

PAPER

## An adaptive deep convolutional neural network for rolling bearing fault diagnosis

To cite this article: Wang Fuan *et al* 2017 *Meas. Sci. Technol.* **28** 095005

View the [article online](#) for updates and enhancements.

### Related content

- [Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network](#)  
Hongkai Jiang, Xingqiu Li, Haidong Shao et al.
- [Rolling bearing fault diagnosis using an optimization deep belief network](#)  
Haidong Shao, Hongkai Jiang, Xun Zhang et al.
- [An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network](#)  
Weiwei Qian, Shunming Li, Jinrui Wang et al.

### Recent citations

- [A novel gas turbine fault diagnosis method based on transfer learning with CNN](#)  
Zhong Shi-sheng *et al*
- [A Self-Adaptive 1D Convolutional Neural Network for Flight-State Identification](#)  
Xi Chen *et al*
- [Fault diagnosis for distillation process based on CNN-DAE](#)  
Chuankun Li *et al*

The Sixteenth International Conference on Condition Monitoring and Asset Management

# CM2019

Tuesday 25 to Thursday 27 June 2019

The Principal Grand Central Hotel, Glasgow, UK

Abstracts and papers should be submitted online at: <https://mc.manuscriptcentral.com/cm2019-mfpt2019>

Still accepting abstracts

# An adaptive deep convolutional neural network for rolling bearing fault diagnosis

**Wang Fuan, Jiang Hongkai, Shao Haidong, Duan Wenjing  
and Wu Shuaipeng**

School of Aeronautics, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China

E-mail: [jianghk@nwpu.edu.cn](mailto:jianghk@nwpu.edu.cn)

Received 21 November 2016, revised 3 February 2017

Accepted for publication 20 April 2017

Published 16 August 2017



## Abstract

The working conditions of rolling bearings usually is very complex, which makes it difficult to diagnose rolling bearing faults. In this paper, a novel method called the adaptive deep convolutional neural network (CNN) is proposed for rolling bearing fault diagnosis. Firstly, to get rid of manual feature extraction, the deep CNN model is initialized for automatic feature learning. Secondly, to adapt to different signal characteristics, the main parameters of the deep CNN model are determined with a particle swarm optimization method. Thirdly, to evaluate the feature learning ability of the proposed method, t-distributed stochastic neighbor embedding (t-SNE) is further adopted to visualize the hierarchical feature learning process. The proposed method is applied to diagnose rolling bearing faults, and the results confirm that the proposed method is more effective and robust than other intelligent methods.

**Keywords:** rolling bearing, adaptive deep convolutional neural network, feature learning, particle swarm optimization, fault diagnosis

(Some figures may appear in colour only in the online journal)

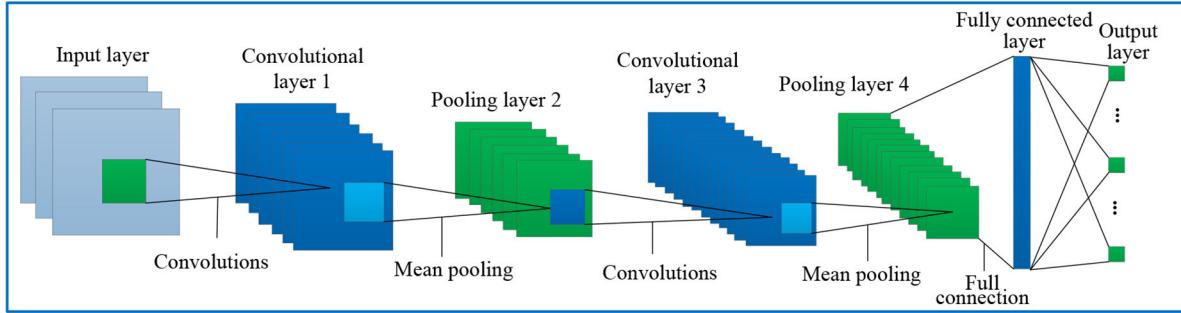
## 1. Introduction

Rolling bearings usually act as key parts in rotating machinery, such as electrical locomotive, gas-turbine and aircraft engines [1–4]. If faults are generated in these rolling bearings, they will greatly affect the operating conditions of rotating machinery and even lead to machine breakdown. Therefore, conducting research on rolling bearing fault diagnosis is of great significance. In recent years, many fault diagnosis methods based on vibration signal analysis have been proposed for rolling bearing fault diagnosis [5–10].

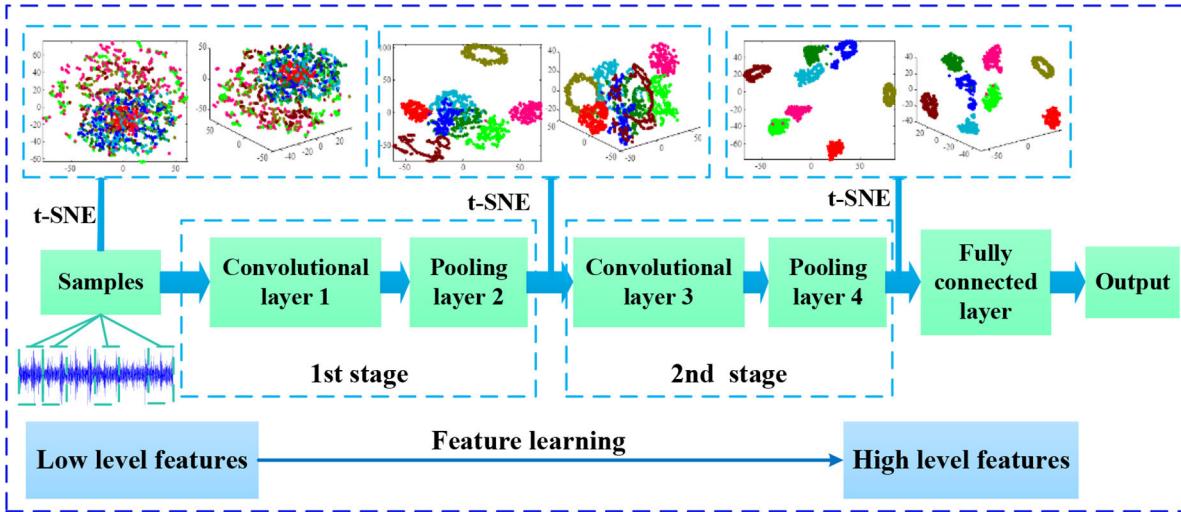
With the rapid development of intelligent technology, an intelligent fault diagnosis method is becoming a hot topic in the field of fault diagnosis. Liu *et al* proposed a bearing fault diagnosis method based on a time-frequency atoms-driven support vector machine (SVM) [11]. Zhang *et al* applied multivariable ensemble-based incremental SVM to rolling bearing fault diagnosis [12]. Qu *et al* proposed a novel fault diagnosis method based on dual-tree complex wavelet packet transform and multiple classifier fusion [13]. Hong *et al* proposed a real-time fault diagnosis method based on SVM [14]. Bin *et al*

proposed a fault diagnosis method based on empirical mode decomposition feature extraction and neural network [15]. However, the existing intelligent fault diagnosis methods usually need manual feature extraction. It is because of this that they use shallow learning models, such as artificial neural network (ANN) [15, 16] and SVM [11–14], which lack a powerful feature learning ability. Therefore, they rely heavily on manual experience or domain expertise, which limit the application of these intelligent methods to fault diagnosis problems. As a consequence, it is of great importance to develop deep learning models for rolling bearing fault diagnosis.

Geoffrey Hinton first proposed a deep learning theory in 2006 [17]. Unlike traditional shallow learning methods, deep learning uses deep architecture and has good feature learning ability. The main deep learning models contain deep convolutional neural networks (CNN) [18], deep belief networks (DBN) [19] and deep auto-encoders (DAE) [20]. Recently, researchers have applied deep learning to fault diagnosis and obtained good performance. Jia *et al* proposed a novel fault diagnosis method based on DAE [21]. Lu *et al* proposed a fault diagnosis based on deep learning and applied it to



**Figure 1.** The deep CNN model used in this paper.



**Figure 2.** Visualization of hierarchical feature learning process with t-SNE method.

diagnose rotary machinery faults [22]. Gan *et al* proposed a hierarchical diagnosis network based on deep learning and applied it to rolling bearing fault diagnosis [23]. Shao *et al* applied DBN to rolling bearing fault diagnosis [24]. Shao *et al* proposed a deep feature fusion method based on deep learning for rotating machinery fault diagnosis [25].

Among the three main models, deep CNN is the first successful deep learning model [18]. By using a deep architecture with multiple processing layers, deep CNN can directly learn the suitable internal features from input data for classification or prediction. In the last few years, deep CNN has been widely applied to pattern recognition problems, such as speech recognition [18], image recognition [19] and fault diagnosis [26, 27]. While applied to different pattern recognition problems, deep CNN models are usually designed manually [28–32]. Due to the complex working conditions, such as different health conditions and different load and speed, rolling bearing vibration signals are very sophisticated and contain different characteristics. Therefore, it is a tremendous challenge to design a proper deep CNN model manually.

In order to effectively diagnose rolling bearing faults and reduce dependence on manual experience and domain expertise, a novel adaptive deep CNN approach is proposed in this paper. Firstly, a deep CNN model is initialized to get rid of manual feature extraction. Then, in order to enable the deep CNN to automatically adapt to the different rolling bearing vibration signals, particle swarm optimization (PSO) [33, 34]

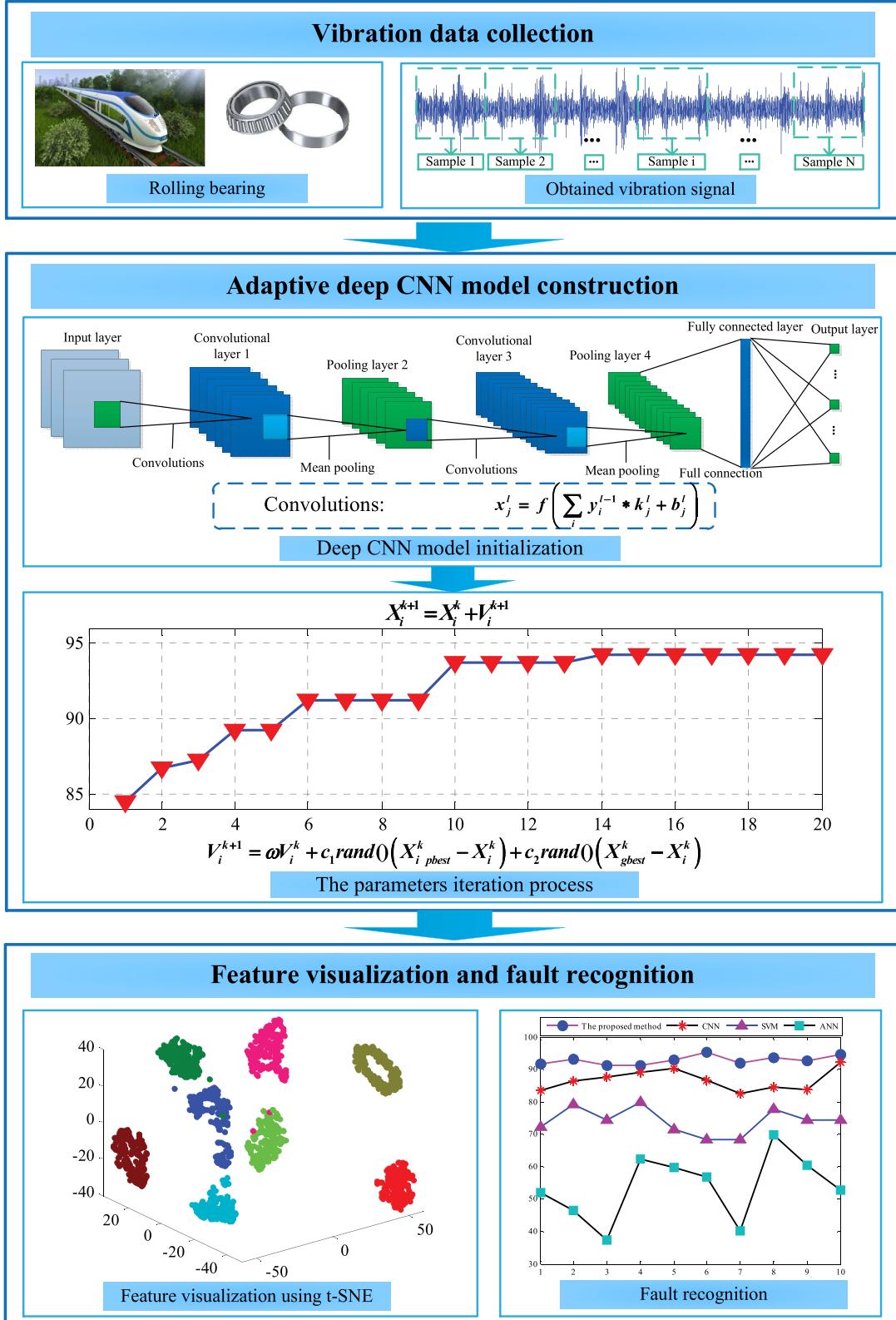
is applied to determine the deep CNN key parameters. Finally, t-distributed stochastic neighbor embedding (t-SNE) [35] is applied to visualize the high-dimensional features in stages of a hierarchical feature learning process.

The contributions of the proposed method are summarized as follows. Firstly, the proposed method can achieve sufficient diagnosis performance without manual feature extraction. Secondly, the proposed method can automatically adapt to the characteristics of vibration signals when applied to different rolling bearing fault diagnosis problems. The proposed method is applied to rolling bearing fault diagnosis, and the results show that the proposed method has good adaptability and robustness.

The rest of this paper is organized as follows. In section 2, the basic theory of deep CNN is briefly introduced. Section 3 describes the proposed method in detail. In section 4, experimental signals are analyzed by the proposed method. The engineering application of the proposed method is presented in section 5. The general conclusion is given in section 6.

## 2. The basic theory of deep CNN

The deep CNN model employs multiple processing layers to process the input data [18]. The main feature learning layers adopted in deep CNN contain convolutional layers and pooling layers. In a convolutional layer, a series of trainable kernels are used to convolve with the input feature maps and then pass through the activation function to generate the

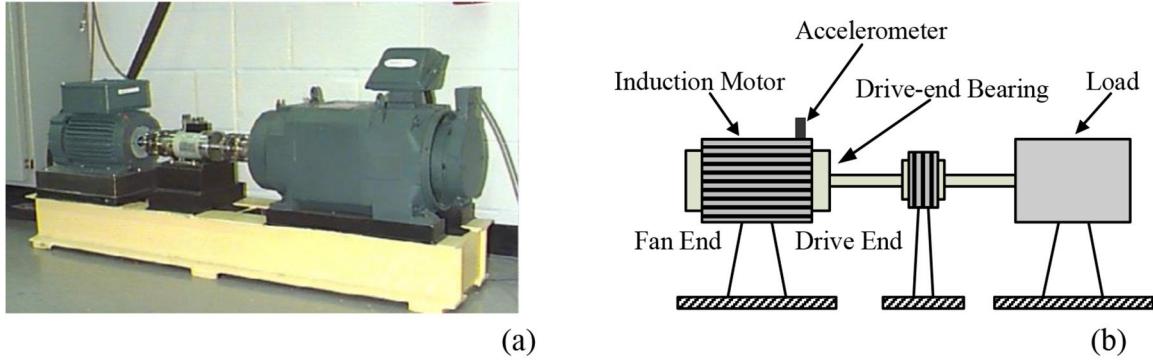


**Figure 3.** The proposed method for rolling bearing fault diagnosis.

output feature maps. Considering a convolutional layer is the  $l$ th layer of deep CNN, the computation is as follows:

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

where  $x_j^l$  denotes the  $j$ th feature map generated by the  $l$ th layer,  $x_i^{l-1}$  denotes the  $i$ th input feature map,  $k_{ij}^l$  denotes the  $j$ th kernel connected with  $i$ th input feature map,  $f(\cdot)$  is an activation function,  $b_j^l$  denotes the bias corresponding to  $j$ th kernel, and  $*$  denotes the 2D convolution operation.



**Figure 4.** The rolling bearing experimental setup. (a) Experiment picture; (b) schematic illustration.

**Table 1.** The description of experiment 1 and experiment 2.

Experiments	Input data	Dimension
Experiment 1	Raw vibration data (without any signal preprocessing)	400(20 * 20)
Experiment 2	The vibration signals are decomposed into 8 frequency band signals and the most sensitive frequency bands are selected as input data	400(20 * 20)

**Table 2.** Rolling bearing fault sample distribution.

Rolling bearing condition	The size of training/testing samples	Label
Normal condition	100/100	1
0.1778 mm/inner race fault	100/100	2
0.1778 mm/roller fault	100/100	3
0.1778 mm/outer race fault	100/100	4
0.3556 mm/roller fault	100/100	5
0.3556 mm/outer race fault	100/100	6
0.5334 mm/inner race fault	100/100	7
0.5334 mm/roller fault	100/100	8

Each convolutional layer is followed with a pooling layer. The pooling layers are used to reduce the resolution of the input feature maps and improve the robustness of deep CNN [28]. In a pooling layer, each feature map from the previous layer is divided into sub-regions with a size of  $k \times k$ , and the mean value or the maximum value is the output of these sub-regions. If the input feature map is the size  $N \times N$  and all the sub-regions are non-overlapping, then the output feature map is the size  $\frac{N}{k} \times \frac{N}{k}$ .

### 3. The proposed method

In this paper, an adaptive deep CNN is proposed for rolling bearing fault diagnosis. This section includes four parts: deep CNN model initialization, deep CNN parameter optimization, visualization of hierarchical feature learning process, and the summarized procedure of the proposed method.

#### 3.1. Deep CNN model initialization

In order to get rid of manual feature extraction, the deep CNN model is initialized for automatic feature learning. The deep CNN model used in this paper is shown in figure 1, which consists of an input layer, convolutional layer C1, pooling layer

**Table 3.** Parameters of PSO in experiment 1.

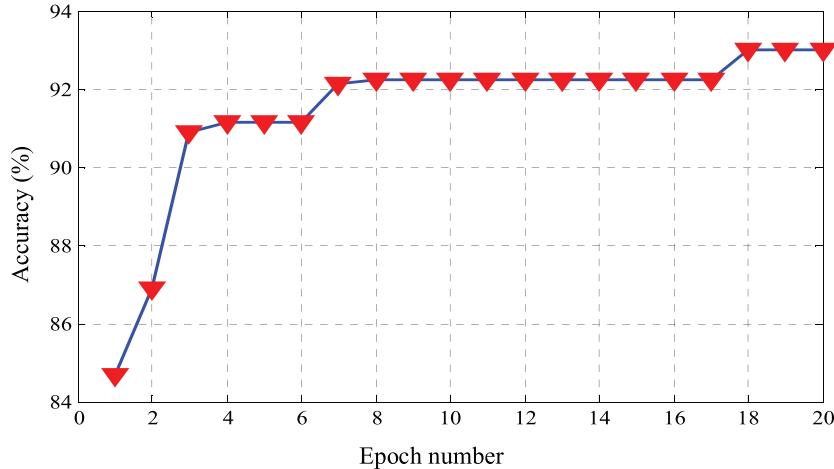
Description	Symbol	Value
Population size of PSO	$n$	25
Iteration number	$N$	20
The dimensionality of particles	—	3
Acceleration constants	$c_1, c_2$	$c_1 = 2, c_2 = 2$
Inertia weight of PSO	$\omega$	0.5
Learning rate of deep CNN	$\eta$	0.8482 (output from PSO)
Kernels in C1 layer	$m_1$	10 (output from PSO)
Kernels in C3 layer	$m_2$	13 (output from PSO)

P2, convolutional layer C3, pooling layer P4, a fully connected layer and an output layer. The C1 layer contains  $m_1$  kernels and the C3 layer contains  $m_2$  kernels, the size of the kernels in the convolutional layers is  $5 \times 5$ , the leaning rate is  $\eta$ , the pooling layers are mean-pooling layers, and the sub-regions in the pooling layers are non-overlapping with a size of  $2 \times 2$ . The activation function is a sigmoid  $\sigma(x) = 1/(1 + e^{-x})$ . The fully connected layer contains 100 processing units.

#### 3.2. The parameter optimization of deep CNN

The key parameters  $m_1$ ,  $m_2$  and  $\eta$  of deep CNN are usually determined by experience and repeated experiments [32], which greatly limit the application of deep CNN to diagnose the faults of different rolling bearings. PSO is a global search method inspired by the behavior of animal groups in nature. Because of its good performance, the PSO algorithm has been widely applied to optimize non-linear, continuous functions in the last few years [33, 34].

In order to enable the deep CNN to automatically adapt to different characteristics of rolling bearing vibration signals, PSO is applied to determine the key parameters of the deep CNN model. The inputs of PSO are as follows: the

**Figure 5.** The iteration process of PSO in experiment 1.

population size, the dimensionality of particles, iteration number, acceleration constants and inertia weight. The outputs of PSO are the optimal parameters. The general procedure is as follows:

**Step 1:** Set the PSO parameters, iteration number  $N$ , the particle population size  $n$ , acceleration constants  $c_1$  and  $c_2$ .  
**Step 2:** Initialize the position and velocity of particle swarms and the position of each particle corresponding to a group of key parameters of deep CNN.

**Step 3:** Train the initialized deep CNN using the parameters value corresponding to particle 1 to particle  $n$ .

**Step 4:** Evaluate each particle using the fitness value (classification accuracy), then select the best particle position of the swarm  $X_{\text{gbest}}^k$  and the best position of each particle  $X_{i\text{pbest}}^k$ .

**Step 5:** Update the position  $X$  and velocity  $V$  of each particle

$$X_i^{k+1} = X_i^k + V_i^{k+1} \quad (2)$$

$$\begin{aligned} V_i^{k+1} = & \omega V_i^k + c_1 r \text{ and } ()(X_{i\text{pbest}}^k - X_i^k) \\ & + c_2 r \text{ and } ()(X_{\text{gbest}}^k - X_i^k) \end{aligned} \quad (3)$$

where  $\omega$  denotes the inertial weight,  $c_1$  and  $c_2$  are acceleration constants.

**Step 6:** Finish the iteration process and design the adaptive deep CNN model using the optimal parameters while the current iteration number is  $N$ . Otherwise, return to step 3 and repeat the optimization process.

### 3.3. Visualization of hierarchical feature learning process

Raw rolling bearing vibration signals collected from different working conditions are very complex and are hard to be identified; they can be considered as low-level features. To accurately diagnose the rolling bearing faults, it is essential to learn the internal suitable fault features from the raw rolling bearing vibration signals. With the proposed adaptive

**Table 4.** The identification accuracy of experiment 1 and experiment 2.

Methods	Classification accuracy (%)	
	Experiment 1 (raw vibration data)	Experiment 2 (preprocessed data)
The proposed method	92.84	99.71
Deep CNN	86.64	93.72
SVM	74.06	88.04
ANN	53.80	67.97

deep CNN method, low-level features are transformed into high-level features layer by layer, which can promote the precision of fault diagnosis. However, it is difficult for us to evaluate the hierarchical learning process, because the learned features are high-dimensional. Therefore, the t-SNE [35] method is applied to visualize the high-dimensional features in each stage, as shown in figure 2. The t-SNE method visualizes features by giving each sample a location in a 2D or 3D map. With the 2D and 3D maps corresponding to each stage, we can evaluate the hierarchical feature learning process.

### 3.4. General procedure of the proposed method

The flowchart of the proposed method is shown in figure 3, and the general procedure is summarized as follows:

**Step 1:** The vibration signals of the rolling bearing are collected by the data acquisition system.

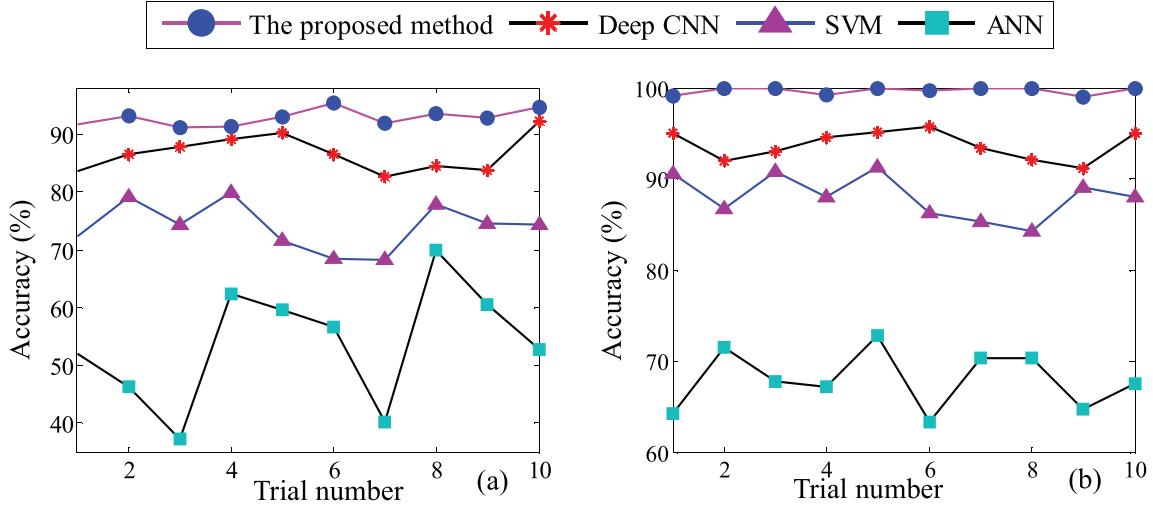
**Step 2:** The vibration signals are divided into training samples and testing samples.

**Step 3:** The deep CNN model is initialized.

**Step 4:** The adaptive deep CNN model with PSO is designed.

**Step 5:** The adaptive deep CNN is applied to learn the features of testing samples and then the learned features are used to diagnose rolling bearing faults.

**Step 6:** The fault diagnosis result is outputted and the hierarchical feature learning process visualized.



**Figure 6.** Diagnosis results of the ten trials. (a) For **experiment 1**; (b) for **experiment 2**.

**Table 5.** The average testing time of **experiment 1** and **experiment 2**.

Methods	Average testing time (s)	
	Experiment 1 (raw vibration data)	Experiment 2 (preprocessed data)
The proposed method	1.239	1.204
Deep CNN	1.105	1.079
SVM	0.223	0.219
ANN	0.305	0.296

#### 4. Experimental verification

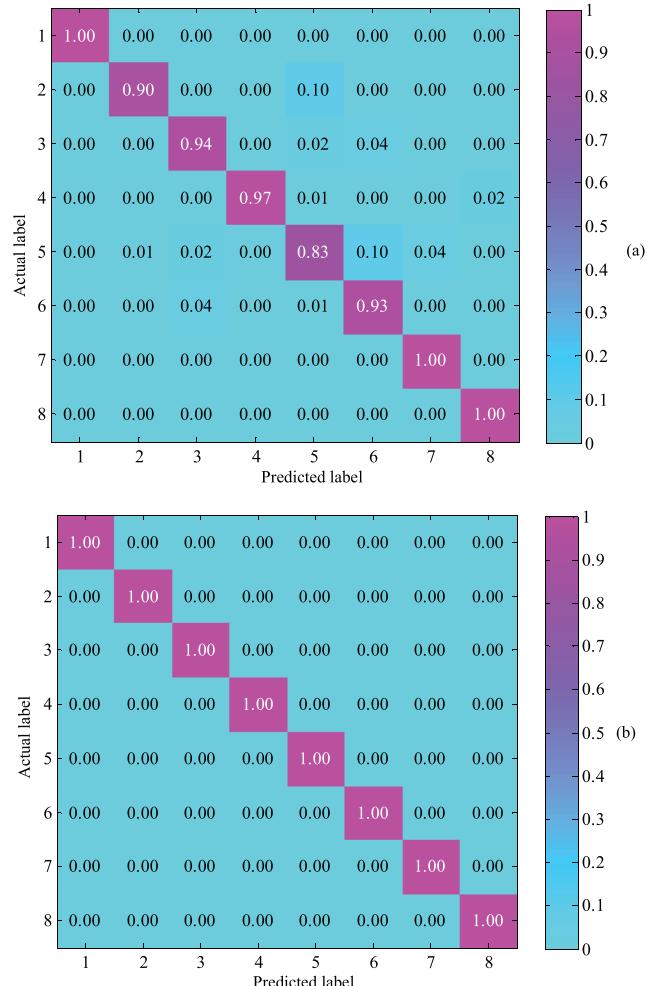
To verify the effectiveness of the proposed method, the experimental signals collected from an experimental setup are analyzed with the proposed method.

##### 4.1. Rolling bearing test rig

The experimental signals are from the Case Western Reserve University [36]. As shown in figure 4, the experimental setup is driven by a motor. Figure 4(a) is an experiment picture and figure 4(b) is a schematic illustration. The faults are seeded on the drive-end bearings. As shown in figure 4(b), accelerometers are attached to the drive end of the motor housing and the sampling frequency is 12 kHz. The vibration signals are collected from different health conditions, which include: (1) normal condition, (2) with outer race fault, (3) with inner race fault, and (4) with roller fault.

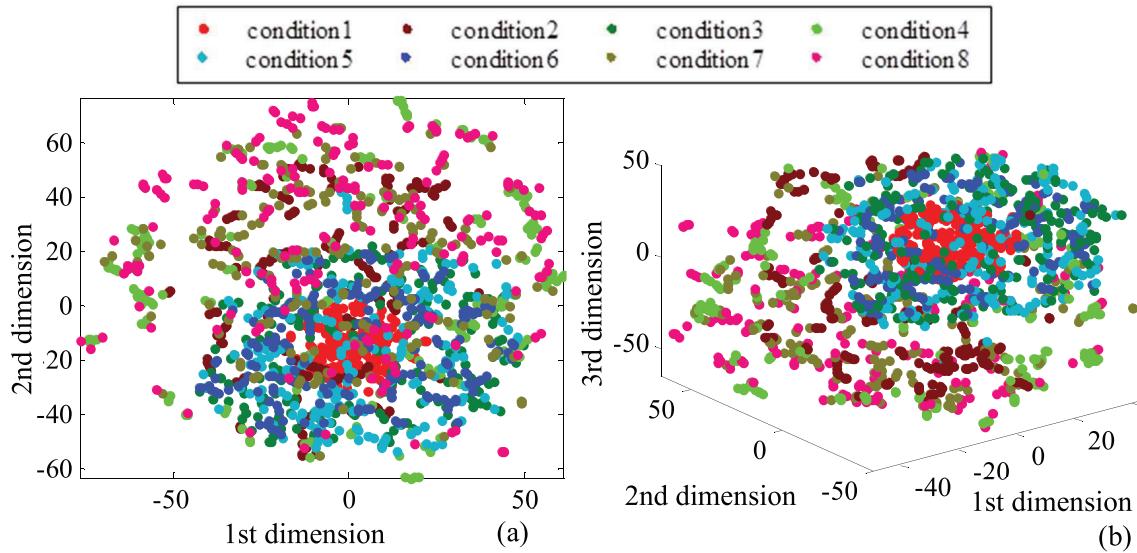
In this case study, four methods are used for rolling bearing fault diagnosis, including the proposed method, deep CNN, SVM and ANN. As shown in table 1, two experiments are considered for comparison. The purposes of the two experiments are described as follows:

- **Experiment 1:** The raw vibration data is directly input to the four methods, without any signal preprocessing and manual feature extraction.

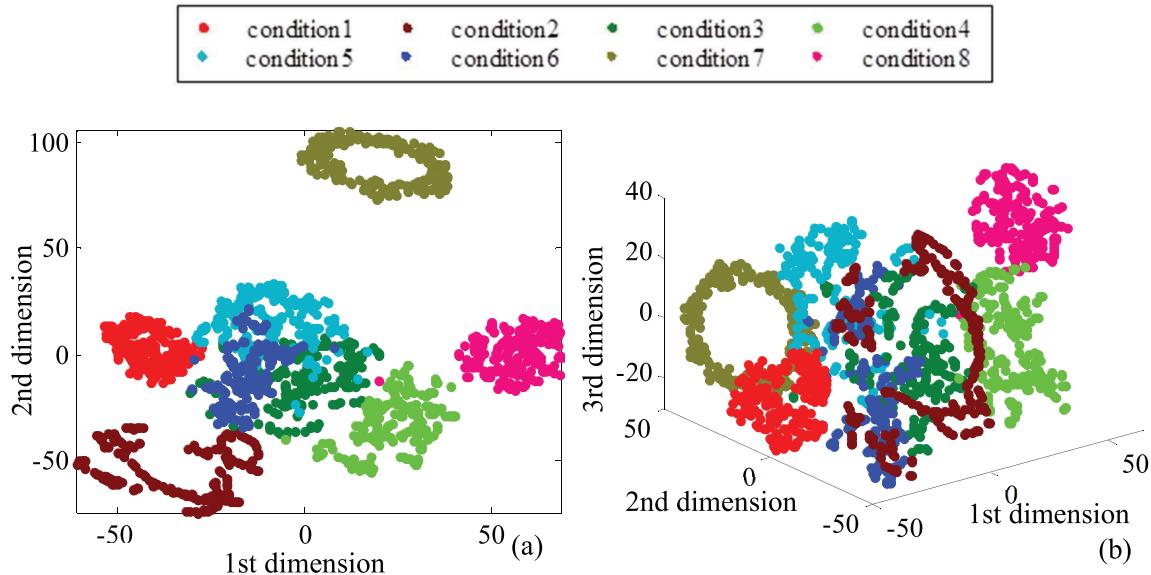


**Figure 7.** The confusion matrix using the proposed method. (a) **Experiment 1**; (b) **Experiment 2**.

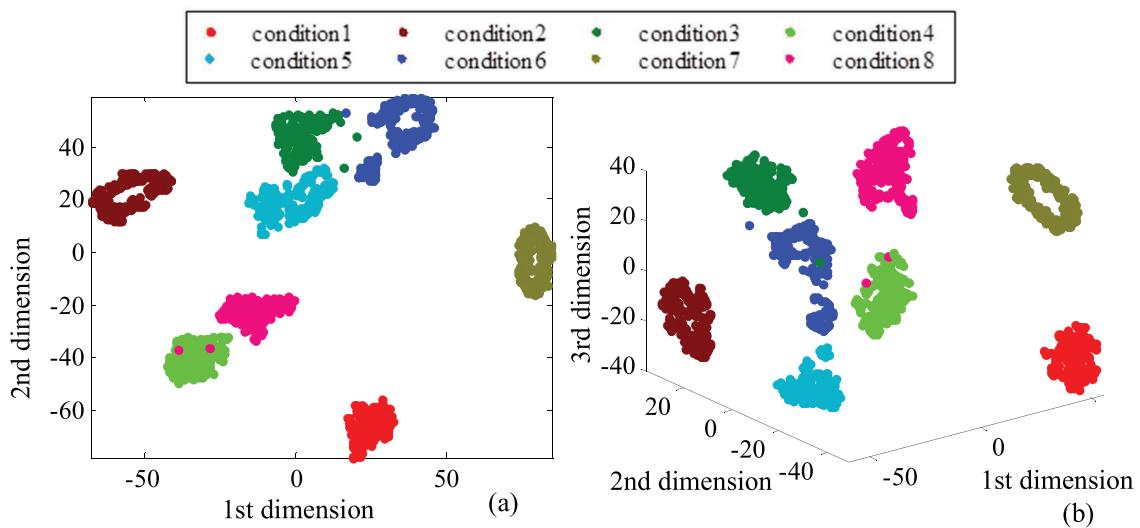
- **Experiment 2:** Three-layer wavelet packet (*db4*) is adopted to decompose the vibration signals. For each vibration signal, the most sensitive frequency band signal with the highest energy is selected. Then, the selected



**Figure 8.** The visualization of raw data features in input layer. (a) 2D embedding; (b) 3D embedding.



**Figure 9.** The visualization of features learned in P2 layer. (a) 2D embedding; (b) 3D embedding.

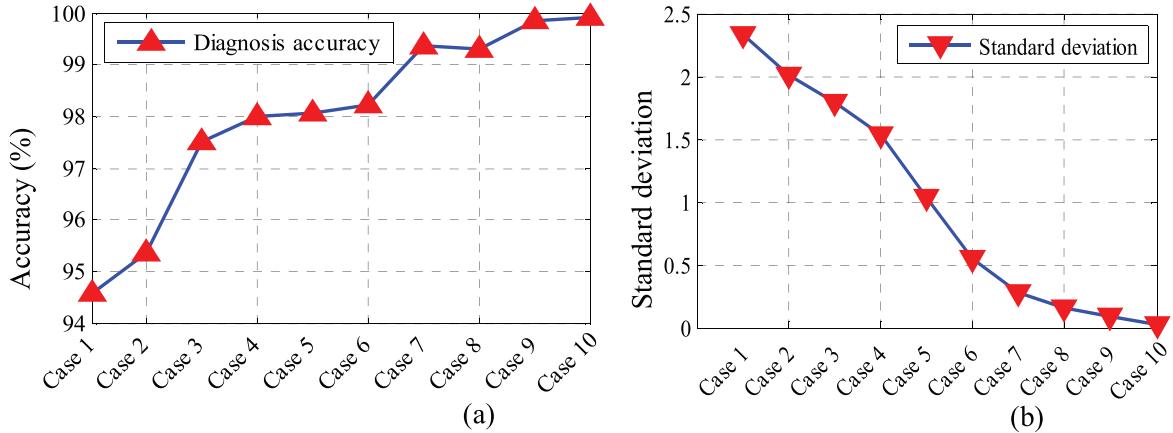


**Figure 10.** The visualization of features learned in P4 layer. (a) 2D embedding; (b) 3D embedding.

**Table 6.** Results of experiment 3 with the proposed method.

Case	Normal samples	Inner/outer race fault sample	Roller fault samples	Total samples	Fault percentage (%)	Diagnosis result (%)
Case 1	600	100/100	22	822	27.01	94.57 ± 2.33
Case 2	600	200/200	44	1044	42.53	95.36 ± 2.01
Case 3	600	300/300	66	1266	52.61	97.49 ± 1.79
Case 4	600	400/400	88	1488	59.68	97.99 ± 1.53
Case 5	600	500/500	112	1712	64.95	98.07 ± 1.03
Case 6	600	600/600	133	1933	68.98	98.23 ± 0.55
Case 7	600	700/700	156	2156	72.17	99.35 ± 0.28
Case 8	600	800/800	178	2378	74.77	99.29 ± 0.15
Case 9	600	900/900	200	2600	76.92	99.84 ± 0.09
Case 10	600	1000/1000	222	2822	78.74	99.93 ± 0.02

Note: the format of the diagnosis result is average accuracy ± standard deviation.



**Figure 11.** Results of experiment 3 with the proposed method. (a) Average identification accuracy; (b) standard deviation.

frequency band signals are divided into samples for fault diagnosis.

The key point is as follows: the goal of this paper is to develop an effective fault diagnosis method and to get rid of manual feature extraction. Therefore, we focus on **experiment 1**, i.e. the raw vibration data are directly used as input features.

As shown in table 2, eight working conditions are considered for **experiment 1** and **experiment 2**. The faults range in diameter from 0.1778 mm to 0.5334 mm. The rotating speed is 1797 rpm. Each working condition has 200 samples and each sample contains 400 sampling data points. Specifically, 50% of samples are randomly selected to train the deep CNN model, and the rest of them are used to test the performance.

#### 4.2. Diagnosis and results

In **experiment 1**, the four methods are explained in detail as follows:

(1) **The proposed method:** the maximum iteration is 300.

The three key parameters are determined by PSO, the C1 layer has 10 kernels, the C3 layer has 13 kernels and the learning rate is 0.8482. The parameters of PSO are shown in table 3 and the iteration process is shown in figure 5.

- (2) **Deep CNN:** C1 layer has 6 kernels, C3 layer has 12 kernels and the learning rate is 0.5. The maximum iteration is 300.
- (3) **ANN:** the number of units in the hidden layer is 200, the learning rate and momentum are 0.1 and 0.05. The maximum iteration is 500. All the parameters are determined by experience and repeated experiments.
- (4) **SVM:** the RBF kernel is applied, the penalty factor is 3 and the radius of the kernel function is 1. All the parameters are determined by experience and repeated experiments.

The average identification accuracies of ten trials in **experiment 1** and **experiment 2** are shown in table 4. Figure 6 shows the accuracy of each trial. The average testing time of the four methods in **experiment 1** and **experiment 2** is shown in table 5 (Core i5, 8 GB memory). The confusion matrixes with the proposed method in **experiment 1** and **experiment 2** are shown in figure 7. The ordinate axis of the confusion matrix denotes the actual label of samples, and the horizontal axis denotes the predicted label of samples. Each confusion matrix is  $8 \times 8$  in size. Considered entries in the matrix are denoted as  $C_{mn}$ , which denotes the ratio of samples actually from the  $m$ th working condition and predicted as the  $n$ th working condition. The ratio values from 0 to 1 are denoted with different colors, as shown in the color bar. Therefore,  $C_{ii}$

**Table 7.** Comparison with several other methods.

Method	Signal preprocessing or manual feature extraction	Considered working conditions	Diagnosis accuracy (%)
The proposed method (raw data)	Without any signal preprocessing and manual feature design	8	92.84
The proposed method (preprocessed data)	Preprocessed with wavelet ( <i>db4</i> ) and no manual feature extraction	8	99.71
[12]	Multivariable feature extraction	7	96.22
[38]	19 features	4	95.8
[39]	Wavelet leaders multifractal features	10	88.9

**Figure 12.** Electrical locomotive rolling bearing test rig.

( $i = 1, 2 \dots 8$ ) denotes the identification accuracy of samples from the  $i$ th working condition.

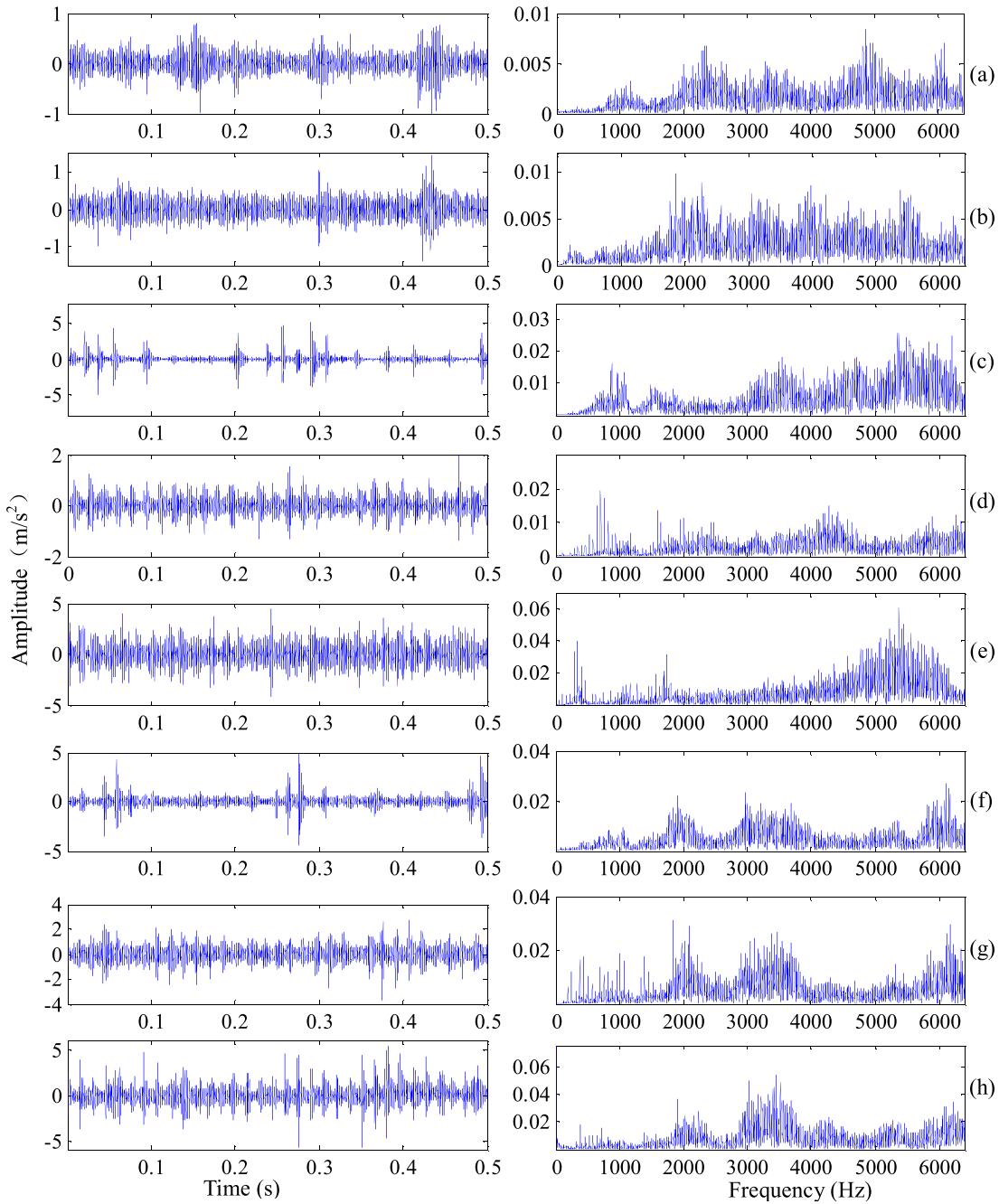
The results show that: (1) the proposed method and the deep CNN have better performance than the SVM method and ANN method. The main reason is that the proposed method and deep CNN are deep learning models, and both of them adopt deep architecture. As a result, they can automatically learn the internal suitable fault features from input data and provide accurate diagnosis results, i.e. no manual feature extraction is needed. (2) Although the proposed method and deep CNN can identify the rolling bearing faults, the proposed method has higher classification accuracy and more robust performance than deep CNN. The reason is that the proposed method is an adaptive deep CNN and has higher adaptability than deep CNN. When applied to diagnose rolling bearing faults, the proposed method can automatically determine the main parameters of deep CNN and adapt to the characteristics of vibration signals. (3) In **experiment 2**, the proposed method achieved 99.71% diagnosis accuracy. The key point is that no manual feature extraction is performed. It is worth noticing that the proposed method can provide sufficient performance just with simple signal preprocessing in this case study, which confirms the superiority of the proposed method.

As shown in table 5, there is an obvious shortcoming of the proposed method: the computation time with the proposed method is much more than that of ANN and SVM. Take **experiment 1** as an example: the testing time of the proposed method is as much as four times that of ANN and six times that of SVM. The reason is that the proposed method and deep CNN are based on deep learning models, which contain more

**Table 8.** The description of rolling bearing vibration signals for **experiment 4**.

Rolling bearing condition	Rotating speed (rpm)	Size of training/testing samples	Label
a. Normal Condition	490	150/100	1
b. Slight rub fault in the inner race	498	150/100	2
c. Slight rub fault in the roller	531	150/100	3
d. Slight rub fault in the outer race	490	150/100	4
e. Serious flaking fault in the outer race	480	150/100	5
f. Compound faults (outer race and roller)	521	150/100	6
g. Compound faults (inner and outer races)	525	150/100	7
h. Compound faults (inner and outer races and roller)	549	150/100	8

hidden layers than ANN and SVM. Therefore, when samples are input to the proposed method and deep CNN, the computation is more complex and need more time, which may limit the application of the proposed method. This paper aims to develop a new method to get rid of manual feature extraction, and to reduce the dependence on manual experience and domain expertise. Therefore, we focus more on a theoretical investigation and are concerned less with computation efficiency. Nevertheless, with the development of the deep learning theory and hardware, it will be more convenient



**Figure 13.** Rolling bearing vibration signals and their spectra for each health condition.

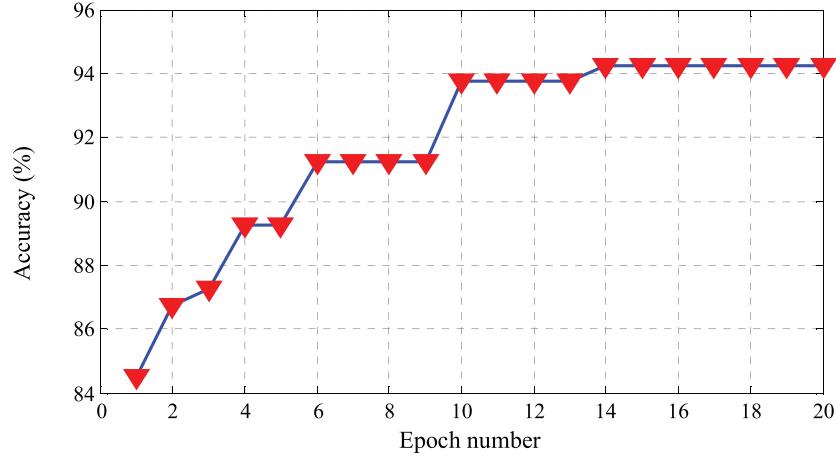
to build deep models, and the computation will thus be more efficient.

#### 4.3. Visualization of hierarchical feature learning process

In order to evaluate the feature learning process of the proposed method, the high-dimensional features learned in each stage are visualized with the method illustrated in section 3.3. To take **experiment 1** as an example: the features in **experiment 1** are visualized in figures 8–10, in which every point represents a sample and the axes represent the t-SNE dimensions. The 2D embedding figures use the first two dimensions of t-SNE features to visualize

**Table 9.** Parameters of PSO in experiment 4.

Description	Symbol	Value
Population size of PSO	$N$	20
Iteration number	$N$	25
The dimensionality of particles	—	3
Acceleration constants	$c_1, c_2$	$c_1 = 2, c_2 = 2$
Inertia weight of PSO	$\omega$	0.5
Learning rate of deep CNN	$\eta$	0.3944 (output from PSO)
Kernels in C1 layer	$m_1$	5 (output from PSO)
Kernels in C3 layer	$m_2$	10 (output from PSO)



**Figure 14.** The iteration process of PSO in **experiment 4**.

each sample. Similarly, the 3D embedding figures use the first three dimensions of t-SNE features to visualize each sample. Every point represents a sample and the axes represent the t-SNE dimensions. The axis values are the computation values of the t-SNE features.

As shown in figure 8, the raw data featured in the input layer correspond to 8 types of rolling bearing health conditions scattered in disorder, and it is impossible to identify them. Figure 9 shows that the features learned in the P2 layer are easier to identify than the raw data features. However, it is difficult to completely distinguish them. Figure 10 shows that the features learned in the P4 layer can be well identified. The results show that the low-level raw signals featured in the input layer are transformed into high-level features layer by layer, which can promote the precision and robustness of classification. Therefore, the results show that the proposed method has good feature learning ability and can adaptively learn the essential fault features from the raw vibration signals.

#### 4.4. Influence of other percentages of faults

In **experiment 1** and **experiment 2**, the percentage is the same for the samples from each working condition. However, much previous research shows that about 90% of rolling bearing faults occurs in the inner race and outer race [37]. Therefore, to study the influence of other percentages of faults, **experiment 3** is carried out according to this fault ratio.

As shown in table 6, **experiment 3** contains 10 cases and the details are follows: (1) in each case, the number of normal samples is always set to 300; around 90% of faulted samples consist of inner race faults and outer race faults, and the rest of the faults are roller faults; (2) for each condition, half of the samples are randomly selected as training samples and the rest are used as testing samples.

Two key points of **experiment 3** need to be pointed out: (1) the goal of this section is to verify the robustness of the proposed method to the other percentages of faults, thus, we directly use the raw vibration data as input features, i.e. the input samples of **experiment 3** are without any signal

**Table 10.** The accuracy of experiment identification.

Methods	Identification accuracy (%)	Average testing time (s)
The proposed method	94.18	1.205
Deep CNN	89.09	1.297
SVM	74.46	0.256
ANN	42.87	0.359

preprocessing and manual feature extraction; (2) we only focus on the fault types, i.e. inner race fault, outer race fault and roller fault, and fault diameters are not considered. Therefore, **experiment 3** is a 4-class pattern recognition problem.

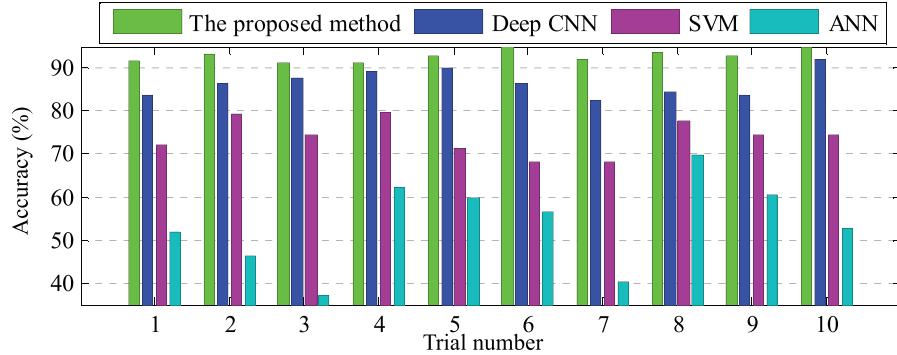
The proposed method is applied to analyze the samples in each case. The average identification accuracy and standard deviation of each case are shown in figure 11.

The results show that: (1) accuracies for case 1 to case 10 are all over 94%, which confirms that the proposed method is robust to the other percentages of faults; (2) with the number of faulted samples increasing, the accuracies increase to 99.93% and the standard deviation values decrease to 0.02%. This is because the proposed method is based on deep learning, which needs massive samples to enhance classification performance. Therefore, with more samples the accuracies increase and the standard deviation value decreases. Consequently, the proposed method is robust for the other percentages of faults and massive samples are useful for enhancing the performance of the proposed method. Certainly, more research should be carried out to study the robustness of the proposed method.

#### 4.5. Comparison with several other methods

To show the advantages of the proposed method, several methods using the same data are compared as shown in table 7.

The results of **experiment 1** and **experiment 2** show that the proposed method can achieve 92.84% identification accuracy without any signal preprocessing. When the vibration data is simply preprocessed, the proposed method can achieve 99.71% identification accuracy without any manual feature extraction.



**Figure 15.** Diagnosis results of the ten trials using the four methods.

In [12], Zhang *et al* proposed a method based on multivariable ensemble-based incremental SVM, which needs multivariable feature extraction. In [38], the author extracted 19 features and achieved 95.8% diagnosis accuracy. In [39], the author applied wavelet leaders multifractal features to diagnose rolling bearing fault diagnosis. It is worth noticing that the methods in [12, 38, 39] need manual feature extraction or feature selection, which greatly depend on manual experience and domain expertise.

We can conclude that, comparing with some traditional fault diagnosis methods, no manual feature extraction is needed for the proposed method. This can greatly reduce the dependence on manual experience and domain expertise.

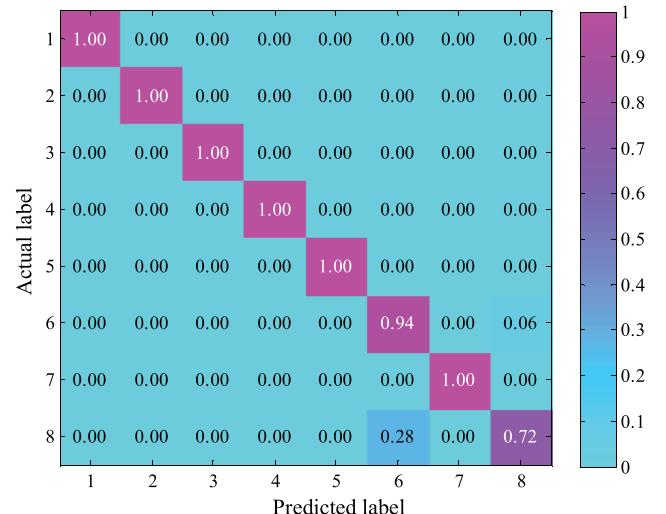
## 5. Engineering application

In this case, the proposed method is applied to analyze rolling bearing vibration signals collected from an electrical locomotive setup.

### 5.1. Data description

Because of its high security and good cost performance, the electrical locomotive is one of the most important vehicles in modern society. Rolling bearings are key parts in electrical locomotives and usually work in heavy-load and varying working conditions. As a result, rolling bearings are susceptible to failure, which can greatly influence the security of electrical locomotives. In this section, the proposed method is applied to electrical locomotive rolling bearing fault diagnosis. Figure 12 shows the electrical locomotive rolling bearing test rig. The vibration signals are collected with accelerometers, and the sampling frequency is 12.8 kHz.

**Experiment 4** is carried out to further verify the effectiveness of the proposed method. The goal of this paper is to reduce the dependence on manual experience and domain expertise, thus the samples in **experiment 4** consist of raw vibration data without any signal preprocessing and manual feature extraction. As shown in table 8, eight working conditions are considered. Time domain figures of the vibration signals and their spectrum are shown in figure 13. Each signal is divided into 250 samples and each sample contains 400 data points.



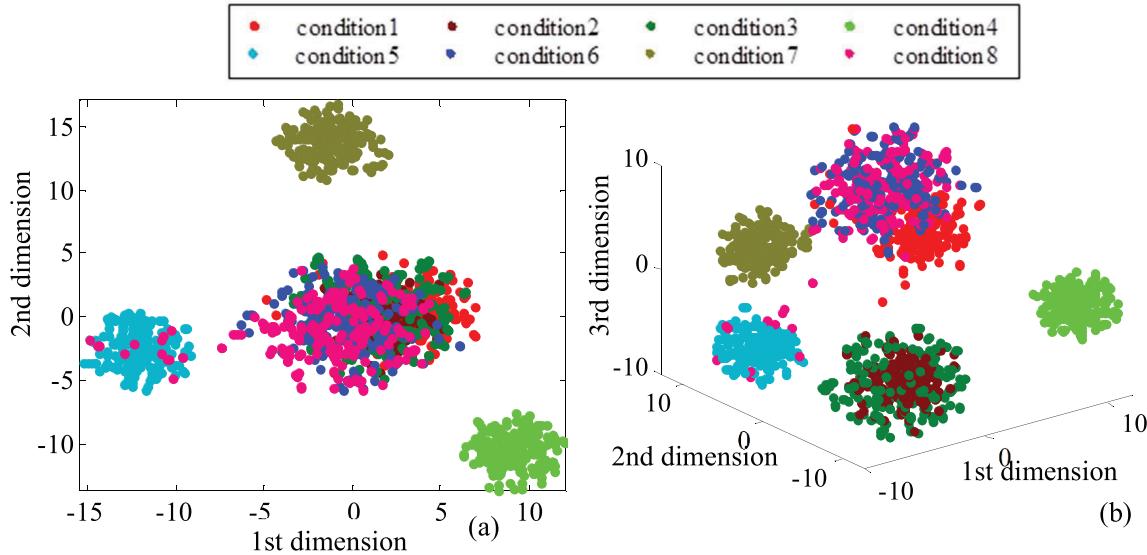
**Figure 16.** The confusion matrix of testing samples using the proposed method.

### 5.2. Results and analysis

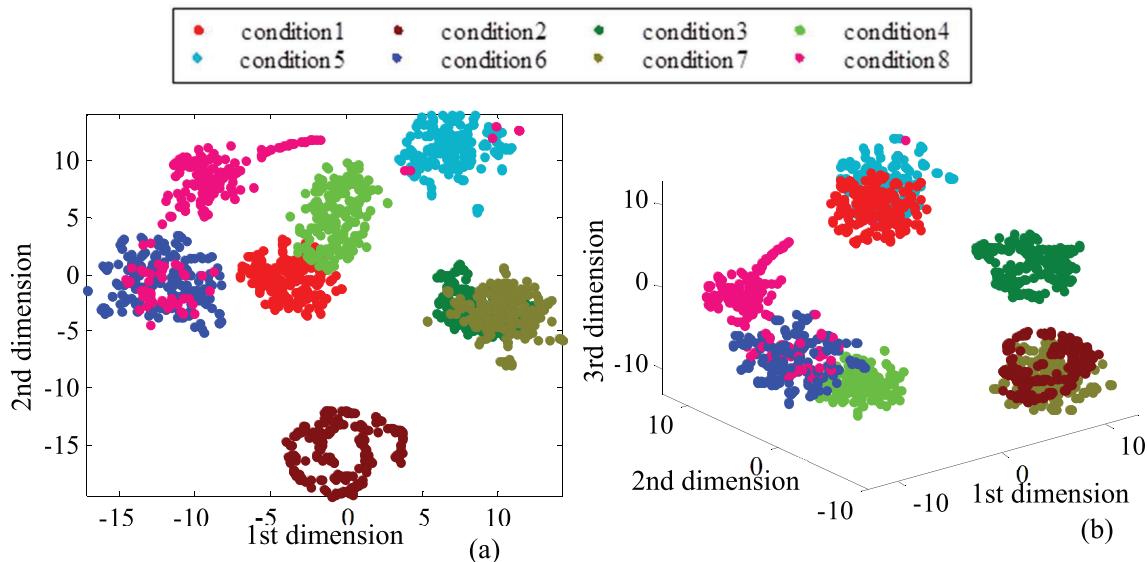
Four methods implementing the proposed method are applied to analyze the same data in **experiment 4**. The detailed explanations of the four methods are as follows:

- **The proposed method:** The maximum iteration is 200. The three key parameters are determined by PSO. C1 layer has 5 kernels, C3 layer has 10 kernels and the learning rate is 0.3944. The parameters of PSO are shown in table 9 and the iteration process is shown in figure 14.
- **Deep CNN:** C1 layer has 6 kernels, C3 layer has 12 kernels and the learning rate is 0.5. The maximum iteration is 200.
- **ANN:** the hidden layer has 200 units, the learning rate and momentum are 0.05 and 0.1, the maximum iteration is 500. All the parameters are determined by experience and repeated experiments.
- **SVM:** the RBF kernel is applied, the penalty factor is 10 and the radius of the kernel function is 0.03. All the parameters are determined by experience and repeated experiments.

The average identification accuracy of ten trials is shown in table 10, and figure 15 shows the accuracy of ten trials. The



**Figure 17.** The visualization of raw data features in the input layer. (a) 2D embedding; (b) 3D embedding.



**Figure 18.** The visualization of features learned in the P2 layer. (a) 2D embedding; (b) 3D embedding.

confusion matrix of testing samples for the tenth trial using the proposed method is shown in figure 16.

As shown in table 10, the proposed method achieves 94.18% identification accuracy without any signal preprocessing and manual feature extraction, which is better than deep CNN, SVM and ANN. Figure 16 shows that the proposed method can completely identify all the working conditions, besides the 6th and 8th working conditions. The reason is that the 6th and the 8th working conditions are compound faults that have complex vibration signals and are very difficult to identify.

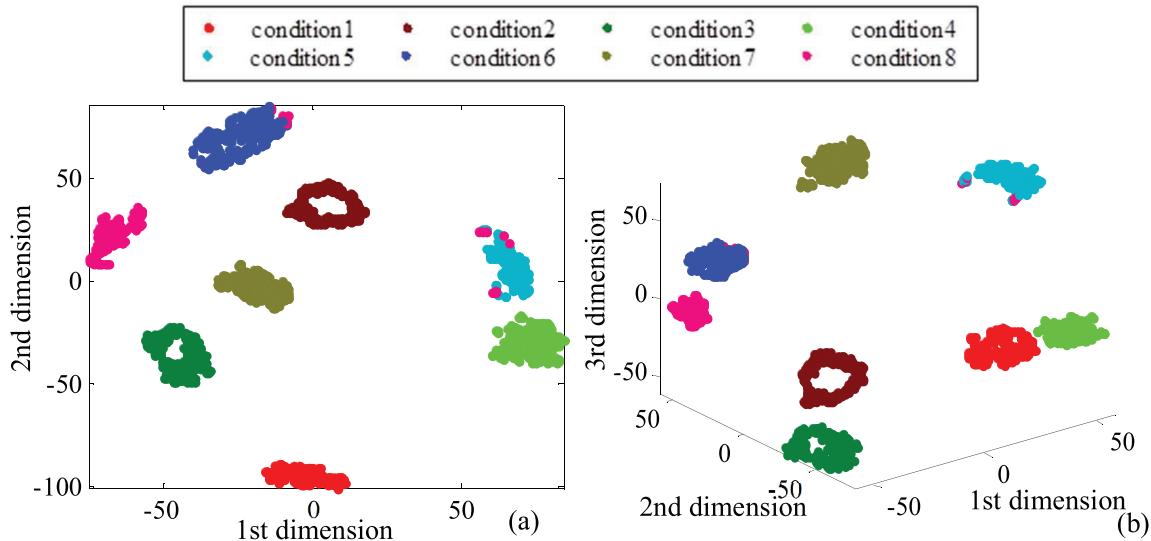
We can conclude that: (1) the proposed method and the deep CNN have higher precision than the ANN method and SVM method. This further confirms the superiority of deep learning methods over shallow learning methods, such as the ANN method and SVM method; (2) the proposed method has higher and more robust identification accuracy than deep CNN, which further confirms that a deep CNN with fixed key

parameters is insufficient to adaptively deal with different rolling bearing signals. Therefore, the proposed method has great advantage on adaptability; (3) more research is needed to improve the computation efficiency and performance of the proposed method when applied to identify the compound faults.

### 5.3. Visualization of hierarchical feature learning process

In this section, the high-dimensional fault features are visualized with the t-SNE method [35] similar to in section 4. The results are shown in figures 17–19.

As shown in figure 17, it is completely impossible to identify the raw data features in input layer. Figure 18 shows that the features learned in P2 layer corresponding to several fault categories are overlapped with each other. Figure 19 shows that the features learned in P4 layer are well clustered category to category, which are easy to identify. The results



**Figure 19.** The visualization of features learned in the P4 layer. (a) 2D embedding; (b) 3D embedding.

**Table 11.** The detailed description of methods in **experiment 5**.

Methods	Detailed description
The proposed method (Deep learning model)	The fully-connected layer has 100 processing units and iteration number is 300. Other parameters are determined with PSO
Standard DBN (Deep learning model)	The architecture is 400-200-300-200-9. The learning rate and momentum are selected as 0.1 and 0.5, respectively
Standard DAE (Deep learning model)	The architecture is 400-250-350-250-9. The learning rate and momentum are selected as 0.3 and 0.1, respectively

**Table 12.** The results of **experiment 5**.

Methods	Average accuracy (%)	
	Training samples	Testing samples
The proposed method	99.89	94.18
Standard DAE	99.03	86.27
Standard DBN	98.85	85.66

further confirm that the proposed method has good feature learning ability and can adaptively learn the essential fault features from raw vibration signals.

#### 5.4. Diagnosis performance comparison of different deep learning models

Our research group has proposed fault diagnosis methods based on DBN [24] and DAE [25]. In order to compare the diagnostic performance of different deep learning models, **experiment 5** is carried out in this section. As shown in table 11, standard DBN and DAE are applied to analyze the same data.

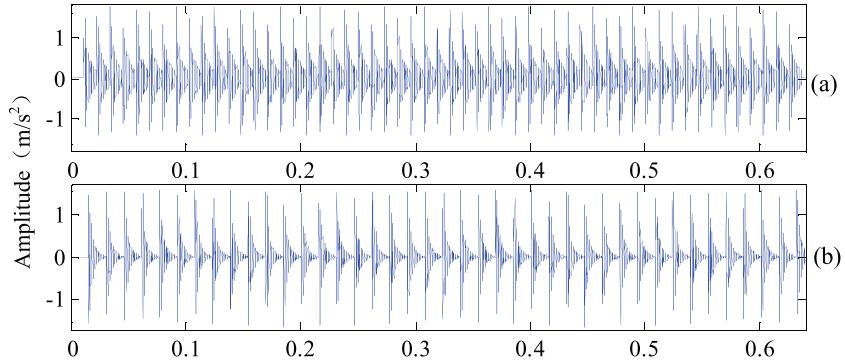
In **experiment 5**, the three methods are applied to analyze the data in **experiment 4**, which is shown in table 8. The average accuracies of ten repeated trials are shown in table 12. The diagnosis results on training samples and testing

**Table 13.** The results of **experiment 6**.

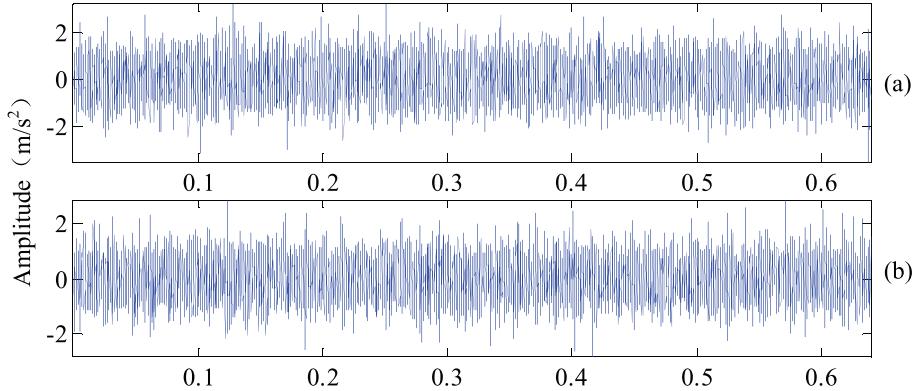
Case	SNR (%)	Average accuracy (%)			
		The proposed method	Deep CNN	SVM	ANN
Case 1	0	100	100	100	75
Case 2	20	100	100	86.67	68.33
Case 3	40	100	78.75	65	50
Case 4	50	91.25	71.67	58.33	46.67
Case 5	60	88.33	62.50	55	40
Case 6	80	81.67	56.67	53.33	31.67

samples are given in the second column and the third column, respectively.

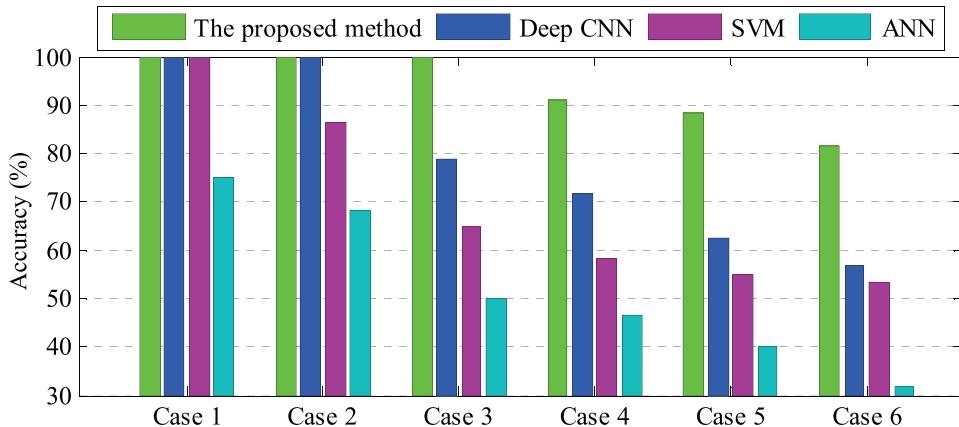
As shown in table 12, the accuracies of the three methods for training samples are nearly the same, while the accuracies for testing samples vary greatly. Therefore, the proposed method has much better testing performance than the other two deep learning methods. This is mainly because the proposed method is an adaptive deep CNN and can automatically adapt to the characteristics of different vibration signals, while the architectures of standard DAE and standard DBN are designed manually and cannot automatically adapt to the signals. Thus, the proposed method has better adaptability and higher diagnosis precision. It will be interesting to carry out more research to investigate the performance of deep learning models when applied to fault diagnosis.



**Figure 20.** Simulation signals without noise. (a) Inner race fault; (b) outer race fault.



**Figure 21.** Simulation signals combined with noise (SNR = 40%). (a) Inner race fault; (b) outer race fault.



**Figure 22.** The results of experiment 6.

### 5.5. Influence of signal noise

In order to research the robustness of the proposed method to signal noise, **experiment 6** is carried in this section. The vibration signals of the inner race fault and outer race fault are simulated according to [24]. As shown in table 13, the signal-to-noise ratios (SNR) for five cases vary from 20% to 80%. The simulation signals without noise are shown in figure 20 and the signals combined with noise (SNR = 40%) are shown in figure 21. Unlike [24], each signal in **experiment 6** contains 32 000 data points and other parameters are set as same as [24].

The signals are directly used as input features without any signal preprocessing and manual feature extraction. Each signal is divided into 80 samples and each sample contains

400 data points. Half of the samples are randomly selected as training samples and the rest are used to test the performance.

For comparison, the proposed method, deep CNN, SVM and ANN are applied to analyze the same data. The average identification accuracies of ten repeated trials for the four methods are shown in table 13 and figure 22.

The results show that the proposed method has the best performance among the four methods. In case 4 to case 6, the accuracy of the proposed method decreases to 81% with the increase of SNR, but the proposed method still performs better than other methods. This further confirms the advantage of the proposed method on adaptability. Meanwhile, the results show that the proposed method may perform worse when the analyzed signals contain heavy noise, and therefore

more research is needed to improve the performance of the proposed method when applied to diagnosis problems with heavy noise.

## 6. Conclusion

In this paper, a novel adaptive deep CNN is proposed for rolling bearing fault diagnosis. Firstly, with a hierarchy feature learning architecture, the proposed method can automatically learn the essential fault features from the input data layer by layer, which can eliminate the need for manual feature extraction. Secondly, when applied to different rolling bearing fault diagnosis problems, the proposed method can automatically determine the main parameters of the deep CNN model to provide sufficient diagnosis results.

The proposed method is applied to diagnose rolling bearing faults. The results show that the proposed method is more effective and robust than other intelligent methods. Compared with other methods, the proposed method can achieve desirable diagnosis accuracy with a minimum of signal preprocessing, which is significant to reduce the dependence on manual experience and domain expertise. Meanwhile, more work is needed to improve the computation efficiency and the robustness of the proposed method to heavy noise. Our research group will continue to focus on this topic in the future.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 51475368) and Shanghai Engineering Research Center of Civil Aircraft Health Monitoring (no. GCZX-2015-02) and Seed Foundation of Innovation and Creation for Graduate Students in Northwestern Polytechnical University (no. Z2017064).

## References

- [1] Jiang H, Xia Y and Wang X 2013 Rolling bearing fault detection using an adaptive lifting multiwavelet packet with a dimension spectrum *Meas. Sci. Technol.* **24** 125002
- [2] Rauber T W, de Assis Boldt F and Varejão F M 2015 Heterogeneous feature models and feature selection applied to bearing fault diagnosis *IEEE Trans. Ind. Electron.* **62** 637–46
- [3] Zhang S et al 2016 Bearing fault diagnosis based on variational mode decomposition and total variation denoising *Meas. Sci. Technol.* **27** 075101
- [4] Wang H et al 2016 Compressed sensing of roller bearing fault based on multiple down-sampling strategy *Meas. Sci. Technol.* **27** 025009
- [5] Zhao M, Tang B and Tan Q 2015 Fault diagnosis of rolling element bearing based on S transform and gray level co-occurrence matrix *Meas. Sci. Technol.* **26** 085008
- [6] Jiang H, Li C and Li H 2013 An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis *Mech. Syst. Signal Process.* **36** 225–39
- [7] Wang Y et al 2016 Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: a review with applications *Mech. Syst. Signal Process.* **66–7** 679–98
- [8] He Y, Huang J and Zhang B 2012 Approximate entropy as a nonlinear feature parameter for fault diagnosis in rotating machinery *Meas. Sci. Technol.* **23** 045603
- [9] Hu B and Li B 2016 A new multiscale noise tuning stochastic resonance for enhanced fault diagnosis in wind turbine drivetrains *Meas. Sci. Technol.* **27** 025017
- [10] He W et al 2015 Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform *Mech. Syst. Signal Process.* **54–5** 457–80
- [11] Liu R et al 2016 Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis *Mech. Syst. Signal Process.* **75** 345–70
- [12] Zhang X, Wang B and Chen X 2015 Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine *Knowl.-Based Syst.* **89** 56–85
- [13] Qu J, Zhang Z and Gong T 2016 A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion *Neurocomputing* **171** 837–53
- [14] Hong K et al 2015 A method of real-time fault diagnosis for power transformers based on vibration analysis *Meas. Sci. Technol.* **26** 115011
- [15] Bin G F et al 2012 Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network *Mech. Syst. Signal Process.* **27** 696–711
- [16] Lashkari N, Poshtan J and Azgomi H F 2015 Simulative and experimental investigation on stator winding turn and unbalanced supply voltage fault diagnosis in induction motors using artificial neural networks *ISA Trans.* **59** 334–42
- [17] Hinton G, Osindero S and Teh Y 2006 A fast learning algorithm for deep belief nets *Neural Comput.* **18** 1527–54
- [18] LeCun Y, Bengio Y and Hinton G 2015 Review: deep learning *Nature* **521** 436–44
- [19] Yu D and Deng L 2011 Deep learning and its applications to signal and information processing *IEEE Signal Process. Mag.* **28** 145–54
- [20] Lore K G, Akintayo A and Sarkar S 2017 LLNet: a deep autoencoder approach to natural low-light image enhancement *Pattern Recognit.* **61** 650–62
- [21] Jia F et al 2016 Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data *Mech. Syst. Signal Process.* **72–3** 303–15
- [22] Lu C et al 2017 Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification *Signal Process.* **130** 377–88
- [23] Gan M, Wang C and Zhu C 2016 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.* **72–3** 92–104
- [24] Shao H et al 2015 Rolling bearing fault diagnosis using an optimization deep belief network *Meas. Sci. Technol.* **26** 115002
- [25] Shao H et al 2017 An enhancement deep feature fusion method for rotating machinery fault diagnosis *Knowl.-Based Syst.* **119** 200–20
- [26] Sun J, Xiao Z and Xie Y 2017 Automatic multi-fault recognition in TFDS based on convolutional neural network *Neurocomputing* **222** 127–36
- [27] Janssens O et al 2016 Convolutional neural network based fault detection for rotating machinery *J. Sound Vib.* **377** 331–45
- [28] Ji S et al 2013 3D convolutional neural networks for human action recognition *IEEE Trans. Pattern Anal. Mach. Intell.* **35** 221–31

- [29] He K *et al* 2015 Spatial pyramid pooling in deep convolutional networks for visual recognition *IEEE Trans. Pattern Anal. Mach. Intell.* **37** 1904–16
- [30] Yang W *et al* 2016 DropSample: a new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition *Pattern Recognit.* **58** 190–203
- [31] Shen W *et al* 2017 Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification *Pattern Recognit.* **61** 663–73
- [32] Ijjina E P and Mohan C K 2016 Hybrid deep neural network model for human action recognition *Appl. Soft Comput.* **46** 936–52
- [33] Zhang B, Xu C L and Wang S M 2016 An inverse method for flue gas shielded metal surface temperature measurement based on infrared radiation *Meas. Sci. Technol.* **27** 074002
- [34] Zhang E *et al* 2016 Sound quality prediction of vehicle interior noise and mathematical modeling using a back propagation neural network (BPNN) based on particle swarm optimization (PSO) *Meas. Sci. Technol.* **27** 015801
- [35] van der Maaten L and Hinton G 2008 Visualizing signals using t-SNE *J. Mach. Learn. Res.* **9** 2579–605
- [36] Lou X and Loparo K A 2004 Bearing fault diagnosis based on wavelet transform and fuzzy inference *Mech. Syst. Signal Process.* **18** 1077–95
- [37] Bently D and Bently Nevada Co. 1989 Predictive maintenance through the monitoring and diagnostics of rolling element bearings *Appl. Note* **44** 2–8
- [38] Li W, Zhang S and He G 2013 Semisupervised distance-preserving self-organizing map for machine-defect detection and classification *IEEE Trans. Instrum. Meas.* **62** 869–79
- [39] Du W *et al* 2014 Wavelet leaders multifractal features based fault diagnosis of rotating mechanism *Mech. Syst. Signal Process.* **43** 57–75