Geometric Features Fusion Method based deep Learning for fault Classification

Funa Zhou[1], Po Hu[1*], Shuai Yang[1]

1. School of Computer and Information Engineering, Henan University, Kaifeng, China
E-mail: zhoufn2002@163.com, hupo4210@163.com(*Corresponding,Author), yangshuai2711@163.com

Abstract: During the operation of mechanical equipment, various types of faults need to be monitored in real time. Some frequency faults can not be well detected in the time domain, which makes accurate real-time detection of mechanical equipment impossible. To solve this problem, a fault diagnosis method based on DNN fusion of geometric features is proposed. By introducing the geometric features of the original data, the problem of the time domain detection being inaccurate when the frequency class faults exist in similar amplitudes is solved. The original, slope and curvature data are respectively constructed into an automatic encoder model to fully excavate the implied frequency information of the data. The original, slope and curvature data are respectively constructed into an automatic encoder model to fully excavate the implied frequency information of the data. The experimental results of the rolling bearing show that the method proposed in this paper can detect the frequency fault well.

Key Words: DNN; Fault diagnosis; DAE; Feature fusion

1. INTRODUCTION

Rolling bearings are the core components of rotating machinery and are used in a wide range of industrial process control applications. Due to the bearing in use process will appear different degree of wear phenomenon, if not timely monitoring of the bearing failure will result in serious losses, and even endanger the lives of the operator[1-4]. Therefore, bearing fault diagnosis in the actual production process is very necessary.

At present, the methods of mechanical equipment fault diagnosis mainly include qualitative model-based method, quantitative model-based method and data-driven method [5-7]. Methods based on both qualitative and quantitative models require a deep understanding of specific diagnostic objects and the ability to abstract accurate mathematical models. These methods are too dependent on the prior knowledge of experts to make them less standardized. Data-driven methods include statistical feature extraction and deep learning methods [8-11]. The method based on statistical feature extraction can only detect the fault and can not classify the fault. In recent years, due to the rise of deep learning methods, deep learning based methods are widely used in the fault diagnosis of complex systems, such as convolutional neural networks, recurrent neural networks and deep neural networks.

In the bearing fault diagnosis, the vibration signal is the easiest to collect. And the vibration signal is more sensitive to bearing failure. Therefore, vibration signals are widely used in bearing fault diagnosis. However, due to the non-linearity, instability, high latitude and the large amount of noise pollution, the failure characteristics of mechanical equipment make it impossible to accurately diagnose mechanical equipment [12-14]. Some scholars put forward the use of signal processing feature extraction methods combined with machine learning methods for mechanical equipment failure classification. Widodo and Yang extract frequency-domain features as SVM data sources to detect mechanical problems [13-17]. When the number of samples is small and the signal is non-stationary, Yu et al. Propose a method of rolling bearing fault classification using a combination of SVM and EMD methods [10]. Hu et al. extracted the energy of each node of Wavelet Packet Transform (WPT) from the vibration signal as the characteristic parameter of bearing fault diagnosis, which greatly preserved the time-frequency characteristic of the characteristic information. The combination of WPT and SVM Fault diagnosis, improve the accuracy of fault diagnosis [18-21]. Based on the non-stationary characteristics of vibration signals of rolling bearings, Wang et al. Used WPT to de-noise the collected signals and extract the energy characteristics of wavelet bands of each frequency band as the input characteristics of Artificial Neural Network (ANN) Learning classification ability and self-organization ability to bearing fault classification and diagnosis [22-26]. Yang and Tang et al. Proposed a method that combined the expert system with the BP neural network (BPNN), which fully utilized the advantages of the expert system and ANN and successfully detected the bearing failure [27]; Jiang et al. Proposed a method using a combination of high-order cumulants with BPNN. This method uses high-order statistics as the eigenvector to improve the accuracy of BPNN in bearing fault diagnosis [28,29]. However, in these studies, SVM and BPNN have many shortcomings as a shallow learning method: the essence of SVM is a dichotomizer, which has low learning efficiency in many classification and large sample problems, how to choose suitable kernel function and scale parameter Often need to experience, SVM method can not be real-time monitoring

This research was supported in part by the Natural Science Fund of China (Grant No. U1604158).

and diagnosis, can not meet the current real-time monitoring of machinery and fault diagnosis requirements; ANN also has some congenital defects: (1) ANN as a shallow neural network, The convergence speed is slow, easily fall into the local optimum, can not well characterize the signal characteristic information; (2) learning complex non-linear data has the disadvantage of low efficiency and low classification accuracy. In summary, as a shallow learning method, SVM and BPNN have not been able to effectively extract features under high-dimensional non-steady-state data [30].

Deep learning is a promising feature extraction tool and has aroused widespread concern of scholars [21,28-30]. Compared with shallow learning, deep learning can perform well feature extraction and in-depth study on nonlinear big data [31,32]. Depth neural network (DNN) adopts unsupervised layer-by-layer greedy training algorithm and global parameter tuning based on BP algorithm, which can not only avoid local optimal problems, but also solve the restriction of the number of labels and samples. Hinton and Salakhutdinov first proposed the theory of depth learning in 2006. This theory is used to construct Deep Neural Network (DNN), which can form a more abstract high-level representation by combining nonlinear transformations of multiple layers and combining lower-level features, so that a learning system can discover the distribution of data without depending on artificial feature selection Expression of features, and learn complex expression function [33]. Due to its excellent feature extraction capabilities, it quickly attracted the attention of experts in the field of fault classification. Lu et al. used the good feature extraction ability of deep neural network and used it to diagnose the bearing fault successfully solved the problem that the traditional feature extraction method can not find the unknown type fault timely and effectively [34]; Jia et al. Used deep neural network to detect rolling Shaft bearing health status [35]; Gan et al. Proposed a fault diagnosis method based on the characteristics of bearing faults. By constructing a two-layer deep neural network, this method can not only accurately locate the fault location of the bearing, but also effectively excavate the fault size of the bearing at the same location . Deep learning, as one of the most popular and hot machine learning methods in the world today, has revolutionized artificial intelligence. However, our application of deep learning is still in its infancy, and there are many aspects that can be improved in the application process. For example, the deep learning method is hard to be real-time for the mechanical equipment frequency fault. Although the frequency fault can be well detected in the frequency domain, it can not guarantee the real-time performance of the detection. In the time domain, the detection frequency class Failure is very difficult. In order to ensure the real-time nature, the real-time detection of the equipment is needed in practice. Therefore, it is necessary to detect the frequency fault in the time domain effectively.

In order to solve the above problems, this paper presents a fault diagnosis method based on DNN fusion of geometric features. The main innovation is to solve the problem of inaccurate time-domain detection when frequency class faults have similar amplitudes and frequent zero crossings by introducing the geometric features of the original data. The original, slope and curvature data are respectively constructed into an automatic encoder model, so as to fully excavate the implied frequency information of the data. Then, the obtained data features are fused and the feature dimension is increased to make the classification model training more accurate. First calculate the slope of the original data, curvature and other data that can represent the frequency characteristics of the value. The second step is to construct three DNN networks for training of each feature. The third step is to fuse the obtained features. The final step is to use the already trained network to diagnose the new samples in real time.

The remaining part of this paper is organized as follows: Section 2 introduces theory of deep learning; Section 3 a fault online diagnosis method based on DNN fusion of geometric features is proposed; Section 4 the validity of the proposed fault diagnosis method is obtained through experiments and simulation analysis; Sections 5 is the conclusion.

2. THEORY OF DEEP LEARNING

Deep learning is an unsupervised learning method. By multi-layer nonlinear transformation, low-level features combined to form more abstract high-level representation, making the learning system can not rely on artificial feature selection, to find the distributed representation of the data, and to learn complex expression functions. Deep learning adopts unsupervised learning to pre-train DNN layer by layer, which helps DNN effectively excavate the fault features in mechanical signals. Then, DNN is fine-tuned by supervised learning to optimize the expression of DNN's fault features and make them have monitoring and diagnostic capabilities. In this paper, DNNs are pre-trained by stacking autoencoders (AE).

2.1 Auto-Encoder

Auto-Encoder is a three-layer unsupervised neural network. Including coding network and decoding network in two parts, as shown in Fig.1, the main input layer, hidden layer and output layer. The input data of AE is the same as the output target. The input data in high-dimensional space is converted into the encoding vector in low-dimensional space through the encoding network, and the encoded vector in low-dimensional space is reconstructed back to the original input data through the decoding network. Since the input signal can be reconstructed at the output layer, the encoded vector becomes a characterization of the input data.

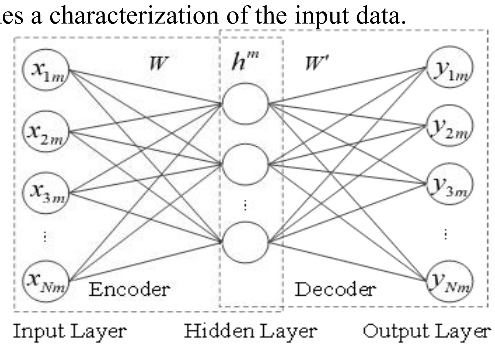

Fig.1 The model of Auto-encoder

Given an untagged data set $\{x_{nm}\},(n=1,2,...,N;m=1,2,...,M)$ consisting of N observation features or variables, each observation variable has M samples. the encoding and decoding process of auto-encoder can be expressed as:

$$h^m = f(x_{nm}) = \sigma_1(Wx_{nm}+b) \quad (1)$$

$$y_{nm} = g(h^m) = \sigma_2(W^T h^m + d) \quad (2)$$

Where σ_1 is the activation function for the encoding network, W is the weight of the input layer and the hidden layer, b is the bias vector of the encoding network, h^m is the activation value of the hidden layer, which is the features of input data x_{nm} ; σ_2 is the activation function of decoding network, W^T is connected to the hidden layer and output layer weights, d is the bias vector for the decoding network, y_{nm} is the network output, which is the reconstruction value of input x_{nm} . σ_1 and σ_2 is Sigmoid function. The Sigmoid function can be depicted as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (3)$$

The essence of training AE is to train and optimize the network parameters W and W^T . In order to make the output y_{nm} as close as possible to the input x_{nm} , we need to optimize the network training parameters. Describe the proximity between input and output by minimizing the reconstruction error $J(x,y;W,b)$ .The reconstruction error as shown in Equation 4:

$$J(x,y;W,b) = \frac{1}{m}\|y-x\|^2 \quad (4)$$

In training process, gradient descent method is used for AE parameter optimization, the concrete update form of network parameters can be formulated as formula (5) - (6).

$$W_l = W_l - \alpha \frac{\partial}{\partial W_l} J(x,y;W,b), l=1,2 \quad (5)$$

$$b_l = b_l - \alpha \frac{\partial}{\partial b_l} J(x,y;W,b), l=1,2 \quad (6)$$

Where α is the learning rate, $\frac{\partial}{\partial W_l} J(x,y;W,b)$ and $\frac{\partial}{\partial b_l} J(x,y;W,b)$ can be calculated by back propagation algorithm.

DNN can be simply seen as a multi-hidden layer neural network in which multiple AE layers are stacked. The bottom-up unsupervised learning method is used to extract features layer-by-layer and fine-tune the entire network with supervised learning methods . So that the DNN can extract the most essential characteristic attribute of some state of the object from the original data data. DNN structure shown in Fig. 2.

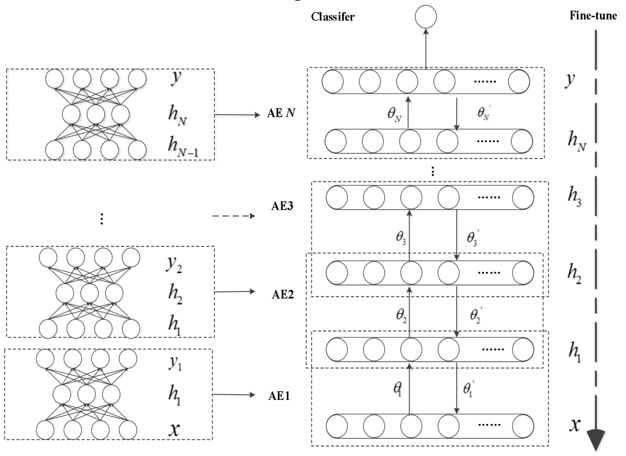

Fig.2 The structure of DNN

First, the DNN network is pre-trained by the unsupervised layer-by-layer greedy training algorithm. First, given an untagged input data set x as the input to the coding network to practice the first automatic encoder $AE1$, the coded vector h_1 is obtained. Training parameter W_1 is obtained by setting x as the output of $AE1$. h_1 is then used as the input to the second automatic encoder $AE2$ and trained on the network parameter W_2 of $AE2$. h_2 as hidden layer data of $AE2$ can be seen as a characteristic representation of $AE2$. This process is repeated to obtain the hidden layer feature h_N of the Nth automatic encoder AEN and the corresponding network training parameter W_N .

Second, add a classifier to the top of the DNN network. The DNN pre-training process is completed through unsupervised training layer by layer, and the layers of feature information are extracted. However, DNN at this time does not have the classification function, in order to achieve the output classification function, but also need to add a DNN classifier on the top. In this paper, we use the Softmax classifier as the output layer of DNN. We use feature h_N of the last hidden layer and the labeled data $L=(1,2,...,k)$ to train the Softmax classifier. The probability $p(x_i = L_m | x_m)$ of each type can be calculated by the following hypothetical function:

$$h_\theta(x_i) = \begin{bmatrix} p(L_i=1|x_i;\theta) \\ p(L_i=2|x_i;\theta) \\ \vdots \\ p(L_i=k|x_i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \vdots \\ e^{\theta_k^T x_i} \end{bmatrix} \quad (7)$$

$$label = a\operatorname*{rg\,max}_{j=1,2\cdots,k}\{p(L_i=j|x_i;\theta),\} \quad (8)$$

Where, θ is the model parameter of Softmax. The model parameters can also be optimized by the same means to minimize the cost function. The cost function of Softmax classifier can be defined as follows:

$$J(\theta) = -\frac{1}{M}[\sum_{i=1}^{M}\sum_{j=1}^{k} 1\{L_i = j\} \log \frac{e^{\theta_j^T x_i}}{\sum_{m=1}^{k} e^{\theta_m^T x_i}}] \quad (9)$$

Where $1\{\bullet\}$ is indicate function.

Finally, we fine-tune the parameters . For the accuracy of feature extraction and output layer classification, the entire DNN training parameters are supervised finely with a finite number of sample tags and a back propagation algorithm, and the fine tuning process is completed by minimizing the reconstruction error $E(\theta)$. The process of parameter update is as follows:

$$E(\theta) = \frac{1}{M}\sum J_\theta(y_{m,}L_m;\theta) \quad (10)$$

$$\theta = \theta - \alpha \frac{\partial E(\theta)}{\partial \theta} \quad (11)$$

where y_m represents the actual output value, θ is a parameter set generated from the whole network training, $\theta = \{\theta_1, \theta_2, ..., \theta_N\}$, back propagation algorithm is used to update the network parameter θ , and α is the learning rate in the process of deep learning. The fne-tuning process uses the labeled data to improve the performance of DNN.

2.2 DNN-Based Classification

In order to accurately use the DNN model to extract the essential characteristics of the health status of mechanical equipment from the input samples, the following steps are required: First, the data of the collected vibration signals should be preprocessed. Therefore, the corresponding slope and curvature values of the original time-domain vibration signal are obtained to reflect the frequency information of the original signal. Secondly, the preprocessed data is used as the DNN model input unsupervised layer pre-training to extract the time-frequency characteristics of the mechanical equipment health status. Finally, based on the finite number of samples of the sample, the entire network is fine-tuned using the backpropagation algorithm to update the entire network parameters θ . In this way we can carry out an effective classification and diagnosis of the health status of machinery and equipment. Mechanical equipment fault diagnosis is usually divided into two processes: training process and testing process. The preprocessed datasets are divided into training data and testing data. The training data is used to construct and train DNN model, and the training parameters θ are obtained. The training parameters θ and test data are used to verify the diagnostic results of the constructed model. The number of misclassifications is taken as a reference index for the classification accuracy of DNN. DNN for mechanical system troubleshooting detailed steps shown in Figure 3.

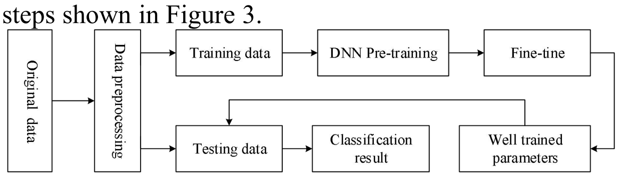

Fig.3 DNN based fault classification diagram

3. DEEP LEARNING FAULT DIAGNOSIS MODEL BASED ON GEOMETRIC FEATURES

There are many frequency system faults in the international system. For the processing of such faults, the majority of the original data is Fourier transformed to obtain the frequency domain information and then processed, which will result in the failure to achieve real-time detection. In this paper, the slope and curvature data of the original data are obtained to increase the time domain Signal frequency information. The corresponding self-encoders are respectively constructed to extract the features of the original data, the slope data and the curvature data, respectively, and then the features are fused and finally classified. The network structure shown in Figure 4.

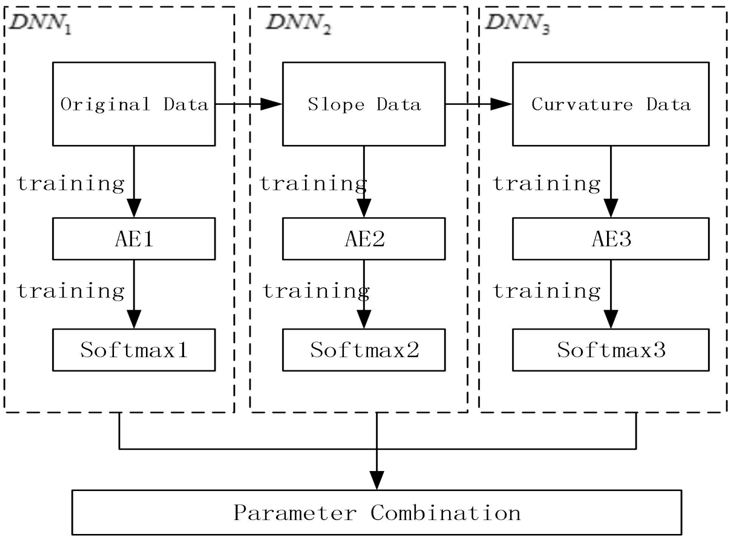

Fig.4 Network structure

The detailed steps of frequency fault monitoring are as follows:

Step 1: Calculate the slope and curvature data for the raw data.

In this step, we calculate the slope curvature data corresponding to the raw data for each moment. Used to get frequency class information in the time domain.

Step 2: Create DNN network of raw data

We establish a DNN model to extract the fault characteristics of the original data respectively. Feature extraction process is described in detail as follows:

(1) Construct a DNN1 network with N hidden layers, and initialize DNN1 training parameters.

$$[\mathrm{Net}_1, \mathrm{Tr}_1] = \mathrm{Feedforward}(\theta_{11}^{'}; H_{11}, H_{12}, ..., H_{1N}; \mathrm{S}_1) \quad (12)$$

Among them $\theta_{11}^{'} = \{W_1, b_1\}$, W_1 is the weight matrix, b_1 is bias vector $H_{11}, H_{12}, ..., H_{1N}$ is the number of DNN1 hidden layer neurons. Network configuration saved in Tr_1 . S_1 representing raw training data set. The number of input neurons of DNN1 is expressed by using (13) M_{11}.

$$M_{11} = size(\mathrm{S}_1) \quad (13)$$

The parameters of DNN1 are initialized by (14) - (15)

$$W_1 = rand(H_{11}, M_{11}) \quad (14)$$

$$b_1 = zeros(H_{11}, 1) \quad (15)$$

Applying unsupervised layer-by-layer feature extraction to datasets by training DNN1 for parameters

$$h_1 = f_{\theta_{11}^{'}}(\mathrm{S}_1) = \sigma(W_1 \cdot \mathrm{S}_1 + b_1) \quad (16)$$

$$h_N = f_{\theta_{1N}^{'}}(h_{N-1}) \quad (17)$$

A softmax classifier is added at the top of DNN1, and a limited set of labels are used for the reverse adjustment of DNN1 and the update of training parameters through (18) - (19).

$$E(\theta_1^{'}) = \min \frac{1}{m} \sum J_{\theta_1^{'}}(Y_1, \rho_1; \theta_1^{'}) \quad (18)$$

$$\theta_1^{'} = \theta_1^{'} - \alpha_1 \frac{\partial E(\theta_1^{'})}{\partial \theta_1^{'}} \quad (19)$$

$\theta_1^{'} = \{\theta_{11}^{'}, \theta_{12}^{'}, ..., \theta_{1N}^{'}, \theta_{1(N+1)}^{'}\}$, $\theta_{1(N+1)}^{'}$ can be calculated by (7) - (8), m is the number of samples. Y_1 indicates the output of DNN_1 , α_1 is the learning rate of the reverse trimming process.

(2) Using the trained DNN_1 to classify the faults, the probability $\mathrm{S}_1^{'}$ of each test sample $\mathrm{S}_1^{'}$ can be calculated from the tested samples and the trained network Net_1 , and then the type of the test samples will be output by using Equation(20).

$$\mathrm{Mode}(m) = \arg\max_b \mathrm{P}(Y_1(m) = b \mid \mathrm{S}_1^{'}(m); \theta_1^{'}; \mathrm{Tr}_1) \quad (20)$$

$\mathrm{Mode}(m)$ is the modal type of the sample, and indicates the fault category of the m th test sample. Fault classification labels obtained by real fault tags and network tests Calculate the number of misclassifications of error classification:

$$e_1 = size(S\_miss) \quad (21)$$

$size$ is the size of the feature dataset and S_miss is the wrongly categorized dataset.

$$S\_miss = \{x_m \mid \arg_{x_m} Mode(m) \neq Label(m)\} \quad (22)$$

Step 3 : Establish a DNN network for slope data

(1) In order to obtain the frequency domain information of the original data, we construct a second hierarchical model whose training data is slope data, S_2 represents the DNN_2 training set.

$$[\mathrm{Net}_2, \mathrm{Tr}_2] = \mathrm{Feedforward}(\theta_{21}^{'}; H_{21}, H_{22}, ..., H_{2N}; S_2) \quad (23)$$

The DNN_2 parameter initialization mechanism and step 1 the same.

Training DNN_2 to get the network parameters, similarly, detailed calculation process reference (16) - (19).

(2) The network parameters obtained through training

$$[\mathrm{Net}_2, \mathrm{Tr}_2] = \mathrm{Feedforward}(\theta_{21}^{'}; H_{21}, H_{22}, ..., H_{2N}; \mathrm{S}_2) \quad (24)$$

Based on the training, slope data can be diagnosed by equation (25).

$$F_2(m) = \arg\max_c P(Y_2(m) = c \mid S_2^{'}(m); \theta_2^{'}; \mathrm{Tr}_2) \quad (25)$$

By comparing with the actual label to calculate the number of error classification e_2

Step 4: Establish a DNN network for curvature data

In order to further extract the slope characteristics of the data, a third DNN network is designed. Build a third deep network Net_3 , S_3 is a training dataset of DNN_3 · $S_3^{'}$ is a test dataset .The parameter training process is similar to step 2, in $S_3^{'}$,the fault type of the mth sample can be calculated by equation (26).

$$F_3(m) = \arg\max_{d} P(Y_3(m) = d \mid S_3^{'}(m); \theta_3^{'}; \mathrm{Tr}_3) \quad (26)$$

Step 4: Feature fusion.

In this paper, the model of DNN which is established by the original, slope and curvature data respectively can fully extract the fault components of the frequency class contained in the fault signal. In order to make the final diagnosis accuracy higher, we must make full use of the trained network to extract the frequency and time domain features of the data. Feature fusion is to recombine the high-level features extracted by these three automatic encoders. Then use the trained softmax classifier to classify the recombined features. The combination of features is as Fig. 5.

The method of deep neural network proposed in this paper is shown in Fig.6.

4. EXPERIMENT AND ANALYSIS OF BEARING FAULTS

IRolling bearings play a crucial role in rotating machinery, bearing health will directly affect the reliability and stability of the entire system. In this paper, the rolling bearing is used as an experimental platform to verify the effectiveness of the fault diagnosis method for geometric feature fusion. The proposed method is compared with the neural network method without feature fusion.

4.3 Description of Experimental Platform

The experimental data set used in this paper is collected by Henan University of Technology。Acceleration sensor is used in the experiment to collect the vibration signal of motor drive end as experimental data of bearing fault diagnosis. In this experiment, the acceleration sensor is used to collect the motor-driven vibration signal of 0hp, the sampling frequency is 48kHz. There are four types of fault diameter :(1)0.007、(2)0.014、(3)0.021 (4) normal condition. Fourier transform is used to preprocess the vibrations signal to get 4500 training samples and 4500 test samples.The experimental platform shown in Fig.5.

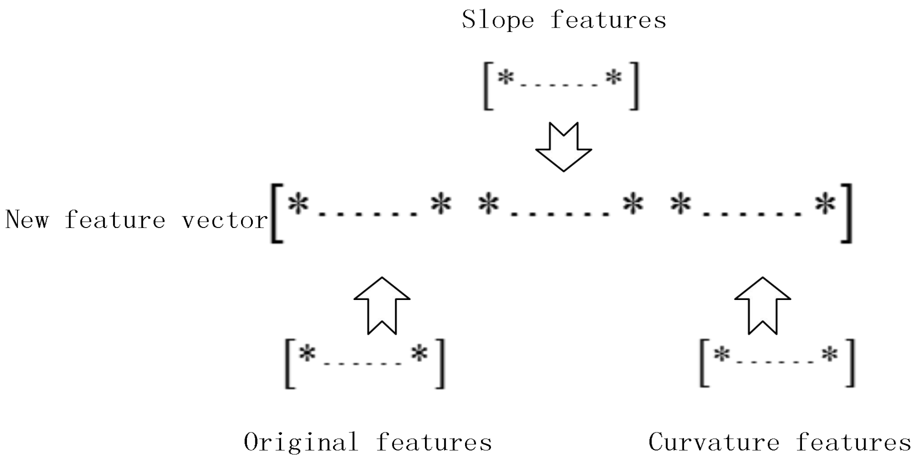

Fig.5 New feature vector

4.3 Description of Experimental Platform

This paper is mainly to solve the problem of frequency-type fault detection in the time domain is difficult to effectively, when the assignment of similar fault signal to the depth of the traditional learning method is difficult to effectively identify different fault types. Therefore, this paper selects three sets of fault data and a set of normal data, the three sets of fault data are the inner fault type, and the load is 0, the difference is the fault diameter is 0.007,0.014,0.021, the data type shown in Table 1. The four data types are selected because these data types have similar amplitudes and frequent zero crossings, so that it is difficult to effectively distinguish the types of faults based on the time-domain signals only while ensuring that other operating conditions are consistent. To ensure real-time and can not be transformed in the frequency domain, this paper proposes to calculate its slope curvature data to increase its frequency domain features. Time domain signal shown in Fig.7 .

4.4 Analysis of results

Our proposed deep learning based on geometric features is applied to bearing fault diagnosis. There are 4,500 samples under each data type and 4 different data types to characterize the frequency fault types of rotating mechanical systems. In order to reduce the influence of randomness, Experiment repeated 10 times. In this paper, DNN's pre-training initialization parameters are shown in Table 1.

Tab.1 DNN model parameters

Training parameter	DNN_1	DNN_2	DNN_3
Hidden layers	6	4	3
Number of neurons	500/250/100/50/20	500/200/50/20	500/200/50
Max number of epochs	1000	800	800
Learning rate	0.01	0.02	0.01

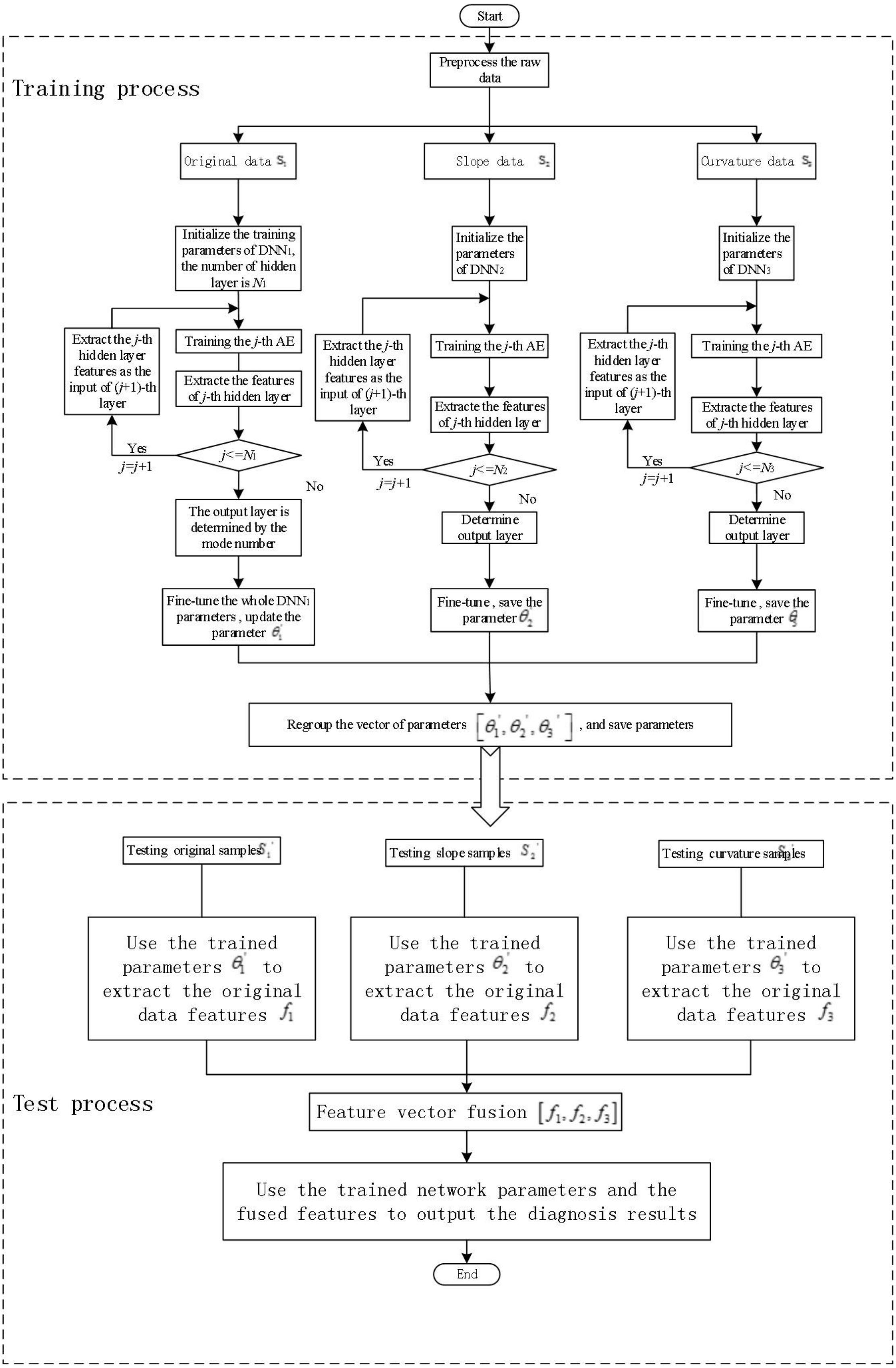

Fig.6 DNN Fault Diagnosis Flowchart Based on Geometric Features

DNN training uses a stochastic gradient descent method, and the maximum number of iterations of DNN in each layer is 1000,800,800 times respectively. Compared with the simulation results without feature fusion, the proposed method is validated. It can be seen from Table 2 that the diagnostic accuracy of the proposed method in frequency class and other faults is higher than that of ordinary neural networks and traditional deep neural networks DNN.

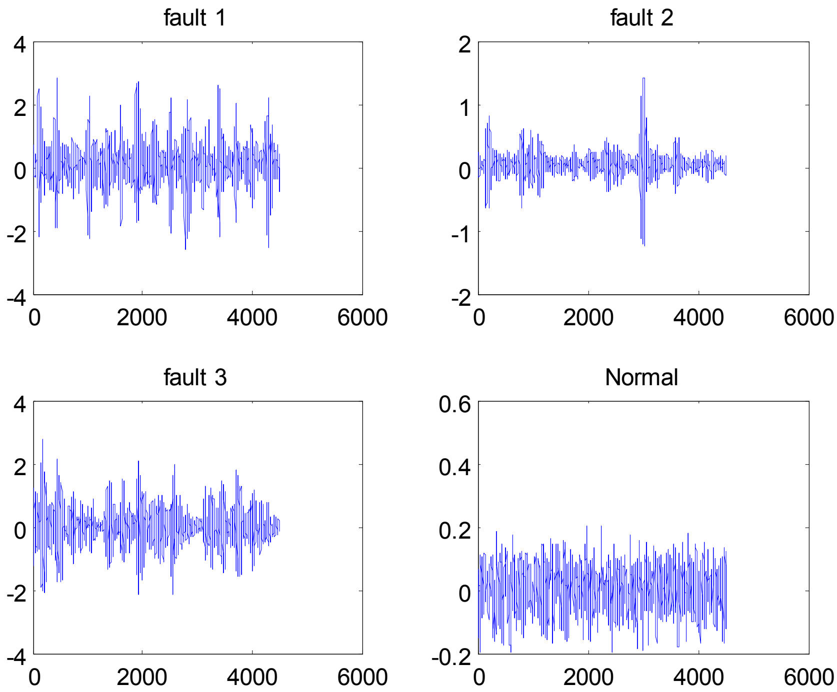

Fig 7 Observation of original signals

Tab.2. The classification results by the proposed model

Method	Diagnostic accuracy
GFFDNN	98.96
DNN	91.52
BP	85.56

5. CONCLUSION

In this paper, aiming at the problem that frequency class fault can not be effectively diagnosed in the time domain, a fault online diagnosis method based on DNN fusion of geometric features is proposed. The main innovation is to solve the problem of inaccurate time-domain detection when frequency class faults have similar amplitudes and frequent zero crossings by introducing the geometric features of the original data. The original, slope and curvature data are respectively constructed into an automatic encoder model, so as to fully excavate the implied frequency information of the data. Then, the obtained data features are fused by features, and the feature dimension is increased to make the classification model training more accurate. In addition, this method solves the problem of time-domain fault classification difficult, which makes the frequency-domain fault detection in the time domain can be effectively detected, which makes real-time fault diagnosis in the industry more practical value can be achieved, which will help the machinery Equipment health status testing. The experimental platform of rolling bearing verifies the effectiveness of the proposed method.

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, 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 and B. S. Dhillon, "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