ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

Application of Deep Neural Network in Fault Diagnosis

Tong Sun

Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China.

Abstract

Rolling bearing is a key component of electromechanical equipment, which plays a vital role in the normal and stable operation of electromechanical equipment. Artificial neural network is widely used in the field of fault diagnosis because it with their nonlinear mapping capabilities can extract fault features from input data. But when the artificial neural network is a kind of shallow learning, its ability to express complex problems is limited. Therefore, this paper proposes a deep neural network (DNN) method for fault diagnosis. This method can extract more advanced and abstract fault features from massive data because of its powerful feature extraction capabilities. Experiments show that deep neural networks have better feature learning and classification performance in the field of fault diagnosis.

Keywords

DNN; Fault Diagnosis; Features; Rolling Bearing.

1. Introduction

With the rapid development of industry, the requirements for electromechanical equipment are getting higher and higher. If a certain part of the mechanical equipment is subjected to unknown interference during operation, it will cause damage to the mechanical equipment, which may lead to economic losses or even more serious consequences for the enterprise. Therefore, the fault diagnosis of the equipment is particularly important. Due to the complexity of the structure and mechanism of the equipment itself, the uncertainty of the parameters and structure, and the dynamic time-varying characteristics, the faults of these equipment are usually characterized by complexity, uncertainty, relevance, and hierarchical characteristics , so diagnosis of early, weak and compound faults of large complex key equipment is more demanding and difficult than the diagnosis of conventional equipment [1]. Traditional fault diagnosis is mostly done by professional technicians and diagnosis experts, so the user's experience and professional knowledge are extremely important. At the same time, due to the complexity of the equipment and the high degree of automation, the amount of data that needs to be analyzed is also very large, so the efficiency of fault diagnosis must be improved [2-7].

Fault diagnosis methods of rolling bearing can be divided into the following three categories: model-based methods, knowledge-based methods and data-driven methods. The model-based fault diagnosis method mainly starts from the dynamic theory to find the cause of the fault, but requires professionals to be familiar with the working principle of the electromechanical equipment, and analyzes based on the mathematical theory through the mathematical model of the fault object. The knowledge-based fault diagnosis method judges the location of the fault based on the empirical knowledge of domain experts, which has industry limitations. The data-driven method does not require precise mechanism models and domain expert knowledge, and only uses various data mining techniques to obtain potentially useful information in the data to diagnose the operating status of the equipment. It is increasingly favored by experts in the engineering field. Machine learning includes shallow learning and deep learning. Shallow learning includes SVM, artificial neural network and extreme learning

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

machine [8-9]. But when the number of samples is too large, the ability of shallow learning for fault diagnosis is limited, which limits its application in the field of fault diagnosis. Compared with shallow learning, deep learning emphasizes the depth of the model and the importance of highlighting feature learning. By building a multi-layer nonlinear network structure to extract its deeper features from the input data, it has been widely used in fault diagnosis.

The existing deep learning fault diagnosis methods can be divided into the following four categories: methods based on convolutional neural networks, methods based on long and short-term memory neural networks, methods based on deep belief networks, and methods based on stacked autoencoders. Convolutional neural network (CNN) extracts the features in the image by stacking the convolutional layer and the pooling layer. The last layer is a fully connected layer, but it cannot perform real-time fault diagnosis [10-11]. The Deep Belief Network (DBN) is a stack of multiple restricted Boltzmann machines (RMB) to achieve a deep structure, but the initialization process of DBN is complicated and the amount of calculation is large [12]. The LSTM extracts features from the data through the gate structure. The forget gate is used to discard useless information, the update gate determines which information needs to be updated, and the output gate determines the output of the LSTM, but the computational complexity is high [13-14]. A deep neural network (DNN) based on an autoencoder can be stacked by multiple autoencoders (AE). The hidden layer of the previous auto-encoder is used as the input of the next auto-encoder, and features are extracted layer by layer. The method of supervision fine-tunes the network. It can directly process one-dimensional time series signals, eliminate redundant information in the data, and has a simple structure and easy implementation.

The other parts of the article are structured as follows: the second part is an introduction to deep learning; the third part is based on the establishment of a DDN fault diagnosis model; the fourth part is the experimental results and analysis; the fifth part is a conclusion.

2. The theory of deep learning

2.1 Autoencoder

In 2006, Hinton proposed an autoencoder. Autoencoder is an unsupervised learning neural network. It is divided into two stages of decoding and encode. The hidden layer features learned from the input data are called encode, and the new features learned to reconstruct the original data is called decoding [15]. The autoencoder is divided into three layers: output layer, hidden layer and output layer. The number of neurons in the input layer and output layer is the same. The structure of the autoencoder is shown as in Figure 1.

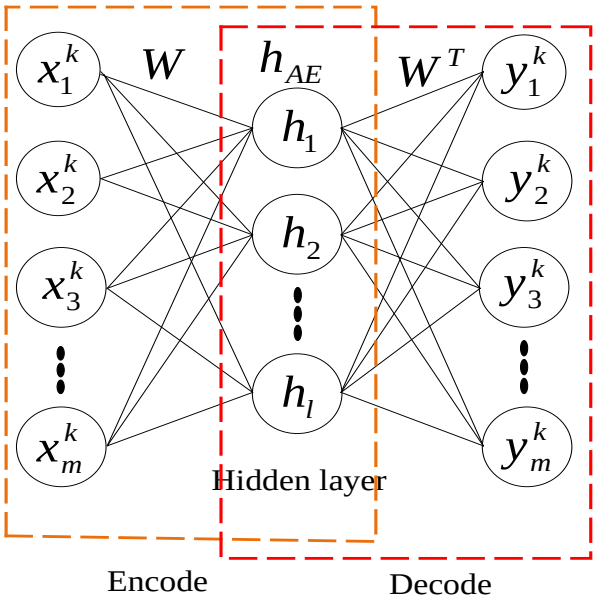

Figure 1. Structure of the autoencoder

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

Given an unlabeled dataset $\{x_m^k\}$, where m is the number of input neurons of the autoencoder and k is the sample size of the dataset. The coding process of the autoencoder is shown in the formula (1):

$$h = f_{\theta_1}(x^k) = relu(Wx^k + b) \tag{1}$$

where f_{θ_1} is activation function of encode. W is weight between the input layer and the hidden layer of autoencoder. b is bias of encode, and $\theta_1 = \{W, b\}$ is parameter set of the encode. $relu$ is the activation function, as shown in the formula (2):

$$relu(x) = \begin{cases} x, x > 0 \\ 0, x < 0 \end{cases} \tag{2}$$

Decode is the reverse process of encode. The hidden layer is used as input for decode.The process is shown in the formula (3):

$$y^k = g_{\theta_1}(h) = relu(W^T h + b^T) \tag{3}$$

Where g_{θ_1} is the activation function of the decode process. W^T is weight between hidden layer and output layer of autoencoder. ${\theta_1}^T = \{W^T, b^T\}$ is the parameter set of the decode, and y^k is reconstructed value of input after decode.

In order to make the output of the autoencoder as equal to the input as possible, its parameters need to be optimized. The error $loss_1(x^k, y^k; \theta_1, \theta_1^T)$ between input and output can be used to optimize the parameter . The reconstruction error of the autoencoder is shown in the formula(4):

$$loss_1(x^k, y^k; \theta_1, \theta_1^T) = \frac{1}{K}\|y^k - x^k\| \tag{4}$$

The parameters of the auto-encoder can be updated by gradient descent. The update process is shown in the formula:(5) and(6):

$$W = W - lr \cdot \frac{\partial}{\partial W} loss_1(x^k, y^k; \theta_1, \theta_1^T), W^T = W^T - lr \cdot \frac{\partial}{\partial W^T} loss_1(x^k, y^k; \theta_1, \theta_1^T) \tag{5}$$

$$b = b - lr \cdot \frac{\partial}{\partial b} loss_1(x^k, y^k; \theta_1, \theta_1^T), b^T = b^T - lr \cdot \frac{\partial}{\partial b^T} loss_1(x^k, y^k; \theta_1, \theta_1^T) \tag{6}$$

Where lr is learning rate.

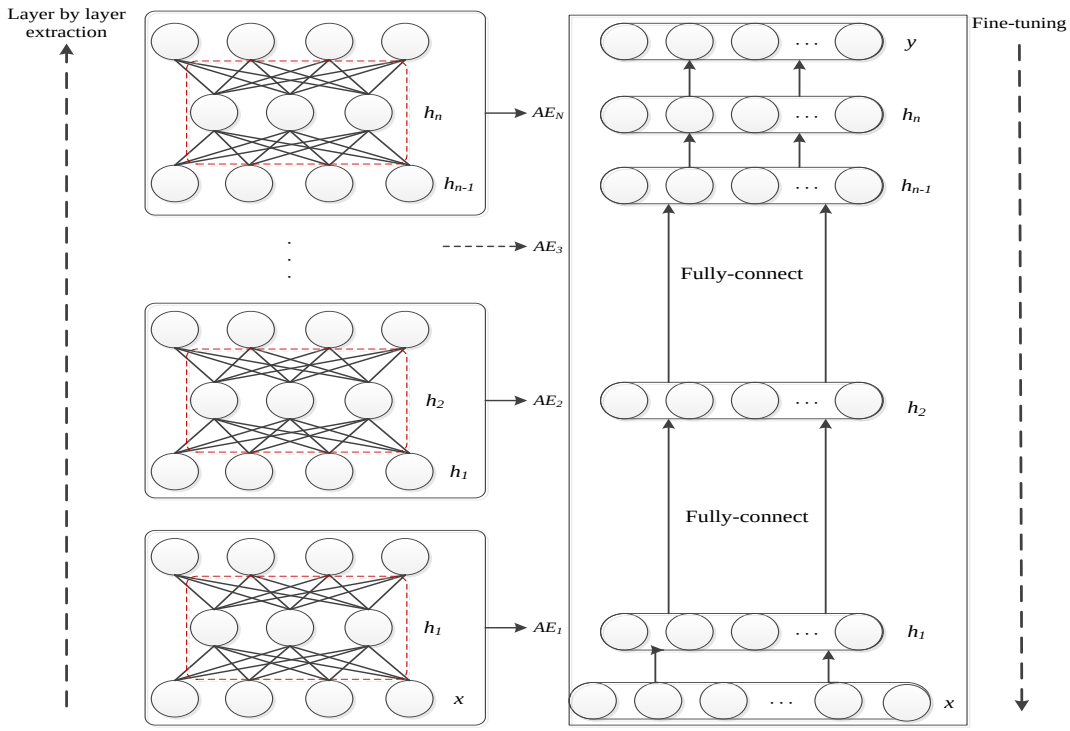

Figure 2. Construction process of DNN

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

2.2 Deep neural network

DNN can be constructed by stacking multiple autoencoders.Bengio Y proposed a stacked autoencoder based DNN model by stacking multiple autoencoders to extract high-order features of input data layer by layer. The specific process of stacked autoencoder is as follows: given the input, train the first autoencoder. The hidden layer output of the first autoencoder is used as the input of the second autoencoder.Repeat the above method, use gradient descent method to train all autoencoders. Then use a supervised method to train to fine-tune the network.The structure of DNN is shown in Figure 2.

3. Fault diagnosis based on deep neural network

The DNN algorithm based on autoencoder is divided into the following steps:

(1) Take sample data X as input data, and use W_1 and b_1 to get hidden layer output of the first autoencoder. The formula (7) is:

$$h_1 = f(W1X + b1) \tag{7}$$

(2) As input data h_1, use W_2 and $b2$ to get the output of second hidden layer, as shown in the formula (8):

$$h_2 = f(W2h_1 + b2) \tag{8}$$

(3) The hidden layer output of (N-1) autoencoders is used as the input of the Nth autoencoder, as shown in the formula(9):

$$h_N = f(WNh_{N-1} + bN) \tag{9}$$

(4) h_N will be used as the input of the softmax classifier, and the labeled data will be used as the training data of the DNN to fine-tune the model parameters through the gradient descent algorithm.

(5) Use test data as the input of the trained DNN to obtain fault classification results.

4. Fault diagnosis based on deep neural network

4.1 Experimental data

The experiment uses the rolling bearing data set of Case Western Reserve University [16] for training to verify the effectiveness and feasibility of the proposed method. Experimental platform of rolling bearing is shown in Figure 3. Use an accelerometer to collect vibration data. The accelerometer uses a magnetic base to attach to the platform to collect vibration signals and then transmit them to a computer. In this paper, the bearing data with a sampling frequency of 12KHz is divided into 6 categories: normal state, inner ring fault, ball fault and 3 different outer ring faults with respect to the load zone position.

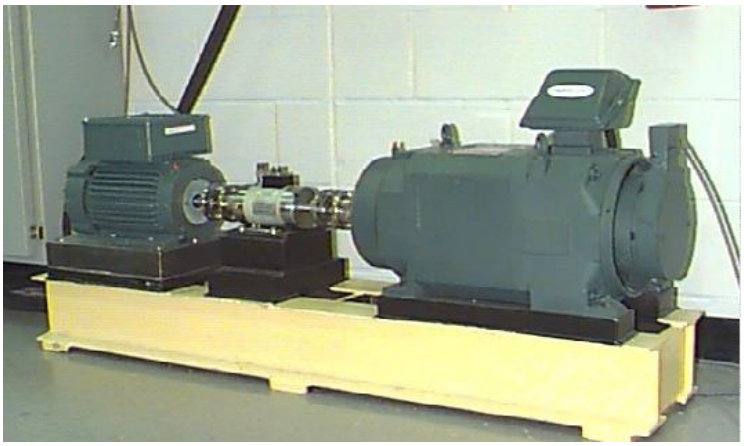

Figure 3. Experimental platform of rolling bearing

Table 1 is the description of the experimental data. The training samples are 100, 200 and 300 respectively, and the corresponding test samples are also 100, 200 and 300.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

Table 1. Experimental data

Fault type	Training samples	Training samples	Category
normal	100/200/300	100/200/300	0
inner race fault	100/200/300	100/200/300	1
ball fault	100/200/300	100/200/300	2
outer race fault 1	100/200/300	100/200/300	3
outer race fault 2	100/200/300	100/200/300	4
outer race fault 3	100/200/300	100/200/300	5
total	600/1200/1800	600/1200/1800	6

4.2 Analysis of experimental results

The neural network structure in this article is five layers, in which the number of neurons in the input layer is 400, the number of neurons in the first hidden layer, the second hidden layer and the third hidden layer are 200, 100, and 50 respectively, and then added Softmax, the number of neurons in the output layer is determined by the number of fault classes, so the number of neurons in the output layer is 6. The activation function used in the forward process is the Relu function. The weight initialization of DNN is random. Compare the network parameter settings of the experiment. In BPNN, the network structure is 400-200-50-6, and the activation function used is the Relu function. When there are two AE stacks, the network parameters are set to 400-200-100-4, and the activation function used is the Relu function. The parameters of network during DNN training are shown in Table 2.

Table 2. Parameters of network

Model of network	Nunber of AE	Number of neurons	Learning rate
BPNN	-	400-200-50-6	0.2
DNN2	2	400-200-100-6	0.2
DNN3	3	400-200-100-50-6	0.2

Table 3. Fault diagnosis result with sample number of 600

Model of network	BPNN	DNN2	DNN3
Test accuracy	69.50%	75.96%	83.93%

Table 4. Fault diagnosis result with sample number of 1200

Model of network	BPNN	DNN2	DNN3
Test accuracy	77.54%	93.25%	95.63%

Table 5. Fault diagnosis result with sample number of 1800

Model of network	BPNN	DNN2	DNN3
Test accuracy	81.89%	95.62%	96.39%

It can be seen from Table 3 to Table 5 that the fault diagnosis accuracy of each model is continuously increasing with the increase of samples. Comparing columns 3 and 4 of Tables 3 to 5, it can be seen that the fault diagnosis accuracy of DD3 is higher than that of DD2. This shows that with the increase of hidden layers, the accuracy of fault diagnosis also increases accordingly. Under different samples, the fault diagnosis accuracy of DD2 and DD3 is higher than that of BPNN.

Figure 4 to Figure 6 are the fault diagnosis results of models when number of samples is 600.The red star is the result of the fault diagnosis of each sample, and the blue circle is each real label.When the red star coincides with the blue circle, the fault diagnosis result is correct.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

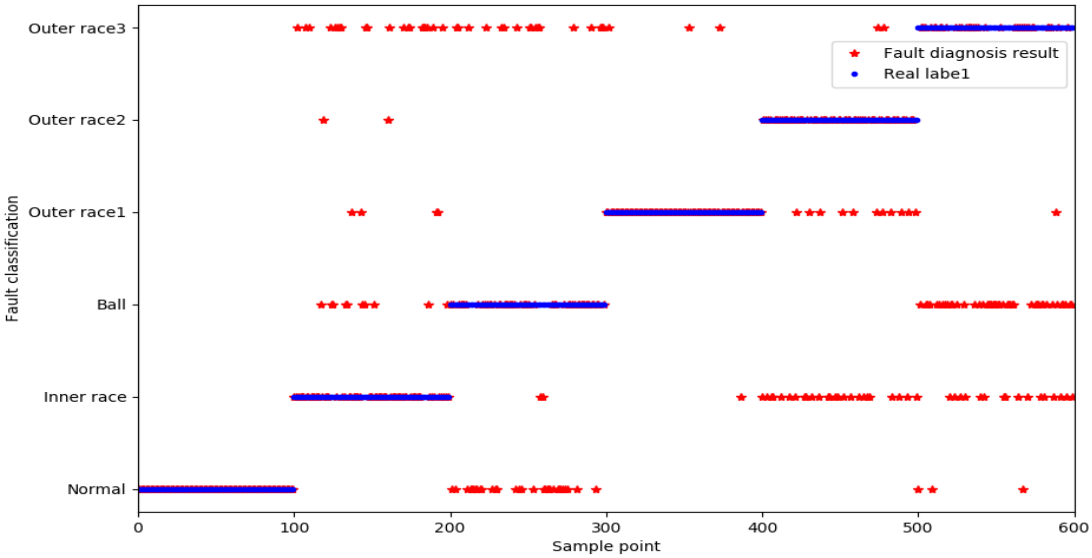

Figure 4. Fault diagnosis results of BPNN

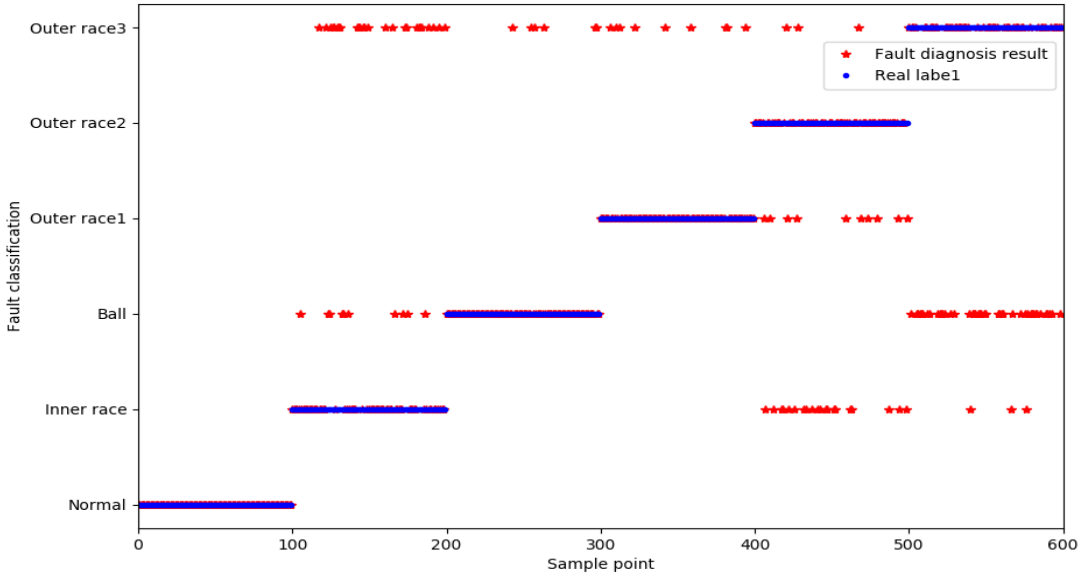

Figure 5. Fault diagnosis results of DD2

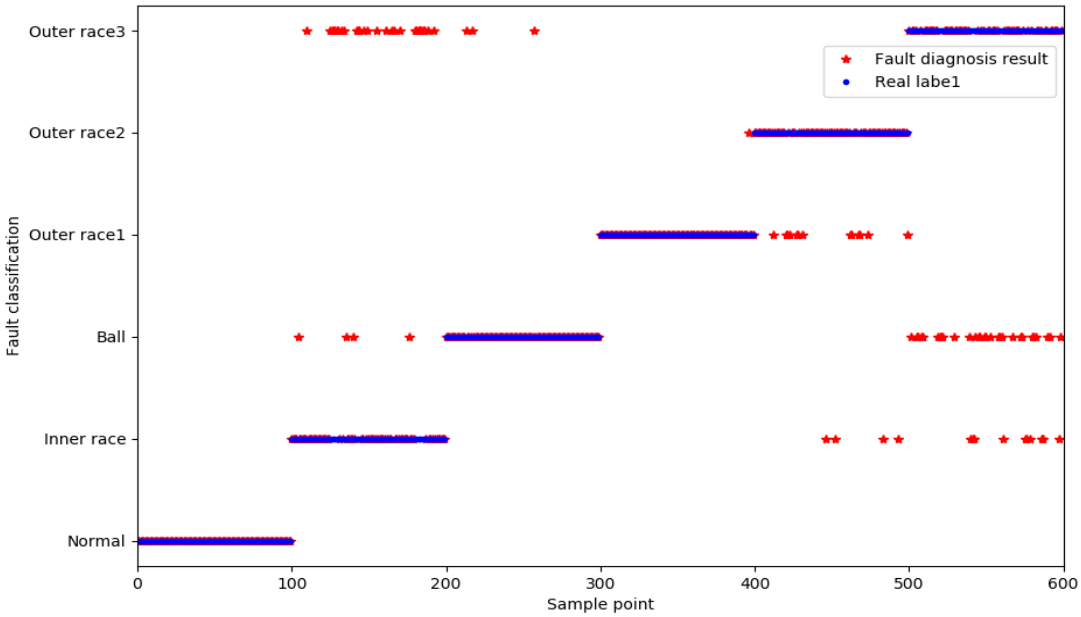

Figure 6. Fault diagnosis results of DD3

ISSN: 2414-1895 DOI: 10.6919/ICJE.202107_7(7).0010

5. Conclusion

This paper uses a deep neural network with autoencoder to diagnose faults in bearing data, and compares the influence of the number of layers and the number of samples on the results of fault diagnosis. Compared with the BPNN,DNN has more powerful feature extraction capabilities than the shallow learning, and has a good performance on the fault diagnosis results.

References

[1] Lei Y G. Research on Hybrid Intelligent Technique and Its Applications in Fault Diagnosis, 2007.

[2] Wong M L D, Jack L B, Nandi A K. Modified self-organising map for automated novelty detection applied to vibration signal monitoring [J]. Mechanical Systems and Signal Processing, 2006,20: 593 -610.

[3] Lei Y G, He Z J,Zi Y Y. Application of an intelligent fault diagnosis method to rotating machinery[J]. Expert Systems with Applications,2009,36: 9941 -9948.

[4] Jardine A K S, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J].Mechanical Systems and Signal Processing,2006,20: 1483 -1510.

[5] Lei Y G, He Z J, Zi Y Y, et al. New clustering algorithm based fault diagnosis using compensation distance evaluation technique[J].Mechanical Systems and Signal Processing,2008,22: 419 -435.

[6] Yang W X. Establishment of the mathematical model for diagnosing the engine valve faults by genetic programming[J].Journal of Sound and Vibration,2006,293: 213 -226.

[7] Lei Y G, He Z J, Zi Y Y, et al. A new approach to intelligent fault diagnosis of rotating machinery [J]. Expert Systems with Applications,2008,35: 1593-1600.

[8] Shu-Ying Li,Lei Xue. Motor's Early Fault Diagnosis Based on Support Vector Machine[J]. IOP Conference Series: Materials Science and Engineering, 2018, 382(3).

[9] Arjun Kumar Shah,Ashish Yadav,H. Malik. EMD and ANN based intelligent model for bearing fault diagnosis[J]. Journal of Intelligent & Fuzzy Systems, 2018, 35(5).

[10] Xia M, Li T, Xu L, Liu L, et al. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks, in IEEE/ASME Transactions on Mechatronics, 2018,23(1):101-110.

[11] Wen L, Li X, Gao L and Zhang Y.A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method, in IEEE Transactions on Industrial Electronics,2018,65(7):5990-5998.

[12] Shao H, Jiang H, Zhang X, et al. Rolling bearing fault diagnosis using an optimization deep belief network[J]. Measurement Science and Technology, 2015,26(11):115-123.

[13] Yang R, Huang M, Lu Q and Zhong M.Rotating Machinery Fault Diagnosis Long-short-term Memory Recurrent Neural Network,IFAC-PapersOnLine,Volume 51, Issue 24,2018,51(24):228-232.

[14] Yu L, Qu J, Gao F, Tian Y and Emiliano Mucchi.A Novel Hierarchical Algorithm for Bearing Fault Diagnosis Based on Stacked LSTM[J].Shock and Vibrationvol,2019:1-10.

[15] 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.

[16] Bearing data Centre, Case Western Reserve University, Available: http://csegroups.case.edu/ bearingdatacenter/home.