Bearing Intelligent Fault Diagnosis Based on Convolutional Neural Networks

Jing An

School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, China

Peng An

BGP INC., China National Petroleum Corporation, Zhuozhou 072750, Hebei, China

Abstract—**The traditional intelligent identification method requires a complex feature extraction process and much diagnosis experience, considering the characteristics of one dimension of bearing vibration signals, a new method of intelligent fault diagnosis based on 1-dimensional convolutional neural network is presented. This method automatically extracts features from frequency domain signals and avoids artificial feature selection and feature extraction. The proposed method is validated on bearing benchmark datasets, these datasets are collected in different fault location, different health conditions and different operating conditions. The result shows that the proposed method can not only adaptively obtain representative fault features from the datasets, but also achieve higher diagnosis accuracy than the existing methods.**

Keywords—**Fault diagnosis, convolutional neural network, one-dimensional, feature extraction.**

I. INTRODUCTION

With the continuous development of computing, sensing and communication technology, modern industrial system presents a new development trend to in the direction of large-scale and complicated, the data reflecting the system running mechanism and the status shows "big data" features of mass, multimode, uncertainty, emergence, multi-source heterogeneity and low value density.

Traditional data-driven fault diagnosis and intelligent identification methods mainly relies on expert experience or signal processing technology of manual extraction and selection of fault features. These signal processing methods include Hilbert transform, stochastic resonance, Wigner-Ville distribution, the wavelet transform, Fourier transform, and Motor current signature analysis, etc. However, these advanced signal processing methods cannot adapt to fault diagnosis demand of big industrial data and usually require expert experience of the rotating machinery and the fault signals. Many machine learning methods, especially neural networks have the advantages of self-learning and adaptive and have been developed [1-3] for machine fault diagnosis.

In recent years, deep neural networks (DNNs) have developed rapidly in academia and industry, and it shows unique advantages and potential in feature extraction and pattern recognition. Compared with the traditional methods, DNNs have been implemented in many applications with outstanding performances, e.g., computer vision, natural language processing, speech recognition, bioinformatics, even heart diseases. In addition, some relevant studies can be found in [4-5]. Because of the superior capability of DNNs, more and more researchers are attracted to study machine fault diagnosis recently. At present, the most widely used benchmark dataset for fault diagnosis is the bearing dataset provided by Case Western Reserve University [6]. It is worth mentioning that the diagnosis accuracy obtained in the following research literature and the present paper are all derived from this dataset. Jia et al. [7] extracts frequency-domain features of the source vibration signals and presents a five-layer SAE model for rotating machinery fault diagnosis, which achieves 99.74% diagnosis accuracy. Sun et al. [8] proposes a sparse autoencoder-based DNN method to diagnose the faults of induction motor. Discriminant features are extracted adaptively and used for fault diagnosis through a sparse autoencoder, which reaches 97.61% diagnosis accuracy. Xia et al. [9] develops a stacked denoising autoencoder approach for fault diagnosis and achieves high performance with a large amount of unlabeled data and a small amount of labeled data, which achieves 97.88% diagnosis accuracy.

Compared with the common DNNs with all fully connected layers, convolutional neural networks (CNNs) are constructed with the unique capability to maintain original information regardless of shift, scale and distortion, the invariance is achieved by local receptive fields, shared weights and spatial sub-sampling. It can be easier to train a CNN with less computational resource and less time. Chen et al. [10] presents a CNN-based method which achieves high fault classification accuracy for gearbox with source signals. One-dimensional (1-D) signals are converted into a square matrix as input of the CNN model and achieves 93.79% diagnosis accuracy. Based on the discrete Fourier transform of the raw vibration signals, Janssens et al. [11] develops a three-layer CNN approach to diagnose the bearing fault and reaches 93.61% diagnosis accuracy. Guo et al. [12] proposes a deep CNN method with

hierarchical adaption for bearing fault classification in four health conditions. The approach converts 1-D source signals into square matrix as the input, which achieves 99.7% diagnosis accuracy with a slower convergence. Lu et al. [13] proposes a hierarchical convolutional network of intelligent fault diagnosis of rolling bearing and achieves 92.60% diagnosis accuracy. Zhang et al. [14] converts the 1-D source signal into a 2-D image, CNN is used to train to acquire discriminative features for fault diagnosis, which obtains 99.95% fault recognition rate. Sun et al. [15] proposed a convolutional discriminative feature learning method for induction motor fault diagnosis, and achieve 99.36% fault recognition rate. Xia et al. [16] presents a multi-sensor fusion CNN-based approach, during the training process, the approach considers both spatial and temporal information of the source signals from multiple sensors. The method achieves higher and more robust diagnosis accuracy, up to 99.41% fault recognition rate.

Since the original sampling data of bearing vibration signal is 1-D periodicity, this paper attempts to solve the fault diagnosis problem by using 1DCNN method directly. The approach of implementing CNN on the 1-D signal is simple and intuitive, and has already been used in speech recognition [17], which treats the signal as an image whose height is equal to 1. The main contributions of this paper can be summarized as the follows: (1) A novel stable fault diagnosis method is presented by using 1DCNN model with high fault diagnosis accuracy. (2) Comprehensive experiments and detailed analysis on bearing datasets have been performed. (3) Representative features of the spectrum are extracted adaptively through the 1DCNN model, without the need of manual feature selection and feature extraction. The method can be applied to diagnose other types of mechanical faults without too much prior knowledge.

The rest of the paper is organized as follows. In Section 2, the basic theory of CNN is introduced. Then, in Section 3, the details of the proposed 1DCNN model are described. Experimental results and analysis are illustrated in Section 4. At last, concluding remarks are provided in Section 5.

II. THE BASIC THEORY OF CONVOLUTIONAL NEURAL NETWORK

A. Convolutional layer

A standard convolutional neural network consists of input layer, convolution layer, pooling layer(sub-sampling), fully connected layer and output layer. In a convolutional layer, the feature maps on the upper layer are convolved with a convolution kernel, and then an activation function is used to obtain an output feature map of the next layer. Each output map may be a combination of the values of multiple input maps. The input of a CNN is X, in this paper, H^l represents the feature maps of the 1 layer of the CNN($H^0 = X$). Suppose H_j^l denotes the jth feature map of the lth layer. The production process of H_j^l is as follows:

$$H_j^l = f(\sum_{i=1}^K H_i^{l-1} * W_{ij}^{(l)} + b_j^l) \qquad (1)$$

where $W_{ij}^{(l)}$ represents the weight of the kernel, which connects the ith feature map of the l-1th layer with the jth feature map of the lth layer. * represents the computation of convolution. H_i^{l-1} denotes the ith feature map of the l-1th layer. $i = 1,2,...K$ is the index of the input feature maps, $j = 1,2,...J$ is the index of the output feature maps. And then a bias matrix b_j^l is given to each output map. At last, a nonlinear activation function f(.) is used to obtain the jth feature map of the lth layer H_j^l. Sigmoid function, hyperbolic tangent function, radial basis function, and rectified linear unit(ReLU) etc. are the standard nonlinear activation functions that are often used.

B. Pooling layer

The pooling layer usually follows the convolution layer and samples the features according to the following sampling rules. The pooling layer can not only reduce the feature, but also keep the scale invariant characteristics of the feature in a sense. For a pooling layer, N input maps produces N output maps with a smaller size. Suppose H_j^l represents the jth feature map of the lth layer, the production process of H_j^l is as follows:

$$H_j^l = f(\beta_j^l down(H_j^{l-1}) + b_j^l) \qquad (2)$$

where $down(.)$ represents the max-pooling function or average pooling function. β_j^l represents the deviation in the multiplier for the jth feature map of the lth layer. b_j^l denotes the additive bias for the jth feature map of the lth layer.

After passing through multiple convolutional layers and pooling layers, the CNN can classify the extracted features by the fully connected network, and obtain the probability distribution Y, which is defined as:

$$Y(i) = P(L = l_i | H_0; (W, b)) \qquad (3)$$

where l_i represents the ith label category. The essence of the CNN is that the original matrix H^0 is mapped to a new mathematical model of expression Y, through multiple layers of data transformation or dimension reduction. The training objective of a CNN is to minimize the loss function $L(W, b)$. The commonly used loss functions include mean square error function, cross entropy loss function, negative logarithm likelihood function, etc.

Take 1-D CNN as an example, the structure diagram of convolutional layer and pooling layer is as shown in Fig.1 (2-D CNN and 3-D CNN in expand), the left is the input of a convolutional layer, the middle is the convolutional layer, the right is the pooling layer. Set the convolution kernel size of the convolutional layer is 1×3, the sliding step is 1, the kernel size of the pooling layer is 1×2, the sliding step is 2.

C. Softmax layer

As a generalization of logistic regression, the $Softmax$ regression is often applied on multiclass classification. The number of categories of machine health conditions determines the dimension of the output layer. $Softmax$ regression evaluates the probability that each input sample belongs to each category. Given a training dataset of k classes with m samples $(x^{(i)}, y^{(i)})$, where $x^{(i)} \in R^n$ and $y^{(i)} \in \{1,2,...,k\}$. The probability is given by

$$P(y^{(i)} = j | x^{(i)}; \theta) = \frac{1}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \dots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (4)$$

where $j = 1 , 2 , \dots . k$. $\theta = [\theta_1, \theta_2, \dots \dots \theta_k]$ is the parameters of the $Softmax$ regression model and $\frac{1}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}$ normalizes the distribution so that the summation of the probability is one. We define the cost function of $Softmax$ regression as

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^{m} \sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}} \right] \quad (5)$$

where $1\{.\}$ is the indicator function, which is 1 if the condition is true and 0 otherwise. The parameter θ is updated by minimizing the cost function $J(\theta)$ over the training dataset.
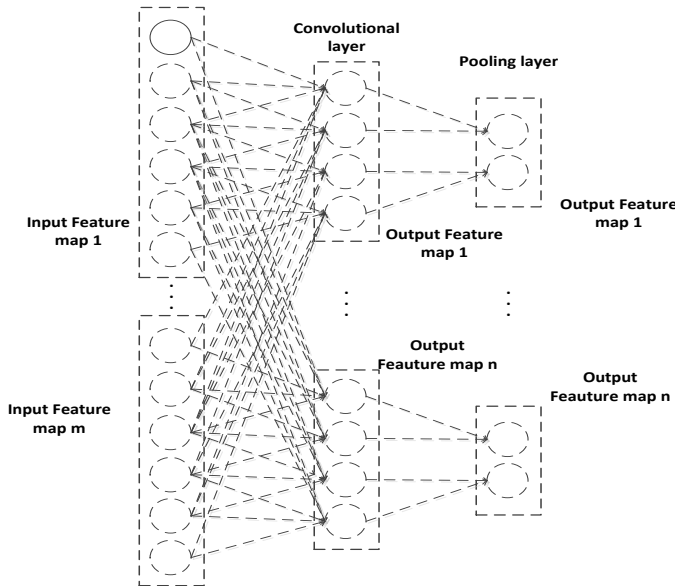

Fig.1. Structure diagram of convolutional layer and pooling layer.

III. FAULT DIAGNOSIS FRAMEWORK BASED ON 1DCNN

Different from image data, vibration signals are commonly 1-D, it is more reasonable to handle vibration signals with 1DCNN. Therefore, the bearing intelligent fault diagnosis method based on 1DCNN is presented. A flowchart of the proposed method is shown in Fig. 2. First of all, according to the approximate number of samples collected from each turn, the raw signals are stacked into 2D input matrix. Secondly, fast Fourier transformation (FFT) is implement on input matrix. Then the samples are divided into three parts, training dataset, validation dataset and testing dataset, respectively. By minimizing the error between the predicted label and the actual one, the training dataset is used to train the initialized 1DCNN model. Then the validation dataset is used to adjust the super parameters of the model. Finally, using the testing dataset, we can evaluate the generalization capability of the trained model and get the diagnosis results of health conditions.
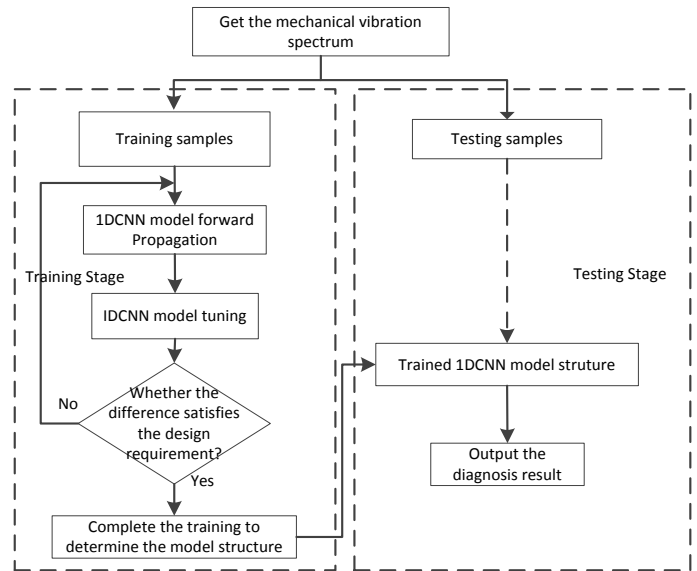

Fig.2. Flowchart of the proposed fault diagnosis approach.

A detailed structure of the 1DCNN-based fault diagnosis model is shown in Fig. 3. It includes an input layer, two convolutional layer, two pooling layer, a fully connected layer and an output layer. The training data is preprocessed as the input matrix X of the 1DCNN model. The input matrix X is convolved by K_1 filters of size $c_1 \times 1$. And then the ReLU operation is used to form the K_1 feature maps. A pooling layer S_1 is followed to subsample by K_1 filter of size $s_1 \times 1$, it can capture K_1 feature maps by using Equation (2). Followed by similar operations including a convolutional layer C_2 and a pooling layer S_2, a fully connected layer and a softmax layer are added in the end to aim to output the machine health conditions.

The cross-entropy is used for loss function of our 1DCNN model. Assuming that $p(x)$ denotes the target distribution and $q(x)$ represents the estimated distribution, the cross-entropy between $p(x)$ and $q(x)$ is defined as:

$$H(p, q) = -\sum_x p(x) \log q(x) \quad (6)$$

Learning rate indicates the speed of the parameter to the optimal value process. Selecting the appropriate learning rate can make the gradient descent optimization algorithm get better performance. To minimize the loss function, in this paper, the Adadelta adaptive adjustment learning rate is used to train our 1DCNN model. Adadelta can avoid manual adjustment of learning rate and adjust automatically in the process of learning. In order to make the 1DCNN model more robust, mini-batch gradient descent is adopted to obtain a balance between the update speed and the update frequency, also it can reduce the convergence and volatility and make the update more stable.

Dropout is used in the training process to control overfitting. The core idea is to randomly drop units (along with their connections) from the neural network during training, but the weight will be retained because it may be activated during the next transmission. Each neuron is discarded with probability p, and the probability $1 - p$ is retained. It is noted that we only use the dropout in the training phase of the model and close the dropout in the test phase,

which is more conductive to the feature extraction and classification capability of the 1DCNN model.
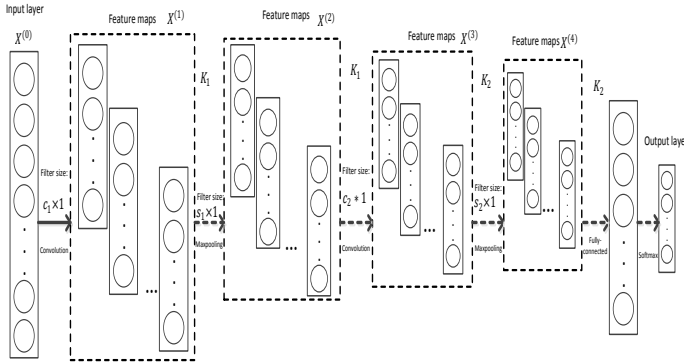

Fig.3. Architecture of 1DCNN-based fault diagnosis model.

IV. EXPERIMENT

A. Data description

The bearing fault data used for experimental validation were obtained from the Bearing Data center of Case Western Reserve University [4]. The diagnosis target is to identify the different faults location with different degrees of severity. The data were collected from a motor driving mechanical system under four different loads (0-3hp) and three different locations: the fan end, the drive end and the base, and the sampling frequency includes 48kHz and 12kHz. There are four fault types of the bearing, normal condition(N), outer race fault(OF), inner race fault(IF), and roller fault(RF). Each type of fault type has three levels of severity with fault diameters of 0.007 inch, 0.014 inch and 0.021 inch, respectively.

In this paper, vibration signals with different fault locations and different health conditions are selected, which are 48kHz of the drive end bearing fault data and 12kHz of the fan end, the drive end and the base bearing fault data, respectively. Two examples are designed, the detailed description for the datasets are described in Table 1 and Table 2. In Table 1, dataset A consists of 2000 samples of 10 bearing health conditions at 1 hp and 1772 rpm, the composition of dataset B and C are similar to A. Each health condition has 200 samples and each sample includes 2400 data points. FFT is implemented on each sample to get the 2400 Fourier coefficients. Because of the coefficients are symmetric, we use the first 1200 coefficients in each sample. Dataset D contains 10 bearing health conditions under loads of 1-3 hp, the same health condition under different loads are labeled as one class, so each condition has 600 samples and dataset D contains 6000 samples.

Table 1 Description of 48kHz drive end bearing datasets

Datasets	Load (hp)	The number of samples	Fault type	Fault diameter (mils)	Classification label
A/B/C/D	1/2/3/1-3	200/200/200/600	N	0	0
		200/200/200/600	RF07	7	1
		200/200/200/600	RF14	14	2
		200/200/200/600	RF21	21	3
		200/200/200/600	IF07	7	4
		200/200/200/600	IF14	14	5
		200/200/200/600	IF21	21	6
		200/200/200/600	OF07	7	7
		200/200/200/600	OF14	14	8
		200/200/200/600	OF21	21	9

In Table 2, under loads of 0, 1, 2 and 3 hp of the drive end, dataset A_1, B_1, C_1 and D_1 include 10 bearing health conditions respectively. Each health condition has 50 samples and each sample includes 2400 data points. Therefore, dataset A_1, B_1, C_1 contain 500 samples, respectively. Fast Fourier transformation is implemented on each sample to get the 2400 Fourier coefficients, and the first 1200 coefficients are used in each sample. Dataset E_1 includes 10 bearing health conditions under loads of 0-3 hp with 200 samples for each condition. So dataset E_1 contains 2000 samples. A_2-E_2 represents the dataset partition of the fan end. A_3-E_3 represents the dataset partition of the base. The partition mode of the fan end and the base are the same as the drive end. Dataset E_{all} is composed of E_1, E_2 and E_3.

Table 2 Description of 12kHz bearing datasets

Datasets		Load (hp)	The number of samples	Fault type	Fault diameter (mils)	Classification label
The drive end	$A_1/B_1/C_1$ /D_1/E_1	0/1/2/3/0-3	50/50/50/50/200	N	0	0
			50/50/50/50/200	RF07	7	1
			50/50/50/50/200	RF14	14	2
			50/50/50/50/200	RF21	21	3
			50/50/50/50/200	IF07	7	4
			50/50/50/50/200	IF14	14	5
			50/50/50/50/200	IF21	21	6
			50/50/50/50/200	OF07	7	7
			50/50/50/50/200	OF14	14	8
			50/50/50/50/200	OF21	21	9
The fan end	$A_2/B_2/C_2$ /D_2/E_2	0/1/2/3/0-3	50/50/50/50/200	N	0	0
			50/50/50/50/200	RF07	7	1
			50/50/50/50/200	RF14	14	2
			50/50/50/50/200	RF21	21	3
			50/50/50/50/200	IF07	7	4
			50/50/50/50/200	IF14	14	5
			50/50/50/50/200	IF21	21	6
			50/50/50/50/200	OF07	7	7
			50/50/50/50/200	OF14	14	8
			50/50/50/50/200	OF21	21	9
The base	$A_3/B_3/C_3$ /D_3/E_3	0/1/2/3/0-3	50/50/50/50/200	N	0	0
			50/50/50/50/200	RF07	7	1
			50/50/50/50/200	RF14	14	2
			50/50/50/50/200	RF21	21	3
			50/50/50/50/200	IF07	7	4
			50/50/50/50/200	IF14	14	5
			50/50/50/50/200	IF21	21	6
			50/50/50/50/200	OF07	7	7
			50/50/50/50/200	OF14	14	8
			50/50/50/50/200	OF21	21	9

B. The parameter setting of 1DCNN model

In order to obtain significant fault features and better robustness, in this paper, the 1DCNN model is proposed to realize the adaptive extraction of mechanical fault data and intelligent identification of mechanical health states.

Considering that the computation and the reconstruction efficiency, the 1DCNN model is constructed as follows: input-C1-S1-C2-S2-F-output. Note that the convolutional layers and pooling layers are appear alternately. Assume that an input feature map of size $a \times 1$, a convolutional kernel of size $c_1 \times 1$, and a pooling region of size $s_1 \times 1$. C_1 denotes the sizes of feature maps in the convolutional layer and S_1 denotes the sizes of feature maps in the pooling layer, they could be calculated as the following:

$C_1 = c \times 1 = (a - c_1 + 1) \times 1$
$S_1 = (c/s) \times 1$

The main parameters of the 1DCNN model are listed in Table 3 and Table 4.

Table 3 Parameters of 1DCNN model

Model parameters	Input layer	Convolutional layer 1	Pooling layer 1	Convolutional layer 2	Pooling layer 2	Output layer
Number of feature maps	1	32	32	64	64	10
Size of feature maps	1200×1	1196×1	598×1	596×1	298×1	
Size of filter		5×1	2×1	3×1	2×1	
Stride		1	2	1	2	

Table 4 Other parameters

Parameters / Values	Train Epochs	Batch size	Dropout rate	Activation function
	30	128	0.25, 0.5	ReLU

C. Comparative analysis result of the datasets of example 1 of the proposed method with other deep learning methods

At the input layer, 70% of the samples are randomly selected for pretraining the 1DCNN model and fine-tune the parameters of the entire network, and the remaining 20% of samples are used for performance testing, and 10% of samples are used as validation dataset. In order to reduce the influence of random factors, the experiment of each dataset was repeated 10 times. The diagnosis of dataset A-D by the proposed model are detailed in Fig. 4. The comparison methods includes the common 2DCNN method proposed by Lu et al. [11], and the SAE method proposed by Jia et al. [5]. The implementation detail and parameter settings of the two methods can be found in [11][5].

The diagnosis results are counted and analyzed, as shown in Table 5. To illustrate, the training accuracy of the 2DCNN algorithm of Table 5 is not written because it is not given in the original article. It can be seen from the Table 5 that in the experiment of the proposed method, the diagnosis accuracy is stable at a high level, and the training accuracy of each dataset is 100%. The average test accuracy is 99.88% to 100%, with a low standard deviation of below 0.089%. It is noteworthy that although dataset D contains a lot of samples for 10 health conditions under different loads, the training and testing accuracy of dataset D are stable at a higher level. This shows that the proposed approach is able to diagnose the bearing faults without being affected by load fluctuation and more samples lead to better diagnostic performance.

The diagnosis accuracy of 2DCNN method [11] is lower than that of the 1DCNN method, and the diagnosis accuracy is between 92.60% and 97.78%, while the standard deviation is between 0.49% and 1.17%. Compared with 1DCNN method, the training accuracy of SAE method [5] is between 99.85% and 100%, while the training standard deviation is below 0.08%, and the testing accuracy is between 99.61% and 99.95%, while the test standard deviation is below 0.22%. It can be seen that the SAE method reaches a high level also, but slightly less than 1DCNN method.

The above results show that the proposed method not only has higher classification accuracy, but also the diagnosis result is very stable. It can completely exclude condition interference and diagnose the 10 health conditions under different fault location and different fault modes of rolling bearing.

To further validate the diagnosis ability of the model, we conduct another experiment with different sample proportions of training dataset, testing dataset and validation dataset used in training and testing of the classification performance of the proposed method. The ratio of training and test samples sizes are 7:2:1, 5:4:1, 3:6:1, respectively. The training and testing accuracies in dataset A are depicted in Fig.5. The diagnosis results are counted and analyzed, as shown in Table 6.
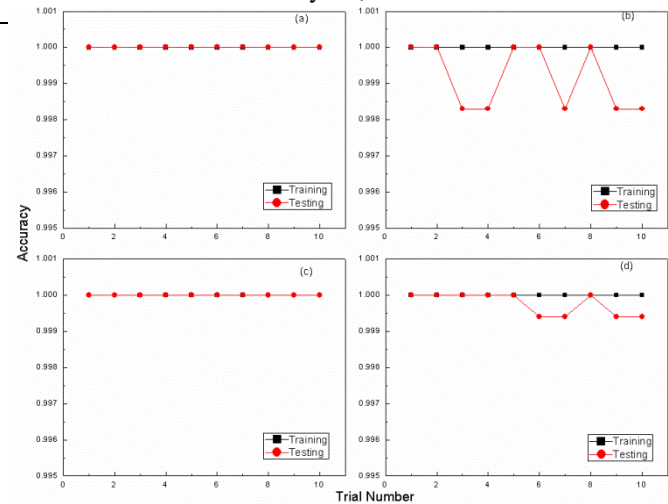

Fig.4. Ten diagnosis results of bearing datasets of 48kHz of the proposed method: (a) *Dataset* A, (b) *Dataset* B, (c) *Dataset* C, (d) *Dataset* D.

Table 5 Diagnosis results of bearing datasets

Datasets	The proposed method (%)		Common 2DCNN method (%)		SAE method (%)	
	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
A	100	100		96.48 ± 0.76	100	99.95 ± 0.06
B	100	99.92 ± 0.09		97.78 ± 0.49	99.94 ± 0.05	99.61 ± 0.21
C	100	100		96.97 ± 0.88	99.94 ± 0.08	99.74 ± 0.16
D	100	99.98 ± 0.03		92.60 ± 1.17	99.85 ± 0.02	99.68 ± 0.22

From the diagnosis results, when the sample ratio is 7:3, the average training accuracy is 100%, and the average test accuracy is 100%. When the sample ratio is 5:5, the average training accuracy and the average test accuracy are all 100%. When the sample ratio is 3:7, the average training accuracy is 100%, and the average test accuracy is 99.979% with a standard deviation of 0.0472%. When the ratio of training samples to test samples was 7:3 and 5:5, the training accuracy and test accuracy are the same, both are 100%, when the ratio was 3:7, the test accuracy is a little bit lower. It shows that the more training samples, the more general distribution of the samples, the better the generalization performance of the training model, the higher the accuracy. Note that when the number of training samples decreases, the classification rates for 1DCNN method has a slightly fall accordingly, but due to the proposed method has the capability of learning highly complexity relationships between the fault features and the

health conditions, it still shows greater diagnosis accuracy in most of the experiments.
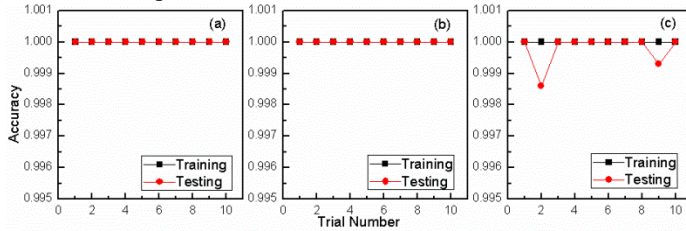

Fig.5. Classification results with different training samples: (a) 7:3, (b) 5:5, (c) 3:7.

Table 6 Classification results analysis with different training samples
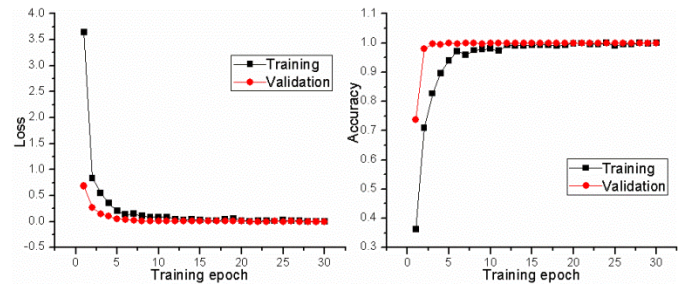
Dataset A	The proposed method (%)	
	Training accuracy	Testing accuracy
7:3	100	100
5:5	100	100
3:7	100	99.979±0.0472

Further, we choose a better experiment from three proportion samples, respectively, and compare loss and accuracy of training process in each experiment, as shown in Fig.6. It is noticed that whether the loss curve convergence or the accuracy curve convergence is the fastest in the ratio of 7:3, the ratio of 5:5 is second, and the ratio of 3:7 is the slowest.

The diagnosis results are counted and analyzed, as shown in Table 7. It can be seen that in loss value, whether the average training loss value or the average test loss value, 7:3 ratio is the smallest, 5:5 ratio is the second, and the ratio of 3:7 is the largest, increasing in turn. In terms of accuracy value, it decreases in turn, which further verifies that the more training samples, the more generalization ability and robustness of the training model, the higher the Accuracy. From another aspect, it is shown that the advantage of deep learning is more obvious when the number of samples is processed. However, as a whole, the diagnosis results are still very high, which is because the 1DCNN model has the ability of highly adaptive learning of high order complex characteristics of vibration signal data.
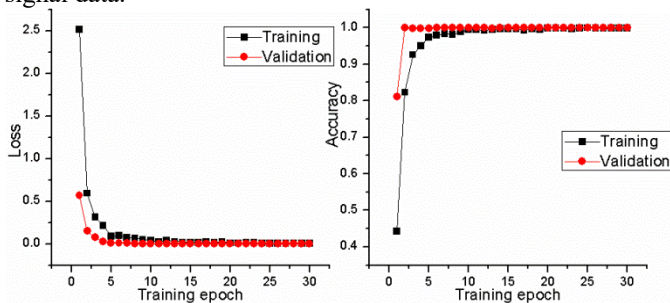
(a) the tenth trial of 7:3

(b) the ninth trial of 5:5
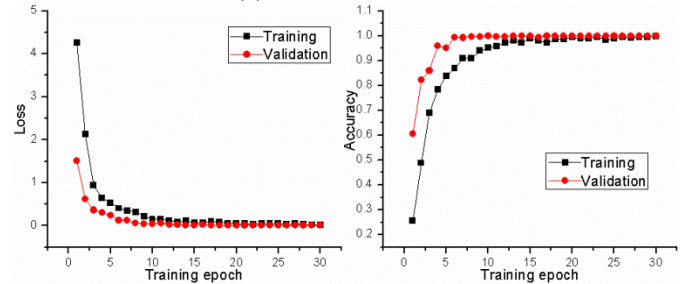
(c) the first trial of 3:7

Fig.6. Loss curves and accuracy curves of the training process of the proposed method.

Table 7 Results analysis of loss and accuracy of the training process

Dataset A ratio	Training Loss	Validation Loss	Training Accuracy (%)	Validation Accuracy (%)
7:3	0.15 ± 0.46	0.03 ± 0.11	96.62±10.5	99.35±3.4
5:5	0.22 ± 0.67	0.05±0.13	94.92±12.6	99.01±4.8
3:7	0.37 ± 0.84	0.12 ±0.3	91.13±16.58	97.18±8.03

Time is another important factor to consider when using deep learning algorithm to solve classification problems. The computer system used in the experiment is 64 bit Win10, and the installed memory (RAM) was 8GB. The experiment is conducted based on the CPU mode of Keras 2.0. The dataset A-C is relatively small, and each epoch runs for about 2 seconds. Dataset D has a large amount of data, and each epoch runs for about 6 seconds. The model runs about 10 epochs and begins to converge. Compared with other deep learning algorithms, the running time is greatly reduced.

D. Comparative analysis result of 2DCNN method with multi-sensor fusion

In 2DCNN method with multi-sensor fusion, the implementation detail and parameter settings of which can be found in [16], the vibration signals which have 10 different health states of the fan end, the drive end and the base at sampling frequency of 12kHz are fused. The dimension of each sample is 3*1200. In order to compare the proposed method with 2DCNN method with multi-sensor fusion, design the datasets in Table 2 of example 2. The drive end data is merged to form the dataset E_1, and the fan end data is merged to form the dataset E_2, and the base data is merged to form the dataset E_3. E_1, E_2, E_3 are merged to form the dataset E_{all}. 1DCNN model parameters are set in Table 3 and Table 4. Obviously, the dataset E_1, E_2, E_3 represent single sensor datasets of the fan end, the drive end and the base, respectively. E_{all} represents the dataset of multi-sensor fusion.

The diagnosis results of the proposed method for dataset $E_1, E_2, E_3, and\ E_{all}$ are shown in Fig.7. The training and test accuracy of 1DCNN method and 2DCNN method with multi-sensor fusion are compared as shown in Table 8.
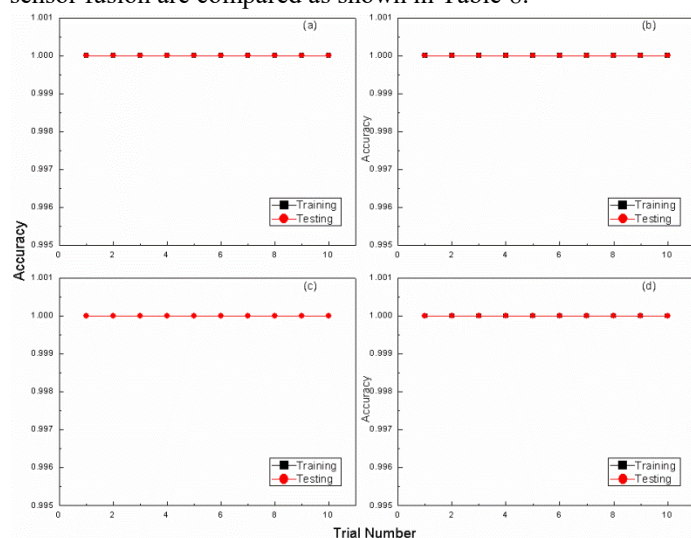

Fig.7. Ten diagnosis results of four bearing datasets of 12kHz of the proposed approach: (a) $Dataset\ E_1$, (b) $Dataset\ E_2$, (c) $Dataset\ E_3$, (d) $Dataset\ E_{all}$.

Table 8 Comparison of two fault diagnosis methods

Datasets	The proposed method (%)		2DCNN method with multi-sensor fusion (%)	
	Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
E_{all}	100	100	99.94±0.06	99.83±0.13
$E_1/E_2/E_3$	100	100	98.18±0.69	97.58±1.4

The experiment is repeated ten times for the health diagnosis of the dataset E_1, E_2, E_3, E_{all}. For dataset E_{all}, the training accuracy and test accuracy of 10 trials of the proposed method are all 100%. The accuracy of 2DCNN method with multi-sensor fusion is slightly lower than the proposed method, the average training accuracy is 99.94%, and the test accuracy is 99.83%. Also for single-sensor dataset E_1, E_2, E_3, The training accuracy and test accuracy of 10 trials of the proposed method are all 100%, while the average training accuracy is 98.18% and the average test accuracy is 97.58% and the standard deviations are below 1.4% using 2DCNN method with multi-sensor fusion, which means the proposed method can not only identify bearing fault types effectively and stably, but also distinguish fault severities.

E. Visualization of learned Representation

One of the fundamental reasons for the high accuracy of the proposed method is that the fault features can be adaptively extracted from the spectrum of mechanical vibration signals. To demonstrate that the proposed method is able to learn effective and discriminative representations for vibration signals automatically, "t-SNE" technique is adopted, which is an effective approach to visualize high-dimensional data. With this technique, data samples can be mapped from their original feature space into a two or three-dimensional space map. In this paper, t-SNE technology is used to map the high-

dimensional features of the model into two dimensional features and visualize them. Due to the output layer of the 1DCNN model is used as the classifier, the feature number 128 of the fully connected layer is used as the sample feature extracted from the model. The four bearing datasets A-D in example 1 are reduced from 128 dimension to 2 dimension and are visualized, as shown in Fig.8. It can be seen that in each bearing dataset, the features of the same condition are clustered well, while features of different conditions are separated well, so more than 99.9% of the diagnosis accuracy is achieved. The above results show that the proposed method can be applied to extract features that are beneficial to the diagnosis according to the characteristics of the mechanical spectrum and the diagnosis task.
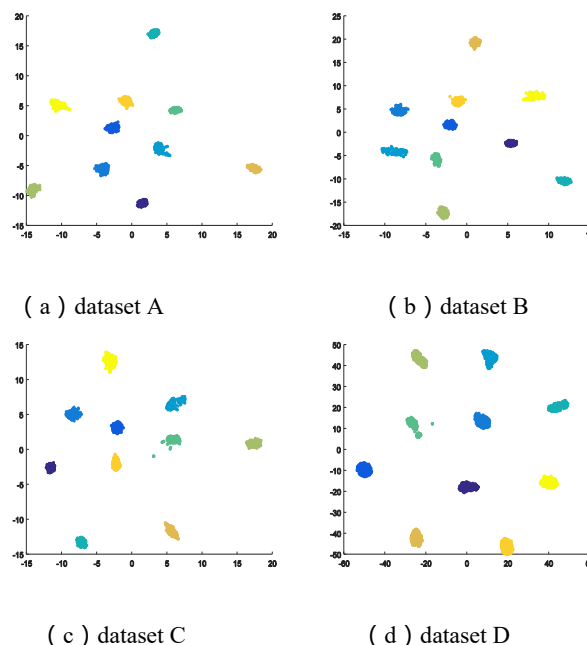

(a) dataset A   (b) dataset B

(c) dataset C   (d) dataset D

Fig.8. Feature visualization of 1DCNN model based on t-SNE technology.

V. CONCLUSION

In view of the characteristic of one-dimensional of the vibration signals, a novel fault feature extraction and diagnosis method based on 1DCNN is presented in this paper. This method can be applied to extract fault feature from the health conditions signal spectrum automatically, instead of relying on expert knowledge. Compared with other state-of-the-art methods, the results show that the diagnosis accuracy is higher and the diagnosis performance is higher and more reliable. Further research is mainly in the following two aspects: (1) Since the dataset is relatively easy to learn, next step is to increase the classification difficulty by adding different noise levels, and compare the accuracy between different methods. (2) The samples of different health states in this paper are labeled, in future studies, diagnosis for mixed tags or unlabeled training data will be performed as a challenging direction.

REFERENCES

[1] Lei Y G., "Opportunities and challenges of machinery intelligent fault diagnosis in big data era", *Journal of Mechanical Engineering*, vol. 54, no. 5, pp. 94-104, 2018.

[2] Zhang S, Zhang S B, Wang B N, et al., "Deep learning algorithms for bearing fault diagnostics - A comprehensive review", *IEEE Access*, vol. 8, pp.29857-29881, 2020.

[3] Hoang D T, Kang H J., "A survey on deep learning based bearing fault diagnosis", *Neurocomputing*, vol. 335, pp. 327-335, 2019.

[4] Kanaan L, Haydar J, Samaha M, et al., "Intelligent Bus Application for Smart City based on LoRa Technology and RBF Neural Network", *WSEAS Transactions on Systems and Control*, vol 15,pp. 725-732, 2020.

[5] Gochhait S, Rimal Y, Pageni S,"The Comparison of Forward and Backward Neural Network Model – A Study on the Prediction of Student Grade", *WSEAS Transactions on Systems and Control*, vol. 16, pp. 422-429, 2021.

[6] Case Western Reserve University Bearing Data Center [Online: http://csegroups.case.edu/bearingdatacenter/home. 2012.

[7] Jia F, Lei Y, Lin J, et al., "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data", *Mechanical Systems and Signal Processing*, vol. 72-73, pp.303-315, 2016.

[8] Sun W J, Shao Sy, Zhao R, et al., "A sparse auto-encoder-based deep neural network approach for induction motor faults classification", *Measurement*, vol. 89, pp.171-178, 2016.

[9] Xia M, Li T, Liu L Z, et al., "Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder", *IET Science, Measurement & Technology*, vol. 11, no. 6, pp. 687-695, 2017.

[10] Chen Z Q, Li C, Sanchez R V., "Gearbox fault identification and classification with convolutional neural networks", *Shock and Vibration*, vol. 2015, pp. 1-10, 2015.

[11] Janssens O, Slavkovikj V, Vervisch B, et al., "Convolutional neural network based fault detection for rotating machinery", *Journal of Sound and Vibration*, vol. 377, pp. 331-345, 2016.

[12] Guo X J, Chen L, Shen C Q., "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis", *Measurement*, vol. 93, 490-502, 2016.

[13] Lu C, Wang Z Y, Zhou B., "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification", *Advanced Engineering Informatics*, vol. 32, pp. 139-151, 2017.

[14] Yuan H L, Zhang W, Peng G L, et al., "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input", *MATEC Web of Conferences*, vol. 95, 13001, 2017.

[15] Sun W J, Zhao R, Yan R Q, et al., "Convolutional discriminative feature learning for induction motor fault diagnosis", *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1350-1359, 2017.

[16] Xia M, Li T, Xu L, et al., "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks", *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, 101-110, 2018.

[17] Palaz D, Collobert R, Doss M M., "Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks", *Computer Science*, pp. 1-5, 2013.