A novel multi-mode fault classification method based on deep learning

Funa Zhou1, Yulin Gao1,\*，Chenglin Wen2,\*

1.School of Computer and Information Engineering, Henan University, Kaifeng, China

2. School of Automation, Hangzhou Dianzi University, Hangzhou, China

[gaoyulinhn@163.com, wencl@hdu.edu.cn (\* Corresponding](mailto:%20gaoyulinhn@163.com,%20wencl@hdu.edu.cn%20(*%20Corresponding) Author)

**Abstract**: Due to the problem of load varying or environment changing, machinery equipment often operates in multi-mode. The data feature involved in the observation often varies with mode changing. Mode partition is a fundamental step before fault classification. This paper proposes a multi-mode classification method based on deep learning by constructing a hierarchical DNN model with the first hierarchical specially devised for the purpose of mode partition. In the second hierarchical, different DNN classification models are constructed for each mode to get more accurate fault classification result. For the purpose of providing helpful information for predictive maintenance, an additional DNN is constructed in the third hierarchical to further classify a certain fault in a given mode into several classes with different fault severity. The application to multi-mode fault classification of rolling bearing fault shows the effectiveness of the proposed method.

**Key words**: deep neural network; fault classification; multi-mode; machine learning; AutoEncoder

# 1. Introduction

Rolling bearing is a very pivotal component in rotating machines, which are widely used in large-scale automated industrial equipment. Mechanical failure caused by rolling bearings may cause abnormality of the rotating machinery system, resulting in huge economic losses, and even cause some unnecessary casualties [1-5]. Therefore, timely and precisely classification is critical for bearing monitoring.

The methods for mechanical equipment fault classification can be divided into qualitative model based method, quantitative model based method and data-driven based method [6, 7]. Qualitative model and quantitative model based methods require precise mathematical model or a large amount of expert knowledge of the system, which will inevitably limit its application in fault classification field. In the recent two decades, data-driven method is widely used in fault detection of complex system. Instead of much more prior knowledge, data-driven approach can detect fault only through the measured data of the complex system [8-11]. The most common used data-driven fault classification methods are statistical feature extraction based methods and machine learning based methods. However, the method based on statistical feature extraction can only realize fault detection and it is unable to realize fault classification. For fault classification, we had better use machine learning method such as Support Vector Machine (SVM), Artificial Neural Network (ANN), and so on.

In the field of mechanical system fault classification, because of the sensitivity of vibration spectrum to equipment failure, vibration signals is usually used as the data source for fault classification of mechanical equipment. Due to mechanical equipment’s characteristics of non-stable, non-linear, large-scale, high-dimensional, and noise polluted, it is usually very difficult for precise fault feature extraction which is the most critical factor of the accuracy of mechanical equipment monitoring [12,13]. Some scholars have put forward some feature extraction methods that combine signal processing technology with machine learning method for fault classification of mechanical equipment. Achmad Widodo et al. extract the frequency-domain feature as the data source of SVM to detect the machinery fault [14]. When the number of samples is small and the signals are non-stationary, Cheng et al.proposed a bearing fault classification method by combining SVM and Empirical Mode Decomposition (EMD) [10]. Hu et al.extracted the energy of each wavelet packet transform (WPT) node as the pre-extracted feature to develop a combined WPT-SVM based method for more accurate bearing fault classification [15]. Wang et al. also used WPT to extract non-stationary characteristics of the bearing’s vibration signal as the pre-extracted feature of ANN [16]. The method uses the nonlinear learning classification ability and self-organizing ability of ANN to classify and diagnose bearing faults. Yang et al. proposed a method combining expert system and back propagation neural network (BPNN) [17]. This method makes full use of the advantages of expert system and ANN to successfully detect the bearing failure. Since bearing vibration signals are susceptible to Gaussian noise, Jiang et al. used high level statistics as the feature vector of BPNN to improve the performance of BPNN in bearing fault classification [18]. However, SVM and BPNN share the shortcomings of shallow learning method: SVM is a two-classifier, and it is inefficient in multi-classification especially in the case when the sample number of observation is very large. Selecting the appropriate kernel function and scale parameter usually need a wealth of experience. ANN also suffers many defects, such as: (1) ANN has a slow convergence rate and can easily converge to the local optimum (2) ANN is ineffective in feature learning of complex non-linear data and usually results in poor classification accuracy. In summary, SVM and BPNN as the shallow learning methods could not well extract the data feature involved in the high-dimensional unsteady data [19]. With the load varying, bearing can work in different steady state, which is called “multi-mode” phenomenon. Current research work on machine learning based classification didn’t take multi-mode problem into account.

For multi-mode process, the data feature of each mode is different [20], but current research on bearing fault classification usually regard it as a single mode for simplicity of data processing which will result inaccurate classification result since feature extracted is inaccurate [21-23]. Therefore, mode partition should be implemented before fault feature extraction of a separate mode for accurate feature extraction. Zhang et al. proposed an improved k-means clustering algorithm based on existing modal partition method [20]. Song et al. studied the issue to distinguish stability mode from transition mode without the number of modes known in advance [24]. Zhao et al. separated multiple modalities according to the diversity analysis in operational phases, and established online monitoring method along multiple batch directions [25]. Zhang et al. used modal subspace separation method to deal with multi-mode monitoring problems [26]. By using various characteristics of the subspace, different mode can well be separated, which can provide chance for more accurate multi-mode fault classification.

Unfortunately, mode partition and corresponding fault monitoring method for a certain multi-mode processes is only specially developed for a specific industrial process [20, 24-27]. It is required to develop a more universal method. Deep learning is a promising ubiquitous feature extraction tool which has attracted widely attention by scholars from various fields [21, 28-30]. Comparing to shallow learning, deep learning can well process the feature extraction and the issue of non-linear big data by constructing a deep network [31, 32]. Through the unsupervised layer-by-layer greedy training algorithm and BP-based global parameter fine-tuning, Deep Neural Network (DNN) can not only avoid the local optimization problem, but also solved the problem of limitation in number of labeled samples and the limitation in generalization ability. Deep learning method was firstly proposed by Hinton et al. in 2006 [22]. In view of its excellent feature extraction capabilities, it also attracts the attention of fault classification experts. Lu et al. successfully used the better feature extraction ability of deep neural network to diagnose the bearing fault [33]. The proposed method overcomes the shortcomings that the traditional feature extraction method could not discover the unknown type fault timely and effectively. Lei et al. used deep neural network to monitor the failure of bearings [34]. Gan et al. proposed a fault classification method based on hierarchical neural network [11]. By constructing a two-layer neural network, the method not only could locate the position of bearing fault, but also effectively mine the fault size of the bearing in the same position. Deep learning, as one of the most popular machine learning method, has brought a subversive revolution to the field of artificial intelligence. However, application about the deep learning is still in infancy, during the application process, there are also many issues demand to improve. For example, the data in Ref. [11] are derived from a single mode, without considering the multi-mode observation caused by load varying problem. Therefore, it cannot fully extract the fault feature involved in the observation of different mode which is essential for the accuracy of multi-mode fault classification.

To solve the above mentioned problems, this paper presents a multi-mode fault classification method based on deep learning. First, a DNN model is constructed, the trained network is used to mode partition; Then, a new set of DNN are constructed for observation data of each mode, and the trained networks are used to determine which component is failure to implement fault location recognition; Finally, for a certain fault in a given mode, another DNN is constructed for classify those observation data with different fault size.

The remainder of this paper follows as: Section 2 overviews the theory of deep learning. Section 3 develops a multi-mode fault classification method based on DNN by hierarchically constructing DNN models with different purpose. In Section 4, effectiveness of the proposed multi-mode fault classification method is demonstrated by experiments analysis. Section 5 concludes this paper.

# 2. Theory of deep learning

Deep learning is a method based on unsupervised feature learning. We use deep learning theory to construct DNN. DNN training process consists of two steps: (1) Using the unsupervised learning algorithm to pre-train the network layer by layer, which is helpful for DNN to efficiently mine features from raw data; (2) Back propagation algorithm is used to fine-tune the parameters of the whole network, optimizing the performance of DNN to mine raw feature. In this paper, DNN is pre-trained by multi- stacking AutoEncoder(AE).

## 2.1. AutoEncoder

AutoEncoder is an unsupervised machine learning structure, and it can be viewed as a three-layer forward artificial neural network, as shown in Fig.1. It consists of the input layer, the hidden layer and the output layer. AutoEncoder is a very special neural network with single hidden layer, whose output is equal to the input. An AutoEncoder network parameters can be adjusted by repeated training process, such that the reconstructed output is an approximation with high accuracy of the input. AutoEncoder is composed of two parts: Encoder and Decoder. The encoder network encodes the input data from the high-dimensional space into low-dimensional space, then the low-dimensional space data is mapped into high-dimensional space through decoder network which realized the reconstruction process from output to input. Therefore, the low-dimensional space data can be used as the characteristic representation of the input data.



**Fig.1.** The model of AutoEncoder

Given an unlabeled datasetconsisting ofobservation features or variables, each observation variable hassamples. The encoder network encodes the sampleto the hidden activate valuewith an activation function. The encoder process is described as follows:

 (1)

where is the encoder function, Sigmoid functionis usually taken as the activation function in the encoder process. *W* is the weight matrix of the network between input layer and the hidden layer, *b* is the bias vector generated by the encoder network, is the connection parameter between the input layer and the hidden layer. The Sigmoid function can be depicted via Eq.(2)

 (2)

Similarly, for the decoder network, the feature matrixobtained from encoder network is used to reconstructthrough the decoder network such that the reconstructedis equal to the input. The decoder process is described as follows:

 (3)

where is the decoder function, is the activation function of the decoder process, represents the weight matrix between the hidden layer and the output layer of the network , *d* is the bias vector generated by the decoder process.

The essence of AE training process is to optimize the network parameters  and . In order to make the outputas close as possible to the input, we characterize the degree of approximation between input and output by minimizing the reconstruction error , The optimization process is described below:

 (4)

In each training process, the gradient descent method is used to update the training parameter  and of the AE network. The processes of network parameter update are as follows:

 (5)

 (6)

where represents the learning rate,partial derivativesand can be calculated with back propagation algorithm.

DNN can be simply viewed as a multi-hidden layers neural network formed by stacking many AutoEncoders. This model uses the bottom-up method of unsupervised learning, extracting the features layer by layer. Then supervised learning method is applied to fine tune the whole network parameters, which can extract the most essential characteristics from original signals. The structure of DNN is shown in Fig.2.



**Fig.2.** The structure of DNN

First of all, pre-train the DNN by using the unsupervised layer-by-layer greedy training algorithm. Firstly, the first AutoEncoder AE1 is trained by giving an unlabeled datasetas the input of encoder network. The encoded featureis the hidden layer of AE1. The training parameteris obtained by designing the uniqueas the output of AE1.Then, useas the input of the second AutoEncoder(AE2) and train AE2 to acquire the network training parameter.  is the hidden layer of AE2 which can be viewed as the characteristics of AE2. After that, chooseas the input of the third Autoencoder (AE3). Repeat the process to get the hidden layer featuresof the *N*th AutoEncoder(AE *N*) and the corresponding network training parameter.

Secondly, a classifier is added in the top layer of DNN. The feature information is extracted by using the unsupervised learning method in the pre-training process of DNN. However, DNN does not have the ability of [classifying](http://www.baidu.com/link?url=t-pS0vlBPfgx-8oFr7h6YdsybOxNE6mbmAVadFLG3ACufFpBLpfNcEDKShsJq2mRLG0QO6sCDR_A8UgVb_EN3PgQyxxHY_LNrGBoFYZlmI3), a classifier should be added in the top of DNN. In this paper, Softmax classifier is used as the output layer of DNN. We suppose the training dataset is , the label is, and the probability for each category can be calculated via the following hypothesis function:

 (7)

Whereis the model parameter of Softmax. Similarly to the AE model, in order to guarantee the performance of the classifier, the classifier model parameter is trained by minimizing the cost function. The cost function of Softmax training process is shown in Eq. (8), where the top network parameteris obtained from minimizing the.

 (8)

Finally, fine-tune. In order to guarantee the accuracy of feature extraction and the classification effectiveness of output layer, the whole DNN training parameters are fine-tuned by using a supervise algorithm of back propagation with some limited number of sample labels. The process of fine-tuning is completed by minimizing the reconstruction error.The procedures for parameter update are as follows:

 (9)

 (10)

whererepresents the actual output value, is a parameter set generated from the whole network training,, Back propagation algorithm is used to update the network parameter, is the learning rate in the process of deep learning. The fine-tuning process uses the labeled data to improve the performance of DNN.

## 2.3. DNN based classification

In order to accurately extract the essential characteristics of the mechanical equipment health conditions by DNN modeling, the following steps are required. Firstly, the original vibration signals should be pre-processed. Since frequency domain signals is more sensitive to mechanical equipment faults, so the original time-domain signals are converted into frequency-domain signals in the first step. Secondly, use the preprocessed data as the input of the DNN model to extract features of mechanical equipment health conditions with unsupervised layer-by-layer pre-training. Last but not least, the whole network parametercan be updated by using the back-propagation algorithm to fine-tune the DNN structure when limited number of labeled samples is available. In this way we can get an effective feature extraction result for fault classification. The preprocessed datasets are divided into training data and testing data. The training data is used to construct DNN model to obtain the training parameter, and the testing network initialized with training parameteris used to verify its effectiveness. Misclassification rate is used as an accuracy indicator of the DNN-based fault classification method. Detailed steps of DNN for mechanical system fault classification are shown in Fig.3.



**Fig.3.** Framework offault classification based on DNN

# 3. Multi-mode fault classification model based on deep learning

There are a number of multi-mode processes in practical system. For multi-mode process, the potential feature extracted from the observation of each steady mode also varies. So it is necessary to separate the observation into several operation modes for accuracy data feature extraction.

Therefore, mode partition is a fundamental step before fault classification. In this paper, this problem is solved by constructing a hierarchical DNN model with the first hierarchical specially devised for the purpose of mode partition. By this means, it can make an effective mode partition for multi-mode process, which can increase the accuracy of DNN-based fault classification. Framework of three-layer DNN is showed in Fig.4.



**Fig.4.** A hierarchical DNN framework of three-layer

The detailed steps for multi-mode fault classification are as follows.

**Step1. Mode partition.**

In this step, we focus on building a DNN model to determine the mode label of each sample. The whole datasets are used as the input of the multi-mode classification model. The mode partition process can be illustrated in detail as follows:

(1) Construct a new *DNN*1 withhidden layers AE descripted in Eq.(11), and initialize the training parameters of *DNN*1.

 (11)

where,**is the weight matrix,**is the bias vector. are the numbers of hidden layer neurons in *DNN*1. The network configuration can be represented by.denotes the training dataset. Usingin Eq.(12) to represent the number of neurons in the input layer of *DNN*1.

 (12)

The parameters of *DNN*1 can be initialized via Eqs.(13)-(14).

 (13)

 (14)

(2) Training of *DNN*1 to obtain the net parameter.Unsupervised layer-by-layer feature extraction based on the training datasetis implemented to the *N*-level AE defined in Eq.(11).

 (15)

 (16)

Add a softmax classifier on the top of *DNN*1. Limited number of training labels set is used to fine-tune *DNN*1 and update the training parametervia Eqs.(17)-(18).

 (17)

 (18)

where, withcalculated by Eqs.(7)-(8),is the number of samples . denotes the output of *DNN*1, is learning rate in fine-tuning process.

(3) Mode partition uses the trained *DNN*1. Once test sampleis obtained，compute the probability of each test sample via the trained. Then use Eq.(19) to divide the test sample into different modes:

 (19)

where;is the mode type of sample. **denotes the mode label of thetest sample.

Compare the mode partition label with the actual mode label to determine the misclassification number as Eq.(20)

 (20)

Whereis the operation to characterize the size of a set,  is the misclassification set defined by Eq.(21)

 (21)

**Step2. Fault source location.**

For a certain mode partitioned in Step1, We can further locate the fault source. The procedure in Step2 is analogous to Step1, which is described below:

(1) According to the mode partition result, we build the second hierarchical of the model which comprises a set ofDNNs, denotes the training dataset in *DNN*2.

 (22)

parameter initialization mechanism of *DNN*2 is same as Step1.

(2) Training of *DNN*2 to obtain the net parameters. Similarly, detailed calculation process can refer to Eqs.(15)-(18).

(3) Determine the fault location by the trained**.

 (23)

The test datasetis used to predict the unknown fault locations based trained**.Assume that each mode hasdifferent fault locations, fault location label for the sample of the mode can be calculated with prediction formula via Eq.(24).

 (24)

Compute the misclassification numberof the  mode. And then the misclassification of this classification step can be computed via Eq.(25)

 (25)

**Step3.** Fault severity recognition.

In order to identify the fault severity, the third hierarchical is devised with the intention to distinguish the fault severity. Construct the third deep network,is the training dataset in *DNN*3, andis the test dataset. Parameter training process is similar to Step2.The severity classification label of the sample in can be determined by Eq.(26)

 (26)

Compute The misclassification number for a given fault in a certain mode can be computed via Eqs.(27)-(28).

 (27)

 (28)

whereis the misclassification number of the fault location in the  mode,  is the misclassification number of all  modes, and is the misclassification number in this step.

**Step4**. Accuracy computation of the whole multi-mode classification network.

In this paper, the classification accuracy of the hierarchical DNN is measured by the numbers of misclassification. The final accuracy is calculated by the ratio of the total number of the misclassification to the total number of samples. The procedure of calculation is as follows:

 (29)

Combing with Eqs.(26)-(28), the final accuracy of the proposed multi-mode fault classification based on DNN can be formulated as Eq.(30)

 (30)

whereis the number of total samples, the flow chart of the proposed multi-mode fault classification method based on three-layer DNN is depicted in Fig.5.



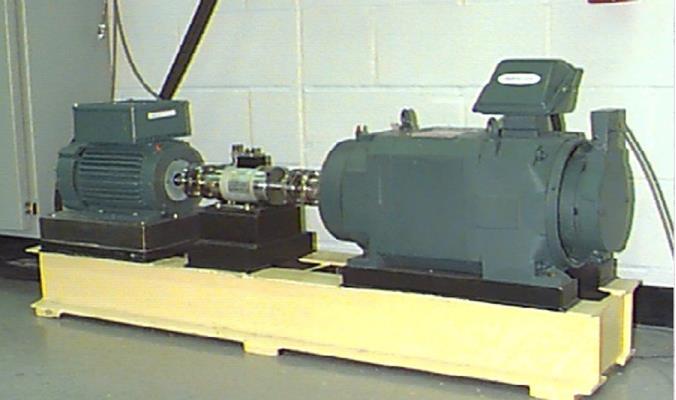
**Fig.5.** Flow chart of multi-mode classification based on DNN

# 4. Application to rolling bearing fault classification

Rolling bearings play an important role for rotating machinery. The health condition of the bearing directly affects the reliability and stability in the whole system. Rolling bearing as the experimental platform is used to verify the effectiveness of the hierarchical DNN multi-mode fault classification method, and the performance of the proposed method is compared with the traditional method such as DNN, BPNN, SVM, hierarchical BPNN and hierarchical SVM, which is listed in detail in Section 4.3.

## 4.1. Experimental platform

The experimental datasets are obtained from the Case Western Reserve University Bearing Data Center in the United States [35]. The experimental platform is shown in Fig.6. It can be seen that the experimental platform consists of a 2hp motor, a power meter, an electronic controller, a torque sensor and a load motor. The vibration signals of the drive end of the motor are collected by the acceleration sensor as the experimental datasets for bearing fault classification. In this experiment, using acceleration sensor to collect the vibration signals with the load 0hp, 1hp, 2hp and 3hp, respectively, and the sampling frequency is 48kHz. There are four types of bearing health condition, (1) Normal condition; (2) Inner race fault; (3) Outer race fault; (4) Roller fault. The sizes of the bearing fault were: 0.007mm, 0.014mm, 0.021mm, respectively.



**Fig.6.** Experimental platform for acquiring the vibration signals of rolling bearing

## 4.2. Data description

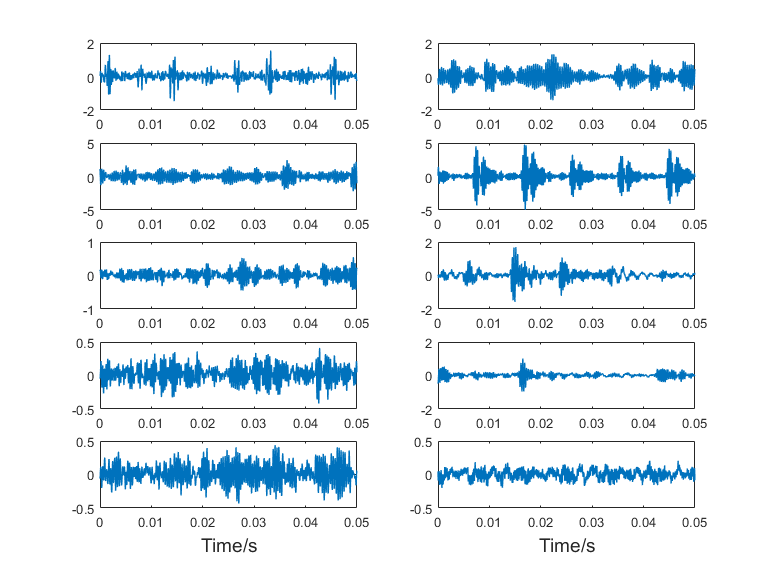
In this case, we collect the vibration signals of the bearing drive end at different loading. The data set collected contains 4 kinds of modes, the motor load is 0hp, 1hp, 2hp, 3hp, respectively, 4 modes are shown in Tab.1. In each mode, there are four states of inner race fault, outer race fault, roller fault, and normal; 3 different fault sizes in each fault state, that is to say 10 different fault types in a single mode. This paper selects 200 samples in each fault type, each sample contains 2048 observation points .100 samples are randomly selected as the training data, and the other 100 samples as the testing data. We use Fast Fourier Transform (FFT) for each sample to get 2048 Fourier coefficients. Because of the symmetry of the Fourier coefficients, so we take the first 1024 coefficients as the new samples, that is to say the dataset contains 8000 samples. In order to compare the proposed method of hierarchical network with single-layer network and explore the effect of different sample numbers on network. For a given mode, the sample number of each DNN is listed in Tab.2. In addition, we present the original time-domain waveforms of the 10 fault types in mode1under A, as shown in Fig. 7.

**Tab.1.** Four modes of rolling bearing

|  |  |  |
| --- | --- | --- |
| Mode | Load (hp) | Rotating speed(rpm) |
| Mode1 | 0 | 1797 |
| Mode2 | 1 | 1772 |
| Mode3 | 2 | 1750 |
| Mode4 | 3 | 1730 |

**Tab.2.** Data description of in dataset for a given mode

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data set for different fault location | Data set for different fault severity | Sample number  More /Fewer | Fault type | Fault size /mm |
|  |  | 200/100 | Normal(N) | 0.000 |
|  |  | 200/100 | Inner race fault (IF) | 0.007 |
|  | 200/100 | Inner race fault (IF) | 0.014 |
|  | 200/100 | Inner race fault (IF) | 0.021 |
|  |  | 200/100 | Outer race fault (OF) | 0.007 |
|  | 200/100 | Outer race fault (OF) | 0.014 |
|  | 200/100 | Outer race fault (OF) | 0.021 |
|  |  | 200/100 | Roller fault (RF) | 0.007 |
|  | 200/100 | Roller fault (RF) | 0.014 |
|  | 200/100 | Roller fault (RF) | 0.021 |



**Fig.7.** Observation of original signals corresponding to 10 fault types

## 4.3. Results of fault classification

The proposed hierarchical DNN structure is applied to bearing fault classification, there are 8000 samples, 4 different modes, 4 fault positions in each mode, totally 40 health conditions in dataset A. The health conditions of rotating machinery system under multi-mode, multi-condition, multi-fault type and large sample data are simulated which demonstrated the performance with the proposed method. To reduce the effect of randomness, the experiment was repeated 20 times. In this paper, the initialized parameters in the DNN pre-training process as shown in Tab.3.

**Tab.3.** DNN model parameters

|  |  |  |  |
| --- | --- | --- | --- |
| Training parameter |  |  |  |
| Hidden layers | 5 | 4 | 3 |
| Number of neurons | 512/400/300/200/100 | 512/400/200/100 | 512/256/100 |
| Max number of epochs | 500 | 300 | 300 |

The network training uses stochastic gradient descent method, on each hierarchical the maximum number of iterations of DNN is 500,300,300, respectively. Simulation of three tradition methods: BPNN, SVM and DNN are compared with simulation of the proposed multi-mode fault classification approach to verify its effectiveness. In addition, hierarchical BPNN(HBPNN), hierarchical SVM(HSVM) are also compared with hierarchical DNN(HDNN). BPNN uses the gradient descent method to update the network weights and bias parameter, one-to-one train mechanism is used to train a SVM with radial basis. The training mechanism of HBPNN and HSVM are the same as HDNN.

Tab.4 Compares the fault classification accuracy in time domain and frequency domain. It can be seen from Tab.4 that rotation machinery fault is more sensitive in frequency domain. So we use FFT as a tool to preprocess the original data.

Tab.5 compares the fault classification result after mode partition. It can be seen from line 2 and line 3 that HDNN can obtain more accurate classification either for fault source location or for fault severity recognition which tells us that mode partition is a critical step in multi-mode fault classification.

The hierarchical model for the case of BPNN and SVM also confirm this conclusion. Comparing line 2 with line 4 and line 6, we can see that HDNN is significantly superior to other hierarchical machine learning models because of the fact that HDNN can get better mode partition accuracy which is shown in Tab.6. On the other hand, we can draw another conclusion that the performance of traditional BPNN method is superior to the traditional SVM method in the large sample case, but the accuracy of HSVM is higher than that of HBPNN due to the fact that SVM does well in small sample learning.

In order to demonstrate the performance of the proposed multi-mode classification method, the hierarchical machine learning methods are employed in this paper. As can be seen from Tab.6, the accuracy of mode partition with proposed HDNN method can reach 99.96%, we can naturally find that the performance of HDNN is superior to HBPNN and HSVM in mode partition procedure.

**Tab.4.** Accuracy of classification in time domain and frequency domain

|  |  |  |
| --- | --- | --- |
| Method | Time domain data | Frequency domain data |
| HDNN | 99.96 | 80.65 |

**Tab.5.** Fault severity classification result comparison after mode partition

|  |  |  |
| --- | --- | --- |
| Method | Accuracy of fault  classification | Accuracy of fault severity  recognition |
| HDNN | 99.79 | 99.52 |
| DNN | 97.06 | 96.38 |
| HSVM | 82.82 | 77.00 |
| SVM | 65.74 | 58.40 |
| HBPNN | 81.28 | 71.68 |
| BPNN | 68.11 | 62.42 |

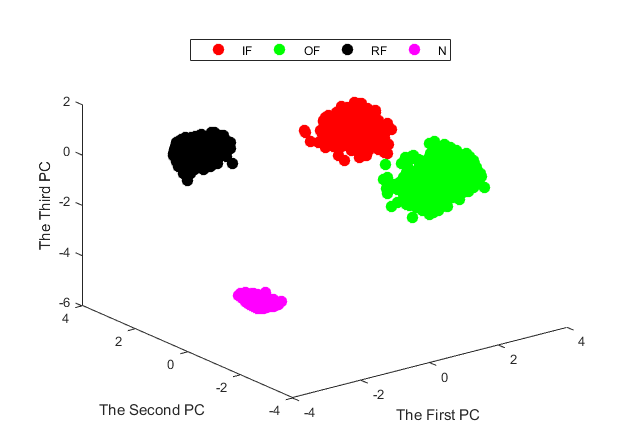
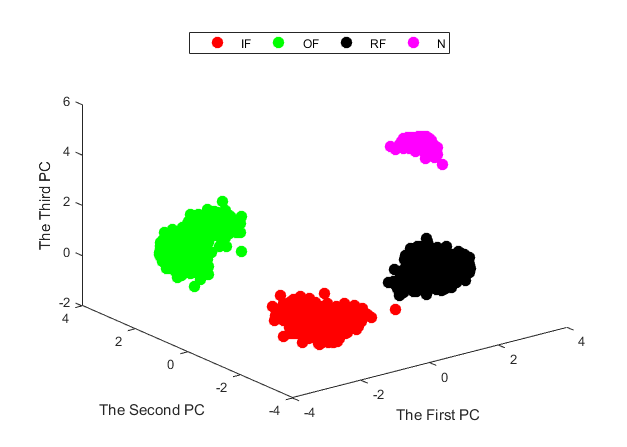
**Tab.6.** The classification results by the second hierarchical of the proposed model

|  |  |
| --- | --- |
| Method | Mode partition |
| HDNN | 99.96 |
| HBPNN | 90.45 |
| HSVM | 89.73 |

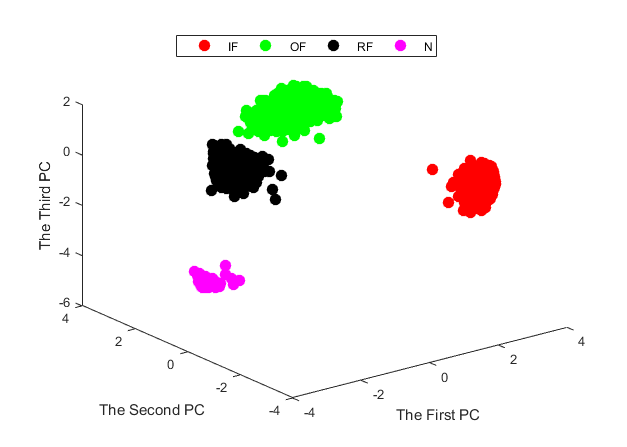
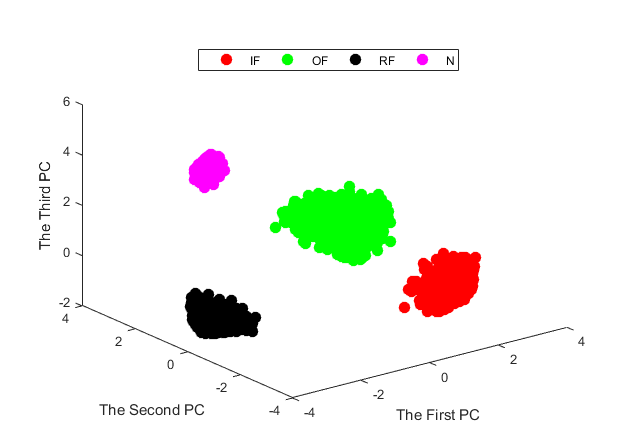
In view of the excellent performance of the proposed multi-mode classification method, we found that the performance of classification was influenced by accurate feature extraction. In order to verify the effectiveness of HDNN based feature extraction method, scatter plots of the feature extracted are demonstrated in Fig.8-Fig.10. As shown in Tab.3, in each training process, the number of neurons in the last hidden layer is 100, that is to say, the feature dimension is 100, which is too large to be visualized. Therefore, PCA is used as a data compression tool to reduce the feature dimension. In this paper, we use the first three key principal components to plot the scatter chart of the fault source location feature extracted by HDNN, as shown in Fig.8. Fig.8 is the scatter plots for fault feature extracted by HDNN after mode partition, while Fig.9 shows the scatter plots for fault feature extracted by DNN without mode partition. From Fig.9 and Fig.10, we can see some fault features are overlapped, which result in an unsatisfactory fault classification result.

Fig.10 is the scatter plot of the feature extracted for different modes. We can see from Fig.10 that HDNN does well in multi-mode fault feature extraction which will greatly affect the accuracy of the successive fault classification.

In summary, the proposed multi-mode classification method can accurately extract the different fault features based on its strong non-linear characterization ability.



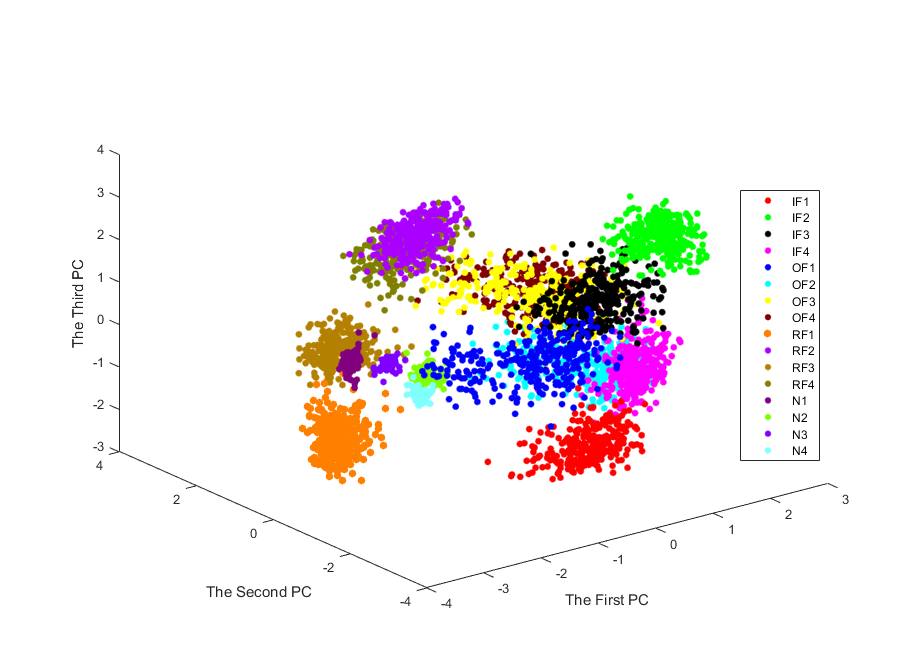
(a) (b)



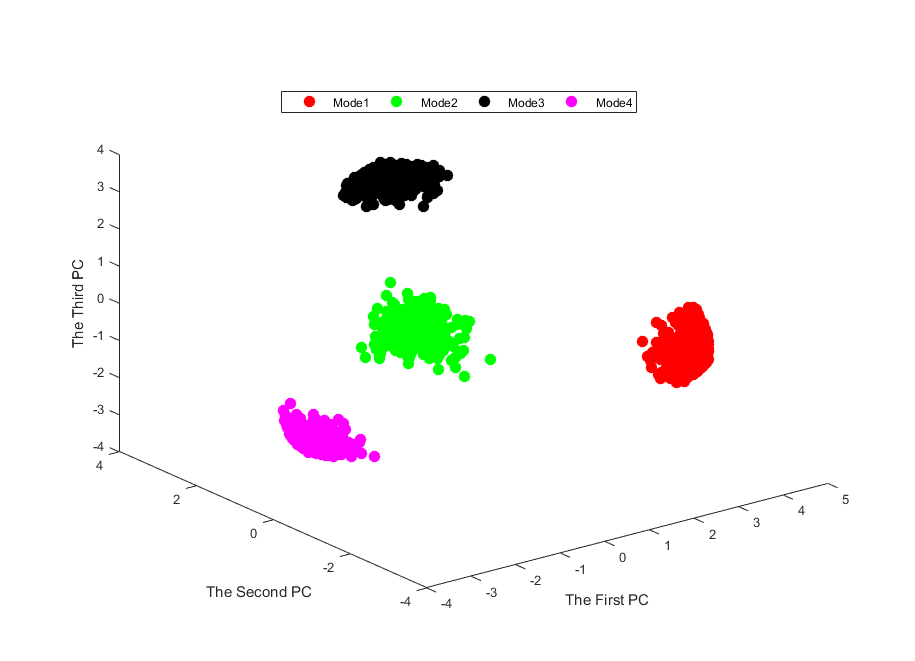
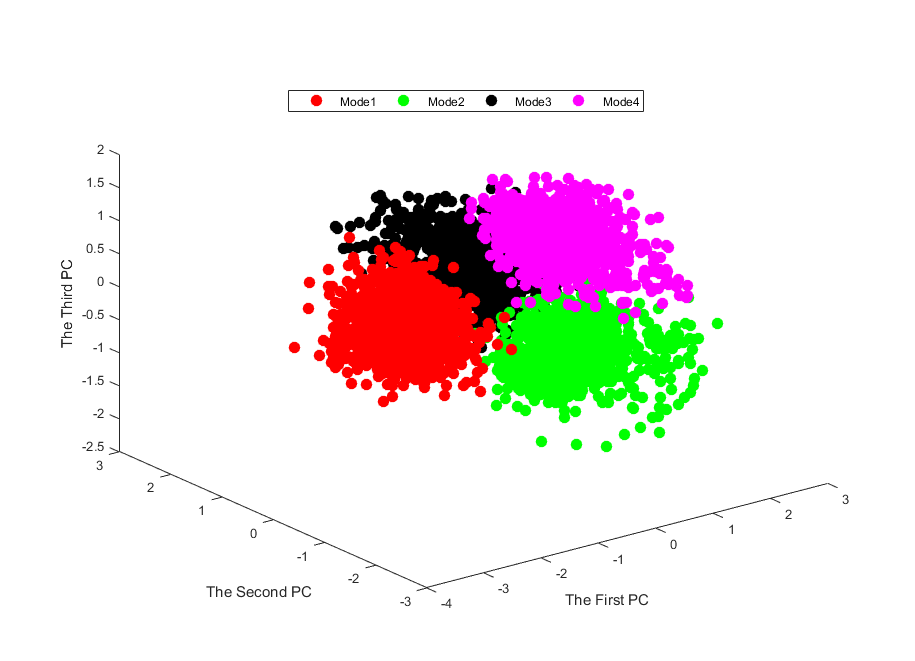
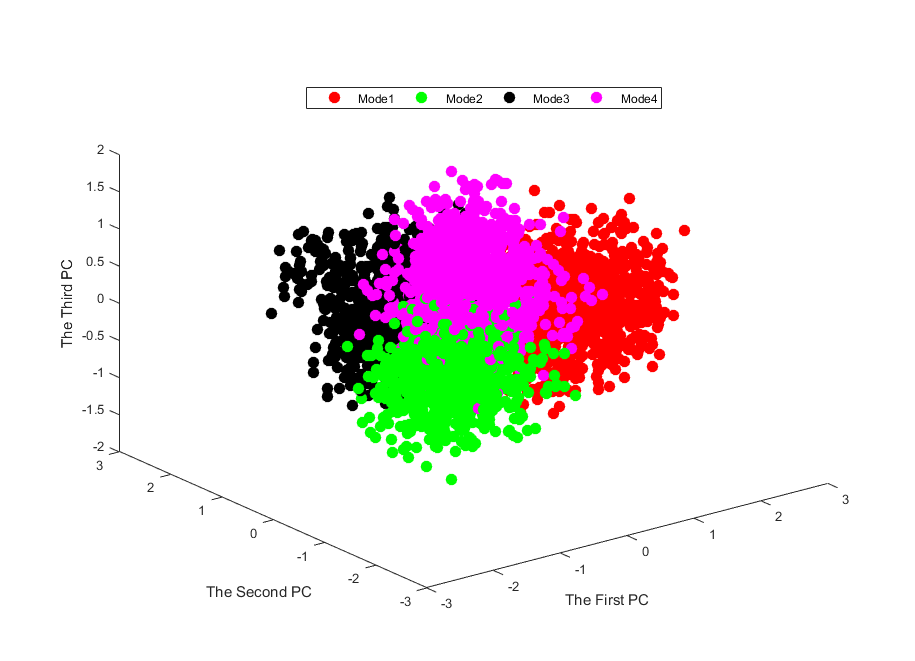
(c) (d)

**Fig.8.** Scatter plots of principal components for the feature of fault classification; (a)-(d) represent

four modes: corresponding to Mode1, Mode2, Mode3, Mode4, respectively



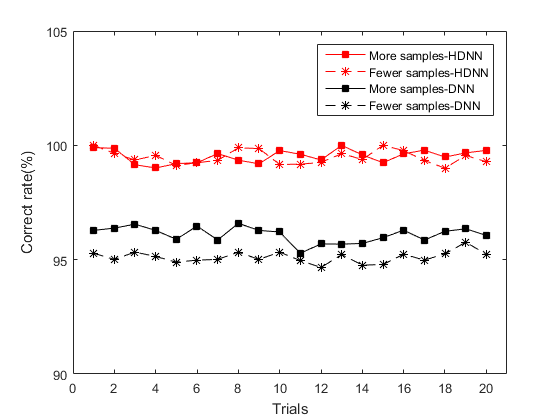
**Fig.9.** Scatter plots of principal components for fault features with traditional DNN method

1. (b) (c)

**Fig.10.** Scatter plots of principal components for mode partition features;(a)-(c) represent

mode partition result corresponding to HDNN,HBPNN,HSVM, respectively



**Fig.11.** Robust of the fault classification method to the sample number of train data with 20 trials

In general, efficiency of the fault classification method is affected by sample number of the train data. Fig.11 displays the fault classification accuracy of DNN and HDNN in two cases. Red line denotes the classification accuracy of the case when more samples are used as the train data. Black line denotes the classification accuracy of the case when fewer samples (only 1/2 of the first case) are used as the train data. In addition, the line with “\*” is the simulation result of HDNN and the line with “□” is the simulation result of traditional DNN.

From Fig. 11，it can be clearly seen that: 1) Fault classification accuracy of HDNN does not vary much for the two cases, while the fault classification accuracy of DNN is greatly affected by the number of train data used; 2) In both case, fault classification accuracy of HDNN is much better than DNN. So we can come to the conclusion that HDNN is a more robust fault classification for multi-mode bearing fault classification in the case when fewer number of train data available.

# 5. Conclusions

In this paper, a novel multi-mode fault classification method based on DNN is developed. The main idea is to construct a hierarchical DNN model with the first hierarchical specially devised for the purpose of mode partition. The second hierarchical model comprising a set of DNN is devised to extract feature separately of different modes and precisely diagnose the fault source. Another set of DNN is devised to distinguish the severity of a certain fault in a given mode, which is helpful for predictive maintenance of the machinery equipment. Rolling bearing is the experiment platform to verify the efficiency of the proposed method.

## Acknowledgment

This research was supported in part by the Natural Science Fund of China (Grant No. U1604158, U1509203 and 61333005) and Technical Innovation Talents Scheme of Henan Province (Grant No. 2012HASTIT005).

# References

[1] D. H. Zhou, Y. Liu and X. He, "Review on fault classification techniques for closed-loop systems," *Acta Automatica Sinica*, vol.39, no. 11, pp. 1933-1943,2013.

[2] F. N. Zhou, J. H. Park and Y. J. Liu, "Differential feature based hierarchical PCA fault detection method for dynamic fault," *Neurocomputing* ,vol.202, pp. 27-35, 2016.

[3] C. Li, RV. Sanchez, G. Zurita, M. Cerrada, D. Cabrera and R. E. Vásquez , "Multimodal deep support vector classification with homologous features and its application to gearbox fault classification," *Neurocomputing*, vol.168, pp.119-127, 2015.

[4] X. He, Z. Wang, Y. Liu and D. H. Zhou, "Least-Squares Fault Detection and Classification for Networked Sensing Systems Using A Direct State Estimation Approach," in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1670-1679, Aug. 2013.

[5] D. Zhao, D. Shen and Y. Q. Wang, "Fault classification and compensation for two-dimensional discrete time systems with sensor faults and time-varying delays, " *Int. J. Robust Nonlinear Control*, doi: [10.1002/rnc.3742](http://dx.doi.org/10.1002/rnc.3742), 2017.

[6] H. Li and D.Y. Xiao, "Surver on data driven fault classification methods", *Control and Decision*, vol. 26, no. 1, pp. 1-9+16, 2011.

[7] R. M. An and Y. Gao, "Spacecraft fault classification based on hierarchical neural network," *Spacecraft environment engineering*, vol. 30, no. 2, pp. 203-208, 2013.

[8] F. N. Zhou , C. L. Wen, Y. B. Leng and Z. G. Chen, "A data-driven fault propagation analysis method", *Journal of Chemical Industry and Engineering(China)*, vol. 61, no. 8, pp. 1993-2001,2010.

[9] H. Q. Ji, X. He and D. H. Zhou, "On the use of reconstruction-based contribution for fault classification, "*Journal of Process Control*, vol. 40, pp. 24-34, 2016.

[10] D. J. Yu, M. F. Chen, J. S. Cheng and Y. Yang, "A fault classification approach for rotor systems based on empirical mode decomposition method and support vector machines, " *Proceedings of the Chinese society for electrical engineering*, vol. 26,no. 16, pp. 162-167, 2006.

[11] M. Gan, C. Wang and C. A. Zhu, "Construction of hierarchical classification network based on deep learning and its application in the fault pattern recognition of rolling element bearings" *Mechanical Systems and Signal Processing*, vol.72-73, pp. 92–104, 2016.

[12] G. F. Bin, J. J. Gao, [X. J. Li](http://www.sciencedirect.com/science/article/pii/S0888327011003207) and [B. S. Dhillon](http://www.sciencedirect.com/science/article/pii/S0888327011003207), "Early fault classification of rotating machinery based on wavelet packets—Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711,2012.

[12] A. Widodo and BS. Yang, "Support vector machine in machine condition monitoring and fault classification," *Mechanical Systems and Signal Processing*, vol. 21,no. 6, pp.2560-2574, 2007.

[13] D. Zhao, Z. P. Lin and Y. Q. Wang, "Integrated state/disturbance observers for two-dimensional linear systems," in *IET Control Theory & Applications*, vol. 9, no. 9, pp. 1373-1383, 2015.

[15] Q. Hu, Z. J. He, Z. S. Zhang and Y. Y. Zi , "Fault classification of rotating machinery based on improved wavelet package transform and SVMs ensemble ," *Mechanical Systems and Signal Processing* ,vol. 21, no.2, pp. 688–705, 2007.

[16] L. Y. Wang, W. G. Zhao, Y. Liu, "Rolling Bearing Fault Classification Based on Wavelet Packet- Neural Network Characteristic Entropy," Advanced Materials Research, Vols. 108-111, pp. 1075-1079, 2010.

[17] Y. Yang and W. Tang, "Study of remote bearing fault classification based on BP Neural Network combination," *2011 Seventh International Conference on Natural Computation*, pp. 618-621,Shanghai, 2011.

[18] L. Jiang, Q. Li, J. Cui and J. Xi, "Rolling bearing fault classification based on higher-order cumulants and BP neural network," *The 27th Chinese Control and Decision Conference (2015 CCDC)*, pp. 2664-2667, Qingdao, 2015.

[19] T. Kuremoto, S. Kimura, K. Kobayashi and M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, 2014.

[20] S. M. Zhang, F. L. Wang, S. Tan and S. Wang, "A fully automatic onine mode identiflcation method for multi-mode processes," *Acta Automatica Sinica*, vol. 42, no.1, pp.60-80, 2016.

[21] Jürgen Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[22] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol.313, pp. 504-507, 2006.

[23] H. Ze, A. Senior and M. Schuster, "Statistical parametric speech synthesis using deep neural networks," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 7962-7966, Vancouver, BC, 2013.

[24] B. Song, S. Tan and H. B. Shi, "Key principal components with recursive local outlier factor for multimode chemical process monitoring," *Journal of Process Control*, vol.47,pp.136–149, 2016.

[25] L. P. Zhao, C. H. Zhao and F. R. Gao, "Inter-batch-evolution-traced process monitoring based on inter-batch mode partition for multiphase batch processes," *Chemometrics and Intelligent Laboratory Systems*, vol.138, pp.178–192, 2014.

[26] Y. W. Zhang, C. Wang and R. Q. Lu, "Modeling and monitoring of multimode process based on subspace separation," *Chemical Engineering Research and Design*, vol.91, no.5, pp.831–842, 2013.

[27] F. N. Zhou, C. L. Wen, T. H. Tang and Z. G. Chen, "DCA based multiple faults classification method," *Acta Automatica Sinica*, vol.35, no.7, pp. 971-982, 2009.

[28] P. Tamilselvan and P. F.Wang, "Failure classification using deep belief learning based health state classification," *Reliability Engineering and System Safety*, vol. 115, pp. 124–135, 2013.

[29] R. B. Huang, C. Liu, G. Q. Li, and J. L. Zhou, "Adaptive Deep Supervised Autoencoder Based Image Reconstruction for Face Recognition," Mathematical Problems in Engineering, vol. 2016, Article ID 6795352, 14 pages, 2016.

[30] H. M. Liu, L. F. Li, and J. Ma, "Rolling Bearing Fault Classification Based on STFT-Deep Learning and Sound Signals," Shock and Vibration, vol. 2016, Article ID 6127479, 12 pages, 2016.

[31] P. L. Wang , C. J. Xia, "Fault detection and self-learning identification based on PCA-PDBNs," *Chinese Journal of Scientific Instrument*, vol.36, no. 5, pp.1147-1154, 2015.

[32] R. Pang, Z. B. Yu,  W. Y. Xiong and H. Li , "Faults recognition of high -speed train bogie based on deep learning ," *Journal of Railway Science and Engineering*, vol.12, no.6, pp. 1283-1288, 2015.

[33] C. Lu，Z.Y. Wang, W. L. Qin and J. Ma, "Fault classification of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Processing*, vol.130, pp.377-388, 2017.

[34] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent classification of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol.72-73, pp. 303–315, 2016.

[35] Bearing data Centre, Case Western Reserve University, Available:

[*http://csegroups.case.edu/bearingdatacenter/home*](http://csegroups.case.edu/bearingdatacenter/home)