good accuracy stability and robustness. The method based on 14 time-domain statistical features combined with SVM has relatively high diagnostic results, but its diagnostic accuracy and stability are slightly lower than that of the proposed method, however it has obvious advantages compared with the other two methods. The diagnostic accuracy of single layer BPNN is relatively low, and the fluctuation is obvious. The diagnostic accuracy of deep layer BPNN is higher than that of single layer BPNN, but the fluctuation is greater.

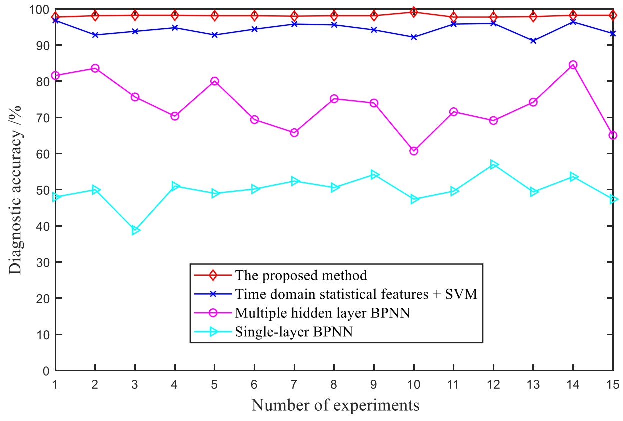


FIGURE 9. Diagnostic accuracy analysis curve of the proposed method and common fault diagnosis method.

The average diagnostic accuracy and the corresponding standard deviation of 15 experiments of the four methods are shown in Table 2. It can be seen that the average diagnostic accuracy of the proposed method is the highest, and the average diagnostic accuracy can reach 98.12%. The method of single layer BPNN is the lowest, and the average diagnostic accuracy is only 49.91%. From the aspect of diagnosis stability, the method proposed in this paper has the highest stability, and its standard deviation is 0.33%. The stability of deep BPNN is the worst, and the standard deviation of diagnosis accuracy is as high as 6.99%.

TABLE 2. Mean and standard deviation of diagnostic accuracy.

experimental method	Average diagnostic accuracy/%	Standard deviation of diagnostic accuracy/%
The proposed method	98.12	0.33
Statistical features combined with SVM	94.38	1.68
Multiple hidden layer BPNN	73.38	6.99
Single layer BPNN	49.91	4.06

Based on the above experimental results, it can be concluded that the proposed method is superior to the commonly used intelligent fault diagnosis methods. The reason is that the rolling bearing vibration signals of different health status categories have their own characteristics. The method proposed in this paper takes advantage of the fact that the vibration signals only have the highest matching degree with the double sparse dictionary atoms obtained from the training set of their own type signals. By eliminating redundant atoms with low contribution rate in sparse decomposition of over complete dictionary, the dictionary model becomes more simplified. The dimension of sparse representation feature coefficient of the original vibration signal with length of 1024 is only 80. Because the feature coefficient is a sparse representation feature vector, and there are only 10 sparse characteristic points, a small amount of data can show the obvious characteristic information of the signal different from the other types. In addition, this paper uses the DBN with strong ability of information mining as classifier, which has a high ability of differential information recognition. High diagnostic accuracy can be achieved by learning sparse representation of feature signals. However, the statistical characteristics based on time domain are often difficult to accurately and comprehensively reflect the characteristics of the original signal itself, so the average diagnostic accuracy of SVM method is slightly lower. Because the deep BPNN uses the back-propagation algorithm to train the network, the whole network has poor performance in stability, generalization and other aspects, so the diagnostic accuracy and stability are not high enough. Due to the limitation of the shallow structure, the single layer BPNN has limited ability to learn the features of high-dimensional original signals, which leads to its low diagnosis accuracy.

D. THE ROLE OF DOUBLE SPARSE COMPREHENSIVE DICTIONARIES IN FAULT DIAGNOSIS

Compared with the DBN method for fault diagnosis when the original signal is used directly, the sparse representation eigenvector of the original vibration signal under the double sparse dictionary is used as the data for deep neural network learning in proposed method.

Its most obvious advantage is to reduce the input data dimension of DBN model, reduce the complexity of neural network training, and greatly save the overall training time of the network. In order to illustrate the classification effectiveness of the proposed method, one sample is randomly selected from all types of fault state sample signals and its sparse representation feature signal under the comprehensive dictionary is obtained for feature separability, as shown in Fig. 10.

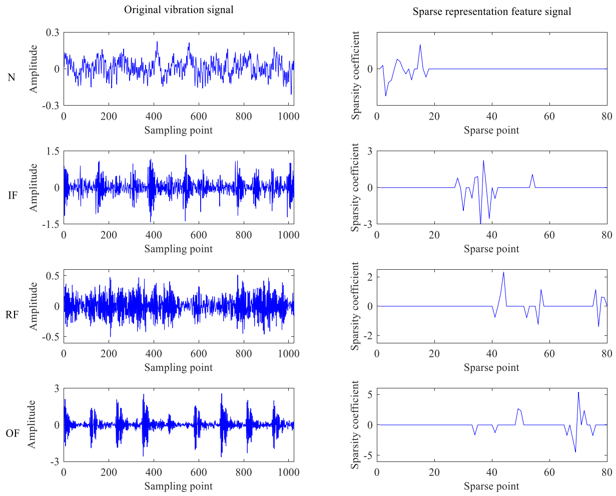


FIGURE 10. The comparison between the proposed method and the original signal classification results using DBN.

It can be seen that the sparse representation feature signal obtained by double sparse dictionary decomposition has many advantages over the original signal. Firstly, compared with the original vibration signal, the dimension of the sparse representation feature signal is reduced from 1024 to 80, and the signal dimension is compressed by more than ten times. Secondly, the original signal has a numerical value at each sampling point, but only 10 sparse points in the sparse representation feature signal have sparse values, and the rest positions are all zero. Because the dimension of sparse representation features is greatly reduced and most of the values of sparse representation features are zero, the fault diagnosis model is greatly simplified and the training parameters are much less, which greatly reduces the calculation of subsequent neural network training and testing, and improves the efficiency of network training and diagnosis. Finally, it can be seen from the figure that the sparse feature signal of each type of fault signal after sparse decomposition has more unique and obvious characteristics than the disordered form of the original signal, and its sparse point distribution is obviously different from other types of signals. In sparse decomposition, each type of signal is only decomposed by the vast majority of atoms in the dictionary trained by the same type of signals. It can be seen from Fig. 9 that most of the normal signal sparse points are distributed between 0 and 20, and most of the inner ring fault signal sparse points are distributed between 20 and 40, and so on. The other signals have the same distribution characteristics. These characteristics of the sparse decomposition signal under the double sparse dictionary model are of great benefit to the training and recognition efficiency of DBN. The results of fault diagnosis using the proposed method and directly using the same length of original signal combined with DBN are shown in Fig. 11. The average diagnosis results, corresponding standard deviations, training time of 15 experiments are shown in Table 3.

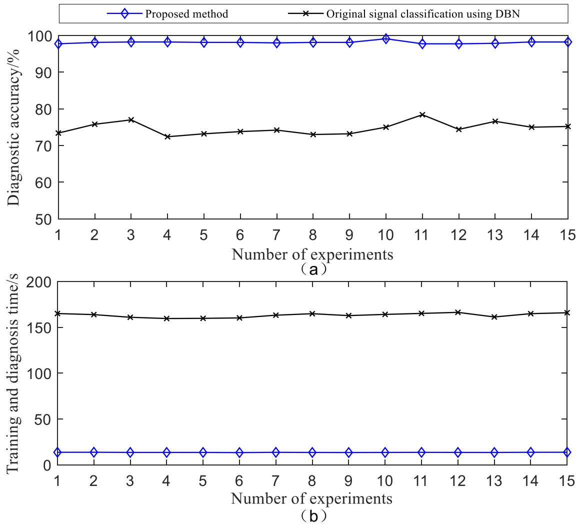


FIGURE 11. The comparison between the proposed method and the original signal classification results using DBN. (a) Diagnostic accuracy comparison; (b) DBN training time comparison.

TABLE 3. The result of proposed method and DBN classification using the original signal directly.

Test subject	Average diagnostic accuracy/%	Standard deviation of accuracy/%	Average training time/s
The proposed method	98.12%	0.33%	13.52
Original signal classification	84.38%	1.68%	163.18

It can be seen from Fig. 11 and Table 3 that the diagnosis accuracy of the method proposed in this paper is higher than that of using time domain vibration signal directly, and the accuracy of fault diagnosis is relatively stable, and the network training time is greatly reduced while the diagnosis accuracy is ensured. However, when DBN is used to classify the original time-domain vibration signals directly, the diagnosis results take longer and fluctuate greatly due to the high dimension of the original data. It can be seen that the double sparse dictionary model proposed in this paper simplifies the feature complexity between different types of signals, greatly improves the diagnostic accuracy and reduces the network training and testing time by sparse decomposition of the original signal.

E. FURTHER EXPERIMENTAL VERIFICATION

In order to further verify the effectiveness of the method proposed in this paper, the bearing experimental data of the University of Cincinnati [42] are tested for fault classification and identification. The data acquisition test rig of IMS is shown in Fig. 12. Four bearings were installed on a shaft. The rotation speed was kept constant at 2000 RPM by an AC motor coupled to the shaft via rub belts. A radial load of 6000 lbs is applied onto the shaft and bearing by a spring mechanism. All bearings are force lubricated. Rexnord ZA-2115 double row bearings were installed on the shaft. PCB 353B33 High Sensitivity Quartz ICP accelerometers were installed on the bearing housing. Data collection was facilitated by NI DAQ Card 6062E. The sampling frequency is 20 kHz.

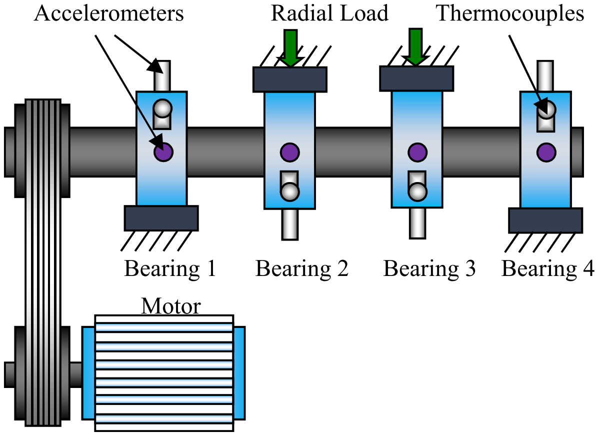
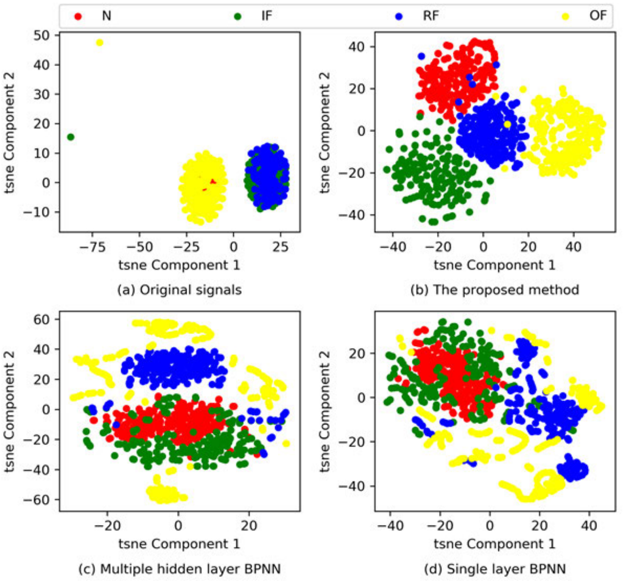


FIGURE 12. The data acquisition test rig of IMS.

In the experiment, the vibration signals from the acquisition channels which produce four types of fault conditions in the process of bearing operation are used as the experimental objects. The early fault signals are intercepted, and the sample length of each signal is 1024. Each type of fault data set contains 700 samples, 500 of which are used for training model and 200 for testing.

After different methods are used to learn features, t-SNE is also used to obtain the two-dimensional projection of feature effect as shown in Fig. 13. Obviously, the proposed method is still superior to other methods in feature learning.

The confusion matrices of fault classification effect of different methods is drawn, as shown in Fig.14. We can see from Fig.14 that the classification performance of the proposed method is still better than the other three methods. The diagnostic results of 15 experiments with four different methods are shown in Fig. 15. It can be seen that the results of 15 experiments of the method proposed in this paper have high diagnostic accuracy, and the diagnostic result curve is approximately a straight line, which shows that the method has good accuracy stability and robustness. The average diagnostic accuracy and the corresponding standard deviation of 15 experiments of the four methods are shown in Table 4. It can be seen that the average diagnostic accuracy of the proposed method is the highest, and the average diagnostic accuracy can reach 95.63%. The method of single-layer BPNN is the lowest, and the average diagnostic accuracy is only 44.38%. From the aspect of diagnosis stability, the method proposed in this paper has the highest stability, and its standard deviation is1.11%. The stability of deep BPNN is the worst, and the standard deviation of diagnosis accuracy is as high as 5.77%. The experimental results further verify the proposed method has better fault feature learning ability and fault diagnosis effect.



(a) Original signals

(b) The proposed method

(c) Multiple hidden layer BPNN

(d) Single layer BPNN

FIGURE 13. Two dimensional projection of feature learning effect of different methods.

TABLE 4. Mean and standard deviation of diagnostic accuracy.

experimental method	Average diagnostic accuracy/%	Standard deviation of diagnostic accuracy/%
The proposed method	95.63	1.11
Statistical features combined with SVM	88.38	2.02
Multiple hidden layer BPNN	69.21	5.77
Single layer BPNN	44.38	3.91

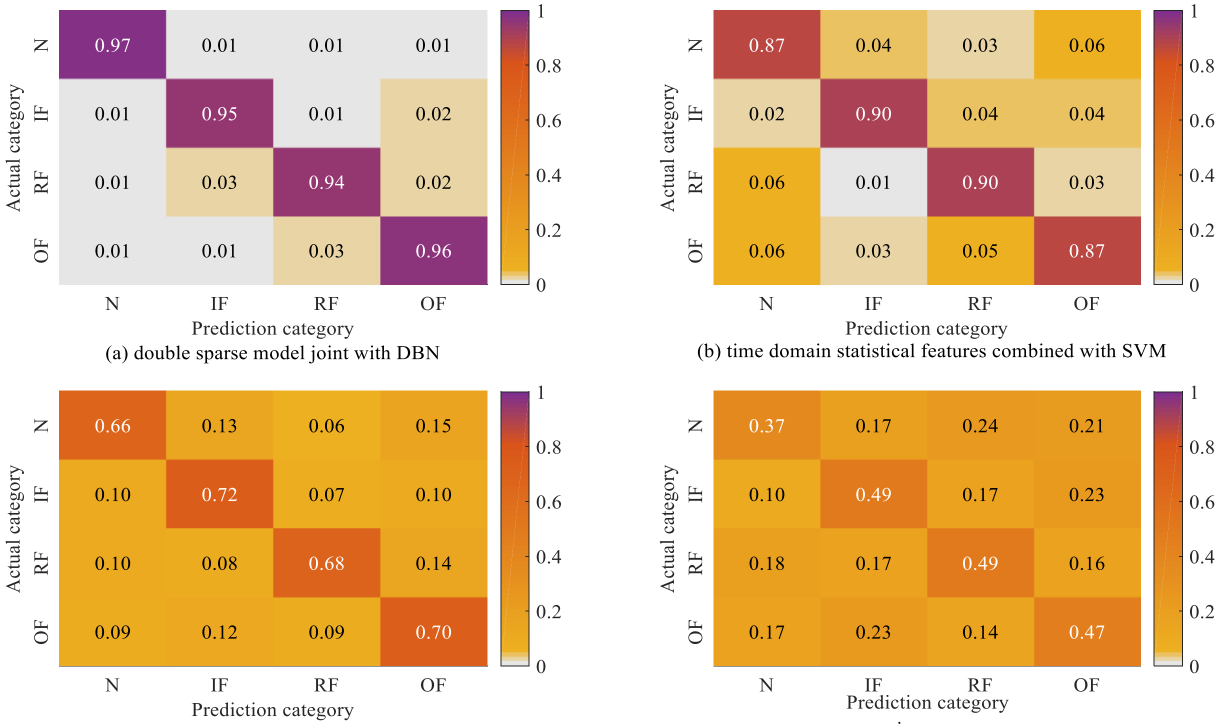


FIGURE 14. The confusion matrix between the proposed method and the common intelligent fault diagnosis method. (a) Classification effect based on double sparse model joint with DBN; (b) Classification effect of time domain statistical features combined with SVM; (c) Classification effect of deep BPNN; (d) Classification effect of single BPNN.

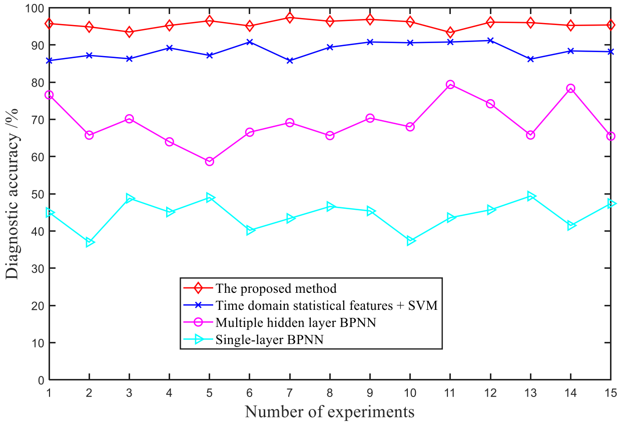


FIGURE 15. Diagnostic accuracy analysis curve of the proposed method and common fault diagnosis method.

V. CONCLUSION

In order to solve the problems of intelligent fault diagnosis, such as the traditional feature extraction process depends on prior knowledge and expert diagnosis experience, the low accuracy of fault diagnosis and the high time-consuming of diagnosis process when large data is processed, a novel fault diagnosis method based on double sparse dictionary learning joint with DBN is proposed in this paper. Firstly, the double sparse sub-dictionaries of all types of fault signals are obtained based on double sparse dictionary learning model. Then, the atoms of the double sparse sub-dictionary are simplified by eliminating the atoms with low usage frequency, and a more concise double sparse comprehensive dictionary is got, which makes the dimensions of sparse representation features are compressed as much as possible. Finally, the fault diagnosis is carried out by using the simplified sparse representation of feature signals combined with DBN. The experimental results show that the sparse decomposition eigenvectors, obtained from the sparse decomposition of the original signal under the double sparse comprehensive dictionary, can be used in fault diagnosis of rolling bearing fault signals. Compared with the traditional intelligent fault diagnosis methods, the method proposed in this paper has ideal and stable diagnosis accuracy, and greatly reduces the training time of DBN. In the future, we will use this method to further study the identification of different fault causes of the same fault type, and compare it to different deep models.

REFERENCES

[1] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: A review," *Measurement*, vol. 48, pp. 292–305, Feb. 2014.

[2] L. Niu, H. Cao, and X. Xiong, "Dynamic modeling and vibration response simulations of angular contact ball bearings with ball defects considering the three-dimensional motion of balls," *Tribol. Int.*, vol. 109, pp. 26–39, May 2017.

[3] L. Cui, Y. Zhang, F. Zhang, J. Zhang, and S. Lee, "Vibration response mechanism of faulty outer race rolling element bearings for quantitative analysis," *J. Sound Vibrat.*, vol. 364, no. 3, pp. 67–76, Mar. 2016.

[4] M. Yakout, M. G. A. Nassef, and S. Backar, "Effect of clearances in rolling element bearings on their dynamic performance, quality and operating life," *J. Mech. Sci. Technol.*, vol. 33, no. 5, pp. 2037–2042, May 2019.

[5] Z. Zhao, X. Yin, and W. Wang, "Effect of the raceway defects on the nonlinear dynamic behavior of rolling bearing," *J. Mech. Sci. Technol.*, vol. 33, no. 6, pp. 2511–2525, Jun. 2019.

[6] S. Yin, G. Wang, and H. R. Karimi, "Data-driven design of robust fault detection system for wind turbines," *Mechatronics*, vol. 24, no. 4, pp. 298–306, Jun. 2014.

[7] X. Liang, M. J. Zuo, and Z. Feng, "Dynamic modeling of gearbox faults: A review," *Mech. Syst. Signal Process.*, vol. 98, pp. 852–876, Jan. 2018.

[8] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.

[9] Y. Lei, F. Jia, D. Kong, J. Lin, and S. Xing, "Opportunities and challenges of machinery intelligent fault diagnosis in big data era," *J. Mech. Eng.*, vol. 54, no. 5, pp. 94–104, Mar. 2018.

[10] X. Yu, F. Dong, E. Ding, S. Wu, and C. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2018.

[11] X. Wang, Y. Zi, and Z. He, "Multiwavelet denoising with improved neighboring coefficients for application on rolling bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 25, no. 1, pp. 285–304, Jan. 2011.

[12] B. Samanta and C. Nataraj, "Use of particle swarm optimization for machinery fault detection," *Eng. Appl. Artif. Intell.*, vol. 22, no. 2, pp. 308–316, Mar. 2009.

[13] P. Liang, C. Deng, J. Wu, Z. Yang, and J. Zhu, "Intelligent fault diagnosis of rolling element bearing based on convolutional neural network and frequency spectrograms," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–5.

[14] J. Cai and Y. Xiao, "Bearing fault diagnosis method based on the generalized s transform time–frequency spectrum de-noised by singular value decomposition," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 233, no. 7, pp. 2467–2477, Apr. 2019.

[15] Y. Wang, Y. Si, B. Huang, and Z. Lou, "Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008-2017," *Can. J. Chem. Eng.*, vol. 96, no. 10, pp. 2073–2085, Oct. 2018.

[16] Y. Si, Y. Wang, and D. Zhou, "Key-performance-indicator-related process monitoring based on improved kernel partial least squares," *IEEE Trans. Ind. Electron.*, early access, Feb. 13, 2020, doi: 10.1109/TIE.2020.2972472.

[17] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[18] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural network," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.

[19] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019.

[20] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.

[21] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, Oct. 2017.

[22] 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," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[23] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.

[24] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vibrat.*, vol. 377, pp. 331–345, Sep. 2016.

[25] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017.

[26] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[27] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[28] Q. Lian, B. Shi, and S. Chen, "Research progress on dictionary learning models, algorithms and applications," *Acta Auto. Sin.*, vol. 41, no. 2, pp. 240–260, Feb. 2015.

[29] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[30] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[31] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[32] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[33] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[34] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 1993, pp. 40–44.

[35] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.

[36] B. Olshausen and D. Field, "Natural image statistics and efficient coding," *Netw., Comput. Neural Syst.*, vol. 7, no. 2, pp. 333–339, May 1996.

[37] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1," *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, Dec. 1997.

[38] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[39] C. A. Johnson, J. Seidel, and A. Sofer, "Interior-point methodology for 3-D PET reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, no. 4, pp. 271–285, Apr. 2000.

[40] J. Guo, B. Shi, C. Lei, X. Wei, and H. Li, "Compressive sensing method for mechanical vibration signals based on double sparse dictionary model," *J. Mech. Eng.*, vol. 54, no. 6, pp. 118–127, Mar. 2018.

[41] CWRU. (2015). *The Case Western Reserve University Bearing Data Center*. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file

[42] *IMS Bearings Dataset*. (2014). [Online]. Available: http://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository

JUNFENG GUO received the M.D. degree from the Lanzhou University of Technology, in 2005, and the Ph.D. degree from Northwestern Polytechnical University, China, in 2008. From 2018 to 2019, he was a Visiting Scholar with the University of Huddersfield, U.K. He is currently an Associate Professor with the School of Mechanical and Electronic Engineering, Lanzhou University of Technology. His research interests include modern testing and advanced control technology, intelligent robot, intelligent complete equipment, and so on.

PENGFEI ZHENG received the B.E. degree in mechanical design manufacture and automation from the Lanzhou University of Technology, Lanzhou, China, in 2016, where he is currently pursuing the M.E. degree in mechanical engineering. His research interests include intelligent fault diagnosis and condition monitoring technology.