International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET229141

Research on Mechanical Fault Diagnosis Algorithm Based on Sound Signal and CNN

Lemei Han, Zhan Wen, Haoning Pu, Wenzao Li
College of Communication Engineering, Chengdu University of Information Technology, Chengdu, Sichuan, China

Article Info
Volume 9, Issue 1
Page Number : 153-160
Publication Issue :
January-February-2022
Article History
Accepted : 08 Feb 2022
Published: 17 Feb 2022

ABSTRACT

Failure diagnosis is of great significance for the timely detection of the safety hazard of the equipment and the guarantee of the normal operation of the production. In fault diagnosis, the way based on the processing of sound signal has the advantages of strong fault sensitivity, easy acquisition, and noncontact measurement, and the way of using neural network provides a more efficient and generally applicable method for fault diagnosis efficiency. For the poor diagnostic accuracy of traditional methods, which requires manual extraction of features and poor general applicability of the model, in this paper, we propose a mechanical failure diagnosis method based on acoustic signals and CNNs. The sound signals were first sampled and features extracted by MFCC, then the data were split into training and test sets in a 6:4 ratio and input to the convolutional neural network. After adjusting the parameters for the comparison experiment, the final experimental model was able to achieve 97.05% test accuracy over 20 training test iterations.
Keywords-CNN, Sound signal processing, MFCC, MIMII data set

I. INTRODUCTION

"Listen to sound, identify faults". This has long been a common method used by humans to detect device failures, especially for experienced people, because of its simplicity and efficiency. Using its features to predict the faults before they occur is of great importance to avoid the sudden failure of equipment, which has a serious impact on the entire industrial production. Traditional workers judge by the sound produced when the machine is running, but this has a lot of drawbacks. Long-term noise work environment damages workers' hearing, and workers' long-term detection efficiency is not high and accuracy is limited. Therefore, the analysis and processing of the collected sound signal can effectively replace the manual method [1].

Fault diagnosis of machines usually uses acceleration sensors to collect vibration signals and then process and analyze them. Sound is also a vibration signal in nature and contains a lot of information, but compared with vibration signal, sound signal has the advantages of fault sensitivity, easy collection, non-contact measurement, etc. [2-3]. For the traditional method of sound signal processing, Liu et al. proposed an improved local mean decomposition (LMD) method, which separates the FM and AM components in the audio message to achieve the gear box composite fault diagnosis [4]. Sun et al. proposed a breaker mechanical fault diagnosis method based on sound characteristics and improved sparse representation classification. The experimental results show that the proposed method can accurately identify breaker mechanical faults and common sound categories in substations [5]. Yang et al. proposed a method to identify mechanical faults of high voltage circuit breakers based on sound signals. By analyzing several characteristic vectors that contribute the most, the diagnosis of mechanical faults of circuit breakers can be achieved. The experimental results show that the proposed method has a higher recognition rate and a lower time cost than the existing feature extraction methods of sound signals. Although these methods are effective and specific, they are not universally applicable. Using a convolution neural network can solve the problem of universality.

Convolutional neural network (CNN) is a common deep learning method. It is a typical multi-layer feed-forward neural network. The end-to-end algorithm structure makes the whole process without manual feature extraction, and most importantly, it is suitable for any mechanical device. Therefore, in order to solve the problem of traditional methods, this paper presents a method of mechanical fault diagnosis based on sound signal and CNN.

II. Fault Diagnosis Method of Machine Based on CNN and Sound Signal

First, the original sound signal is sampled. Considering the large amount of data and the increase of training time of the neural network, MFCC is used to extract features and achieve the effect of dimension reduction. Then the data is divided into training set and test set in a 6:4 ratio. Finally, it is input into the convolution neural network for training and testing the classification results. The specific process is shown in Figure 1.

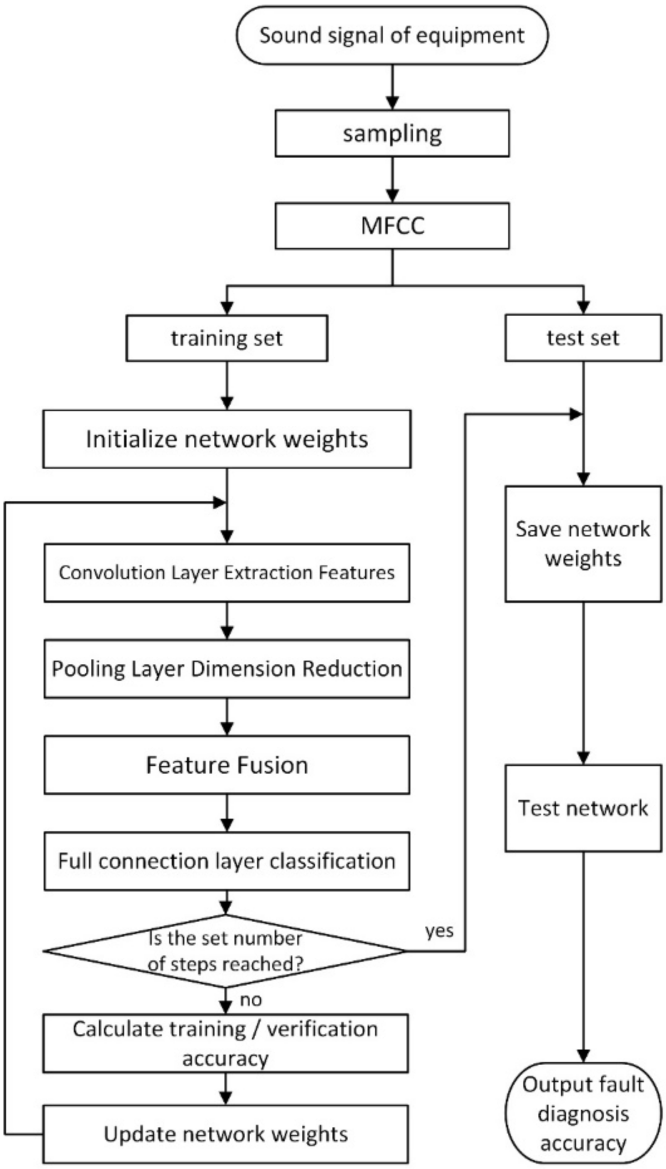

Figure 1. Method Flowchart

A. Mel-Frequency Cepstrum Coefficient

Because human ears have different perceptions of sound at different frequencies of f, the Mel frequency is based on the auditory characteristics of the human ear, and it corresponds non-linearly to the f frequency. Mel Frequency Cepstrum Coefficient (MFCC) is a calculated f-spectrum characteristic based on the relationship between them. MFCC has been widely used in speech recognition. Mel frequency is proposed to describe this perceptual characteristic. It is a common feature parameter [7] in speech-related recognition research. It has a non-linear relationship with frequency f and its conversion formula is [8]:

$$B = 2595 \lg(1 + \frac{f}{700}) \quad (1)$$

or:

$$f = 700\left(10^{\frac{B}{2595}} - 1\right) \quad (2)$$

where B is Mel frequency; F is the frequency.
MFCC uses the relationship between Mel frequency and F to extract a cepstrum coefficient from Mel scale frequency domain, which is a commonly used feature extraction method in speech correlation recognition research. The feature extraction process is as follows:

- The signal $x(t)$ is pre-emphasized through a high-pass filter to enhance its high-frequency component, then the resulting signal is framed and windowed to obtain the time-domain signal $x(n)$.
- Then, the time domain signal $x(n)$ is transformed by Fast Fourier transform (FFT) to obtain the spectrum $X(k)$. The conversion formula is as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{j2\pi nk}{N}}, 0 < k < N-1 \quad (3)$$

- The resulting $X(k)$ passes through a set of Mel frequency filter banks to obtain the Mel spectrum, and logarithmic $S(m)$ is obtained by processing logarithmic energy:

$$S(m) = \ln(\sum_{k=1}^{N-1} |X(k)|^2 H_m(k)), 0 \le m \le M \quad (4)$$

- Finally, MFCC can be obtained by $S(m)$ through Discrete Cosine Transform (DCT).

$$c(n) = \sum_{m=1}^{M-1} S(m) \cos\frac{\pi n(m+0.5)}{M}, 0 \le m \le M \quad (5)$$

The specific flow chart of MFCC is shown in Figure 2 below:

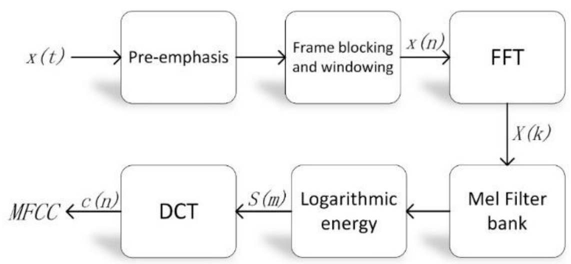

Figure 2. MFCC flowchart

III. Convolution Neural Network

CNN is a typical deep feed-forward artificial neural network, which is inspired by the biological perception mechanism. It is generally composed of convolution layer, pooling layer and fully connected layer, as shown in Figure 3.

The CNN constructed in this paper is a multilayer feature extraction network composed of convolution layer, pooling layer, full connection layer and classification layer. The convolution and pooling layers are arranged alternately, and the output of the convolution layer is normalized batch by batch scaling and shifting, then transferred to the active layer for non-linear processing. The whole network continually learns from the original signals one by one to extract their inherent characteristics, and transforms them into abstract deep features, revealing the essential characteristics of the original signals, and finally joins the SoftMax layer to classify the fault categories.

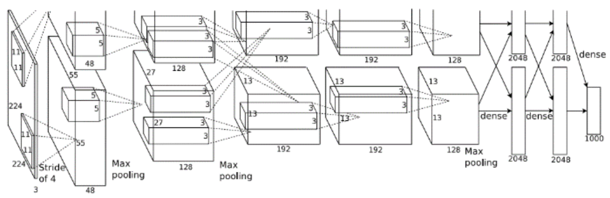

Figure 3. Basic network structure of CNN

1) **Convolution layer: Convolution layer is the core component of CNN. The formula for convolution operation is as follows:**

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} * \omega_{ij}^l + b_j^l) \quad (6)$$

where: superscript l denotes l layer in the corresponding network. i, j denote the sequence number of the feature map in layer l and layer $l-1$, respectively. M_j represents the feature map in layer $l-1$ connected to the jth feature map of layer l; ω_{ij}^l represents the convolution kernel parameters entered by the jth feature map of layer l corresponding to the ith feature map of layer $l-1$; b_j^l means offset; $*$ Represents a convolution operation.

2) Pooling layer: Generally, the pooled layer is located in the middle of the continuous convolution layer. The pooled layer can gradually reduce the size of the expression space, reduce the network parameters and computational load, and also play a role of controlling over-fitting. The pooled layer consists of maximum pooling and average pooling. In recent years, maximum pooling has been widely used because it has been proved to be more effective. Therefore, maximum pooling is used in this paper, and its mathematical description is shown in equation (7).

$$p^{l(i,j)} = \max_{(j-1)W+1 \le t \le jW} \{a^{l(i,t)}\} \tag{7}$$

where: $a^{l(i,t)}$ is the activation value of the tth neuron in the lth layer iframe; W is the width of the pooled area; $p^{l(i,j)}$ is the result of pooling.

3) Full connection layer: Full connection layer classifies the extracted features. The output of the last pooled layer is flattened to a one-dimensional eigenvector, and as input to the fully connected layer, the forward propagation of the fully connected layer is shown in equation (8).

$$n_j^{l+1} = \sum_{i=1}^{n} W_{ij}^l a^{l(i)} + b_j^l \tag{8}$$

where: W_{ij}^l is the weight value between the ith neuron in layer l and the jth neuron in layer $l+1$. n_j^{l+1} is the output value of layer $l+1$ jth neuron; b_j^l is the offset of all the neurons in layer l to the jth neuron in layer $l+1$.

The final output layer uses the SoftMax activation function to transform the input neurons into a probability distribution of sum 1.

B. Experimental Validation

Experimental Environment

This experiment is to configure Python environment under Windows 10 operating system, using PyCharm 2020.2.3 (Community Edition) software, and install a deep learning framework of keras-2.4.3 and TensorFlow-2.5.0.

Experimental Data

This experiment uses MIMII data set (Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection) [9-10]. This dataset truly records the sound of industrial machines under normal and abnormal operating conditions in real factory environments, and the background noise recorded in multiple real factories is mixed with the sound of machines. The dataset contains sound produced by four types of industrial machines, valves, pumps, fans, and slides. Sound is recorded using an 8-channel microphone array at a sampling rate of 16 kHz and 16 bit. Each type of machine consists of seven separate product models (now only four models, the next version will include the remaining three). The data for each model contains normal and abnormal sounds, and the data has three SNR versions of -6dB, 0dB and 6dB. Part of the data is shown in Figure 4 below.

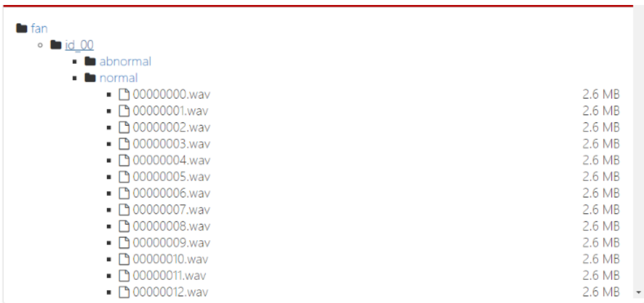

Figure 4. MIMII data set

C. Preprocessing of Experimental Data

Despite the use of publicly available sound datasets online, the number of data is relatively small, and it is still difficult to achieve the amount of data required for training models. Training in-depth learning models on such a small dataset can easily lead to over-fitting (the models perform well on training data, but do not generalize well on test data). Therefore, the data expansion method is used to solve this problem.

Compared with other formats, wav format files can most completely guarantee the sound quality of source files, is lossless, and is more conducive to extracting the characteristics of device failure, but wav format files will occupy more memory (MIMII dataset all data occupies nearly 100 GB of memory), so this paper only selects pump data for experiments. Considering the problems of universality and large difference of category data, the sound data of the four types of pumps are mixed together, and the sound data is divided and processed to increase the amount of data, prevent the network from overfitting, and finally 3648 WAV files are obtained.

D. Selection of Parameters

4) Activation Function: Relu is a rectifier linear unit. It is also one of the most widely used activation functions. The most widely used activation function is first because it is non-linear, which means that the back-propagation algorithm is available. A very good feature of Relu is that it does not activate all the neurons at once. Relu function expression:

$$f(x) = \max(0, x) \tag{9}$$

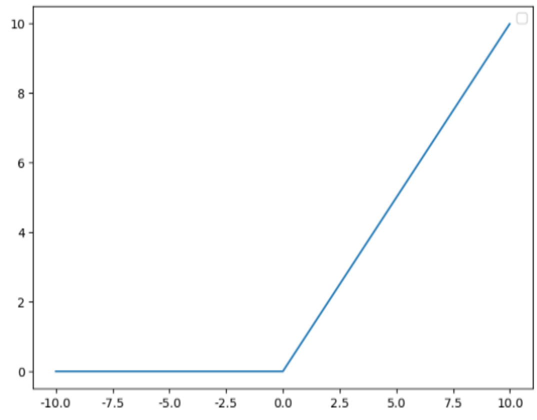

Figure 5. Relu function image

As you can see in Figure 5, the Relu function outputs 0 when the input is negative, and the neuron is not activated. This means that only a few neurons will be activated at the same time, sparse the network and thus very efficient for calculation. The common activation functions are Sigmoid and Tanh, but they have a common disadvantage: when the values of input data are large, the learning speed of the neural network is slow.

5) Loss Function: The cross-entropy loss function is often used in the classification of neural networks. The design of this paper is two classifications, which are expressed as:

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N} - \sum_{c=1}^{M} y_{ic} \log(p_{ic}) \tag{10}$$

where: M: the number representing the category; y_{ic}: Represents the indicator variable (0 or 1), if the category is the same as the category of sample i, it is 1, otherwise 0; p_{ic} : the prediction probability that observation sample I belongs to category C.

6) Optimization Algorithm: Adam optimizer is used in this design. The Adam algorithm is a combination of momentum gradient reduction and RMSprop, which can adaptively change the learning rate, require less resources, and converge faster (find the minimum loss value) to speed up the learning speed and effect of the machine.

IV. Experimental Validation

In this experiment, the average value was obtained after 10 experiments. The specific network adjustment parameters are as follows:

For the adjustment parameters of convolution kernel size, experiments are carried out by setting the convolution kernel size to 4, 8, 16, 32 respectively, while setting other parameters unchanged. The result is shown in Figure 6 below. From the graph, we can see that when other parameters are unchanged and the convolution kernel size is set to 16, it has the highest test accuracy (97.05%) and the smallest loss function value. A comparison of 32 sets of experiments showed that the larger the convolution kernel, the better the final classification. Therefore, the convolution kernel size is selected as 16 during subsequent parameter adjustment.

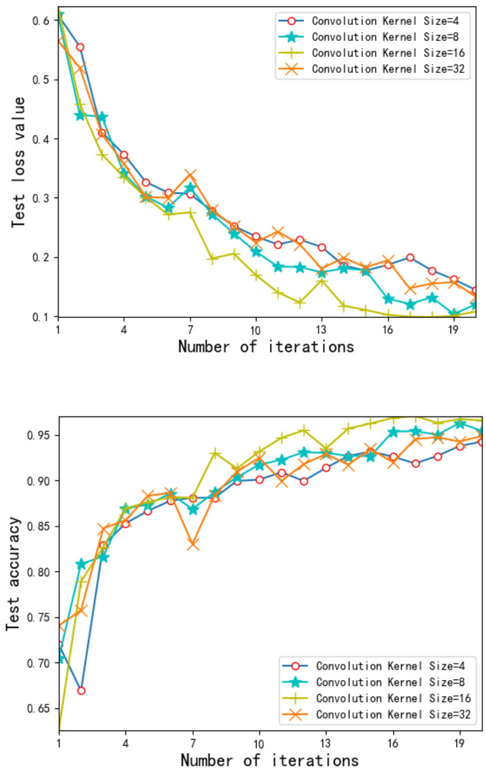

Figure 6. CNN model comparison of different convolution cores

For the adjustment parameters of convolution and pooling layers: set the number of convolution and pooling layers to 1, 2, 3, 4 respectively, while setting other parameters unchanged. As a result, in Figure 7 below, we can see that when the other parameters are unchanged and the number of convolution pooling layers is set to 2, it has the highest test accuracy and the lowest loss function value. Comparing 3 and 4 groups of experiments, we found that the more layers there are, the better the final classification will be. Therefore, the number of convolution pooling layers is chosen as 2 during subsequent parameter adjustment.

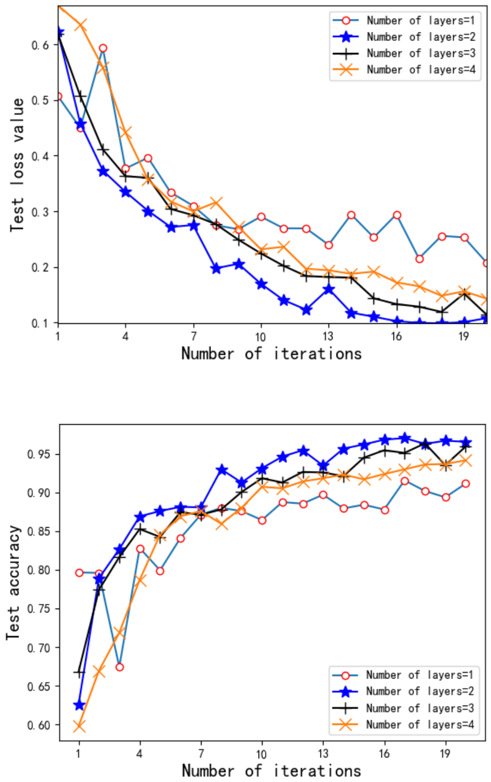

Figure 7. Comparison of CNN models with different convolution pooling layers

For the adjustment parameters of the number of full connection layers: set the number of full connection layers to 1, 2, 3, 4 respectively, while setting other parameters unchanged. As a result, in Figure 8 below, we can find that when the other parameters are unchanged and the number of convolution pooling layers is set to 2, it has the highest test accuracy and the lowest loss function value. Comparing 3 and 4 groups of experiments, we found that the more layers there are, the better the final classification will be. Therefore, in the subsequent adjustment of parameters, the number of full connection layers is selected as 2.

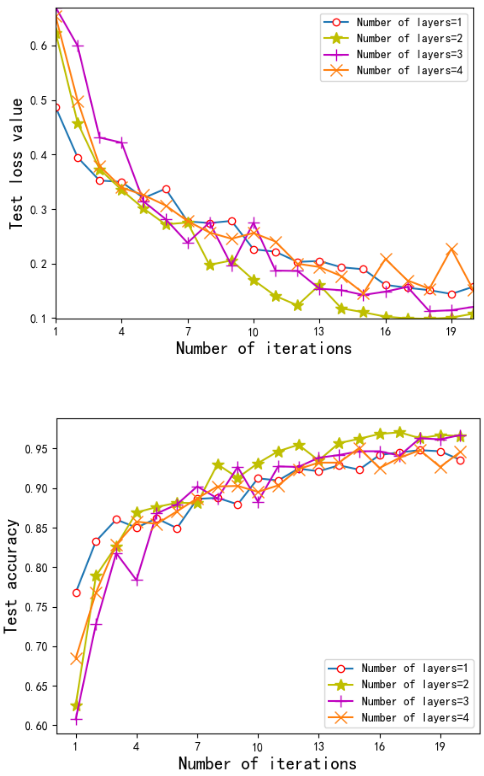

Figure 8. Comparison of CNN models with different number of full connection layers

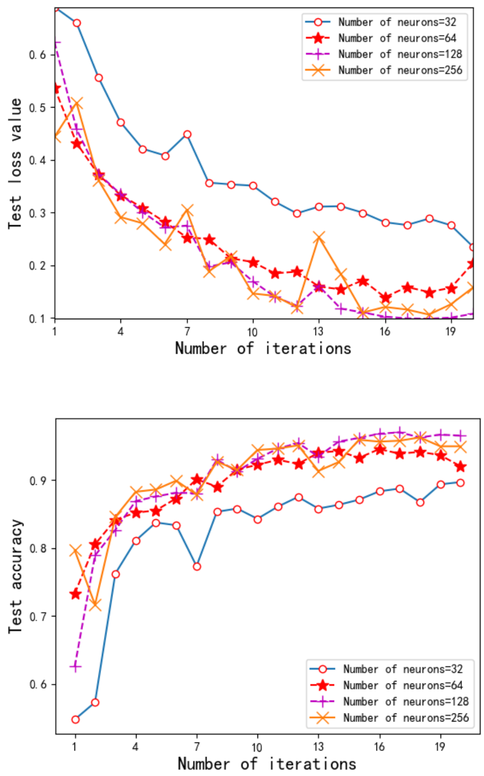

Figure 9. Comparison of CNN models for different numbers of neurons in the full connective layer

For the adjustment parameters of the number of fully connected layers of neurons: experiment was carried out by setting the number of fully connected layers of 32, 64, 128 and 256, respectively, while setting other parameters unchanged. As a result, in Figure 8 below, we can find that when the other parameters are unchanged and the number of convolution pooling layers is set to 128, it has the highest test accuracy and the lowest loss function value. Comparing 256 groups of experiments, it was found that the more neurons there were, the better the final classification was. Therefore, in the subsequent adjustment of parameters, the number of full connection layers is selected as 128.

After adjusting the parameters above, the parameters of the neural network are obtained as follows:

TABLE I. CNN MODEL PARAMETER

category	parameter
Convolution layer	Layer/Convolution Kernel: 2,16
Pooling layer	Layer: 2
Full connection layer	Number of neurons: 128
Activation function	Relu
Dropout layer	0.2
Softmax layer	Classification number is 2
Maximum iteration	20
Batch size	128
Optimizer	Adam

The highest test accuracy of the CNN model in 20 iterations is 97.05%. This shows that the method proposed in this paper is effective and has some application value. And compared with previous experiments, we can find that for the adjustment parameters of the neural network, the larger the parameters, the better, but the more suitable the parameters. Therefore, for the adjustment process, we can only patiently do a few more experiments to verify the quality of the parameters.

V. CONCLUSION

In this paper, a method of mechanical fault diagnosis based on sound signal and CNN is presented. The main method is to sample the collected sound signal, then extract the signal characteristics, classify the data into training set and test set according to 6:4 ratio, and then input it into CNN to classify the result. By comparing and adjusting the parameters for too many sets of data, we find that the parameters of the optimal network model are two layers of convolution, pooling and full connectivity, and the convolution nucleus size is 16 and the number of neurons in the full connectivity layer is 128. In only 20 training test iterations, the test accuracy can reach 97.05%, which indicates the validity of the model. This method can effectively solve the drawbacks of traditional artificial hearing to achieve fault diagnosis, which requires manual feature extraction and poor universal applicability for signal processing. This method only needs normal and abnormal sound data during any mechanical operation, so it has some practical application value.

VI. REFERENCES

[1]. T. Feng and J. Wang et al, Fault diagnosis method for micro-vibration motor based on CNN and time-frequency characteristic map of sound[J]. CHINA MEASUREMENT & TEST, 2019,45(10):120-127.

[2]. J. J. Li, Research and Application of the Fault Diagnosis of Rolling Bearing Based on the Sound Signal [D]. Shijiazhuang Tiedao University,2017.

[3]. W. W. Cai and J. Huang et al, Research on Fault Diagnosis Method for Micro Motor Based on Sound Signal [J]. MACHINE TOOL & HYDRAULICS, 2020,48(23):190-195.

[4]. S. K. LIU and Y. J. WU et al, Wind turbine gearbox fault diagnosis based on sound signal and improved MS-LMD[J]. JOURNAL OF VIBRATION AND SHOCK, 2021, 40(11):11.

[5]. Y. W. Sun and L. G. Luo et al, Mechanical Fault Diagnosis Method of Circuit Breaker Based on Sound Characteristics and Improved Sparse Representation Classification [J/OL]. Power System Technology:1-9[2022-01-10].

[6]. Y. W. Yang and Y. G. Guan et al, Mechanical Fault Diagnosis Method of High Voltage Circuit Breaker Based on Sound Signal [J]. Proceedings of the CSEE, 2018,38(22):6730-6737.

[7]. B?Rvik T , Hopperstad O S , Berstad T , et al. A computational model of viscoplasticity and ductile damage for impact and penetration[J]. European Journal of Mechanics, 2001, 20(5):685-712.

[8]. Sahidullah M , Saha G . A novel windowing technique for efficient computation of MFCC for speaker recognition [J].IEEE Signal Processing Letters , 2013, 20(2):149-152.

[9]. Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi, “MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection,” arXiv preprint arXiv:1909.09347, 2019.

[10]. Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi, “MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection,” in Proc. 4th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE), 2019.

Cite this article as : Lemei Han, Zhan Wen, Haoning Pu, Wenzao Li, "Research on Mechanical Fault Diagnosis Algorithm Based on Sound Signal and CNN ", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 1, pp. 153-160, January-February 2022. Available at doi : https://doi.org/10.32628/IJSRSET229141
Journal URL : https://ijsrset.com/IJSRSET229141