IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE FOR PROGNOSTICS AND HEALTH MANAGEMENT (PHM) USING DISPARATE DATA STREAMS

Received October 21, 2019, accepted November 4, 2019, date of publication November 14, 2019,
date of current version November 26, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2953490*

# Fault Diagnosis of Rotating Machinery Based on Combination of Deep Belief Network and One-dimensional Convolutional Neural Network

**YIBING LI[1,2], LI ZOU[1], LI JIANG[1,2], AND XIANGYU ZHOU[1]**
[1]School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China
[2]Hubei Key Laboratory of Digital Manufacturing, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Li Jiang (ijiang@whut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51705384 and Grant 51875430, in part by the Hubei Province Natural Science Foundation of China under Grant 2019CFB565, and in part by the Fundamental Research Funds for the Central Universities under Grant WUT: 2018IVA022.

**ABSTRACT** The traditional intelligent diagnosis methods of rotating machinery generally require feature extraction of the raw signals in advance. However, it is a very time-consuming and laborious process for extracting the sensitive feature information to improve classification performance. Deep learning method, as a novel machine learning approach, can simultaneously achieve feature extraction and pattern classification. With the characteristics of Deep Belief Network (DBN) and one-dimensional Convolutional Neural Network (1D-CNN) (e.g. learning complex nonlinear, sparse connection and weight sharing), a precise diagnosis method based on the combination of DBN and 1D-CNN is proposed. Firstly, the DBN composed of three pre-trained restricted Boltzmann machines (RBMs) is constructed to achieve feature extraction and dimensionality reduction of the high-dimensional raw data. Secondly, the low-dimensional features extracted by DBN are fed into 1D-CNN for further extracting the abstract features. Finally, Soft-max classifier is employed to identify different operating conditions of rotating machinery. The superiority of the proposed method is validated by comparison with several state-of-the art fault diagnosis methods on two experimental cases. Meanwhile, the proposed method is tested in different background noises and on the imbalanced datasets. The results show that it has higher efficiency and accuracy than the state-of-the art fault diagnosis methods.

**INDEX TERMS** Deep belief network (DBN), one-dimensional convolutional neural network (1D-CNN), rotating machinery, feature extraction, intelligent fault diagnosis.

## I. INTRODUCTION

With the rapid development of science and technology, rotating machinery in modern industry has been moving toward high speed, super precision and high efficiency [1], [2]. After a long-term operating in the complex working environment, the core components of rotating machinery, including gears and bearings, are prone to cause various unperceivable faults. If not detected and managed, these failures may affect the operation of the whole rotating machinery and cause huge economic losses to enterprises [3]–[5]. Therefore, It's urgent for us to develop some advanced diagnosis methods, which can accurately and efficiently detect the potential faults of the key components of rotating machinery [6], [7].

At present, there are many methods used in fault diagnosis of rotating machinery, including oil debris analysis, electrical signature analysis, acoustic emission detection, vibration signal analysis, temperature analysis and so on [8]. In contrast with the other approaches, the vibration signal analysis is more common, and the relevant researches are more mature [9], [10]. Additionally, the vibration signals of rotating machinery usually carry more valuable information.

A complete fault diagnosis method based on pattern recognition consists of three steps: signal preprocessing [11], feature extraction [12] and pattern classification [13]. Each step has a critical impact on the final recognition accuracies of the model [14]. In order to extract more detailed fault

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

information, the time domain, frequency domain and time-frequency domain [8] signal analysis methods are generally employed, including fast Fourier transform [15], wavelet transform [16], empirical mode decomposition [17] and so on. But to some extent, these methods may generate the redundant information. In order to solve the problem, Principal Component Analysis(PCA) and kernel PCA(KPCA) were widely used for feature dimension reduction which can eliminate the redundant information. The intelligent diagnosis methods are newly developed fault diagnosis technologies, among which Artificial Neural Network (ANN) and Support Vector Machine (SVM) are widely used for pattern classification [18], [19]. Although the traditional diagnosis methods can be clearly seen to have made great progress in the field of the fault diagnosis [3], they have two drawbacks:

(1) These methods generally combined with the feature extraction process, and their performances depends on the extracted features, while the selection of the appropriate features relies on professional knowledge and advanced experiences [20].

(2) The shallow structures of these methods based on ANN or SVM are difficult to effectively learn the deep fault information of complex non-linear signals [3].

Deep learning method, as a new field of machine learning, can effectively solve the drawbacks of the traditional diagnosis methods by relying on its deep structure [21]. Deep learning model contains multiple hidden layers, which are used for extracting the deep features of the complex signals. The extracted features are not designed by human engineers, but obtained by self-learning the deep features of input data [22]. Sparse Auto-Encoder (SAE), Long Short-Term Memory (LSTM), Deep Belief Network (DBN) and Convolutional Neural Network (CNN) are effective deep learning methods, and are widely used in fault diagnosis in recent years [23]. Wen *et al.* [20] designed a deep transfer learning based on SAE for fault diagnosis. Zhang *et al.* [24] proposed a subset based on deep auto-encoder for fault diagnosis. Lei *et al.* [25] proposed a method based on LSTM for fault diagnosis of wind turbine. Appiah *et al.* [26] developed a LSTM based automatic feature extraction. Qin *et al.* [27] proposed an improved DBN for fault diagnosis. Tao *et al.* [28] developed a method based on adaptive DBN and time-frequency characteristics of travelling wave. Xie *et al.* [29] proposed an end-to-end model based on improved DBN for fault diagnosis. Wen *et al.* [30] designed a new hierarchical convolutional neural network (HCNN) as the two-level hierarchical diagnosis network. Chen *et al.* [23] developed a fault diagnosis method based on CNN and discrete wavelet transform. Li *et al.* [31] developed a network based on an ensemble CNN and deep neural network (CNNEPDNN). Gong *et al.* [19] designed a modified 2D-CNN for fault diagnosis. Among them, DBN and CNN have attracted wide attention recently. The main attribution is that DBN consists of multiple RBMs and is trained by greedy learning layer by layer, which makes it more possible to learn complex nonlinear characteristics [21]. CNN has the characteristics of weight sharing and sparse connection, so that fewer parameters need to be optimized in the training process [23]. Because of their powerful feature learning ability, these deep learning models can self-learn the deep features of the original signals. In addition, CNN has achieved very successful performance in the field of fault diagnosis, but most researches still use two-dimensional CNN. It is not suitable for learning the spatial characteristics of one-dimensional time series signals [9].

In recent years, one-dimensional CNN(1D-CNN) has been successfully used in the field of fault diagnosis. Wu *et al.* [9] optimized two-dimensional CNN to be 1D-CNN for planetary gearbox fault diagnosis. Huang *et al.* [32] designed a signal status recognition based on 1D-CNN and its feature extraction mechanism analysis. However, high-dimensional signals will lead to excessive network parameters, and then increase the network training time. Fortunately, depending on the number of neurons in each layer, DBN can achieve feature dimensionality reduction of input data [33]. It can greatly simplify the complexity and calculation cost of signal processing.

Thereby, a fault diagnosis method of rotating machinery is developed based on combination of DBN and 1D-CNN. Firstly, a DBN composed of three pre-trained Restricted Boltzmann Machines (RBMs) is constructed to achieve feature extraction and dimensionality reduction of the raw high-dimensional data. Secondly, the low-dimensional features extracted by DBN are fed into 1D-CNN for further extracting the abstract features. Finally, Soft-max classifier is employed to achieve classification of different health conditions.

The paper is organized as follows. Section II described the DBN construction, the structure of traditional CNN, 1D-CNN and the proposed method. In section III, the experimental case for gearing is analyzed, and another case for bearing of the case western reserve university (CWRU) is also used for experimental analysis. Section IV draws the conclusions about the proposed method.

## II. BASIC PRINCIPLE OF THE PROPOSED METHOD

### A. RESTRICTED BOLTZMANN MACHINE TRAINING AND DEEP BELIEF NETWORK CONSTRUCTION

A standard DBN is stacked with many RBMs which is a special form of the Markov random field [29]. The detailed structure of the RBM is showed in Figure 1. Each RBM consists of one visible layer, which is described as $\mathrm{v} = \{v_1, v_2, \ldots, v_i, \ldots, v_n\}^T (v_i \in \{0, 1\})$. One hidden layer is described as $h = \{h_1, h_2, \ldots, h_i, \ldots, h_n\}^T (h_i \in \{0, 1\})$.

The input data is fed into the visible layer which can transform the input data to the hidden layer. The visible layer is associated with the hidden layer through weight connection, and neurons of each layer are not connected to each other.

The energy function of the RBM can be defined as [34]:

$$E(\mathrm{v}, \mathrm{h}|\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} v_i w_{ij} h_j - \sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j \quad (1)$$

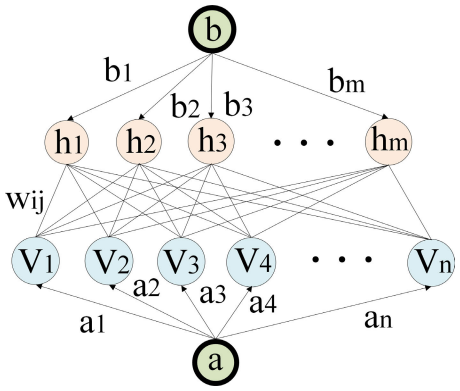



**FIGURE 1.** RBM structure.

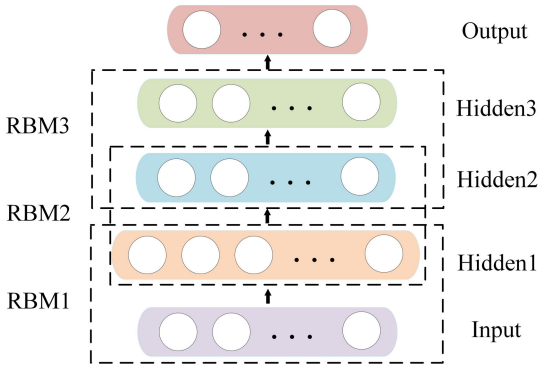



**FIGURE 2.** The construction process of a DBN with three RBMs.

where $\theta = \{w,a,b\}$, $n$ and $m$ indicate the number of the visible neural and the hidden neural, respectively. $i$ and $j$ indicate the $i^{th}$ and $j^{th}$ the neural, $v_i$ and $h_j$ represent the $i^{th}$ visible neural and the $j^{th}$ hidden neural, $a_i$ and $b_j$ is the bias of the $i^{th}$ visible neural and the $j^{th}$ hidden neural, and $w_{ij}$ represents the weight between the $i^{th}$ visible neural and the $j^{th}$ hidden neural.

The joint probability of the visible neurons and the hidden neurons is given by [35]:

$$P(\mathrm{v}, \mathrm{h}|\theta) = \frac{1}{Z(\theta)} \exp(-\mathrm{E}(\mathrm{v}, \mathrm{h}|\theta)) \tag{2}$$

which is the Gibbs distribution of the RBM network [29]. $Z(\theta)$ is the partition function and defined as:

$$Z(\theta) = \sum_{v} \sum_{h} \exp(-\mathrm{E}(\mathrm{v},\mathrm{h}|\theta)) \tag{3}$$

The associated two edge probabilities of the visible neural and the hidden neural are defined as:

$$P(\mathrm{v}) = \frac{1}{Z(\theta)} \sum_{\mathrm{h}} \exp(-E(\mathrm{v}, \mathrm{h}|\theta)) \tag{4}$$

$$P(\mathrm{h}) = \frac{1}{Z(\theta)} \sum_{\mathrm{v}} \exp(-E(\mathrm{v}, \mathrm{h}|\theta)) \tag{5}$$

The conditional probabilities of the visible neural and the hidden neural are given by [21]:

$$P(\mathrm{v}|\mathrm{h}) = \prod_{i} P(v_i|h) \tag{6}$$

$$P(\mathrm{h}|\mathrm{v}) = \prod_{j} P(h_j|\mathrm{v}) \tag{7}$$

The visible neural and hidden neural are independent, so the distributions of the condition probabilities are defined as:

$$P(v_i = 1|\mathrm{h}) = \frac{1}{1 + \exp(-a_i - \sum_{j=1}^{m} w_{ij} h_j)} \tag{8}$$

$$P(h_j = 1|\mathrm{v}) = \frac{1}{1 + \exp(-b_j - \sum_{i=1}^{n} w_{ij} v_i)} \tag{9}$$

The visible layer $v_i$ represents the input data, which is mapped to the hidden layer according to the probabilities in Equation (9). Consequently, it constitutes the first RBM. Simultaneously, it is the input data of the second RBM. Repeating this process to update the parameters, so as to form a feature representation that is more abstract and has more representability than the lower layer. This is DBN's greedy learning strategy of layer by layer. The weight can be updated as [32]:

$$\Delta w_{ij} = \eta(< v_i h_j > - < v_i' h_j' >) \tag{10}$$

where $\eta \in (0, 1)$ refers to the learning rate, $< \cdot >$ [36] indicates the mean over the training data.

The DBN is stacked by the three RBMs. The detailed structure of the DBN is shown in Figure 2. The training process of the DBN can be divided into two steps:

(1) The forward learning process of the stacking RBMs. The input data of each RBM is the output of the lower layer RBM. The learning process belongs to unsupervised learning. This step takes the acquired parameters as the initial values of the next step. In other words, it can provide prior knowledge of the input data for the supervised learning process.

(2) The backward fine-tuning learning process of DBN. The parameters of the model are fine-tuned globally. For multi-classification problems, the Soft-max is generally selected as the classifier at the top layer of the network based on the learned deep feature [37].

### B. STRUCTURE OF TRADITIONAL CONVOLUTION NEURAL NETWORK

CNN is a typical feed-forward neural network, which essentially extracts the features of input data by establishing multiple filters. The process of convolution and pooling can extract the topological structure features contained in the input data. The structure of CNN composes of input layer, hidden layer (convolution layer and pooling layer), and fully connected layer. The typical CNN structure is shown in Figure 3. The hidden layer composes of two alternating convolution layers and pooling layers. In the process of the CNN feature extraction, the features of input data are extracted by alternating convolution layers and pooling layers. The fully connected

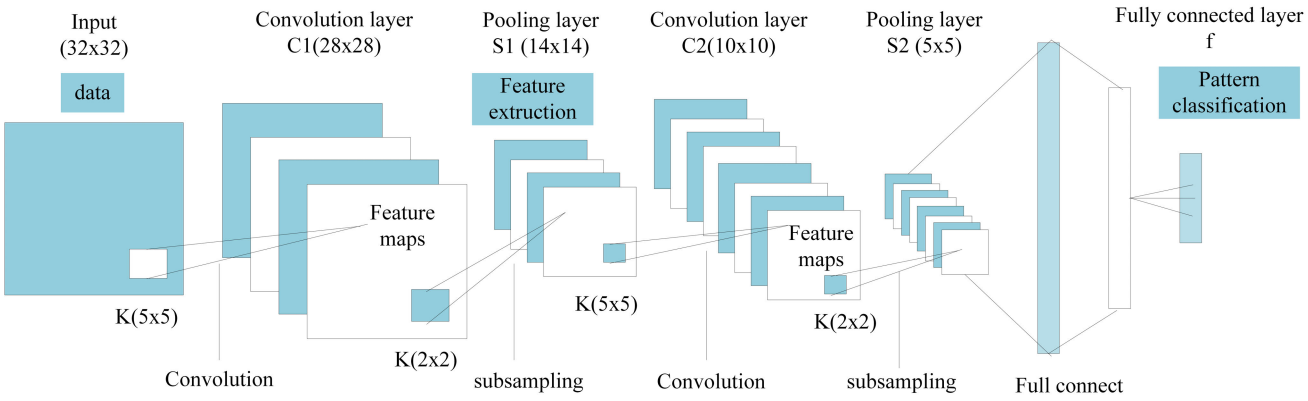



**FIGURE 3.** Typical convolution neural network structure.

layer is adopted to approach the final result. In the convolution layer, the process of convolution operation is taken to the output feature vectors of the previous layer, and then the nonlinear activation function is used to construct the output feature vectors. The output of convolution layer is the convolution result of multiple input features. The fully connected layer of CNN includes the end-to-end connection of the output of the last pooling layer, and then the Soft-max classifier is employed to achieve the multi-classification problem.

The CNN has the characteristics of sparse connection and weight sharing. Among them, sparse connections use spatial topology to establish the spatial relation of incomplete connections between adjacent layers, which can reduce the number of parameters needed to be trained in the model. Weight sharing is used to avoid overfitting. With the number of network layer increasing, the extracted features are constantly abstracted. Finally, the robust features with translation and rotation invariance are obtained from the original input data. In addition, the pooling operation takes advantage of the local characteristics of the data itself, reduces the data dimensions, optimizes the network structure, and improves the robustness of the features.

### C. ONE DIMENSIONAL CONVOLUTION NEURAL NETWORK MODEL

The traditional CNN is mainly used for two-dimensional image recognition. As a one-dimensional time series signal, the vibration signal is correlated with the data points at each time. If it is directly converted into a two-dimensional form, the spatial correlation in the original sequence will be destroyed, and the fault related information may be missing. One-dimensional CNN well deals with one-dimensional signals. The input data of 1D-CNN is one-dimensional, so its convolution kernel adopts one-dimensional structure correspondingly, and the output of each convolution layer and pooling layer is one-dimensional feature vector. Figure 4 shows the optimized 1D-CNN network structure.

The optimized 1D-CNN is composed of input layer, hidden layer (convolution layer and pooling layer) and fully

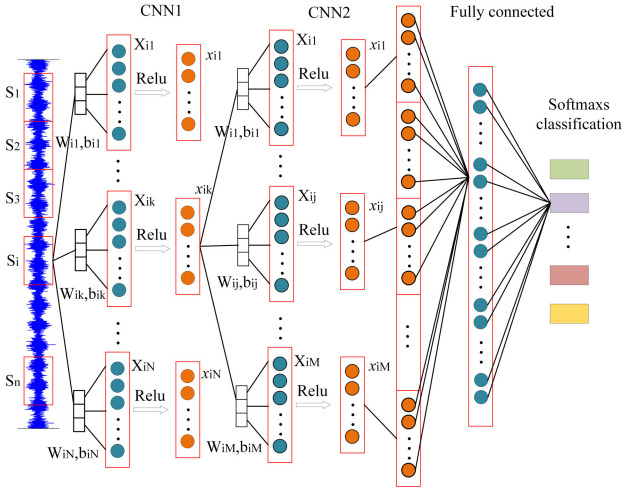



**FIGURE 4.** Optimized 1D-CNN network structure.

connected layer. In the input layer, the original vibration signals are segmented according to the time step. The feature extraction layer consists of convolution layer and pooling layer. Vibration data from the input layer are received, and multiple convolution kernels in the convolution layer are used for extracting the features of the vibration signal. Thus, multiple feature vectors are obtained. The pooling operator can reduce the dimensionality of eigenvectors and improve the robustness of nonlinear features. The hierarchical extraction of nonlinear features of input signals is achieved by alternating convolution pooling layers. The classification layer consists of fully connected layers, in which the first fully connected layer achieves the 'flattening' operation of features. It means that all feature vectors are connected from head to tail to form one-dimensional vectors. The number of neurons in the fully connected layer is equal to that of the health conditions. Soft-max regression classifier is used to achieve the target output category.

The output of CNN includes the output of the full connection and the last pooling layer, and then Soft-max classifier is adopted to achieve fault classification. The model can be

expressed as [38]:

$$O = \begin{bmatrix} P(y=1|x; W_1, b_1) \\ P(y=2|x; W_2, b_2) \\ \cdots \\ P(y=3|x; W_3, b_3) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} \exp\left(W_j x + b_j\right)} \begin{bmatrix} \exp\left(W_1 x + b_1\right) \\ \exp\left(W_2 x + b_2\right) \\ \cdots \\ \exp\left(W_k x + b_k\right) \end{bmatrix} \quad (11)$$

where $O$ is the output of the CNN model, $W$ and $b$ are the weight and bias of each layer neurons.

$[X_{i1}^{l_1}, X_{i2}^{l_2}, \ldots, X_{iN1}^{l_1}]$ can be acquired for the signal $S_i$ after the first layer of the optimized CNN, where:

$$X_{ik}^{l_1} = b_{ik}^{l_1} + conv1D(S_i, W_{ik}^{l_1}) \quad (12)$$

To increase the nonlinear features of the CNN, the activation function is used to required. The output of the first CNN is:

$$x_{ik}^{l_1} = (\text{Relu}(X_{ik}^{l_1}) - \overline{\text{Relu}(X_{ik}^{l_1})}) \downarrow Subconv \quad (13)$$

the pooling process is represented as $X_{ik}^{l_1}$ ($\downarrow Subconv$).

The $X_{ik}^{l_2}$ is the average output of the first convolution layer.

$$X_{ik}^{l_2} = \frac{1}{N_1} \sum_{j}^{N_1} f(x_{ij}^{l_1}) \quad (14)$$

The output of the second convolution layer is:

$$x_{ik}^{l_2} = \text{Relu}(X_{ik}^{l_2} - \overline{X_{ik}^{l_2}}) \downarrow Subconv \quad (15)$$

Soft-max regression model is an extension of the logistic regression model in multi-class classification problems. The identified label y has a vector form, indicating the distribution probability of the current sample category in all possible categories. The loss function of the Soft-max regression model can be expressed as:

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y_i = j\} \lg \frac{e^{\theta_i^T x_i}}{\sum_{i=1}^{k} e^{\theta_i^T x_i}} \right] \quad (16)$$

### D. GENERAL PROCEDURES OF THE PROPOSED METHOD

In recent years, DBN and CNN have been widely used in the field of fault diagnosis. Although the deep learning methods are widely used, they often need to be combined with the traditional feature extraction methods to preprocess the raw data in many paper. Thus, It's still unable to avoid the tedious calculation process, nor can it exert the strong feature learning ability of deep learning. It results in that they do not make full use of the strong automatic learning and adaptive ability of deep learning approaches. Therefore, an adaptive intelligent feature extraction and pattern recognition method is proposed, which can accurately recognize multiple health conditions. The detailed flowchart of the proposed method is shown in the Figure 5.

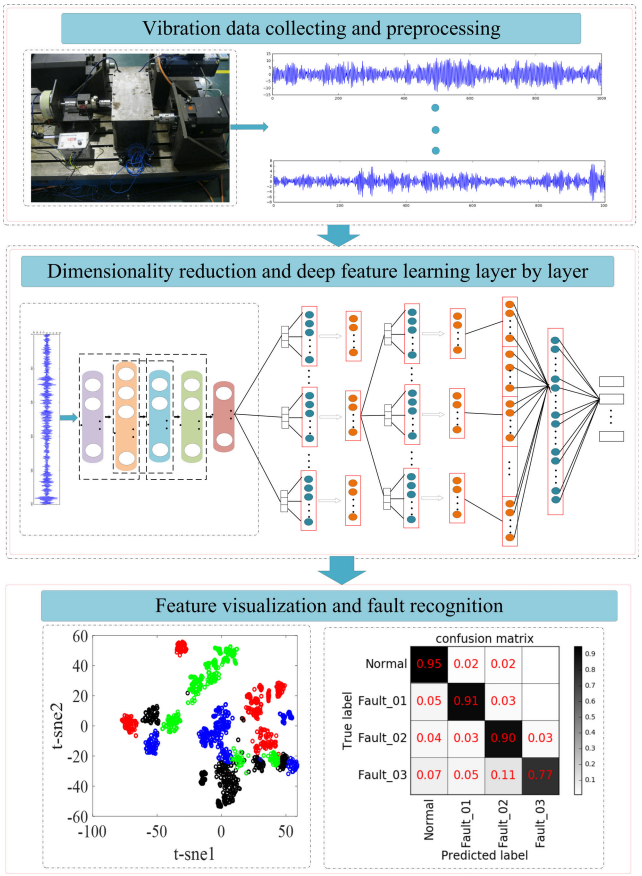
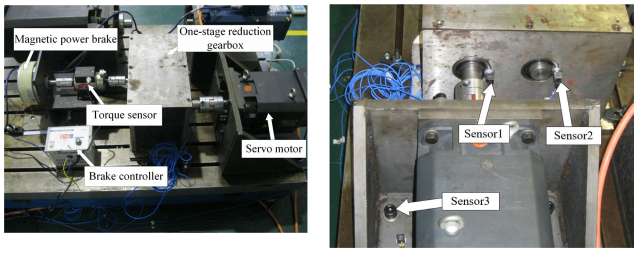



**FIGURE 5.** The flowchart of the proposed method.




(a) (b)

**FIGURE 6.** (a)One-stage reducer platform. (b) Installation of tri-axial accelerometers.

The detailed general process of the proposed method can be followed as:

(1) The vibration signals of the key parts of rotating machinery are acquired by acceleration sensors.

(2) Subsequently, the collected vibration signals are transformed to the data format required by the model (including the sample data and the corresponding labels).

(3) A DBN composed of three pre-trained RBMs is constructed to learn the abstract features of the samples, and then achieve the feature extraction and dimension reduction of the raw data. There is no need to manually extract any features.

(4) The low-dimensional features extracted by DBN are split into training samples and testing samples at a certain rate.

**TABLE 1.** Parameter of the gear set.

| Types | Teeth number | Module (mm) | Teeth width (mm) |
|---|---|---|---|
| Driving gear | 50 | 2 | 20 |
| Driven gear | 80 | 2 | 20 |


**FIGURE 7.** Different health conditions of Gears.

(5) The training samples are fed into 1D-CNN for further feature extraction, which is composed of two convolution layers.

(6) Soft-max classifier is employed to achieve the classification of different health conditions.

(7) The visualization results and the confusion matrix are employed to verify the performance of the proposed method.

## III. EXPERIMENTAL RESULTS

### A. CASE 1: FAULT DIAGNOSIS OF GEARS

#### 1) EXPERIMENTAL SETUP AND DATA ACQUISITION

In order to verify the performance of the proposed method, a series of gear fault diagnosis experiments on one-stage reduction platform were tested.

The platform is made up of five parts: a brake controller a one-stage reduction gearbox, a servo motor, a torque sensor, and a magnetic power brake, as shown in Figure 6(a). In Figure 6(b), tri-axial accelerometers (PCB-356A16) are fixed on the base of the test bed, the gear end of the driving wheel and the gear end of the driven wheel, respectively. They are used to collect vibration signals in the direction of X, Y and Z. Table 1 shows the detailed information of the driven wheel and the driving wheel fixed in the one-stage reduction gear box. In Figure 7, the driving gear, has different radial crack lengths, was used as the monitoring object, including four kinds of health conditions (normal, 1/4 crack, 1/2 crack, and 3/4 crack). The radial crack of gear is created by wire-electrode cutting. The cutting length is defined by the following formula: $l_i = i \times (R_{b1} - r_1)/4, i = 0, 1, 2, 3.$

**TABLE 2.** The information of the experimental platform.

| Types | Value |
|---|---|
| Crack length (mm) | 0,5,10,15 |
| Speed of the gear (rpm) | 300,600,900,1200,1500 |
| Loads applied to the gears (Nm) | 4, 8 |
| Number of segment samples for each working condition | 50 |
| Points of each segment sample | 1000 |

where $R_{b1}$ is the dedendum circle radius, and $r_1$ is the shaft hole radius of the driving gear. Their values are 27.5mm and 47.5mm, respectively. More detailed information about the experimental platform can be referred to [39], [40].

The gears with different crack lengths and working conditions are employed to verify the performance of the proposed method. Table 2 shows the information of the experimental platform.

From Table 2, we can know that the driving gear was set at five speeds. Then, two loads of 4Nm and 8Nm were applied to the driving gears at each rotation speed. Therefore, each crack severity corresponds to 10 operating working conditions. The data of each channel from tri-axial acceleration sensors was acquired by the data recorder (NI PXI-1042) in the experiment platform. The collected data consists of nine channels, and the X direction data of the second sensor is selected as the experimental data. The sampling frequency was set as 5000HZ. Fifties segment samples were acquired for each working condition, and each segment sample contained 1000 sampling points, as shown in Table 2. So that each health condition contained a total of 500 samples under 10 conditions. The program is performed on an i7-4700MQ processor at 2.40 GHz with 8 GB memory.

#### 2) RESULTS AND ANALYSIS

The vibration signals collected from the one-stage reduction gearbox experiment platform were transformed into high-dimensional fault sample dataset ($N \times D(2000\times1000)$, N indicates the sample size, D is the feature dimension). Then the high-dimensional features are fed into the DBN, which includes three RBMs. The architecture of DBN is 1000-800-700-500 when implementing feature extraction and dimension reduction. Thus, the output dimension of the last RBM is 500. Finally, after further feature extraction by 1D-CNN, the extracted feature vectors are input to the Soft-max classifier for fault identification.

To verify the influence of the training sample sizes on the test accuracies, 200 to 1800 samples were randomly selected from the dataset as the training samples for experiments following the Kennard and Stone algorithm [44, 45]. In order to avoid randomness of the test results, the average recognition accuracies and the average training time of the ten trials were used for analysis. Figure 8 shows the experimental results. The recognition accuracies of the training samples were nearly above 99%. The recognition rates of the testing samples were slightly lower than those of the training

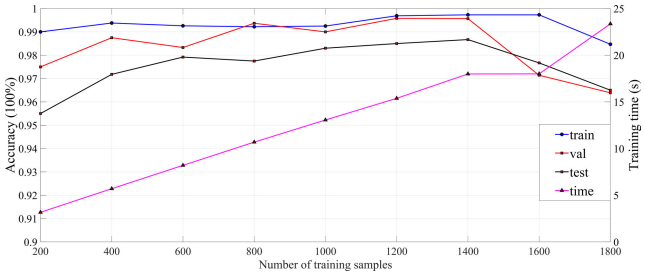



**FIGURE 8.** Classification results for different training samples.

**TABLE 3.** Description of the gearing datasets.

| Health conditions | Training samples/Testing samples | | |
|---|---|---|---|
| | Dataset A | Dataset B | Dataset C |
| Normal | 350/150 | 350/150 | 350/150 |
| Fault_1 | 350/150 | 250/250 | 150/350 |
| Fault_2 | 350/150 | 150/350 | 150/350 |
| Fault_3 | 350/150 | 150/350 | 50/450 |

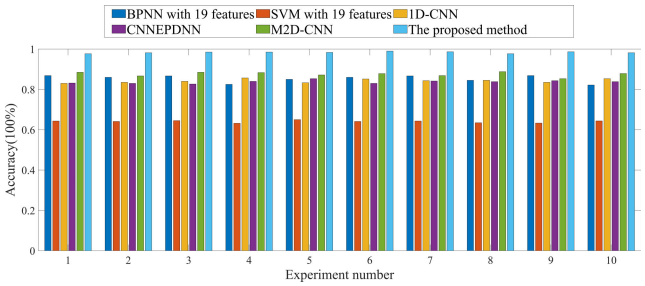



**FIGURE 9.** Ten experiment results of different method.

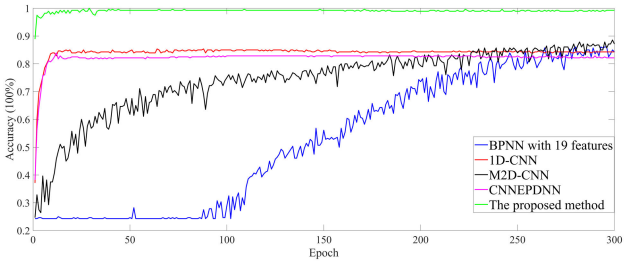



**FIGURE 10.** Trends of accuracy of different methods in each epoch.

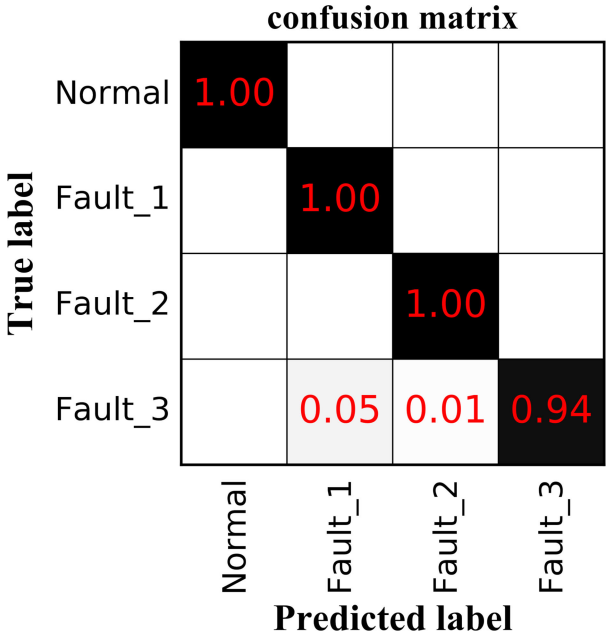



**FIGURE 11.** The result of the classification confusion matrix of the proposed method.

samples, however, they were nearly above 96%. Although the proposed method can ultimately achieve very high classification accuracy, the training sample sizes have some impacts on the recognition performance. When the training sample size is 1400, the proposed method acquires the highest recognition accuracy of the testing samples, and consumes very little training time.

In real working condition, the acquisition of the fault data is often more difficult than that of normal data. For analysis, three different datasets are constituted to verify the performance of the proposed method [43]. Dataset A has the same split ratio for each health condition, it's a balanced dataset. Dataset B and dataset C have different split ratios for each health condition, they are both imbalanced dataset, Table 3 shows the detailed information of the three datasets.

In the experiments, the proposed method is compare with the other five representative methods, including BPNN and SVM with 10 time-domain features and 9 frequency-domain features manually extracted from the raw signals [8], and 1D-CNN [9], CNNEPDNN [31], M2D-CNN [19] with the raw signals.

The 10 trials were used for analyzing the performance of the proposed method on dataset A. The ten detailed experiment results of each method are given in Figure 9 and Table 4. Figure 9 displays that the proposed method has the best recognition accuracies in comparison with the other five methods. From Table 4, we can know that the average recognition accuracy of the proposed method is 98.17%, which overpasses those of the BPNN method with 19 features (86.83%), and overpasses the two SVM approach with 19 features (64.33%), and 1D-CNN (85.33%), CNNEPDNN (83.17%) and M2D-CNN (88.5%). The standard deviation of the proposed method is 0.0039, which is less than those of the other five methods (0.0167, 0.0055, 0.0088, 0.0076 and 0.0102, respectively). Therefore, the proposed method has the characteristics of high accuracy and strong robustness.

Figure 10 illustrates the trends of the accuracy in each epoch through the training process. From Figure 10, the proposed method rapidly converges and builds a precise model, and the performance of the proposed method is the best in comparison with those of the other approaches. We also can find that the highest accuracy value happened when the epoch is nearly twenty. The confusion matrix of the proposed method for one of the ten trails is shown in Figure 11. From Figure 11, we also can know that it shows that the recognition accuracy of the Fault_3 is the lowest, but the whole recognition accuracy is still very high.

**TABLE 4. Detailed experiment results of different methods.**

| Methods | Average fault type recognition accuracy (%) | | | | |
|---|---|---|---|---|---|
| | Normal | Fault_1 | Fault _2 | Fault _3 | Average accuracy (%) ± standard deviation |
| BPNN with 19 features | 96 | 93 | 75 | 83 | 86.83 ± 1.67 |
| SVM with 19 features | 70 | 55 | 83 | 49 | 64.33 ± 0.55 |
| 1D-CNN[9] | 87 | 89 | 84 | 81 | 85.33 ± 0.88 |
| M2D-CNN[19] | 95.3 | 92 | 90 | 76.7 | 88.5 ± 1.02 |
| CNNEPDNN[31] | 87.3 | 81.3 | 83.3 | 80.7 | 83.17 ± 0.76 |
| The proposed method | 100 | 99 | 99 | 94 | 98.17 ± 0.39 |

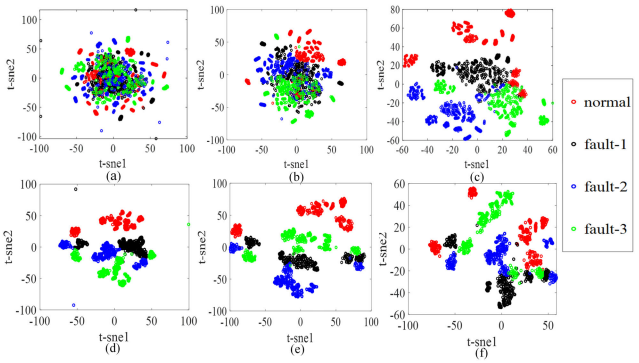



**FIGURE 12. Feature visualization via t-SNE. (a) Raw data features (b) Features in the 1st hidden layer of DBN (c) Features in the 2nd hidden layer of DBN (d) Features in the 3rd hidden layer of DBN (e) Features in the 1st hidden layer of 1D-CNN (f) Features in the 2nd hidden layer of 1D-CNN.**

To better understand the process of feature extraction, the t-SNE technique [44] was used to visualize the learning characteristics of each layer. From Figure 12, we can easily observe that the features with the same fault types are gradually become closer, and the features with the different fault types are more separable. It reflects that the developed method has powerful feature extraction ability. Thus, it makes the classifier easier to identify the different health conditions.

In real application, the gears often operate in complicated environments which brings about strong background noises in the acquired signals. To test the robustness of the designed method, the Gaussian white noises [45] are added to the raw vibration signals with different signal noise ratios (SNRs). The SNR is defined as follows:

$$SNR_{dB} = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) \tag{17}$$

where $P_{signal}$ represents the power of the signal and $P_{noise}$ represents the noise.

The SNRs ranged from −4 to 4dB. A total of 10 experiments were carried out to analyze the impacts of different SNRs. The detailed diagnosis results can be obtained from Figure 13. We can clearly observe that the designed method has stronger robustness than the other seven approaches. The average recognition accuracies of the designed method are over 94% within all considered SNR levels from -4 to 4dB,

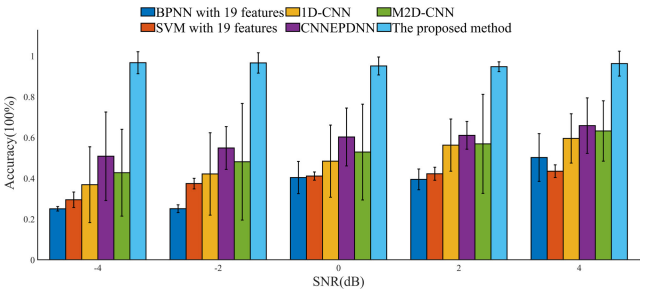



**FIGURE 13. Comparison of accuracy testing on signals with different SNR values.**

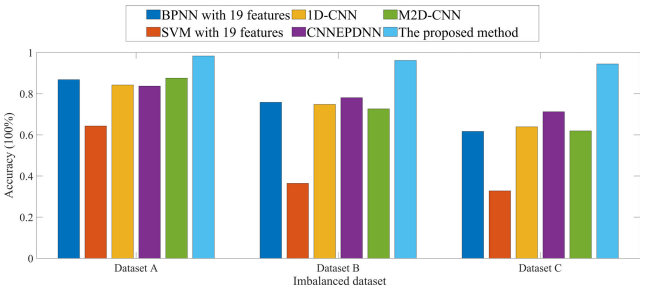



**FIGURE 14. Experiment results of the different datasets with six methods.**

which are much higher than those of the other methods. The length of the vertical line above each bar graph represents the value of the standard deviation. The standard deviations of the designed method are also smaller than those of the other five methods, and have the highest robustness.

Subsequently, the unbalanced datasets are used for further experiments. Figure 14 shows the classification accuracies of the testing samples for the three kinds of datasets using the above five methods. It displays that the designed method achieves the best accuracy (98.33% in dataset A, 96.18% in dataset B, 94.46% in dataset C), and the accuracies of the other methods are less than 87.58% in dataset A, 78.09% in dataset B and 71.23% in dataset C.

To show the detail diagnosis results, the confusion matrices of the proposed method for the two datasets are shown in Figure 15. It can be clearly observed that, the recognition accuracy of each health condition in dataset B is less than that of the dataset A (shown in Figure 11). Thus, there are more samples misclassified. However, the accuracies of each health condition in dataset B are above 93%. Dataset C (the seriously

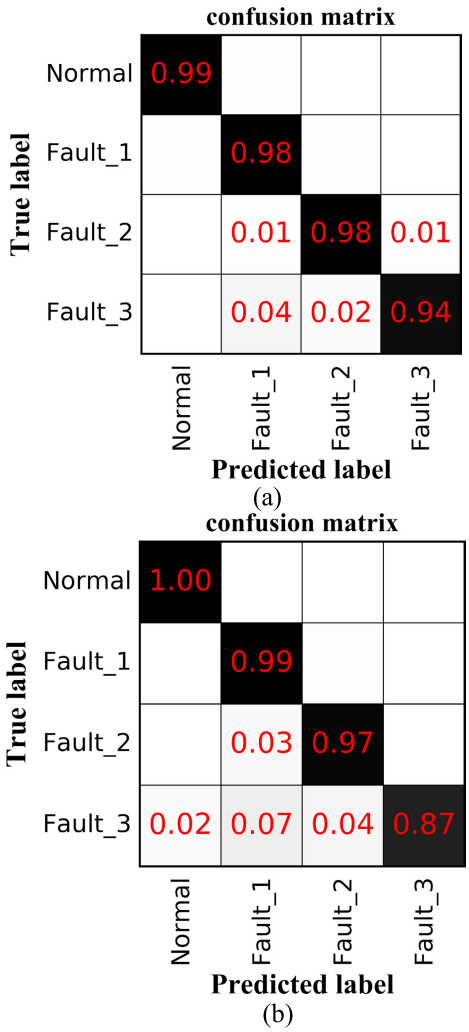



**FIGURE 15.** The result of the classification confusion matrix of the proposed method. (a) Dataset B, (b) Dataset C.

imbalanced dataset) has the fewest training samples of the fault_3 condition. Thus, the recognition accuracy of the fault_3 is the lowest (only 87%). The results indicate that the proposed method has good recognition accuracy for the imbalanced dataset.

### 3) PARAMETER DISCUSSION

In order to achieve a high recognition accuracy of the proposed method, the optimal structure parameters must be selected. Currently, some researchers have done some work on this problem. Such as Smith [46] developed linear learning rate test (LLR Test) to determine the learning rate. However, LLR Test has a low resolution ratio when the learning rate range is large. Wen *et al.* [47] investigated the learning rate estimation of deep learning, but ignored the imbalance of the dataset. The architecture of the deep learning model also can be optimized in the heuristic way, but it is easy to fall into local optimum and time-consuming for complex networks [3]. Thereby, we conform a single ideal just like Ref [48] to determining the 1D-CNN structures. We employ the different optimizers, learning rates, structure types and

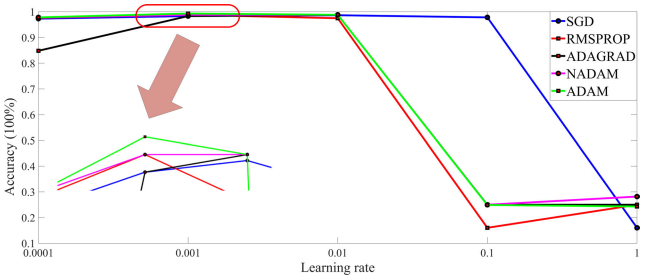



**FIGURE 16.** Classification accuracies for different optimizers and learning rate.

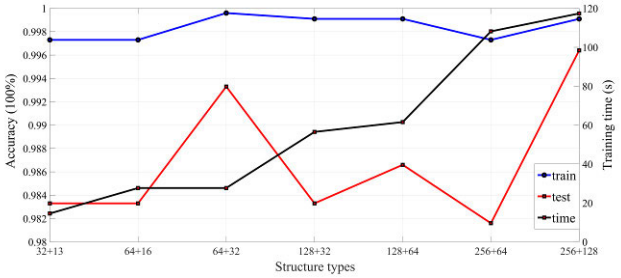



**FIGURE 17.** Classification accuracies and training time for different structure types.

batch sizes to analyze the impact on the recognition accuracy and training time of the proposed method.

Figure 16 shows the experiment results of different optimizers and learning rates. We can clearly observe that the change tendency of the average accuracy as we increase the value of learning rate (from 0.0001 to 1) using SGD, RMSPROP, ADAGRAD, NADAM and ADAM, respectively. The four optimizers excepting SGD can have similar classification results, and all the optimizers can show good classification performance when the learning rate ranges between 0.001 and 0.01. ADAM has the best classification effect as the learning rate is 0.001.

To investigate how the structure type impacts the classification accuracy of the designed method, we conform the rule that the number of convolution kernels of the latter layer is no more than half that of the former layer [48]. Figure 17 shows the detailed diagnosis results. The recognition accuracies of the training samples are keeping a very high values and so are the testing samples. Although the proposed method with the structure type (256,128) has the highest accuracy (99.64%), it consumes more computing time (117.374s). The proposed method with the structure type (64,32) also can achieve high accuracy (99.33%). Moreover, it takes less computing time (27.6916s).

In general, the deep learning model updates the parameters by dividing the training samples into multiple batches to achieve convergence. Therefore, the value of the batch size is important for the classification accuracy of the model. The detail diagnosis results are displayed in Figure 18. We can know that the best experiment result is when the batch size is 200, and just consumes slight more computing time. The parameter selection of the 1D-CNN is shown in Table 5.

The parameters selection of each method are described below:

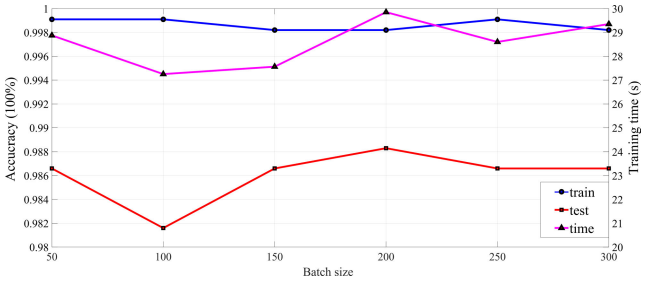



**FIGURE 18.** Classification accuracies and training time for different batch size.

**TABLE 5.** Detailed description of the parameters.

| Parameter description | Value |
|---|---|
| The number of CNN layer | 2 |
| The number of pooling layer | 2 |
| Length convolution kernel | 3×1, 3×1 |
| Pooling size | 3×1, 3×1 |
| The optimizer | ADAM |
| Learning rate | 0.001 |
| Structure of CNN layer | 64, 32 |
| Batch size | 200 |

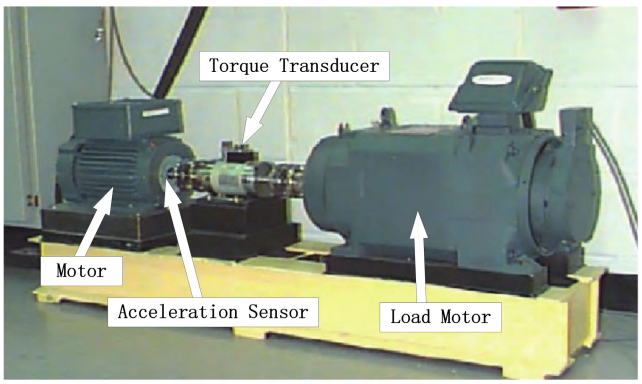



**FIGURE 19.** A rolling bearing fault experiment platform.

BPNN with 19 features: The structure is 20-18-17, and the learning rate is 0.01. These parameters are selected followed the guiding principle and experiences.

SVM with 19 features: The penalty factor is 16, Poly kernel is selected, and the radius of the gamma function is 0.01625.

1D-CNN with raw data: The structure of the 1D-CNN is set to 64,32. The length convolution kernels are set to $3 \times 1, 3 \times 1$. The learning rate is 0.001.

CNNEPDNN: The structure of the CNN is set to 20,40. The length convolution kernels are set to $5 \times 1$, $5 \times 1$. The structure of DNN is set to 20-40-80. The learning rate is 0.0001.

M2DCNN: The structure of the 1D-CNN is set to 16,32,16. The length convolution kernels are set to $2 \times 2$, $2 \times 2$, $6 \times 10$. The learning rate is 0.001.

**TABLE 6.** Description of the bearing dataset.

| Health conditions | Identification label | Training samples/Testing samples Dataset A | Dataset B | Dataset C |
|---|---|---|---|---|
| 0.007_Ball | 1 | 80/40 | 60/60 | 40/80 |
| 0.007_InnerRace | 2 | 80/40 | 60/60 | 40/80 |
| 0.007_OuterRace12 | 3 | 80/40 | 60/60 | 40/80 |
| 0.007_OuterRace3 | 4 | 80/40 | 60/60 | 40/80 |
| 0.007_OuterRace6 | 5 | 80/40 | 60/60 | 40/80 |
| 0.014_Ball | 6 | 80/40 | 40/80 | 40/80 |
| 0.014_InnerRace | 7 | 80/40 | 40/80 | 40/80 |
| 0.014_OuterRace6 | 8 | 80/40 | 40/80 | 40/80 |
| 0.021_Ball | 9 | 80/40 | 40/80 | 40/80 |
| 0.021_InnerRace | 10 | 80/40 | 40/80 | 40/80 |
| 0.021_OuterRace12 | 11 | 80/40 | 40/80 | 40/80 |
| 0.021_OuterRace3 | 12 | 80/40 | 40/80 | 40/80 |
| 0.021_OuterRace6 | 13 | 80/40 | 40/80 | 40/80 |
| 0.028_Ball | 14 | 80/40 | 40/80 | 20/100 |
| 0.028_InnerRace | 15 | 80/40 | 40/80 | 20/100 |
| normal | 0 | 160/80 | 160/80 | 160/80 |

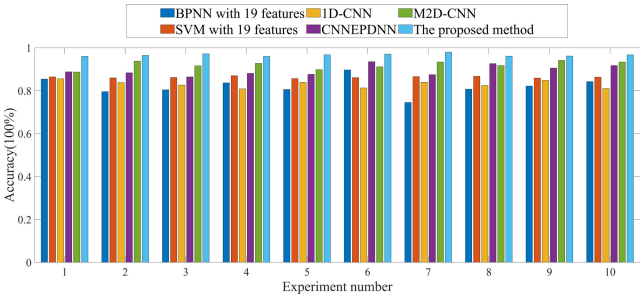



**FIGURE 20.** Ten experiment results of different methods.

### B. CASE 2: CWRU DATA OF BEARING

In order to verify the performance of the proposed method, samples from the open bearing dataset of the CWRU were used for experimental analysis [48]. The experimental platform for bearing failure was shown in Figure 19. A 1491.4w three-phase induction motor shaft (left) generates the driving force and a load motor (right) generates the rated load. A torque sensor's automatic alignment system connects the two parts. Two acceleration sensors are separately installed in the fan end and drive end of the motor respectively to collect vibration signals of the fault bearing. EDM technology was used to acquire different fault degree in inner-race, ball and outer-race of the bearings, respectively. Different fault severities (0.007in, 0.014in, 0.021in and 0.028in) can be acquired.

Sixteen classes health condition are acquired to analyze the performance of the designed method. In order to make the experiment more practical, three datasets with different split ratios are adopted to test the recognition performance of the designed method, as shown in Table 6. Dataset A is an ideal dataset with sufficient training samples for each health condition. Dataset B and dataset C are both imbalanced datasets. Dataset C has fewer serious fault samples than Dataset B. The detailed description of the three datasets are shown in Table 6.

Figure 20 shows the ten detailed experiment results of different methods. It shows that the designed method has higher

**TABLE 7. Experiment results of different methods.**

| Methods | Average recognition accuracy of each health condition (%) | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 0 | Average accuracy(%) |
| BPNN with 19 features | 100 | 100 | 95 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 97 | 100 | 100 | 100 | 100 | 0 | 87.79 ± 3.83 |
| SVM with 19 features | 100 | 100 | 97 | 100 | 100 | 3 | 88 | 100 | 100 | 8 | 5 | 97 | 100 | 100 | 100 | 100 | 86.47 ± 0.38 |
| 1D-CNN[9] | 95 | 100 | 97 | 97 | 100 | 57 | 35 | 97 | 65 | 47 | 65 | 100 | 75 | 65 | 100 | 100 | 82.21 ± 1.31 |
| M2D-CNN[19] | 95 | 100 | 95 | 100 | 100 | 35 | 100 | 98 | 38 | 100 | 98 | 100 | 100 | 100 | 100 | 100 | 91.62 ± 1.69 |
| CNNEPDNN[31] | 95 | 100. | 100 | 100 | 100 | 68 | 7 | 100 | 58 | 45 | 95 | 100 | 83 | 98 | 100 | 100 | 88.82 ± 2.30 |
| The proposed method | 97 | 100 | 97 | 100 | 100 | 100 | 100 | 88 | 75 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 97.50 ± 0.59 |

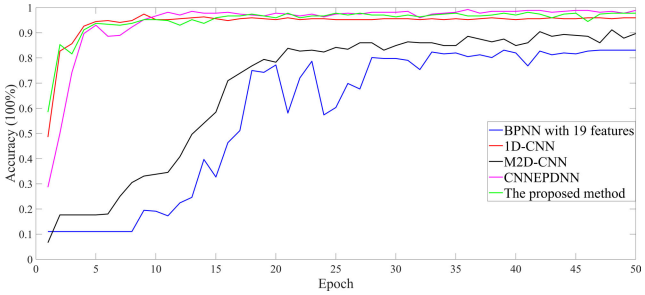



**FIGURE 21. Trends of accuracy of different methods in each epoch.**

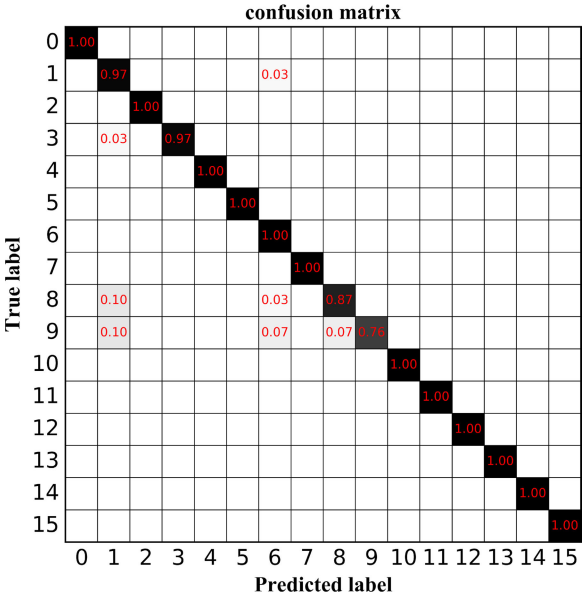



**FIGURE 22. The classification confusion matrix of the proposed method.**

accuracy than the other five methods. In Table 7, the average recognition accuracy of each health condition and the average accuracy of dataset A are also shown. The average recognition accuracy of the testing samples is (97.5%), which is much higher than those of BPNN with 19 features (87.79%), SVM with 19 features (86.47%), 1D-CNN (82.21%), CNNEPDNN (88.82%) and M2D-CNN (91.62%).

Figure 21 shows the evolution of the accuracy of the five methods as we increase the epoch through the training process. From Figure 21, the proposed method rapidly converges

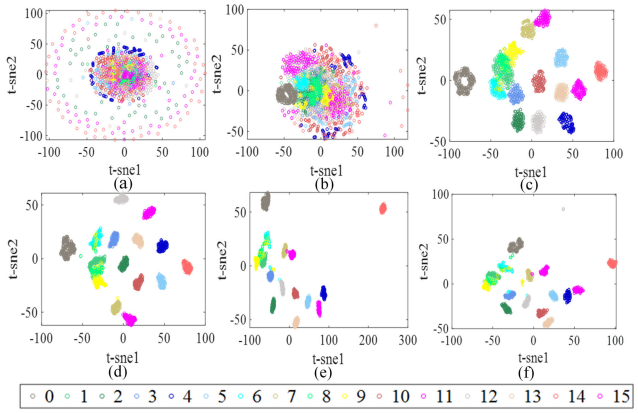



**FIGURE 23. Feature visualization via t-SNE. (a) Raw data feature, (b) Feature in the 1st hidden layer of DBN, (c) Feature in the 2nd hidden layer of DBN, (d) Feature in the 3rd hidden layer of DBN, (e) Feature in the 1st hidden layer of 1D-CNN, (f) Feature in the 2nd hidden layer of 1D-CNN.**

nearly forty iterates and builds a precise model. We also find that the highest accuracy value happened when the epoch is nearly twenty. Figure 22 shows the classification confusion matrix of the designed method for one of the ten trails. From Figure 22, the model has the lowest accuracy (75% for the label 9 (0.021_Ball)), and the second lowest accuracy (88% for the label 8 (0.014_OuterRace6)), but the classification accuracies of the other operating conditions are still very high. The results indicate that the designed method is significantly superior to the other five methods.

In order to understand the effect more intuitively of the designed method on feature extraction and classification, the t-SNE technique is used to visualize the network output of each layer. As shown in Figure 23, different colors represent different health conditions. The darker the color, the smaller the label value of the corresponding health condition. From Figure 23, each layer of the network can extract the abstract features of signals. With the increase of the number of the the networ layer, the points of the same colors are closer together, while the points of different colors are farther seperable. It can be observed that the designed method has good feature extraction ability.

To evaluate the robustness of the designed method, the additive Gaussian white noises are added into the raw

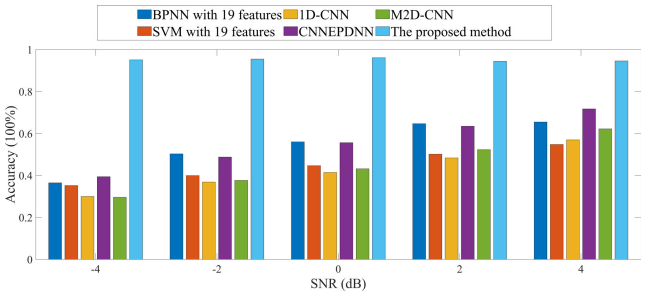



**FIGURE 24.** Comparison of accuracy testing on signals with different SNR values.

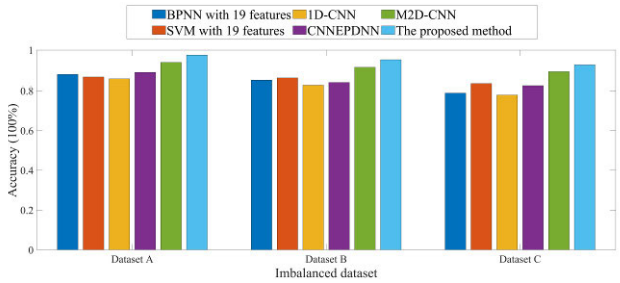



**FIGURE 25.** Experiment result of different datasets with the six methods.

vibration signals, and the SNRs range from -4 to 4dB. The 10 experiments were carried out to analyze the impact of different SNRs, the detailed diagnosis results are displayed in Figure 24. The proposed method has the strongest robustness in comparison with the other five approaches. The average classification accuracies are all over 94.45% within all considered SNR levels from -4 to 4dB, which are much higher than those of the other methods. In addition, when classifying 16 kinds of health conditions, the other methods cannot effectively identify them. Thus, the proposed method can achieve fault diagnosis of more health conditions.

In real-life, the fault samples are often difficult to obtain, so the imbalanced datasets are used for further experiments. From Figure 25, M2D-CNN achieves good performance, and the 1D-CNN has the worst performance. The proposed method also has a very high accuracy in the case of a small amount of fault samples. However, the accuracies of the other approaches are still lower than that of the proposed method.

To show the detail experiment results, the confusion matrices of the classification accuracies of the testing samples for the three kinds of datasets are plotted. Figure 26 shows the classification accuracy of each health condition. In dataset B, the training samples of the fault conditions are fewer than those of the dataset A (shown in Figure 22). There are more misclassified samples, but the accuracies of each health condition are above 93% (except the fault type 6 is 81% and the fault type 9 is 74%). Dataset C (the seriously imbalanced dataset) has the fewest training samples of the 0.028_Ball and 0.028_InnerRace conditions. As shown in Figure 26(b), the accuracy of the label 1 is the lowest as 23%. However, the accuracies of the other health conditions are nearly 100%. The results show that the designed method has the super performance in the imbalanced dataset.

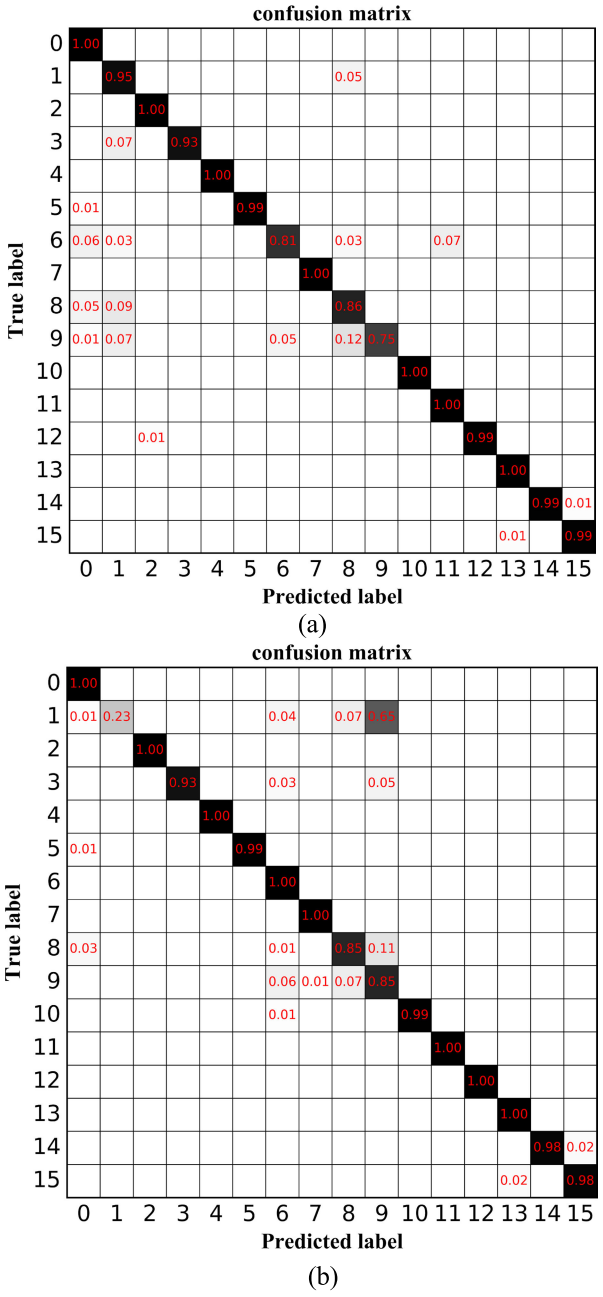



**FIGURE 26.** The classification confusion matrix of the designed method. (a) Dataset B, (b) Dataset C.

## IV. CONCLUSION

A novel method based on DBN and 1D-CNN for intelligent fault diagnosis of rotating machinery is designed. It can be achieved as follows: Firstly, the DBN built by multiple pre-trained RBMs extracts the initial features of the raw data and achieves dimensionality reduction. Secondly, the optimized 1-DCNN is used for further feature extraction. Finally, the extracted abstract features are fed into Soft-max classifier.

Deep learning model contains multiple hidden layers, which are used to extract the deep features of complex signals. The features extracted by these hidden layers are not designed by human engineers, but obtained by self-learning

the abstract features of input data. In the future, the study will be paid more attention to the deep learning method based on the feature extraction. The DBN and CNN will be further studied to improve their performances of fault diagnosis.

## REFERENCES

[1] R. B. Randall, J. Antoni, and S. Chobsaard, "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals," *Mech. Syst. Signal Process.*, vol. 15, no. 5, pp. 945–962, 2001.

[2] X. Liu, R. B. Randall, and J. Antoni, "Blind separation of internal combustion engine vibration signals by a deflation method," *Mech. Syst. Signal Process.*, vol. 22, no. 5, pp. 1082–1091, 2008.

[3] G. Xu, M. Liu, Z. Jiang, D. Soffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, Mar. 2019.

[4] Z. Qiao, Y. Lei, J. Lin, and F. Jia, "An adaptive unsaturated bistable stochastic resonance method and its application in mechanical fault diagnosis," *Mech. Syst. Signal Process.*, vol. 84, pp. 731–746, Feb. 2017.

[5] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[6] D. Abboud, M. Elbadaoui, W. A. Smith, and R. B. Randall, "Advanced bearing diagnostics: A comparative study of two powerful approaches," *Mech. Syst. Signal Process.*, vol. 114, pp. 604–627, Jan. 2019.

[7] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[8] R. Razavi-Far, E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S. H. Kia, and H. Henao, and G.-A. Capolino, "Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6331–6342, Aug. 2019.

[9] C. Wu, P. Jiang, C. Ding, F. Feng, and T. Chen, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network," *Comput. Ind.*, vol. 108, pp. 53–61, Jun. 2019.

[10] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 1–35, Mar. 2016.

[11] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Process.*, vol. 96, pp. 1–15, Mar. 2014.

[12] J. Zheng, J. Cheng, Y. Yang, and S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination," *Mechanism Mach. Theory*, vol. 78, no. 16, pp. 187–200, 2014.

[13] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 92–104, May 2016.

[14] J. Liang, Y. Zhang, J.-H. Zhong, and H. Yang, "A novel multi-segment feature fusion based fault classification approach for rotating machinery," *Mech. Syst. Signal Process.*, vol. 122, pp. 19–41, May 2019.

[15] M. Kircheis and D. Potts, "Direct inversion of the nonequispaced fast Fourier transform," *Linear Algebra Appl.*, vol. 575, pp. 106–140, Aug. 2019.

[16] Q. Cheng, Z. Zhao, C. Tang, G. Qian, and S. Islam, "Diagnostic of transformer winding deformation fault types using continuous wavelet transform of pulse response," *Measurement*, vol. 140, pp. 197–206, Jul. 2019.

[17] Y. Hu, S. Zhang, A. Jiang, L. Zhang, W. Jiang, and J. Li, "A new method of wind turbine bearing fault diagnosis based on multi-masking empirical mode decomposition and fuzzy c-means clustering," *Chin. J. Mech. Eng.*, vol. 32, no. 1, p. 46, 2019.

[18] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, Feb. 2018.

[19] W. Gong, H. Chen, Z. Zhang, M. Zhang, R. Wang, C. Guan, and Q. Wang, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, Apr. 2019.

[20] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2017.

[21] H. Jiang, H. Shao, X. Chen, J. Huang, C. Li, and J. V. de Oliveira, "A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery," *J. Intell. Fuzzy Syst.*, vol. 34, no. 6, pp. 3513–3521, 2018.

[22] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[23] R. X. Chen, X. Huang, L. X. Yang, X. Y. Xu, X. Zhang, and Y. Zhang, "Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform," *Comput. Ind.*, vol. 106, pp. 48–59, Apr. 2019.

[24] Y. Zhang, X. Li, L. Gao, and P. Li, "A new subset based deep feature learning method for intelligent fault diagnosis of bearing," *Expert Syst. Appl.*, vol. 110, pp. 125–142, Nov. 2018.

[25] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renew. Energy*, vol. 133, pp. 422–432, Apr. 2018.

[26] A. Y. Appiah, X. Zhang, B. B. K. Ayawli, and F. Kyeremeh, "Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis," *IEEE Access*, vol. 7, pp. 30089–30101, 2019.

[27] Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3814–3824, May 2019.

[28] J. Tao, Y. Liu, and D. Yang, "Bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock Vib.*, vol. 2016, Aug. 2016, Art. no. 9306205.

[29] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu, "An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis," *IEEE Access*, vol. 6, pp. 63584–63596, 2018.

[30] L. Wen, X. Li, and L. Gao, "A new two-level hierarchical diagnosis network based on convolutional neural network," *IEEE Trans. Instrum. Meas.*, to be published.

[31] H. Li, J. Huang, and S. Ji, "Bearing fault diagnosis with a feature fusion method based on an ensemble convolutional neural network and deep neural network," *Sensors*, vol. 19, no. 9, p. 2034, Apr. 2019.

[32] S. Huang, J. Tang, J. Dai, and Y. Wang, "Signal status recognition based on 1DCNN and its feature extraction mechanism analysis," *Sensors*, vol. 19, no. 9, p. 2018, Apr. 2019.

[33] Y. Wang, Z. Xie, K. Xu, Y. Dou, and Y. Lei, "An efficient and effective convolutional auto-encoder extreme learning machine network for 3D feature learning," *Neurocomputing*, vol. 174, pp. 988–998, Jan. 2016.

[34] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, p. 115002, 2015.

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

[36] R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 872–879, doi: 10.1145/1390156.1390266.

[37] 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 Syst. Appl.*, vol. 41, no. 9, pp. 4113–4122, 2014.

[38] 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.

[39] J. Liu, Y. Hu, Y. Wang, B. Wu, J. Fan, and Z. Hu, "An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis," *Meas. Sci. Technol.*, vol. 29, no. 5, 2018, Art. no. 055103.

[40] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multi-scale network," *IEEE Trans. Ind. Electron.*, to be published.

[41] R. W. Kennard and L. A. Stone, "Computer aided design of experiments," *Technometrics*, vol. 11, no. 1, pp. 137–148, Feb. 1969.

[42] M. Daszykowski, B. Walczak, and D. L. Massart, "Representative subset selection," *Anal. Chim. Acta*, vol. 468, no. 1, pp. 91–103, 2002.

[43] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.

[44] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[45] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1310–1320, Jun. 2017.
[46] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Santa Rosa, CA, USA, Mar. 2017, pp. 464–472.
[47] L. Wen, L. Gao, and X. Li, "A new snapshot ensemble convolutional neural network for fault diagnosis," *IEEE Access*, vol. 7, pp. 32037–32047, Mar. 2019.
[48] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
[49] Bearing Data Center. Accessed: Oct. 20, 2018. [Online]. Available: http://csegroups.case.edu/bearingdatacenter/home

**YIBING LI** received the M.S. and Ph.D. degrees in mechanical manufacturing and automation from the Wuhan University of Technology, in 2003 and 2008, respectively. He is currently a Professor with the School of Mechanical and Electronic Engineering, Wuhan University of Technology. He has published about 20 journal articles. His current research interests include fault diagnosis of mechanical equipment, and intelligent manufacturing and shop scheduling.

**LI ZOU** was born in Huanggang, Hubei, China, in 1994. He received the B.Eng. degree from the School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan, China, where he is currently pursuing the M.S. degree.

His current research interest includes fault diagnosis of mechanical equipment.

**LI JIANG** received the M.S. degree in mechanical design and theory from Wuhan University, Wuhan, China, in 2006, and the Ph.D. degree in mechanical and electronic engineering from the Huazhong University of Science and Technology, Wuhan, in 2013. She is currently a Lecturer with the School of Mechanical and Electronic Engineering, Wuhan University of Technology.

**XIANGYU ZHOU** was born in Jingzhou, Hubei, China, in 1996. He received the B.Eng. degree from the School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan, China, where he is currently pursuing the M.S. degree.

His current research interest includes fault diagnosis of mechanical equipment.