A rolling bearing fault diagnosis model based on WCNN-BiGRU

Qilin Cheng, Bo Peng*, Qiang Li, Siping Liu
School of Information Engineering, Southwest University of Science and Technology, Mianyang, China
*Email address: pengball@swust.edu.cn

2021 China Automation Congress (CAC) | 978-1-6654-2647-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAC53003.2021.9728363

***Abstract*—Aiming at problems such as poor diagnosis effect of simple models and complex structure of hybrid models in the fault diagnosis of rolling bearing vibration signals, this paper establish a fault diagnosis model, which combines WCNN with BiGRU. First, the model extracts feature of the raw data through a one-dimensional wide convolutional network. Then, the reduced-dimensional fault sequence is sent to the BiGRU for further characteristics extracting. Finally, a classifier is used to categorize the fault diagnosis result. A comparison between our model and other models is given on the CWRU dataset. The experimental results show that the model proposed in this paper has a higher fault diagnosis rate than MLP, GRU, VGGNet-11, CNN-GRU, WDCNN, and with a fewer parameters. Besides, the model is applied to the data generated by the PT300 bearing fault simulation platform for fault diagnosis. Results show that the average accuracy of our model on both datasets exceeds 99%.**

Keywords – rolling bearing fault diagnosis;* PT300 bearing fault simulation platform*; WCNN; BiGRU

I. INTRODUCTION

Rolling bearings are one of the indispensable parts in industrial equipment. Bearing failures and serious safety problems are often caused by environmental factors such as high pressure, high temperature, and wear. Therefore, research on bearing fault diagnosis technology is a key content in the field of mechanical fault diagnosis.

Traditional statistical pattern recognition methods mostly use Bayesian criteria and linear and nonlinear discriminant functions for fault diagnosis, but they all have certain limitations. In recent years, scholars have been very enthusiastic about deep learning research. Bearing fault diagnosis has become a key field of applied research in deep learning. However, traditional deep learning methods still exist some problems. Such as, requiring manual feature extraction, complicated model structure, numerous parameters, and long training time. Han [1] presented a new method to extract fault feature, which combinations of local mean decomposition and multi-scale symbolic dynamic information entropy. It can accurately distinguish the faults and normal conditions of rolling bearings, and accurately classify the faults. Khorram [2] proposed an end-to-end CNN+LSTM deep learning method for bearing fault diagnosis. This model can accurately diagnose faults in a relatively short time. Li [3] proposed an end-to-end fault diagnosis method for rolling bearings. LSTM, GRU and 1DCNN are used to construct a deep learning network structure. Experiments have proved that 1DCNN performs best, which has a great significance to the intelligent diagnosis, prediction and health management of bearings. Xu [5] proposed a bearing fault diagnosis method, which combines deep convolutional neural network and random forest ensemble learning. The continuous wavelet transform is used to convert the one-dimensional time-domain vibration signal into a two-dimensional (2D) gray image. But there is still the problem of slow model convergence.

In view of the early fault diagnosis of rolling bearings, Zhang [6] proposed a rolling bearing initial failure detection method based on multi-scale convolutional neural network and gated recurrent unit networks with attention mechanism (MCNN-AGRU), which effectively solved the running status of rolling bearings. Early fault detection issues in monitoring and performance degradation assessment. In order to realize intelligent diagnosis and improve the recognition rate, Song [7] proposed a CNN-based wide convolution kernel neural network, which uses a strategy of widening the convolution kernel to obtain a larger receptive field and improve the network's anti-noise ability.

The bidirectional gated recurrent unit (BiGRU) can fully explore the correlation between time series and improve the model recognition effect. However, when the amount of data is too large, the BiGRU also has the problem of poor extraction of non-linear features of the data and slow model training [8].

Regarding the issue above, this paper combines the characteristics of one-dimensional wide convolutional kernel neural network (WCNN) and BiGRU built a rolling bearing fault diagnosis model that combine WCNN with BiGRU. The raw vibration signal data can directly input into the model for feature extraction. First, the model uses a layer of CNN with wide convolution kernel to extract the short timing characteristics of bearing faults, and a layer of CNN with small kernel to achieve nonlinear mapping. Then, the fault sequence after dimensionality reduction by convolution input into the BiGRU for training and learning. Finally, the “softmax” classifier is used to achieve fault classification and output.

II. THEORY OF CNN AND GRU

A. Convolutional Neural Network

Convolutional neural network is a feed-forward neural network. It has the characteristics of simple structure, weight sharing and pooling down sampling. It can efficiently extract and classify the feature information in image data, and has extremely strong image recognition performance. In recent years, through a lot of research, it has been found that CNN also has advantages in processing time series, and has excellent noise tolerance on time series.

As shown in Figure 1, the input layer, convolutional layer, pooling layer, fully connected layer and "softmax" output layer constitute a simple CNN model. A deeper network structure can be constructed by stacking each part of the convolutional neural network.

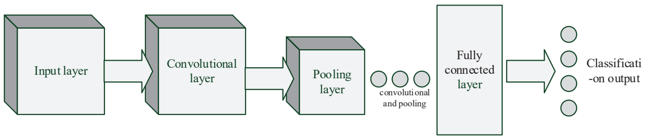

Fig. 1 The structure of convolutional neural network

B. Feature Extraction of Convolutional Layer

The function of the CNN layer is to extract features of things through convolution, and the features extracted by convolution are generally local information. Neurons only need to perceive the part, share the weights of different convolutional layers, and integrate the local information at a higher layer to obtain global information. The feature extraction process can express as in (1):

$$x_{i+1} = W_i \otimes x_i + b_i \tag{1}$$

Among the formula (1), x_i is the input of the i layer; x_{i+1} is the feature result of the convolution calculation; $\otimes$ is the convolution operator; W_i is the weight of the convolution kernel; b_i is the bias value.

In the CNN network, batch normalization used in each layer of convolution to correct the parameters of the network to improve the sparsity in the network training process and prevent the model from overfitting.

C. GRU Network

GRU is a kind of recurrent neural network, which proposed to solve the problems of long-term memory and gradient in back propagation. Compared with LSTM, GRU is easier to train under the same performance, which can greatly improve training efficiency.

GRU has a reset gate and an update gate, a total of two gates. The reset gate acts on the new input information and the previous memory information, and the update gate is used to constrain the input amount of the previous memory information to the current time step. The GRU unit structure is shown in Figure 2.

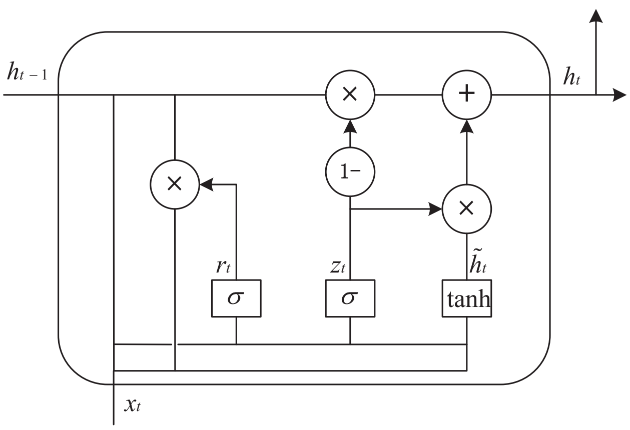

Fig. 2 The structure of GRU

The calculation formulas for the reset gate and update gate are as follows in (2):

$$\begin{gathered} r_t = \sigma(W^{(r)} \bullet [h_{t-1}, x_t]) \\ z_t = \sigma(W^{(z)} \bullet [h_{t-1}, x_t]) \end{gathered} \tag{2}$$

Among the formula (2), r_t is the reset gate, z_t is the update gate, x_t is the input at the time t. h_{t-1} is the information of the $t-1$ time step. $W^{(r)}$ is the weight matrix from the input layer to the update gate. $W^{(z)}$ is the weight matrix from the hidden state to the reset gate.

The hidden state calculation formula is as follows in (3):

$$\begin{gathered} \tilde{h}_t = \tanh(W^{(h)} \bullet [r_t * h_{t-1}, x_t]) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{gathered} \tag{3}$$

Among the formula (3), $W^{(h)}$ is the weight matrix from the input layer to the hidden state, h_t is the information of the t time step, $\tilde{h}_t$ is the candidate information at time t.

A forward GRU network and a reverse GRU network constitute a BiGRU network, which can more fully mine fault characteristic information.

III. WCNN-BiGRU BEARING FAULT DIAGNOSIS MODEL STRUCTURE

One-dimensional convolutional neural networks can be well used to analyze signal data with a fixed length period. Therefore, this paper establishes a WCNN-BiGRU model based on the characteristics of CWRU bearing vibration data. The model structure is shown in Figure 3.

The structural parameters of the model are obtained after many experiments and adjustments. The CNN structure consists of two convolutional pooling layers, as well as a fully connected layer. The batch normalization is used by each convolutional layer to improve training efficiency, and uses L2 regularization to prevent overfitting.

The model's first convolution layer uses a wide convolution kernel to obtain a larger receptive field, which is beneficial to extract sequence features and suppress noise interference [4]. The second convolutional layer uses a small convolution kernel to perform non-linear mapping of features. The maximum pooling method is used to retain the main characteristic information in the pooling layer. A layer BiGRU network is used to learn feature sequence information. In order to make the model effect better, the Adam algorithm is used to optimize the network weights. Finally, the softmax classification function and the fully connected layer are combined to achieve fault classification and output. The final WCNN-BiGRU model's structure parameters is shown in Table 1.

IV. EXPERIMENTAL ANALYSIS

A. Data Preprocessing

This paper uses CWRU's public dataset to verify the model. In the experiment, the dataset is divided into training set, validation set and test set according to the ratio of 7:2:1. The dataset uses sliding sampling method for data enhancement. In the experiment, 2000 samples were selected for each type of failure. There are totally 20000 samples. The 1024 data points are used for fault diagnosis each time, and the final experimental sample information is shown in Table 2. Each type of fault is divided into 1400, 400, and 200 samples in proportion, and labeled with 1 to 10.

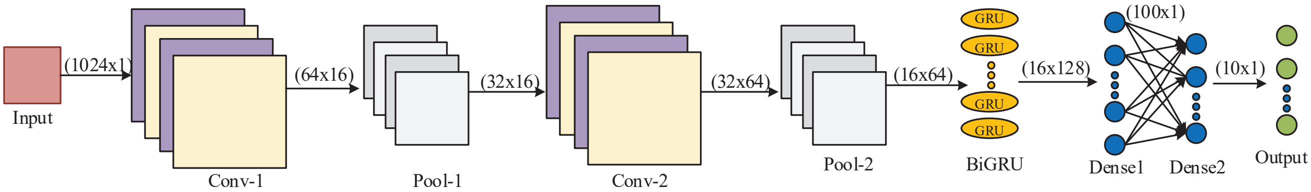

Fig. 3 Model structure

Table 1 Model structure parameters

Network Layer	Kernel Size/ Stride	Neurons/Filters	Activation	Pooling	Input Size	Output Size
Conv-layer 1	64x1/16	16	relu	2	(1024,1)	(64,16)
Conv-layer 2	1x1/1	64	relu	2	(32,16)	(32,64)
BiGRU layer 1	-	64	tanh	-	(16,64)	(16,128)
Dense1	-	100	relu	-	(16,128)	(100,1)
Dense2	-	10	softmax	-	(100,1)	(10,1)

Table 2 Sample information of CWRU experimental dataset

Fault types	Number of samples			Label
	Train set	Validation set	Test set	
Normal	1400	400	200	1
Ball (7mils)	1400	400	200	2
Ball (14mils)	1400	400	200	3
Ball (21mils)	1400	400	200	4
InnerRace (7mils)	1400	400	200	5
InnerRace (14mils)	1400	400	200	6
InnerRace (21mils)	1400	400	200	7
OuterRace (7mils)	1400	400	200	8
OuterRace (14mils)	1400	400	200	9
OuterRace (21mils)	1400	400	200	10

B. Analysis of Experimental Results

The model proposed in this paper is built through the Keras framework, and the computer is configured with CPU lntel(R) Core(TM) i5-8300H CPU @2.30GHz 2.30 GHz, 16G memory. The experiment choose the CWRU's bearing fault vibration data as the experimental dataset. The sampling frequency of the samples is 12KHz, and the samples are selected from the normal state data , the bearing inner race , outer race and ball fault data under the fault sizes of 7 mils, 14 mils and 21 mils.

The experiment selects bearing fault data under 0HP. The maximum number of the model iterations is set to 15, and the batch size is set to 128. The accuracy of the proposed model on the validation set is shown in Figure 4, and the loss value is shown in Figure 5.

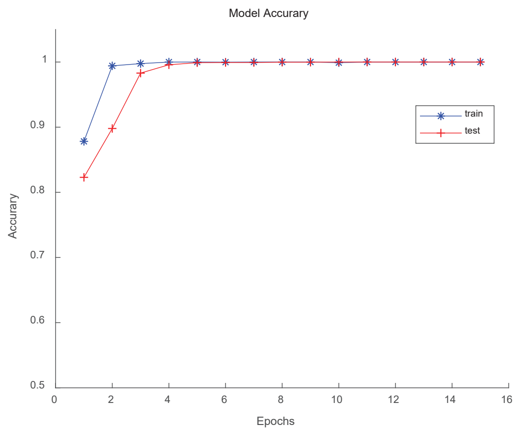

Fig. 4 Model accuracy on validation set

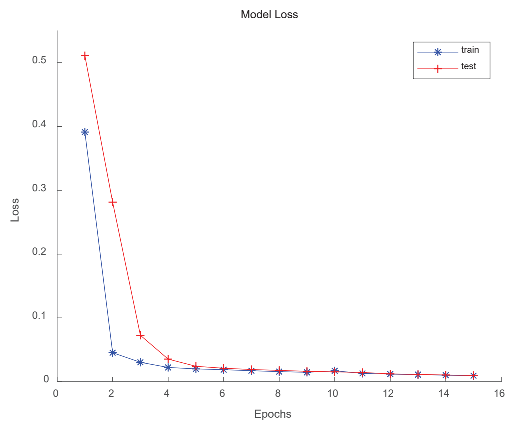

Fig. 5 Model loss on validation set

According to the results, it can be seen that with the number of iterations increases, the accuracy of the model on the validation set quickly converges and close to 1.0, and the target loss value function also quickly converges and close to 0.

In order to verify the performance of the proposed model, the raw data is sent to the model for fault classification verification. The experimental method of this paper refers to [9]. Under the same experimental conditions, the MLP model, VGGNet-11 model, single-layer GRU model, CNN-GRU model, and WDCNN[4] model are used for fault diagnosis and comparative analysis. The test results of each model obtained is shown in Table 3. The accuracy of all models on the validation set is shown in Figure 6, and the loss value is shown in Figure 7.

Table 3 The different model's average calculation results on test set

Model	Accurary	Loss	Total Params
MLP	0.8465	0.6105	5506570
VGGNet-11	0.9835	0.1209	87010186
GRU	0.8510	0.3581	668234
CNN-GRU	0.9819	0.0483	192010
WDCNN	0.9980	0.0330	41286
WCNN-BiGRU	0.9995	0.0124	258278

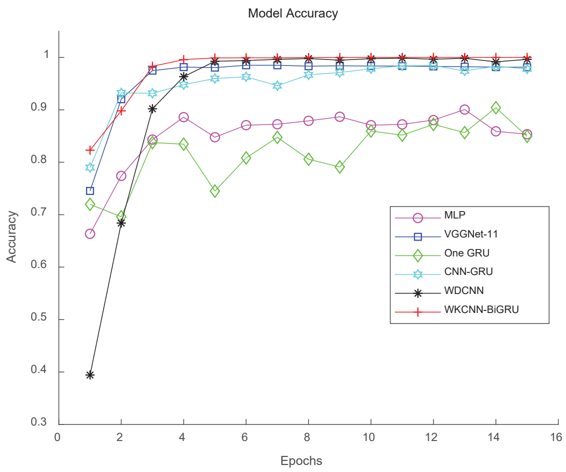

Fig. 6 Comparison of the accuracy values of different models on the validation set

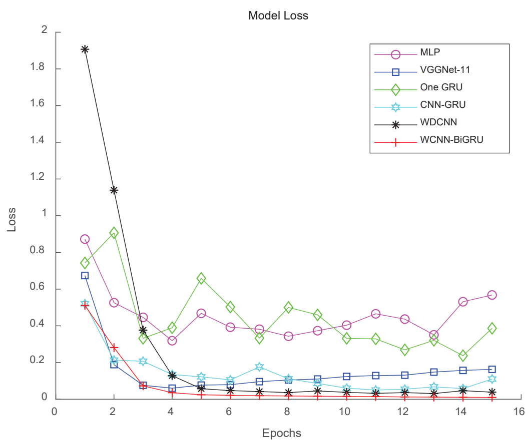

Fig. 7 Comparison of the loss values of different models on the validation set

As can be seen from the figure, WDCNN has the best accuracy among the five comparison models, and it has an accuracy of 99.8%. In addition, the WDCNN model is the most stable, and the stability of other models is slightly worse. The WCNN-BiGRU model proposed in this paper has an accuracy of 99.95% on the test set, which is higher than other comparison models. The WCNN-BiGRU model also has the lowest loss value and the fastest convergence in the other five models compared in experiment.

Table 3 and the results on the validation set show that the WCNN-BiGRU model has fewer parameters and higher fault diagnosis accuracy than the complex models (MLP, VGGNet-11, GRU). Compared with the simple structure model (WDCNN, CNN-GRU), the WCNN-BiGRU model converges faster and has a higher accuracy rate.

V. EXPERIMENTAL ANALYSIS OF PT300 BEARING FAULT SIMULATION PLATFORM

Previously, the WCNN-BiGRU model has verified the feasibility and correctness by using an internationally recognized dataset. Next, the model is used to diagnose the raw data generated by the VALENIAN PT300 bearing vibration simulation platform. The VALENIAN PT300 platform can simulate six states of the bearing, which include: ball fault (FBB), comprehensive fault (FBC), inner race fault (FBI), outer race fault (FBO), trestle fault (FBT) and normal. The platform is shown in Figure 8.

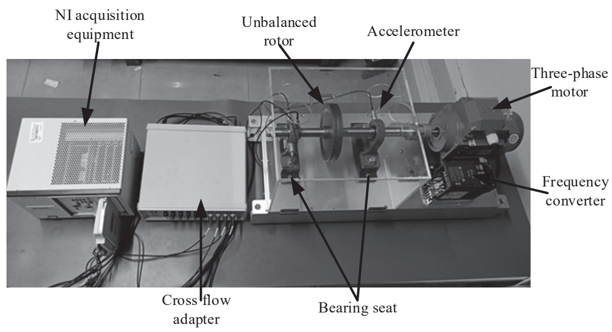

Fig. 8 PT300 bearing vibration fault simulation platform

Refer to the bearing vibration simulation platform of CWRU to configure the test environment parameters. Data collection is performed under the condition of bearing rotating speed of 1780r/h. The data is collected by NI equipment to the host computer, and the sampling frequency is set to 12 KHz. The method of making the dataset is the same as the bearing fault dataset of CWRU. The experiment is carried out under the condition same as the CWRU's dataset. The distribution of the number of samples is shown in Table 4.

Table 4 Sample information of PT300 bearing vibration dataset

Fault types	Number of samples			Label
	Train set	Validation set	Test set	
Normal	1400	400	200	1
FBC	1400	400	200	2
FBI	1400	400	200	3
FBO	1400	400	200	4
FBT	1400	400	200	5
FBB	1400	400	200	6

The model proposed in this paper is applied to this data set for fault classification test. In each experiment, the loss value and accuracy of the fault classification were recorded, and a total of 100 times experiments are carried out. The loss value of the model's fault classification is shown in Figure 9, and the accuracy is shown in Figure 10.

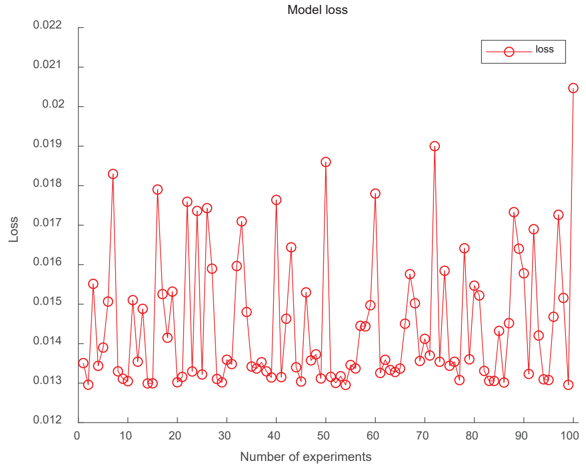

Fig. 9 Model Loss

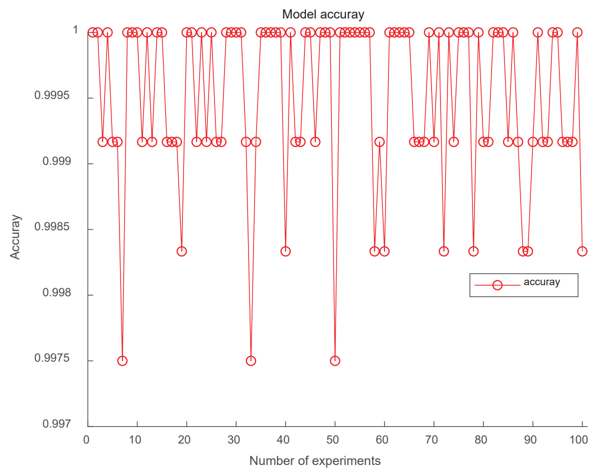

Fig. 10 Model Accuracy

The experimental results show that the classification accuracy of the model exceeds 99.7%, and the loss value is basically kept below 0.02. It indicates that the WCNN-BiGRU model still has high stability and accuracy in actual fault applications.

VI. CONCLUSION

The WCNN-BiGRU model proposed in this paper can directly act on the raw data of bearing vibration signals and has a simple structure. By comparing with the MLP model, VGGNet-11 model, GRU model, CNN-GRU model, WDCNN model on the CWRU dataset, it can be seen that in the case of simple structure, the fault diagnosis accuracy of the WCNN-BiGRU model is higher and the loss value is smaller and the model converges faster.

Finally, our model is applied to the bearing vibration data generated by the PT300 bearing fault simulation platform for fault diagnosis. Under the same experimental conditions, the model proposed in this paper carried out 100 experiments on the dataset collected by the PT300 bearing vibration platform. The results show that the fault classification accuracy of the proposed model exceeds 99.7%, which shows that the model has good reliability and practicability.

VII. ACKNOWLEDGMENT

This paper supports by National Key R&D Program (2019YFB1705100).

VIII. REFERENCES

[1] Han M , Wu Y , Wang Y , et al. Roller bearing fault diagnosis based on LMD and multi-scale symbolic dynamic information entropy[J]. Journal of Mechanical Science and Technology, 2021(2).

[2] Khorram A , Khalooei M , Rezghi M . End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis[J]. Applied Intelligence, 2020(1):1-16.

[3] Li Y , Qiu B , Wei M , et al. Deep Learning based End-to-End Rolling Bearing Fault Diagnosis[C] 2019 Prognostics and System Health Management Conference (PHM-Qingdao). 2019.

[4] Wei Z , Peng G , Li C , et al. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals[J]. Sensors, 2017, 17(3):425.

[5] Xu G , Liu M , Jiang Z , et al. Bearing Fault Diagnosis Method Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning[J]. Sensors, 2019, 19(5).

[6] Zhang X , Cong Y , Yuan Z , et al. Early Fault Detection Method of Rolling Bearing Based on MCNN and GRU Network with an Attention Mechanism[J]. Shock and Vibration, 2021, 2021(3):1-13.

[7] Song X , Cong Y , Song Y , et al. A bearing fault diagnosis model based on CNN with wide convolution kernels[J]. Journal of Ambient Intelligence and Humanized Computing, 2021.

[8] Qiu H , Fan C , Yao J , et al. Chinese Microblog Sentiment Detection Based on CNN-BiGRU and Multihead Attention Mechanism[J]. Scientific Programming, 2020, 2020:1-13.

[9] CHEN Baojia, CHEN Xueli, SHEN Baoming, et al. An Application of Convolution Neural Network and Long Short-Term Memory in Rolling Bearing Fault Diagnosis[J]. Journal of Xi'An Jiaotong University, 2021, 55(6): 28-36.

[10] Mao X T , Zhang F , Wang G , et al. Semi-random subspace with Bi-GRU: Fusing statistical and deep representation features for bearing fault diagnosis[J]. Measurement, 2020:108603.

[11] Deng H , Zhang W X , Liang Z F . Application of BP Neural Network and Convolutional Neural Network (CNN) in Bearing Fault Diagnosis[J]. IOP Conference Series: Materials Science and Engineering, 2021, 1043(4):042026 (9pp).

[12] Han S , Jeong J . An Weighted CNN Ensemble Model with Small Amount of Data for Bearing Fault Diagnosis[J]. Procedia Computer Science, 2020, 175:88-95.