2306. Rolling bearing fault identification using multilayer deep learning convolutional neural network

Hongkai Jiang[1], Fuan Wang[2], Haidong Shao[3], Haizhou Zhang[4]
School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China
[1]Corresponding author
E-mail: [1]*jianghk@nwpu.edu.cn*, [2]*2015200080@mail.nwpu.edu.cn*, [3]*hdshao@mail.nwpu.edu.cn*, [4]*zhanghaizhou666@mail.nwpu.edu.cn*

Check for updates

Abstract. The vibration signal of rolling bearing is usually complex and the useful fault information is hidden in the background noise, therefore, it is a challenge to identify rolling bearing faults from the complex vibration environment. In this paper, a novel multilayer deep learning convolutional neural network (CNN) method to identify rolling bearing fault is proposed. Firstly, in order to avoid the influence of different characteristics of the input data on the identification accuracy, a normalization preprocessing method is applied to preprocess the vibration signals of rolling bearings. Secondly, a multilayer CNN based on deep learning is designed in this paper to improve the fault identification accuracy of rolling bearing. Simulation data and experimental data analysis results show that the proposed method has better performance than SVM method and ANN method without any manual feature extractor design.

1. Introduction

As a key part of rotating machinery, rolling bearing is widely used in modern machine [1, 2]. Its working condition is directly related to whether the equipment can operate normally. Faults in rolling bearings can lead to machine breakdown, and even bring serious economic loss to industries. Therefore, carrying out fault diagnosis researches on rolling bearing is very necessary, and it has been a hot research topic in recent years [3, 4].

In recent years, various fault diagnosis methods have been proposed [5-13]. Yu et al. applied EMD method and Hilbert spectrum to the rolling bearing fault diagnosis [7]. Tian et al. proposed a rolling bearing fault diagnosis method based on LMD-SVD and extreme learning machine [8]. Rauber et al. introduced a method based on heterogeneous feature models to bearing fault diagnosis [9]. Tian et al. applied differential geometry to rolling bearing fault diagnosis [10]. Ma et al. applied softmax regression to the fault diagnosis and health assessment of centrifugal pumps [11]. Among the proposed methods, intelligent fault diagnosis methods based on artificial neural networks [14-16], SVM [13, 17] have been the center of intelligent fault diagnosis researches. However, current intelligent fault diagnosis methods are shallow learning models, which only involve a few hidden layers. As a result, their learning ability is limited and they need careful feature extractor design with manual intervention and domain expertise when applied to multi-calss and complex fault diagnosis researches. These disadvantages greatly limit the application of intelligent fault diagnosis methods, which prompt researchers to focus on deep learning methods.

In 2006, Geoffery Hinton proposed the concept of deep learning [18]. Because of the good performance, a lot of research works based on deep learning methods have been proposed in recent years [19-21]. Among the deep learning models, CNN is the first truly successful deep learning model [22]. CNN is a multilayer model consisted of multiple processing layers and can transform the raw input data into essential internal features layer by layer to improve the classification accuracy, in other words, no careful manual feature extractor design is required for CNN. Because of the good performance of CNN, it is widely applied in pattern identification problems [23-25]

Therefore, this paper proposes a fault identification method based on CNN and applied the method to the rolling bearing fault identification problems.

The diagnosis procedure of the proposed method is as follows. Firstly, vibration signals of rolling bearings under various conditions are obtained by the data acquisition system; secondly, the obtained vibration signals are preprocessed using the method illustrated in this paper; third, a CNN model for rolling bearing fault diagnosis is designed; finally, the designed CNN model is used to diagnose rolling bearing faults.

The main advantage of the proposed method is that the proposed method has excellent feature learning ability and can automatically learn the essential features from vibration data, which greatly increases the classification accuracy of rolling bearing fault diagnosis problems without any manual feature extractor and feature selection design.

The rest of this paper is organized as follows. The deep learning CNN model is briefly introduced in Section 2. The proposed method is described in Section 3. In Section 4, the proposed method is applied to analyze the simulation signal and experimental signal. The conclusions are given in Section 5.

2. Deep learning CNN model

2.1. The architecture of CNN

As a deep learning model with multilayer architecture, CNN relies more on automatic learning and less on careful manual design [21, 22]. The main layers of CNN contain convolutional layers and subsampling layers. The convolutional layers perform as shared-weight extractors and the subsampling layers perform subsampling on the output of the previous convolutional layers.

The inputs of convolutional layers are a set of units from the previous layers. The convolutional layers perform convolution operation on the input maps with a set of trainable kernels. For a convolutional layer in the lth layer in the CNN, the computation is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right), \tag{1}$$

where b_j^l denotes a trainable bias, $f(\cdot)$ denotes the activation function, k_{ij}^l denotes the convolutional kernel, M_j denotes feature map and * denotes discrete convolution operation.

The subsampling layers are designed to reduce the complexity of CNN. In this paper, subsampling layers compute the average values over a neighborhood in each feature map. The computation is as follows:

$$x_j^l = f\left(\beta_j^l down(x_j^{l-1}) + b_j^l\right), \tag{2}$$

where β_j^l denotes the weight vector value, b_j^l denotes a trainable bias parameter. $down(\cdot)$ is a subsampling function.

2.2. CNN training

The CNN model designed in this paper is trained by backprop algorithm, which contains feedforward pass and backpropagation pass [22]. In the feedforward pass, the output of each layer is the input of the next layer. Therefore, the output of each layer will affect the output of the network. The training error is computed according to the squared-error loss function. For a training dataset with N training samples and c classes, the training error E is computed according to the following formula:

$$E = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{c} (z_k^n - y_k^n)^2, \tag{3}$$

where z_k^n is the kth dimension of the nth pattern's target, and y_k^n is the kth output layer unit corresponding to the nth input pattern.

For an ordinary fully connected layer l, the output x^l is as follows:

$$x^l = f(u^l), \quad u^l = W^l x^{l-1} + b^l, \tag{4}$$

where W^l denotes the weight vector and b^l denotes the bias vector.

In the backpropagation pass, the parameters are updated with the training error. δ is the sensitive of each unit with respect to perturbations of the bias b. In this case, because $\partial u / \partial b = 1$, δ can be computed as follows:

$$\delta = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial u} \frac{\partial u}{\partial b} = \frac{\partial E}{\partial u}. \tag{5}$$

For each layer, the weights are updated by adding ΔW^l. ΔW^l is computed as follows:

$$\Delta W^l = -\eta \frac{\partial E}{\partial W^l} = -\eta x^{l-1} (\delta^l)^T, \tag{6}$$

where η denotes the learning rate.

The output layer is a fully connected layer. The sensitive δ^L for the output layer neurons are computed as follows:

$$\delta^L = \frac{\partial E}{\partial b^L} = \frac{\partial E}{\partial u^L} \frac{\partial u^L}{\partial b^L} = \frac{\partial E}{\partial u^L} = f'(u^L) \circ (y^n - z^n), \tag{7}$$

where $\circ$ denotes element-wise multiplication.

For a convolutional layer at the lth layer, the sensitive δ_j^l for jth map is computed as follows:

$$\delta_j^l = \beta_j^{l+1} \left(f'(u_j^l) \circ up(\delta_j^{l+1}) \right), \tag{8}$$

where β_j^{l+1} is the weight of the subsampling layer at layer $l + 1$, $up(\cdot)$ denotes an upsampling operation.

For a subsampling layer at lth layer, the sensitive δ_j^l is computed as follows:

$$\delta_j^l = f'(u_j^l) \circ conv2(\delta_j^{l+1}, rot180(k_j^{l+1}), 'full'), \tag{9}$$

where, the kernel k_j^{l+1} is rotated 180 degrees to make the convolution function perform cross-correction. $conv2$ denotes full 2D convolution operation.

3. The proposed method

Because of the excellent feature learning ability of deep learning CNN, a novel multilayer deep learning CNN method to identify rolling bearing faults is proposed in this paper.

3.1. The preprocessing of rolling bearing vibration signals

The amplitude values of rolling bearing vibration signals vary greatly under various fault

conditions, which will affect the fault identification accuracy. To solve this problem, this paper adopts the following formula to preprocess the obtained rolling bearing vibration signals, and normalize the signal amplitude values to [0, 1]:

$$y = (y_{max} - y_{min}) \times \frac{x - x_{min}}{x_{max} - x_{min}} + y_{min}, \tag{10}$$

where x_{max} and x_{min} are the maximum value and minimum value of the raw data, y_{max} is 1 and y_{min} is 0 in this paper. y denotes the preprocessed signal. In this paper, the maximum value of the preprocessed signal is 1 and the minimum value of the preprocessed signal is 0. The preprocessing operation is beneficial for improving the fault identification accuracy.

The preprocessed data is divided into samples in this paper. The training dataset and test dataset are composed of samples and fault labels.

3.2. Multilayer deep learning CNN for rolling bearing fault identification design

In this paper, rolling bearing fault identification method using multilayer deep learning CNN are designed as follows:

Step 1: Use accelerometers to collect the vibration signals of rolling bearings.

Step 2: Preprocess the collected vibration signals using the method in 3.1 and construct the training dataset and testing dataset.

Step 3: Design the multilayer deep learning CNN model.

Step 4: Train the designed CNN model.

Step 5: Diagnose on the testing dataset using the well trained multilayer deep learning CNN.

The flowchart of the proposed method is described in Fig. 1. In Fig. 1, numpochs is the current training epoch and maxpoches is the maximum training epoch. Firstly, design the multilayer deep learning CNN model and input the preprocessed vibration signals of rolling bearings. Secondly, initialize the multilayer deep learning CNN model. Then, compute the output of networks and back propagate the error to update the weights. Lastly, diagnose on the testing dataset using the trained multilayer deep learning CNN and output the fault identification accuracy.

The CNN designed in this paper consists of input layer, convolutional layer C1, subsampling layer S2, convolutional layer C3, subsampling layer S4, and the output layer. The maps in the input layer are in the size of 20×20. The convolutional layer C1 contains 6 feature maps and the size of convolutional kernels is 5×5. In S2 and S4, feature maps from convolutional layers are divided into sub-regions with the size of 2×2 and the sub-regions are non-overlapping, the mean value of each sub-region is the output. C3 contains 12 feature maps and the size of convolutional kernels is 5×5. The output layer is a softmax classifier.

4. Simulation and experimental validation

In this paper, simulation data and experimental data are used to verify effectiveness of the proposed method.

4.1. Case 1: simulation signal analysis

In this case study, two vibration signals of rolling bearings are simulated. The fault patterns of simulated rolling bearings contain outer fault and inner fault. In fact, the vibration signals are interfered by background signal when rolling bearings are working in rotating machinery. The simulation vibration signal x is described as follows:

$$x = \sum_{i=1}^{N} \left[e^{-\zeta 2\pi f_1(t-i/f_1)} \sin 2\pi f_1(t - i/f_1)\sqrt{1-\zeta^2} \right. \tag{11}$$
$$\left. + e^{-\zeta 2\pi f_2(t-i/f_2)} \sin 2\pi f_2(t - i/f_2)\sqrt{1-\zeta^2} \right] + a \times randn(1,n),$$

where x is composed of two different impulse response signals with carrier center frequencies corresponding to f_1 and f_2 respectively. ζ denotes the damping ratio. A noise signal is added to x and a is the amplitude of the noise signal.
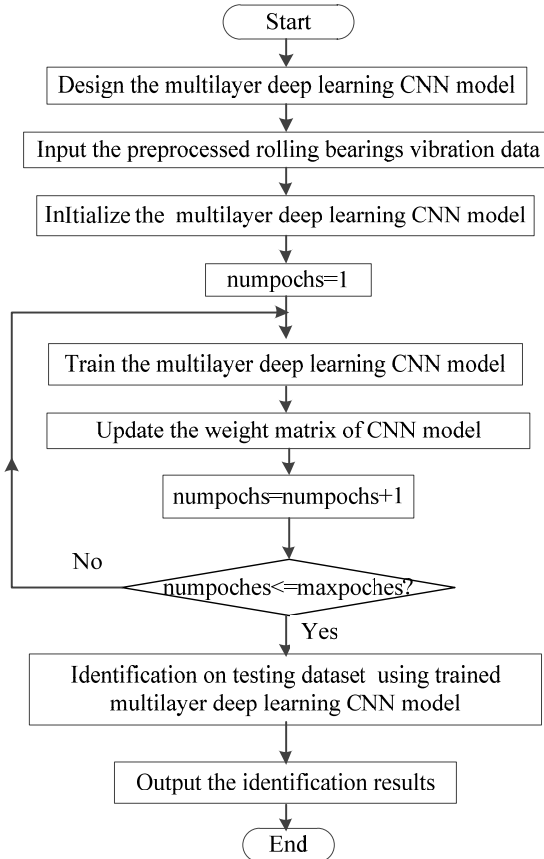

Fig. 1. The flowchart of the proposed method

The characteristic frequency of the inner race fault signal is $f_i = 87.897$ Hz, and N for inner race fault signal is 112. The characteristic frequency of the inner race fault signal is $f_o = 64.819$ Hz, and N for outer race fault signal is 83. The sampling frequency f_s is 12.8 KHz. The simulation time t is 1.28 s and $n = 16384$. The parameter values of the two kinds of rolling bearing faults simulation signals are described in Table 1. Fig. 2 is the time domain figures for inner race fault and outer race fault simulation signals without noise. Fig. 3 is the time domain figures for rolling bearing inner race fault and outer race fault simulation signals combined noise.

Table 1. The parameters value of rolling bearing faults simulation signals

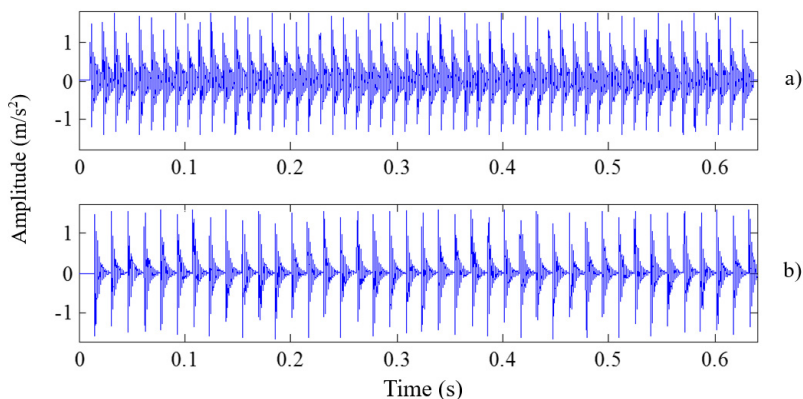| Fault condition	f_1 / Hz	f_2 / Hz	ζ	a	N	f_s	t
Inner race fault	1200	5200	0.02	0.3	112	12.8 KHz	1.28 s
Outer race fault	2000	5200	0.02	0.3	83	12.8 KHz	1.28 s

Fig. 2. The time domain figures for simulation signals without noise: a) inner race fault; b) outer race fault
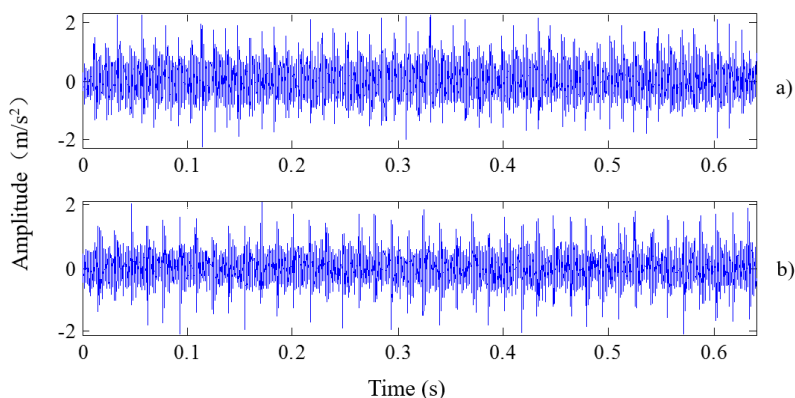

Fig. 3. The time domain figures for simulation signals combined with noise:
a) inner race fault; b) outer race fault

In this case study, the simulation signals are preprocessed according to the method illustrated in Section 3.1. As shown in Table 2, two comparative datasets are designed, including dataset A and dataset B. Dataset A is vibration data without noise and dataset B is vibration data combined with noise. In each simulation signal, the first 16000 data points are equally divided into 30 training samples and 10 testing samples, each sample contains 400 data points. Therefore, each dataset has 60 training samples including 30 inner race fault samples and 30 outer race fault samples and 20 testing samples including 10 inner race fault samples and 10 outer race fault samples.

Table 2. Fault samples distribution for simulation signals

Dataset	Fault condition	Training sample	Testing sample	Label
Dataset A (without noise)	Inner race fault	30	10	1
	Outer race fault	30	10	2
Dataset B (with noise)	Inner race fault	30	10	1
	Outer race fault	30	10	2

In order to verify the effectiveness of the proposed method, the samples are input to the designed CNN model without any manual feature extraction. For comparison, the artificial neural network (ANN) and SVM methods are respectively used to analyze the same datasets without any manual feature extraction. The three methods are explained as follows. (1) The proposed method: the CNN model is designed as illustrated in Section 3.2, the learning rate is 1 and the training epoch is 100. (2) ANN method: the scaled conjugate gradients method is used to train the ANN

model, the learning rate is 0.25, maximum training epochs is 100 and the hidden layer has 400 neurons. (3) SVM method: the RBF kernel is applied, the penalty factor is 0.92 and the radius of the kernel function is 0.44. All the parameters are determined by experience and repeated experiments.

Fig. 4 shows the classification accuracy using the proposed method, SVM method and ANN method. In dataset A, the classification accuracy of training samples based on the proposed method, SVM and ANN is 100 %, 100 % and 96 %, respectively, The classification accuracy of testing samples in dataset A based on the proposed method, SVM and ANN is 100 %, 100 % and 75 %, respectively. In dataset B, the simulation signals are combined with noise, and the classification accuracy of training samples based on the proposed method, SVM and ANN is 100 %, 100 % and 92 %, respectively, the classification accuracy of testing samples in dataset B based on the proposed method, SVM and ANN is 100 %, 75 % and 65 %, respectively.
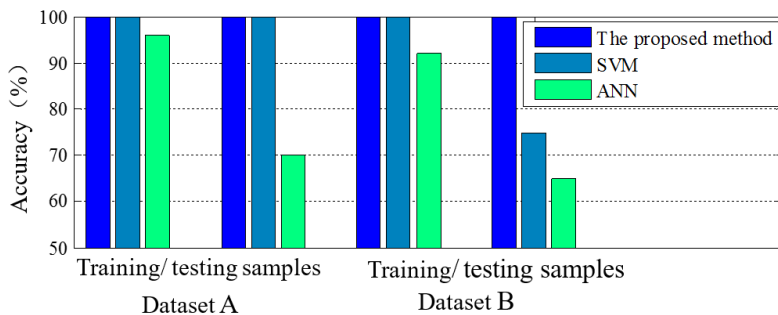

Fig. 4. Identification accuracy of simulation signals

The simulation result confirms that the proposed method has better classification performance than SVM methods and ANN methods, especially when the signals are combined with noise. The proposed method has better feature learning ability.

4.2. Case 2: experimental signal analysis

In this case study, the rolling bearing data are from the electrical engineering laboratory of Case Western Reserve University [12]. As shown in Fig. 5, the test stand is composed of a driving motor, a torque transducer, a dynamometer and control electronic unit. The testing rolling bearings contain four health conditions: (1) normal condition, (2) inner race fault, (3) outer race fault and (4) ball fault. The vibration signals were collected by accelerometers, and the sampling frequency is 12 kHz.
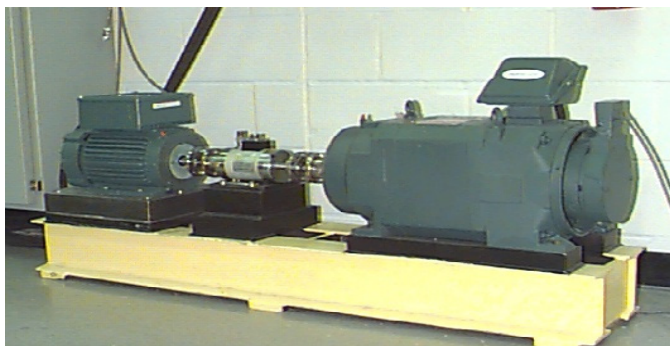

Fig. 5. The test stands of rolling bearings

In this case, 11 vibration signals collected at the speed of 1797 rpm from the drive end

containing four health conditions and varied fault severity are selected to verify the effectiveness of the proposed method. Fig. 6 are the time domain figures of vibration signals of rolling bearings with four health conditions. As Table 3 shows, each vibration signal is preprocessed according to the method illustrated in section 3.1. Each signal is equally divided into 300 samples and each sample contains 400 data points. The training dataset has 2200 (200×11) samples and the testing dataset has 1100 (100×11) samples. The proposed method is used to analyze these samples, and SVM method and ANN method are used to.
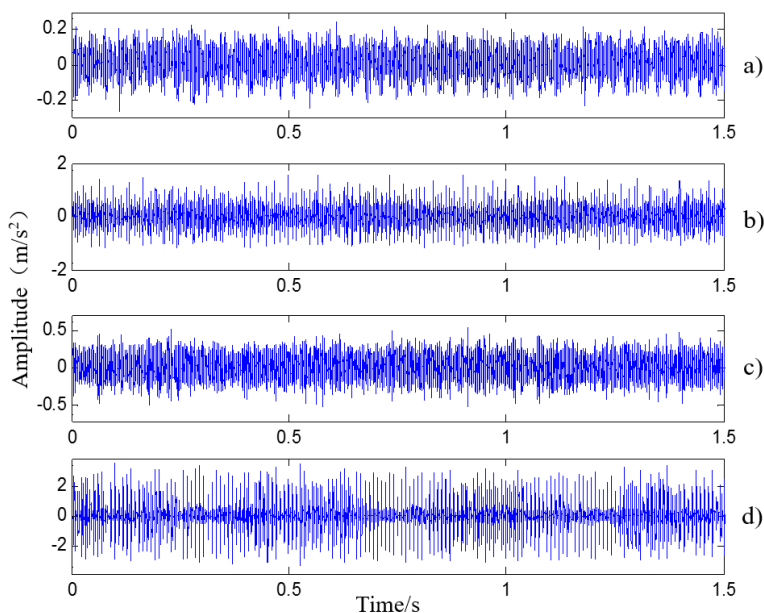

Fig. 6. Time domain figures for rolling bearings vibration signals:
a) normal condition; b) inner race fault; c) ball fault; d) outer race fault

Without any manual feature extraction and any manual feature selection, the samples are directly input to the designed CNN, SVM and ANN. The three methods are explained as follows. (1) The proposed method: the CNN model is designed as illustrated in Section 3.2, the learning rate is 1 and the training epoch is 300. (2) SVM method: the RBF kernel is applied, the penalty factor is 0.50 and the radius of the kernel function is 0.92. (3) ANN method: the scaled conjugate gradients method is used to train the ANN model, the learning rate is 0.05 and maximum training epochs is 500, the hidden layer has 400 neurons. All parameters are determined by experience and repeated experiments.

Table 3. Rolling bearing fault sample distribution

Rolling bearing condition	Training sample	Testing samples	Label
Normal condition	200	100	1
0.007/Inner race fault	200	100	2
0.007/Ball fault	200	100	3
0.007/Outer race fault	200	100	4
0.014/Ball fault	200	100	5
0.014/Outer race fault	200	100	6
0.021/Inner race fault	200	100	7
0.021/Ball fault	200	100	8
0.021/Outer race fault	200	100	9
0.028/Inner race fault	200	100	10
0.028/Ball fault	200	100	11

As shown in Table 4 and Fig. 7, the classification accuracy of training samples based on the proposed method is 98.36 %, and it is much higher than those using SVM method and ANN method, which are 77.27 % and 75.09 %. The classification accuracy of testing samples based on the proposed method is 88.00 %. That is much higher than those based on SVM and ANN, which are 63.18 % and 53.91 %. The proposed method performs much better than SVM and ANN.

Table 4. Identification results of rolling bearing vibration signals

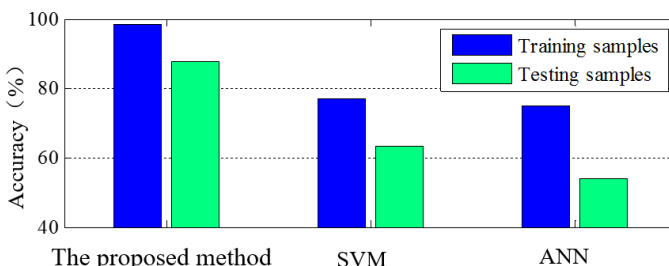| Methods	Identification accuracy	
	Training samples	Testing samples
The proposed method	98.36 %	88.00 %
SVM	77.27 %	63.18 %
ANN	75.09 %	53.91 %

Fig. 7. Classification accuracy of rolling bearing faults

The classification result of the 11 class of samples is shown in the following confusion matrix in Fig. 8. Fig. 8(a) is the confusion matrix for training dataset and Fig. 8(b) is the confusion matrix for testing dataset. The ordinate axis of each confusion matrix is the actual labels of classification and the horizontal axis is the predicted labels.

Although many researches show that SVM and ANN have excellent performance when applied to fault diagnosis, the proposed method performs much better than SVM and ANN. The main reason is that the SVM and ANN are used to analyze the raw data with multiple fault patterns without any feature extractor design and any feature selection in this paper. The current fault diagnosis methods based on SVM ANN require manual feature extractor and feature selection design while the proposed method based on CNN is a multilayer deep learning model, which uses a well-developed trainable topology to replace feature extraction step and can automatically transform the raw input data into suitable internal features to improve the performance. In fact, as Yann Lecun has pointed out, the key aspect of deep learning is that these layers of features are not designed by human engineers: they are learned from data using a general-purpose learning procedure [22].

As a summary, the proposed method is more suitable for the complex fault diagnosis problems without any manual feature extractor and feature selection design than SVM method and ANN method. The experimental result confirms that the proposed method has good classification performance.

5. Conclusions

This paper proposes a multilayer deep learning CNN method for rolling bearing fault diagnosis problems without any feature extractor and feature selection design. The normalization preprocessing method is used to preprocess the rolling bearing vibration signals, and which can avoid the influence of different characteristics of the input data on the identification accuracy.
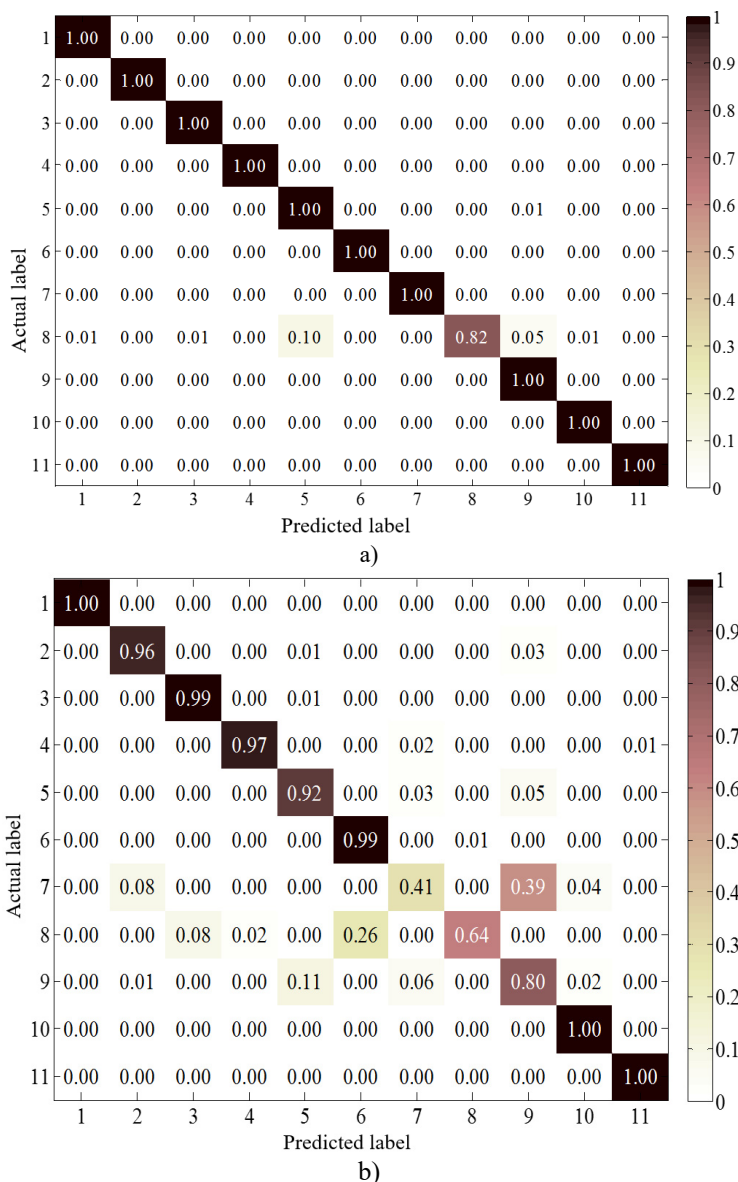
Fig. 8. Result of rolling bearing fault identification: a) training samples; b) testing samples

The proposed method is validated to analyze rolling bearing simulation data and experimental data without any feature extractor and feature selection design. As the analysis result show, the proposed method has better and more robust performance than ANN method and SVM method. The proposed method can automatically learn effective features from vibration signals with high classification accuracy and requires no careful manual intervention. The future study will pay more attention to improve the performance of the proposed method with signal processing methods.

Acknowledgements

References

[1] **Lei Yaguo, Lin Jing, He Zhengjia, et al.** Application of an improved kurtogram method for fault diagnosis of rolling element bearings. Mechanical Systems and Signal Processing, Vol. 25, 2011, p. 1738-1749.

[2] **Rai Akhand, Upadhyay S. H.** A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International, Vol. 96, 2016, p. 289-306.

[3] **Wang Yanxue, Liang Ming** An adaptive SK technique and its application for fault detection of rolling element bearings. Mechanical Systems and Signal Processing, Vol. 25, 2011, p. 1750-1764.

[4] **He Wangpeng, Zi Yanyang, Chen Binqiang, et al.** Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. Mechanical Systems and Signal Processing, Vols. 54-55, 2015, p. 457-480.

[5] **Jiang Hongkai, Li Chengliang, Li Huaxing** An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis. Mechanical Systems and Signal Processing, Vol. 36, 2013, p. 225-239.

[6] **Xiang Jiawei, Zhong Yongteng, Gao Haifeng** Rolling element bearing fault detection using PPCA and spectral kurtosis. Measurement, Vol. 75, 2015, p. 180-191.

[7] **Yu Dejie, Cheng Junsheng, Yang Yu** Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. Mechanical Systems and Signal Processing, Vol. 19, 2005, p. 259-270.

[8] **Tian Ye, Ma Jian, Lu Chen, et al.** Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. Mechanism and Machine Theory, Vol. 90, 2015, p. 175-186.

[9] **Rauber Thomas W., de Assis Boldt Francisco, Miguel Varejão Flávio** Heterogeneous feature models and feature selection applied to bearing fault diagnosis. IEEE Transactions on Industrial Electronics, Vol. 62, 2015, p. 637-646.

[10] **Tian Ye, Wang Zili, Lu Chen, et al.** Bearing diagnostics: A method based on differential geometry. Mechanical Systems and Signal Processing, Vol. 80, 2016, p. 337-391.

[11] **Ma Jian, Lu Chen, Zhang Wenjin, et al.** Health assessment and fault diagnosis for centrifugal pumps using softmax regression. Journal of Vibroengineering, Vol. 16, 2014, p. 1464-1474.

[12] **Lou Xinsheng, Loparo Kenneth A.** Bearing fault diagnosis based on wavelet transform and fuzzy inference. Mechanical Systems and Signal Processing, Vol. 18, 2004, p. 1077-1095.

[13] **Zhang XiaoLi, Wang BaoJian, Chen XueFeng** Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine. Knowledge-Based Systems, Vol. 89, 2015, p. 56-85.

[14] **Hajnayeb A., Ghasemloonia A., Khadem S. E., et al.** Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis. Expert Systems with Applications, Vol. 38, 2011, p. 10205-10209.

[15] **Lashkari Negin, Poshtan Javad, Fekri Azgomi Hamid** Simulative and experimental investigation on stator winding turn and unbalanced supply voltage fault diagnosis in induction motors using artificial neural networks. ISA Transactions, Vol. 59, 2015, p. 334-342.

[16] **Chine W., Mellit A., Malek V., et al.** A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks. Renewable Energy, Vol. 90, 2016, p. 501-512.

[17] **Yin Zuyu, Hou Jian** Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. Neurocomputing, Vol. 174, 2016, p. 643-650.

[18] **Hinton G., Osindero S., Teh Y.** A fast learning algorithm for deep belief nets. Neural Computation, Vol. 18, 2006, p. 1527-1554.

[19] **Kim Sangwook, Choi Yonghwa, Lee Minho** Deep learning with support vector data description. Neurocomputing, Vol. 165, 2015, p. 111-117.

[20] **Schmidhuber Jürgen** Deep learning in neural networks: an overview, Neural Networks, Vol. 61, 2015, p. 85-117.

[21] **Yu Dong, Deng Li** Deep learning and its applications to signal and information processing. IEEE Signal Processing Magazine, Vol. 28, 2011, p. 145-154.

[22] **Yann LeCun, Bengio Yoshua, Hinton Geoffrey** Review: deep learning. Nature, Vol. 521, 2015, p. 436-444.

[23] **Abdel Hamid O., Mohamed A. R., Hui Jiang, et al.** Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, Language Processing, Vol. 22, 2014, p. 1533-1545.
[24] **Wu Haibing, Gu Xiaodong** Towards dropout training for convolutional neural networks. Neural Networks, Vol. 71, 2015, p. 1-10.
[25] **Sainath Tara N., Kingsbury Brain, Saon George, et al.** Deep convolutional neural networks for large-scale speech tasks. Neural Networks, Vol. 64, 2015, p. 39-48.

Hongkai Jiang received Ph.D. degree in Mechanical Engineering School from Xi'an Jiao Tong University, Xi'an, China, in 2006. Now he works in School of Aeronautics, Northwestern Polytechnical University, Xi'an, China. His current research interests include fault diagnosis, health monitoring, and damage identification.
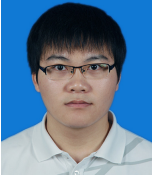
Fuan Wang received Bachelor degree in School of Aeronautics from Northwestern Polytechnical University, Xi'an, China, in 2015. Now he is a graduate student. His current research interests include fault diagnosis, intelligent method.

Haidong Shao received Master degree in School of Aeronautics from Northwestern Polytechnical University, Xi'an, China, in 2015. Now he is a doctoral student. His current research interests include fault diagnosis, intelligent prognosis.

Haizhou Zang received Bachelor degree in School of Aeronautics from Northwestern Polytechnical University, Xi'an, China, in 2015. Now he is a graduate student. His current research interests include fault diagnosis, signal processing.